

PREDICCIÓN PROBABILÍSTICA DE ESTADOS GLUCÉMICOS PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

Alvis Cabrera Tejera

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.



TESIS DOCTORAL

PREDICCIÓN PROBABILÍSTICA DE ESTADOS GLUCÉMICOS
PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL
ANÁLISIS DE DATOS COMPOSICIONALES

ALVIS CABRERA TEJERA

2023



TESIS DOCTORAL

PREDICCIÓN PROBABILÍSTICA DE ESTADOS GLUCÉMICOS
PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL
ANÁLISIS DE DATOS COMPOSICIONALES

ALVIS CABRERA TEJERA

2023

PROGRAMA DE DOCTORADO EN TECNOLOGÍA

Supervisada por:
Josep Vehí Casellas
Aleix Beneyto Tantiña

Memoria presentada para optar al título de doctora por la
Universidad de Girona



TESIS DOCTORAL

PREDICCIÓN PROBABILÍSTICA DE ESTADOS GLUCÉMICOS
PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL
ANÁLISIS DE DATOS COMPOSICIONALES

Presentado en cumplimiento parcial de los
requisitos para obtener el título de doctora
por la Universidad de Girona.

Por:

Alvis Cabrera Tejera

Supervisores:

Josep Vehí Casellas

Aleix Beneyto Tantiña

*A mi Dios, que nunca me ha abandonado.
A mi Alison, por ser mi razón de ser.
A mi esposo y amigo, Ernesto, por su soporte siempre.
A mi madre, Alvis, por sus innumerables esfuerzos.
A mi abuelo, Juan, por confiar en mi.*

AGRADECIMIENTOS

Este logro no habría sido posible sin el invaluable respaldo y guía incondicional de un grupo de personas que me han acompañado a lo largo de este camino. Por tanto, deseo expresar mi más sincera gratitud.

Primeramente agradecer a Dios, por siempre sostenerme. A mi madre Alvis, por todo su esfuerzo y ser mi heroína. A mi abuelo Juan, que me ha apoyado en todas mis aventuras con amor incondicional. Agradezco a mis suegros, quienes desde el primer día, me hicieron sentir como una hija más. A mis hermanos y sobrinos que aunque lejos me motivaban a ser mejor.

A mi amado esposo Ernesto, le expreso mi gratitud y admiración por ser mi apoyo incondicional y por dedicar su tiempo a discutir sobre mi trabajo para ayudarme a mejorar e ir por más. A mi pequeña Alison, por ser nuestro motor impulsor y una bendición en nuestras vidas.

Mi más sincera gratitud por mi supervisor el Dr. Josep Vehí, quien me brindó la oportunidad de realizar el doctorado a pesar de no conocerme. Me ha guiado con sus desafíos, paciencia y motivación para ver más allá. Siempre le estaré agradecida y le admiraré profundamente.

Expresar mi gratitud hacia mi co-supervisor Dr. Aleix Beneyto, quien discutió ideas y compartió su inmensa sabiduría y su escaso tiempo conmigo, por su exigencia y profesionalidad, y por ser un ejemplo a seguir como investigador consagrado.

Asimismo, quiero agradecer al Dr. Josep Antoni Martín-Fernández, quien desde que lo conocí, con su carácter afable y generoso compartió su vasto conocimiento de estadística y CoDa conmigo, siempre dispuesto a ayudar con prontitud en todas mis tareas y nuevas ideas.

Estoy muy agradecida con Iván Contreras, quien no solo nos brindó sus conocimientos, sino también su amistad sincera desde el primer momento, ganándose un pedacito de nuestro corazón con su sencillez. También quiero agradecer a Lyvia Biagi, quien me dedicó su tiempo y apoyo profesional a pesar de estar lejos de la UdG, y por su amistad más allá del trabajo.

Agradecer a todos mis colegas de MiceLab por los intercambios profesionales. A Anna Comas por toda la asistencia en el laboratorio desde mi incorporación y otros procedimientos relacionados con la beca y escuela de doctorado. También a Jorge Bondia por recomendarnos al Profesor Josep Vehí a pesar de no conocernos previamente.

Finalmente, agradecer al Programa de Investigadores en Formación de la Universidad de Girona (IFUdG2019) por hacer posible el desarrollo de toda esta investigación.

*Alvis Cabrera Tejera
Girona, España
mayo, 2023*

LISTA DE PUBLICACIONES

Esta tesis se basa en un compendio de las siguientes publicaciones:

1. Ernesto Estremera*, **Alvis Cabrera***, Aleix Beneyto and Josep Vehí. A simulator with realistic and challenging scenarios for virtual T1D patients undergoing CSII and MDI therapy, *Journal of Biomedical Informatics*, 132 (104141), **2022**, (Estremera et al., 2022).
2. **Alvis Cabrera**, Lyvia Biagi, Aleix Beneyto, Ernesto Estremera, Iván Contreras, Marga Giménez, Ignacio Conget, Jorge Bondia, Josep Antoni Martín-Fernández and Josep Vehí. Validation of a Probabilistic Prediction Model for Patients with Type 1 Diabetes Using Compositional Data Analysis, *Mathematics*, 11(5): 1241, **2023**, (Cabrera et al., 2023).
3. **Alvis Cabrera**, Ernesto Estremera, Aleix Beneyto, Lyvia Biagi, Iván Contreras, Josep Antoni Martín-Fernández and Josep Vehí. Individualized prediction of blood glucose outcomes using Compositional Data analysis, *International Journal of Applied Mathematics and Computer Science*, Enviado.

* Contribuyó igualmente a este trabajo.

El trabajo de investigación que condujo a esta tesis resultó en publicaciones adicionales en revistas y conferencias, que se enumeran a continuación y están ordenadas por fecha de publicación.

Revistas

4. Ernesto Estremera, Aleix Beneyto, **Alvis Cabrera**, Ivan Contreras and Vehi, J. (2023). Intermittent closed-loop blood glucose control for people with type 1 diabetes on multiple daily injections. *Computer Methods and Programs in Biomedicine*, 107568. <https://www.sciencedirect.com/science/article/pii/S016926072300233X>

Posters

1. **Alvis Cabrera**, Josep Vehi, Aleix Beneyto and Ernesto Estremera (2021, June). "Incorporating long-acting insulin into the hovorka model for in silico simulations of MDI therapies in T1D patients". In *Diabetes Technology & Therapeutics* 23: A98-A98. 140 Huguenot street, 3rd flor, New Rochelle, NY 10801 USA: MARY ANN LIEBERT, INC.
2. Ernesto Estremera, Josep Vehi, Aleix Beneyto and **Alvis Cabrera** (2021, June). "Generation of realistic scenarios including insulin variability and mixed meal library". In *Diabetes Technology & Therapeutics* 23: A97-A98. 140 Huguenot street, 3rd flor, New Rochelle, NY 10801 USA: MARY ANN LIEBERT, INC.

ACRÓNIMOS Y ABREVIATURAS

Los siguientes acrónimos y abreviaturas pueden ser encontrados en esta tesis.

Acrónimos y abreviaturas

ADA	American Diabetes Association
AdaBoost	Adaptive Boosting
alr	the additive logratio transformation
ANN	Artificial Neural Network
ARX	Autoregressive model with exogenous input
BDL	Below Detection Limit
BG	Blood Glucose
BGM	Blood Glucose Monitoring
BM	Bayesian-multiplicative
CGM	Continuous Glucose Monitoring
CHO	Carbohydrate
CIBERDEM	Centro de Investigación Biomédica en Red en Diabetes y Enfermedades Metabólicas Asociadas
clr	the centred logratio transformation
CNN	Convolutional Neural Network
CLÍNIC	Hospital Clínic of Barcelona
CoDa	Compositional Data
CRNN	Convolutional Recurrent Neural Network
CSII	Continuous Subcutaneous Insulin Infusion
CSIC	Biomedical Research Institute in Barcelona
CV	Coefficient of Variation
DL	Detection limit
DM	Diabetes Mellitus
DRNN	Dilated Recurrent Neural Network
DSS	Decision Support System
EM	Expectation and Maximization algorithm
eSCAPE	Consorcio Español de Páncreas Artificial y Tecnología de la Diabetes
ESN	Echo State Network
FDA	Food and Drug Administration

Gla-100	Glargine 100 U/ml
Gla-300	Glargine 300 U/ml
GMI	Glucose Management Indicator
GV	Glycemic variability
HbA1c	Glycated Hemoglobin
HLA	Human Leukocyte Antigen
IDDM	Insulin-dependent Diabetes Mellitus
IDF	International Diabetes Federation
IDIPAPS	Institut d'Investigacions Biomèdiques August Pi i Sunyer
IFG	Impaired Fasting Glucose
IVGTT	Intravenous Glucose Tolerance Test
DM	Diabetes Mellitus
ilr	the isometric logratio transformation
Institute ai2	Instituto Universitario de Automática e Informática Industrial
LSTM	Long Short-Term Memory
MATLAB	MATrix LABoratory
MAR	Missing At Random
MCAR	Missing Completely At Random
MDI	Multiple Daily Injections
MEDERI Living Lab	Medical Devices Research & innovation Living Lab
NMAR	Not Missing At Random
NIDDM	Non-insulin-dependent Diabetes Mellitus
ODE	Ordinary Differential Equations
OMS	World Health Organization
PH	Prediction Horizon
PotAGT	Potential abnormality of glucose tolerance
PrevAGT	Previous abnormality of glucose intolerance
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SMBG	Sensor-Augmented Pump
SMBG	Sensor-Augmented Pump
ST	Short-Terms
SVR	Support Vector Regression
TAR	Time Above Range
TBR	Time Below Range
TIR	Time in Range
T1DM	Type 1 Diabetes Mellitus
T2DM	Type 2 Diabetes Mellitus
T1D	Type 1 Diabetes
T2D	Type 2 Diabetes
UCI	Unidades de Cuidados Intensivos
UdG	Universitat de Girona
UPV	Universitat Politècnica de València
VP	Virtual Patient

ÍNDICE

Lista de Publicaciones	v
Acrónimos y Abreviaturas	vii
Lista de Tablas	xiii
Lista de Figuras	xv
Abstract	xvii
Resumen	xix
Resum	xxi
1 Introducción	1
1.1 Diabetes mellitus	1
1.1.1 Complicaciones	4
1.1.2 Hormonas pancreáticas. Insulina y glucagón	6
1.1.3 Análogos de insulina. Terapias para T1D	8
1.1.4 Monitoreo continuo de glucosa y métricas asociadas	11
1.2 Modelos matemáticos para T1D	14
1.3 Algoritmos de predicción para T1D	18
1.4 Análisis Composicional de Datos	21

1.4.1	Principios y operaciones básicas	22
1.4.2	Transformaciones principales	24
1.4.3	Tratamiento de ceros	26
1.4.4	Aplicaciones con datos composicionales	30
1.5	Contexto de investigación y conjuntos de datos	31
1.6	Objetivos	33
1.7	Estructura de la tesis	34
2	Predicción probabilística de estados glucémicos para pacientes con diabetes tipo 1 mediante el análisis de Datos Composicionales	35
2.1	Simulador con escenarios realistas y desafiantes para VPs con T1D que se someten a terapia CSII y MDI	37
2.2	Validación del modelo probabilístico de predicción para pacientes con T1D utilizando análisis de datos composicionales	51
2.3	Predicción individualizada de estado glucémico mediante el análisis de datos composicionales	71
3	Discusión	85
3.1	Análisis de la incorporación de las insulinas de acción prolongada al modelo de Hovorka	86
3.2	Metodología y métricas para la validación del modelo probabilístico de transición	89
3.3	Propuesta de predicción de estados glucémicos en días con alta variabilidad .	91
4	Conclusiones y trabajo futuro	97
4.1	Contribuciones	97
4.2	Trabajos futuros	99
A	Librería de comidas mixtas	103

B	Parámetros y distribuciones para generar la cohorte virtual	115
C	Reporte de Perfil de glucosa ambulatorio (AGP)	121
	Referencias bibliográficas	125

LISTA DE TABLAS

TABLA	Página
1.1 Clasificación histórica de la DM.	3
1.2 Análogos de insulina.	10
1.3 Métricas CGM para atención clínica.	14
1.4 Características de los modelos matemáticos.	18
1.5 Estado del arte de los algoritmos de predicción.	19
1.6 Partición Binaria Secuencial.	26
3.1 Resultados de exactitud (métrica propuesta) después de una validación cruzada de 5 veces para el Paciente 2 (54 combinaciones de períodos).	90
3.2 Resultados de exactitud (métrica propuesta) después de una validación cruzada de 5 veces para el Paciente 8 (más de 200 combinaciones de períodos).	91
3.3 Ejemplo de predicción glucémica para las próximas 2h para el paciente 1, día 3 con 3 y 5 clases	95
3.4 Ejemplo de predicción glucémica para las próximas 2h para el paciente 1, día 49 con 3 y 5 clases	96
B.1 Distribución de probabilidad y límites de los parámetros univariados del modelo de Hovorka.	116
B.2 Parámetros de los pacientes 1 hasta el 24 (parte 1).	117
B.3 Parámetros de los pacientes 1 hasta el 24 (parte 2).	118

LISTA DE TABLAS

B.4	Parámetros de los pacientes 25 hasta el 47 (parte 1).	119
B.5	Parámetros de los pacientes 25 hasta el 47 (parte 2).	120

LISTA DE FIGURAS

FIGURA	Página
1.1 Número de personas con diabetes por cada región según su edad. Adaptado de la IDF (2022). África (AFR), Europa (EUR), Oriente Medio y África del Norte (MENA), América del Norte y el Caribe (NAC), América del Sur y Central (SACA), Región de Asia Sud-Oriental (SEA), Pacífico Occidental (WP).	4
3.1 BG del paciente 1, día 3 (azul) y día 49 (rojo).	92

ABSTRACT

Diabetes mellitus (DM) is a chronic disease that affects a large number of people worldwide, making it a global pandemic. Type 1 diabetes (T1D) accounts for approximately 10 % of all cases. It is characterized by autoimmune destruction of the pancreatic β cells, which are responsible for producing insulin. This loss of β cells leads to a permanent lack of insulin, resulting in an abnormal state of blood glucose (BG) homeostasis known as hyperglycemia. Subsequently, it can cause both chronic microvascular (retinopathy, neuropathy, and nephropathy) and macrovascular (cardiovascular and cerebrovascular diseases) complications, as well as other acute complications.

In the treatment of T1D, exogenous insulin is necessary to reduce BG levels to normoglycemia (70-180 mg/dL), which has been established as the control target. Normalization of BG is the main task of diabetes treatment, attempting to minimize hypoglycemia and hyperglycemia events. The main mechanisms of intervention in glycemic control in people with T1D are insulin administration and dietary adjustment. In both cases, the effectiveness of the treatment is influenced by the delay in insulin absorption and action and food intake. Therefore, the ability to predict the future glycemic profile is essential to helping the patient make decisions and avoid risky situations.

The availability of continuous glucose monitoring (CGM) systems has allowed for the systematic collection of glucose measurements at short intervals (5 or 15 min), generating a large amount of real-time data. This increase in data availability has led to the development of new mathematical prediction models that, along with improved measurement accuracy, allow for more reliable and long-term predictions despite the uncertainty and variability inherent in glucose measurements.

This thesis presents a simulation tool for T1D, where cohorts of virtual patients (VPs) are generated, incorporating models of long-acting insulins to evaluate multiple daily injection (MDI) and continuous subcutaneous insulin infusion (CSII) therapies in challenging and realistic scenarios. A mathematical methodology based on compositional data (CoDa) is proposed to validate a probabilistic model of transition between different categories of periods, providing a novel metric that could be used in any process that needs validation with compositional data. Finally, an individualized model is presented to predict the mean and coefficient of variation (CV) of glucose for the following 2 and 4 hours. From these predictions, the estimated minimum and maximum BG values are calculated. An information system called "traffic light" has been implemented and validated, which updates people on their glycemic status, risks related to hyperglycemia, hypoglycemia, and CV for the next hours.

ABSTRACT

The incorporation of long-acting insulin was evaluated in the MiceLab diabetes simulator, where simulation results were compared with those obtained in clinical trials. The validation and proposal of prediction algorithms were evaluated using sets of measurements from individuals with T1D who use CGM devices. The results are promising and suggest that these models could improve the accuracy of BG prediction, thus contributing to technological advancements and the optimization of therapies to improve the quality of life for individuals with T1D.

RESUMEN

La diabetes mellitus (DM) es una enfermedad crónica que afecta a un gran número de personas en todo el mundo, siendo una pandemia global. La diabetes tipo 1 (T1D) representa aproximadamente el 10% de todos los casos. Se caracteriza por una destrucción autoinmune de las células β del páncreas, que son las encargadas de producir la insulina. Esta pérdida de células β conduce a una falta permanente de insulina, lo que provoca un estado anormal de homeostasis de la glucosa en sangre (BG), conocido como hiperglucemia. A su vez, puede causar complicaciones crónicas tanto microvasculares (retinopatía, neuropatía y nefropatía), como macrovasculares (enfermedades cardiovasculares y cerebrovasculares) y otras complicaciones agudas.

En el tratamiento de la T1D, es necesario recurrir a la insulina exógena para reducir los niveles de BG hasta la normoglucemia (70-180 mg/dl), que se ha establecido como objetivo de control. La normalización de la glucemia es la tarea principal del tratamiento de la diabetes, intentando minimizar los eventos de hipoglucemia e hiperglucemia. Los principales mecanismos de intervención en el control glucémico en personas con T1D son la administración de insulina y el ajuste de la dieta. En ambos casos, la eficacia del tratamiento está influenciada por el retraso en la absorción y acción de la insulina y la alimentación. Por lo tanto, la habilidad para predecir el perfil glucémico en el futuro es esencial para ayudar al paciente a tomar decisiones y evitar situaciones de riesgo.

La disponibilidad de sistemas de monitorización continua de la glucosa (CGM) ha permitido la recopilación sistemática de mediciones de glucemia a intervalos cortos (5 ó 15 min), generando una gran cantidad de datos en tiempo real. Este aumento en la disponibilidad de datos ha dado lugar al desarrollo de nuevos modelos matemáticos de predicción, que, junto con la mejora en la precisión de las mediciones, permiten realizar predicciones más fiables y a largo plazo a pesar de la incertidumbre y variabilidad inherentes a la medición de la glucemia.

En esta tesis se presenta una herramienta de simulación para T1D, donde se generan cohortes de pacientes virtuales (VPs), se incorporan modelos de insulinas de acción prolongada para evaluar terapias de múltiples inyecciones diarias (MDI) y de infusión continua de insulina subcutánea (CSII) en escenarios desafiantes y realistas. Se propone una metodología matemática basada en datos composicionales (CoDa) para validar un modelo probabilístico de transición entre diferentes categorías de períodos, brindando una métrica novedosa que podría utilizarse en cualquier proceso que se desee validar cuyos datos sean composicionales. Finalmente, se presenta un modelo individualizado para predecir la media y el coeficiente de variación (CV) de la glucemia para las siguientes 2 y 4 horas. A partir de estas predicciones, se calculan los

valores mínimos y máximos estimados de BG. Se ha implementado y validado un sistema de información de "traffic light" que actualiza a las personas sobre su estado glucémico, los riesgos relacionados con hiperglucemia, hipoglucemia y CV para las próximas horas.

Se evaluó la incorporación de insulina de acción prolongada en el simulador de diabetes de MiceLab, donde se compararon los resultados en simulación con los obtenidos en ensayos clínicos. La validación y la propuesta de los algoritmos de predicción se evaluaron considerando conjuntos de mediciones de personas con T1D que utilizan dispositivos CGM. Los resultados son prometedores y sugieren que estos modelos podrían mejorar la precisión en la predicción de la BG, contribuyendo así a los avances de la tecnología y la optimización de las terapias para mejorar la calidad de vida de las personas con T1D.

La diabetis mellitus (DM) és una malaltia crònica que afecta un gran nombre de persones a tot el món, sent una pandèmia global. La diabetis tipus 1 (T1D) representa aproximadament el 10 % de tots els casos. Es caracteritza per una destrucció autoimmune de les cèl·lules β del pàncrees, que són les encarregades de produir l'insulina. Aquesta pèrdua de cèl·lules β condueix a una falta permanent d'insulina, fet que provoca un estat anormal d'homeòstasi de la glucosa la sang (BG), conegut com hiperglucèmia. Alhora, pot causar complicacions cròniques tant microvasculars (retinopatia, neuropatia i nefropatia), com macrovasculars (malalties cardiovasculars i cerebrovasculars) i altres complicacions agudes.

En el tractament de la T1D, cal recórrer a l'insulina exògena per reduir els nivells de BG fins a la normoglucèmia (70-180 mg/dl), que s'ha establert com a objectiu de control. La normalització de la glucèmia és la tasca principal del tractament de la diabetis, intentant minimitzar els esdeveniments d'hipoglucèmia i hiperglucèmia. Els mecanismes principals d'intervenció en el control glucèmic en persones amb T1D són l'administració d'insulina i l'ajust de la dieta. En tots dos casos, l'eficàcia del tractament està influenciada pel retard en l'absorció i l'acció de l'insulina i l'alimentació. Per tant, l'habilitat per predir el perfil glucèmic en el futur és essencial per ajudar el pacient a prendre decisions i evitar situacions de risc.

La disponibilitat de sistemes de monitoratge continu de la glucosa (CGM) ha permès la recopilació sistemàtica de mesuraments de glucèmia a intervals curts (5 o 15 min), generant una gran quantitat de dades en temps real. Aquest augment en la disponibilitat de dades ha donat lloc al desenvolupament de nous models matemàtics de predicció, que, juntament amb la millora en la precisió dels mesuraments, permeten realitzar prediccions més fiables i a llarg termini malgrat la incertesa i variabilitat inherents a la mesurament de la glucèmia.

En aquesta tesi es presenta una eina de simulació per a T1D, on es generen cohorts de pacients virtuals (VPs), s'hi incorporen models d'insulines d'acció perllongada per avaluar teràpies de múltiples injeccions diàries (MDI) i d'infusió contínua d'insulina subcutània (CSII) a escenaris desafiadors i realistes. Es proposa una metodologia matemàtica basada en dades composicionals (CoDa) per validar un model probabilístic de transició entre diferents categories de períodes, brindant una mètrica nova que podria utilitzar-se en qualsevol procés que es vulgui validar les dades del qual siguin composicionals. Finalment, es presenta un model individualitzat per predir la mitjana i el coeficient de variació (CV) de la glucèmia per a les 2 i 4 hores següents. A partir d'aquestes prediccions, calculeu els valors mínims i màxims estimats de BG. S'ha implementat i validat un sistema d'informació de "traffic light" que actualitza a les persones sobre el seu estat glucèmic, els riscos relacionats amb hiperglucèmia, hipoglucèmia i CV per a

les properes hores.

Es va avaluar la incorporació d'insulina d'acció perllongada al simulador de diabetis de MiceLab, on es van comparar els resultats en simulació amb els obtinguts en assaigs clínics. La validació i la proposta dels algorismes de predicció es van avaluar considerant conjunts de mesuraments de persones amb T1D que utilitzen dispositius CGM. Els resultats són prometedors i suggereixen que aquests models podrien millorar la precisió en la predicció de la BG, contribuint així als avenços de la tecnologia i l'optimització de les teràpies per millorar la qualitat de vida de les persones amb T1D.

INTRODUCCIÓN

Este capítulo presenta una breve introducción a la diabetes mellitus (DM), las complicaciones, desafíos asociados y las tecnologías actuales que se utilizan para superarlos (Sección 1.1 hasta la Sección 1.3). Luego, la Sección 1.4 resume los principios básicos de la teoría de los datos composicionales (CoDa) y la Sección 1.5 describe el contexto en el que se ha desarrollado este trabajo. Finalmente, se presentan los principales objetivos de esta investigación en la Sección 1.6, y la estructura de este documento en la Sección 1.7.

1.1 Diabetes mellitus

La DM es un trastorno metabólico caracterizado por hiperglucemia crónica debido a defectos en la secreción y/o acción de la insulina. Existen dos tipos principales de DM: la diabetes tipo 1 (T1D), también conocida como diabetes insulino dependiente, causada por una destrucción autoinmune de las células pancreáticas β , lo que resulta en una disminución o ausencia de secreción de insulina. Esto puede causar desviaciones en los niveles de glucosa en sangre (BG) fuera de los rangos normales, lo que aumenta el riesgo de hiperglucemia. La T1D tiene

una predisposición genética y se suele diagnosticar en edad temprana, pero también puede manifestarse en adultos, si la enfermedad autoinmune no ha sido lo suficientemente agresiva para destruir una gran cantidad de células β en la primera infancia (Holt et al., 2017). La diabetes tipo 2 (T2D), también conocida como diabetes no insulino dependiente, es una enfermedad metabólica adquirida, que se caracteriza por la disminución de la secreción de la insulina, un aumento de la resistencia a la insulina o ambos. Se conoce como resistencia a la insulina la disminución en la sensibilidad celular a esta, caracterizada por mayor necesidad para la captación de BG. Como consecuencia, el páncreas se ve obligado a aumentar su producción para compensar esta situación. Factores como el sobrepeso y la falta de actividad física contribuyen a este fenómeno y aumentan el riesgo de desarrollar DM a largo plazo. La resistencia a la insulina es una complicación temprana en la evolución de la enfermedad. La T2D es la forma más prevalente de diabetes en adultos y su aparición está asociada a factores de riesgo relacionados con el estilo de vida, la edad y antecedentes familiares (Holt et al., 2017). La tabla 1.1 muestra el desarrollo de la clasificación histórica de la DM.

Según la Federación Internacional de Diabetes (IDF), se estimó que la prevalencia mundial de diabetes en personas de 20 a 79 años en 2021 fue del 10,5% (536,6 millones de personas), previendo un aumento al 12,2% (783,2 millones) en 2045. Fue similar en hombres y mujeres, y más alta en personas de entre 75 - 79 años. En 2022, hubo 530 000 casos nuevos de T1D diagnosticados en todas las edades, 201 000 de ellos menores de 20 años. Los gastos sanitarios relacionados con la diabetes a nivel mundial se estimaron en 966 000 millones de USD en 2021 y se prevé que alcancen 1 054 000 millones de USD en 2045. La figura 1.1 muestra la cantidad de personas con T1D por grupos de edades por región del mundo. La región de Europa tiene el mayor número de casos, seguida de las regiones de América del Norte y el Caribe (Sun et al., 2022).

Tabla 1.1: Clasificación histórica de la DM.

Año	Clasificación y cambios
década de 1930	El primer Comité de Expertos en DM de la Organización Mundial de la Salud (OMS) consideró que la única clasificación fiable era por edad de inicio y la clasificó en: - enfermedad de inicio juvenil - enfermedad de inicio en la madurez* (Organization et al., 1980).
década de 1970	Con el descubrimiento de los genotipos del antígeno leucocitario humano (HLA) comunes en la diabetes juvenil y el descubrimiento de anticuerpos contra las células de los islotes. Se indicó que los pacientes más jóvenes con diabetes, los cuales requerían terapia con insulina, tenían un trastorno autoinmunitario (Organization et al., 1980).
1979-1980	El Grupo Nacional de Datos sobre la Diabetes propuso dos clases principales de diabetes: - insulino dependiente: diabetes mellitus (IDDM; tipo 1). - diabetes no insulino dependiente (NIDDM; tipo 2) También se definieron dos clases de riesgo: - anormalidad previa de intolerancia a la glucosa (PrevAGT) - anormalidad potencial de tolerancia a la glucosa (PotAGT) (Group, 1979)
1985	Se revisó la clasificación de 1980 y volvió a las descripciones clínicas conservando IDDM y NIDDM, pero omitiendo el tipo 1 y el tipo 2. Se introdujo la DM relacionada con la desnutrición en reconocimiento de un fenotipo diferente que se encuentra particularmente en Asia y África y la tolerancia alterada a la glucosa (IGT) como una clase de alto riesgo (Organization et al., 1985).
2006	La clasificación fue revisada por la OMS, pero no se introdujeron nuevas modificaciones, las principales clases fueron: - Diabetes mellitus tipo 1 (T1DM) - Diabetes mellitus tipo 2 (T2DM) - diabetes gestacional - otros tipos específicos asociados a defectos genéticos, endocrinopatías, entre otros más desarrollado en (Organization et al., 1999a,b).
2010	La IGT se eliminó de la clasificación formal de tipos de diabetes, ya que no se consideró diabetes, pero se mantuvo como un estado de riesgo y la alteración de la glucemia en ayunas (IFG) se introdujo como otro estado de riesgo.

* En ese momento, había muchos otros tipos de fenotipos comunes, tales como: la diabetes frágil, gestacional, pancreática, endocrina, resistente a la insulina y iatrogénica, sin embargo, para la mayoría de los casos, no se conocía claramente su causa.

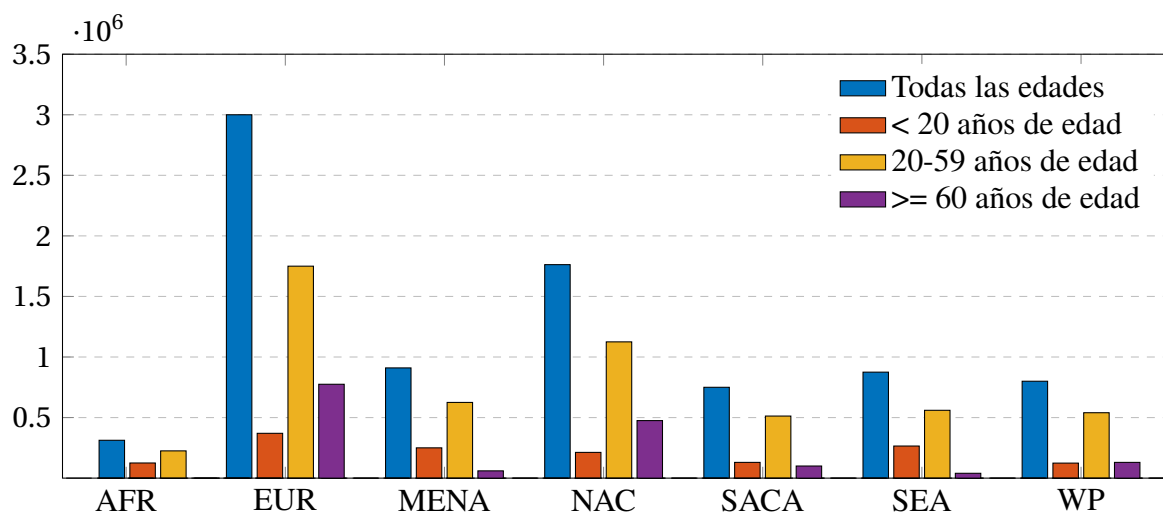


Figura 1.1: Número de personas con diabetes por cada región según su edad. Adaptado de la IDF (2022). África (AFR), Europa (EUR), Oriente Medio y África del Norte (MENA), América del Norte y el Caribe (NAC), América del Sur y Central (SACA), Región de Asia Sud-Oriental (SEA), Pacífico Occidental (WP).

1.1.1 Complicaciones

La ausencia o resistencia a la insulina en personas con DM sin un manejo adecuado, control y seguimiento conducen a hiperglucemia, que puede causar daños a órganos del cuerpo y generar complicaciones. Algunas de las complicaciones a corto plazo para pacientes mal controlados suele ser la hipoglucemia ($BG < 70$ mg/dl), esta puede provocar debilidad o temblores, sudoración, mareos, hambre, nerviosismo y confusión. Si no se trata, puede complicarse y provocar dolor de cabeza, irritabilidad, mala circulación, mala coordinación, y adormecimiento en la boca y la lengua. Una hipoglucemia grave puede producir alteraciones de la conciencia, convulsiones, e incluso puede desembocar en un coma diabético o la muerte, en casos extremos. Tanto así, la hiperglucemia ($BG > 180$ mg/dl), puede causar sed excesiva y hambre, micción frecuente, náuseas, vómitos y falta de energía, con permanencia prolongada puede llevar a una condición conocida como cetoacidosis (Holt et al., 2017). Esta aumenta las probabilidades de los problemas cardiovasculares conduciendo al riesgo de hipertensión arterial, enfermedad arterial periférica y accidente cerebrovascular (Gao and Chilibeck, 2020). También puede dañar

los nervios del cuerpo, desarrollándose la neuropatía, lo que puede causar entumecimiento, hormigueo y dolor en las manos y pies. Además, puede afectar el sistema nervioso autónomo, lo que puede provocar problemas de digestión, micción y disfunción eréctil (Prabhakar et al., 2021).

Las complicaciones a largo plazo incluyen enfermedades microvasculares, que ocasionan trastornos oculares, renales o del sistema nervioso. Las macrovasculares pueden ocasionar afecciones cardíacas, accidentes cerebrovasculares o patologías vasculares periféricas (Collaboration et al., 2010; Li et al., 2019; Bjerg et al., 2019; Palladino et al., 2020; Steinmetz et al., 2021). Se estima que después de aproximadamente 10 a 15 años de duración de la enfermedad, las complicaciones microvasculares y macrovasculares de la diabetes crónica comienzan a tener un impacto, y después de aproximadamente 30 años, las causas cardiovasculares se vuelven cada vez más importantes. La mortalidad relativa por causas cardiovasculares es al menos tan alta para las mujeres como para los hombres (García Jaramillo et al., 2011; Leal Moncada et al., 2013; Holt et al., 2017).

La hiperglucemia crónica es el principal factor de riesgo para el desarrollo de complicaciones en la DM, sin embargo, las fluctuaciones de BG grandes o frecuentes pueden contribuir de forma independiente a estas complicaciones (Kovatchev, 2019; Yoo and Kim, 2020). La variabilidad de la glucemia (GV) se refiere a esta fluctuación de los niveles de la BG y reducirla se ha convertido en una de las prioridades en la investigación de la T1D. Se reconocen dos tipos de GV: la variabilidad intra-paciente, que se refiere a las fluctuaciones en los niveles de glucemia que se producen a lo largo del día en una misma persona, y la variabilidad inter-paciente que indica la falta de reproducibilidad de las curvas entre diferentes personas. Estas variabilidades surgen, entre otras causas, de los diferentes procesos fisiológicos involucrados, como la absorción de insulina subcutánea, los ritmos circadianos de sensibilidad a la insulina, la acción de las hormonas contra-reguladoras, etc (Moscoso-Vasquez et al., 2016).

Existe una considerable cantidad de evidencia clínica que respalda la relación entre la GV y

el desarrollo de complicaciones diabéticas (Sartore et al., 2012; Kovatchev, 2019). En particular, se ha demostrado que el aumento de la GV contribuye de manera independiente al desarrollo de complicaciones microvasculares. En estudios de retinopatía diabética, se ha observado que el aumento de la hemoglobina glicosilada (HbA1c) explica solo un 11% de la variación en el riesgo de desarrollar esta complicación, lo que sugiere que otros factores independientes de la HbA1c deben explicar el 89% restante (Lachin et al., 2008).

Estudios sugieren que una GV significativa, especialmente cuando se acompaña de episodios de hipoglucemia, puede tener un impacto negativo no solo en el desarrollo y progresión de las complicaciones de la diabetes, sino también en otras condiciones médicas que requieren atención en unidades de cuidados intensivos (UCI) (DeVries, 2013). Estudios recientes que han analizado las mediciones de monitoreo continuo de la glucosa (CGM) han mostrado una fluctuación significativa en los valores de BG en niños con T1D, incluso en aquellos con excelentes valores de HbA1c. Este hallazgo sugiere que, además de la HbA1c, la GV puede tener un valor predictivo para el desarrollo de complicaciones (Ayano-Takahara et al., 2015).

1.1.2 Hormonas pancreáticas. Insulina y glucagón

La homeostasis de la glucosa es esencial para el funcionamiento normal del cuerpo humano, hormonas como la insulina y el glucagón juegan un papel importante en este asunto. El proceso de reposición de energía es llevado a cabo mediante reacciones químicas intracelulares que utilizan azúcares y ácidos grasos como combustible, oxígeno y enzimas como catalizadores. La glucosa actúa como principal fuente de energía para las células humanas, especialmente para el cerebro y el sistema nervioso central. El transporte de la glucosa a través del cuerpo es realizado por la circulación sanguínea después de su absorción en el intestino durante la digestión. La concentración de BG se encuentra bajo una estricta regulación fisiológica que depende de la tasa de entrada y salida de glucosa en el torrente sanguíneo. Múltiples fuentes contribuyen a esta regulación, incluyendo la digestión de alimentos, la descomposición del

glucógeno y la glucogénesis. La glucosa proveniente de la ingesta de alimentos se metaboliza y se libera en la sangre, mientras que el glucógeno almacenado en el hígado y los músculos puede ser degradado para liberar glucosa adicional. Además, la glucogénesis permite la producción de glucosa a partir de sustratos no carbohidratados. Estos mecanismos garantizan un equilibrio preciso en los niveles de BG.

Insulina y glucagón

La insulina es una hormona reguladora producida por las células β de los Islotes de Langerhans en el páncreas. Su principal función es regular los niveles de BG, facilitando la entrada de la glucosa en las células del cuerpo para su utilización metabólica. La secreción de insulina se activa en respuesta a los niveles elevados de BG y promueve la captación y almacenamiento de glucosa en los tejidos, como el músculo y el tejido adiposo (Holt et al., 2017). Se considera la principal hormona anabólica del organismo, con una finalidad reguladora en el proceso del metabolismo de los macronutrientes (carbohidratos (CHO), proteínas y ácidos grasos), al promover la absorción de glucosa desde la sangre al hígado, grasa y tejido muscular. Es una proteína formada por dos cadenas polipeptídicas, con un total de 51 aminoácidos. La insulina se secreta en respuesta a niveles altos de BG y de la ingesta de aminoácidos, los cuales tienen lugar después de las comidas. En las personas con T1D y algunas con T2D, el páncreas no es capaz de secretar cantidades suficientes de insulina, lo que conlleva a que requieran de infusión de insulina exógena para sobrevivir (Atkinson et al., 2011; Holt et al., 2017).

El glucagón es otra hormona que consta de 29 aminoácidos y es producida por las células pancreáticas α . Ayuda a aumentar los niveles de azúcar en la sangre y así regular correctamente el metabolismo de la glucosa. Mientras la insulina modera los niveles de glucosa durante las comidas, el glucagón hace lo mismo durante los períodos de ayuno (promueve la producción de glucosa en el hígado durante el ayuno).

1.1.3 Análogos de insulina. Terapias para T1D

Los análogos de insulina son moléculas diseñadas sintéticamente que imitan la estructura y función de la insulina endógena producida por el cuerpo humano. Estos análogos han sido desarrollados con el objetivo de mejorar la eficacia y seguridad de la terapia de insulina utilizada en el tratamiento de la DM. La modificación de la estructura de la insulina natural ha permitido desarrollar análogos con propiedades farmacocinéticas y farmacodinámicas mejoradas. Por ejemplo, algunos análogos de insulina tienen una absorción más rápida después de la inyección, lo que facilita un inicio de acción más temprano y una duración de acción más corta. Otros análogos presentan una absorción más lenta y una duración de acción más prolongada, lo que permite una mejor cobertura de la glucemia a lo largo del día.

Las insulinas ultra-rápidas son una versión más reciente con un inicio de acción más rápido que las insulinas análogas de acción rápida, lo que permite su dosificación al comienzo o durante una comida para controlar mejor los picos de glucosa postprandial. Estudios clínicos han demostrado que éstas tienen una eficacia similar a otras opciones de insulina bolus disponibles, especialmente para reducir la HbA1c. Sin embargo, su uso ha resultado en reducciones significativas en los niveles de glucemia postprandial de 1 y 2 horas (Avgerinos et al., 2021; Wong and Kroon, 2021). A los pacientes tratados con infusión continua de insulina subcutánea (CSII), se han asociado con un mayor tiempo en normoglucesmia, menor tiempo en hipoglucesmia a lo largo de las 24 horas y un mejor control de la glucemia postprandial. Por otro lado, se ha reportado que ha aumentado la probabilidad de cambios no planificados en el set de infusión, sin mejoras considerables en el control (Stamati et al., 2022).

Los análogos de insulina de acción rápida son formulaciones que tienen una acción breve en el cuerpo, lo que permite una rápida liberación de insulina en el torrente sanguíneo. Estos análogos se utilizan para controlar los niveles de azúcar en sangre después de las comidas y se administran antes o durante éstas (Misra and Mathieu, 2020).

Afrezza es una insulina que se suministra a través de inhalación en lugar de inyección

subcutánea. Es una insulina regular humana recombinante que se ha formulado como un polvo seco para ser inhalado en los pulmones, tiene un rápido inicio de acción y un tiempo rápido para alcanzar su acción máxima, pero una duración de acción más corta en comparación con la insulina regular humana o los análogos de insulina de acción rápida, lo que demuestra cómo un cambio en la vía de administración de la insulina regular humana puede afectar su perfil farmacocinético y farmacodinámico (Hirsch et al., 2020).

En 1950, se lanzó la insulina de protamina neutra (NPH, también llamada insulina isofánica), su duración es más larga, aunque no lo suficiente para imitar la liberación basal fisiológica diaria en personas con deficiencia severa de insulina; por lo tanto, generalmente esta debe administrarse dos veces al día para lograr una cobertura basal adecuada. Además, su perfil de tiempo-acción puede diferir entre individuos según factores fisiológicos. Debido a que es un precipitado con protamina y zinc, debe ser resuspendida mediante un suave movimiento de balanceo de 12 a 15 veces antes de la inyección, si no se sigue este procedimiento, puede agregar significativamente a la variabilidad día a día (Lucidi et al., 2015). Dadas las limitaciones de la insulina NPH en satisfacer las necesidades de insulina basal, la biotecnología fue nuevamente empleada para desarrollar el primer análogo de insulina de acción prolongada (Glargine). Este tipo de análogos son formulaciones que tienen una acción lenta en el cuerpo, lo que permite una liberación sostenida de insulina en el torrente sanguíneo. Estos análogos son utilizados para controlar los niveles de BG durante todo el día y se administran, normalmente, una vez al día, en la mañana o en la noche (Sharma et al., 2019; Misra and Mathieu, 2020; Alsofiani et al., 2022). La tabla 1.2 muestra información actualizada de los análogos de insulina.

La administración subcutánea de estos análogos de acción prolongada se emplean para la T1D, en la terapia de múltiples inyecciones diarias (MDI) para cubrir las necesidades de insulina basal del paciente. Varias investigaciones se han centrado en desarrollar modelos matemáticos para evaluar y optimizar terapias con estas insulinas, probar nuevas estrategias de tratamiento a lazo abierto (OL) y lazo cerrado (CL), tanto en plataformas de simulación como

Tabla 1.2: Análogos de insulina.

Tiempo de acción	Tipos	Marcas	Inicio	Pico	Duración (h)
Ultra-rápida	Aspart + rápida ultra-rápida lispro (URLi)	Fiasp	~16–20 min	~1.5–2.2	5–7
		Lyumjev	15–17 min	57 min	4.6–7.3
Acción rápida	Lispro	Humalog Lyumjev	10-15 min	1-1.5 min	3-5
	Aspart	NovoRapid Fiasp	10-15 min	1-2 h	3-5
	Glulisina	Apidra Novolín	10-15 min	1-2 h	3-5
	Regular	Humalin-R Velosulin	~ 30 min	1.5-3.5 h	7-8
	lispro, biosimilar	Admelog	15 min	30–90 min	3–5
	humana inhalada	Afrezza	~12 min	35–55 min	1.5–4.5
Acción intermedia	NPH	Humalin-N Actrapid	1.5-4 h	2.8-13 h	12-18
Acción prolongada	Detemir	Lemevir	2-3 h	6 - 8 h	~ 24
	Glargina 100	Abasaglar Lantus	2-3 h	No pico	~ 24
	Glargina 300	Toujeo	6 h	No pico	24-36
Ultra-lentas	Degludec	Tresiba	2-4 h	No pico	~ 42

en ensayos clínicos (Rosenstock et al., 2000; Raskin et al., 2000; Schiavon et al., 2017, 2019; Visentin et al., 2019; Pala et al., 2019; Lee et al., 2020).

Terapias para la T1D

El avance en los análogos de insulina se ha convertido en un componente crucial del tratamiento de esta enfermedad. Entre las terapias disponibles se encuentra la MDI, que se compone de la administración de insulina basal (de acción intermedia o prolongada), como la insulina isofánica NPH, lispro protamina, glargina, detemir, degludec, entre otras, una o dos veces al día y la administración subcutánea de insulina en bolos prandiales, como la insulina regular, lispro, aspártica, glulisima, entre otras, antes o durante las comidas. El objetivo de este

régimen es imitar la secreción fisiológica de insulina por las células β del páncreas (Brink, 2022).

La terapia de CSII consiste en la administración de pequeñas dosis de insulina basal de forma continua. Es un tratamiento que incluye una bomba de insulina, un reservorio de insulina, un catéter y una cánula. La bomba se alimenta por baterías y suministra insulina basal a un ritmo fijado. La insulina fluye desde el reservorio de la bomba a través de un catéter flexible y se infunde en el tejido subcutáneo mediante una cánula insertada en la zona abdominal. Este tratamiento permite una infusión continua de insulina en el tejido subcutáneo, lo que mejora el control metabólico en pacientes con T1D. La CSII permite una mayor flexibilidad en el tratamiento, ya que los pacientes pueden ajustar las dosis de insulina según sus necesidades individuales, como las variaciones en la actividad física y en la ingesta de alimentos (Pala et al., 2019; Ross and Neville, 2019).

Estudios recientes corroboran que ambos tratamientos tienen ventajas y desventajas. La terapia CSII se asocia con una menor GV de HbA1c en comparación con la MDI (Scott et al., 2019). En (William et al., 2022) se concluyó que ambas terapias tuvieron un impacto similar en la reducción de la HbA1c, el peso, el uso de insulina y el tiempo en rango (TIR). Sin embargo, en otro estudio similar el grupo de CSII tuvo una glucemia promedio más baja pero aumentó la masa corporal significativamente (Babiker et al., 2022). En (Dowling and Maidment, 2023), se evaluaron las diferencias en los síntomas depresivos y el estrés relacionado con la DM entre adultos con T1D que usaban CSII o MDI, concluyendo que el estrés fue significativamente mayor en los usuarios tratados con MDI.

1.1.4 Monitoreo continuo de glucosa y métricas asociadas

Los sistemas de CGM consisten en un sensor subcutáneo que mide continuamente los niveles de glucosa en el líquido intersticial y transmiten los datos a un dispositivo de lectura en tiempo real. Proporcionan información sobre los cambios en los niveles de BG, lo que permite a

los pacientes y médicos tomar decisiones sobre el manejo de la enfermedad. Además, los dispositivos de CGM eliminan la necesidad de pruebas frecuentes de glucemia, que pueden ser inconvenientes para los pacientes, lo que los convierte en una alternativa más eficaz y en una herramienta esencial en el manejo de la T1D. El CGM permite detectar tendencias a largo plazo en los niveles de BG, lo que permite ajustar los tratamientos de manera más precisa. Varios estudios clínicos han demostrado que su uso frecuente se asocia con una reducción en los episodios de hipoglucemia y una mejora en el control glucémico en pacientes con T1D y T2D, así como una disminución en los niveles de HbA1c (Simon et al., 2008; Seyed Ahmadi et al., 2020; Freckmann, 2020). El CGM se ha relacionado con la mejoría en el control de la glucemia y la prevención de episodios de hipoglucemia e hiperglucemia en pacientes con T1D (Benjamin, 2002).

La evolución de las tecnologías asociadas a los sistemas de CGM ha experimentado una mejora progresiva en la precisión de adquisición de mediciones desde los sensores, lo que ha permitido reducir las discrepancias entre las mediciones de glucosa obtenidas mediante CGM y las mediciones de BG capilar, que se considera el estándar de referencia. Se han realizado mejoras en los algoritmos de calibración, la estabilidad de los sensores y la capacidad de detectar y corregir posibles interferencias, como la variabilidad de la perfusión sanguínea o la interferencia de medicamentos (Didyuk et al., 2021).

Para la adquisición de las medidas, los transmisores se conectan al sensor y transmiten los datos a un dispositivo, como un reloj inteligente, un dispositivo de lectura específico o aplicaciones móviles que muestran y almacenan los datos de los sensores. Algunos sistemas están diseñados para integrarse con bombas de insulina, permitiendo a los pacientes ajustar automáticamente su dosis de insulina en función de los niveles de BG. Otros sistemas utilizan algoritmos para analizar los datos de los sensores y proporcionar información y recomendaciones para el manejo de la diabetes, o bien se integran con sistemas de salud electrónicos, lo que permite a los pacientes y médicos acceder a los datos de BG a través de una plataforma en

línea (Rodríguez-Rodríguez et al., 2019; Tyler and Jacobs, 2020; Vettoretti et al., 2020; Tena et al., 2021; Makroum et al., 2022).

Con la adquisición de estas mediciones y la información en tiempo real sobre los niveles de BG, se ha facilitado a los pacientes y a sus médicos tomar decisiones sobre cómo manejar la enfermedad, ajustando el tratamiento en función de los niveles de BG. Estos datos son utilizados para calcular y ajustar los bolos óptimos de insulina. También han sido el pilar en la implementación de sistemas de predicción de la glucemia, mejorando así el control metabólico de los pacientes, evidenciando un incremento en el TIR (Cobry et al., 2010; Walsh et al., 2011; Herrero et al., 2015, 2017; Vallejo-Mora et al., 2017; Noaro et al., 2020; Adolfsson et al., 2020)

Métricas CGM estandarizadas

El control glucémico se evalúa mediante la medición de HbA1c (refleja la glucemia promedio durante aproximadamente 3 meses), el CGM utilizando el TIR y/o el indicador de control de glucosa (GMI) y el monitoreo de glucosa en sangre (BGM). HbA1c es una de las métricas utilizadas para guiar la gestión de la glucemia y un marcador valioso para el desarrollo de complicaciones (Assessment, 2022). Sin embargo, entre sus limitaciones se encuentra que no refleja la variabilidad de la BG. En un estudio realizado con 130 sujetos con T1D, aquellos con un coeficiente de variación (CV) alto (>36%), estuvieron más tiempo en hipoglucemia que aquellos con un CV bajo, a pesar del mismo HbA1c (Toschi et al., 2020). Además, se ha informado que carece de información detallada sobre el control glucémico a corto plazo, y puede verse fácilmente interferida por diversas afecciones clínicas, como anemia, embarazo o enfermedad hepática. Por lo tanto, la HbA1c por sí sola no puede representar el estado glucémico real de un paciente (Yoo and Kim, 2020).

La Asociación Americana de Diabetes (ADA) resume las recomendaciones actuales de práctica clínica de la T1D. Están destinadas a proporcionar los componentes de atención, metas y pautas generales de tratamiento y herramientas para evaluar la calidad de la atención. Su comité de expertos multidisciplinario es responsables de actualizar los estándares de atención

Tabla 1.3: Métricas CGM para atención clínica.

Métrica	Descripción
TAR (tiempo por encima del rango) hiperglucemia nivel 2	$BG > 250$ mg/dl ($>13,9$ mmol/l)
TAR hiperglucemia nivel 1	$181 \leq BG \leq 250$ mg/dl ($10,1-13,9$ mmol/l)
TIR (tiempo en rango objetivo)	$70 \leq BG \leq 180$ mg/dl ($3,9-10,0$ mmol/l)
TBR (tiempo por debajo del rango) hipoglucemia nivel 1	$54 \leq BG \leq 69$ mg/dl ($3,0-3,8$ mmol/l)
TBR hipoglucemia nivel 2	$BG < 54$ mg/dl ($\leq 3,0$ mmol/l)
variabilidad glucémica (% CV) $<36\%$	algunos estudios sugieren que los objetivos de % CV más bajos ($<33\%$) brindan protección adicional contra la hipoglucemia (Care, 2022)

anualmente, o con mayor frecuencia si se justifica (Assessment, 2022). El consenso internacional sobre TIR proporciona orientación sobre las métricas estandarizadas de CGM (tabla 1.3, distinguido en colores las tres principales). El tiempo por debajo del rango objetivo (TBR), se asocia con el riesgo de situaciones microvasculares, y diversas evidencias lo respaldan como un marcador predictivo de complicaciones, otra métrica es el tiempo por encima del objetivo (TAR), ambos son parámetros útiles para la evaluación del régimen y eficacia del tratamiento. Convirtiéndose en métricas importante en el control glucémico, correlacionándose con la HbA1c en la mayoría de los estudios (Vigersky and McMahon, 2019; Yoo and Kim, 2020; Advani, 2020; Wright Jr et al., 2020; Care, 2022).

1.2 Modelos matemáticos para T1D

Los modelos matemáticos desempeñan un papel importante en la investigación y tratamiento de la diabetes al proporcionar herramientas para comprender la fisiología de la enfermedad, guiar la terapia de insulina y prevenir complicaciones a largo plazo. Los enfoques de modelado adoptados ofrecen modelos de diferentes niveles de complejidad. Los científicos y matemáticos han trabajado en el desarrollo de modelos y simulaciones computacionales para representar y analizar la respuesta del sistema glucorregulatorio a diversos estímulos y condiciones fisiológi-

cas, buscando describir la dinámica compleja de las células β pancreáticas y la eficiencia del lazo de retroalimentación glucosa-insulina (Ackerman et al., 1964; Bergman and Cobelli, 1980; Sorensen, 1985; Hovorka et al., 2004; Dalla Man et al., 2007; Wilinska and Hovorka, 2008).

La historia de los modelos matemáticos para la diabetes se remonta a la década de 1960, cuando Ackerman y su equipo introdujeron un modelo pionero en la modelización de los procesos fisiológicos de la glucosa. Se trata de un modelo lineal, que surgió como método de detección de la diabetes, basándose en la tolerancia oral a la glucosa (Ackerman et al., 1964). En ese año, el investigador sueco Claes Hellerström, comenzó a estudiar la fisiología de las células β en el páncreas. En 1967, junto a su colega Bengt Persson publicaron un modelo matemático para describir la secreción de insulina por las células β del páncreas (Matschinsky, 1996). En la década de 1980, el investigador británico John Pickup desarrolló un modelo matemático para describir la respuesta glucémica a la insulina en pacientes con T1D, conocido como el "modelo de control glucémico", se utiliza para predecir los niveles de BG en respuesta a diferentes dosis de insulina y CHO consumidos (Pickup et al., 2011). Estos fueron los primeros pasos en el modelado matemático, a continuación se describen los modelos matemáticos más conocidos del sistema glucosa-insulina ordenados por complejidad. La tabla 1.4 muestra una descripción en cuanto a ecuaciones diferenciales ordinarias (ODE) y sus limitaciones, todos estos modelos son no lineales.

- **Modelo mínimo de Bergman (1979) (Complejidad mínima):** Es un modelo unicompartmental, este se originó para analizar la respuesta corporal a la infusión intravenosa de glucosa, la cual se emplea en el test de tolerancia a la glucosa intravenosa (IVGTT), un procedimiento clínico que mide las concentraciones de glucosa e insulina en sangre antes y después de la infusión, con el objetivo de evaluar el proceso de metabolización en el cuerpo. A partir de estas mediciones, se calcula un índice de acción de la insulina (llamada sensibilidad a la insulina). A pesar de los años y debido a su simplicidad sigue utilizándose en la actualidad, tanto como herramienta clínica para comprender los efectos

compuestos de la secreción de insulina, la sensibilidad a la insulina en la tolerancia a la glucosa y el riesgo de T2D (Bergman and Urquhart, 1971; Bergman et al., 1979; Bergman and Cobelli, 1980; Bondia et al., 2018; Bergman, 2021).

- **Sorensen (1985) (Complejidad alta):** Fue formulado para simular el comportamiento tanto de individuos normales como en personas con diabetes. Cada subsistema divide el cuerpo en múltiples compartimentos, que representan el espacio sanguíneo de uno o varios tejidos biológicos. Específicamente, el cuerpo se divide en seis compartimentos fisiológicos distintos: el cerebro, que representa al sistema nervioso central, el corazón y los pulmones, que representan los volúmenes de mezcla rápida del corazón, los pulmones y las arterias, la periferia, que incluye el músculo esquelético y el tejido adiposo, el intestino, el hígado y los riñones. Se consideran por separado los subsistemas de glucosa e insulina (Sorensen, 1985; Hipszer, 2008; Hernández-Ordoñez and Campos-Delgado, 2008; Bondia et al., 2018).
- **Hovorka (2002) (Complejidad media):** Es en un modelo compartimental que relaciona las concentraciones plasmáticas de insulina como entrada y glucosa como salida. En 2004, Hovorka y colaboradores presentaron un modelo de simulación no lineal, que incorpora submodelos compartimentales de la cinética de insulina subcutánea y glucosa, así como un submodelo de dos compartimentos para la absorción de glucosa en el tracto gastrointestinal. La digestión y absorción de CHO se describe mediante la respuesta al impulso de un modelo lineal de segundo orden, como entrada la cantidad de CHO de la comida y salida la tasa de aparición de glucosa en la sangre. La farmacocinética de la insulina subcutánea se representa mediante un modelo lineal de tercer orden. El modelo describe la masa de glucosa en compartimentos accesibles y no accesibles, y los subsistemas de acción de la glucosa e insulina se interconectan. Aunque la capacidad del modelo para representar la variabilidad intra e inter-paciente es significativa, esta última no es lo suficientemente representativa (Hovorka et al., 2002, 2004; Hovorka, 2006, 2011;

Bondia et al., 2018; Som et al., 2019; Mughal et al., 2023). Este modelo es el núcleo del simulador de Cambridge, con 18 pacientes virtuales (VPs), es un entorno para evaluar el comportamiento de sistemas de administración de insulina para la T1D. Agrega modelos generales de medición de glucosa subcutánea y de bombas de insulina, en este simulador los VPs presentan variabilidad intra e inter-paciente ya que cada sujeto cuenta con un conjunto único de parámetros y algunos de estos son variantes en el tiempo (Plank et al., 2006; Wilinska and Hovorka, 2008; Wilinska et al., 2010). También se han creado otros simuladores basados en este modelo (Resalat et al., 2019).

- **Dalla Man (2007) (Complejidad media/alta):** Es un modelo compartimental donde se produce la utilización independiente de la insulina, que representa el plasma y los tejidos que se equilibran rápidamente y la utilización dependiente de la insulina que representa los tejidos que se equilibran lentamente. El modelo relaciona las concentraciones plasmáticas (glucosa e insulina), con los flujos de glucosa (producción endógena de glucosa, tasa de aparición de glucosa, utilización de glucosa, extracción renal) y flujos de insulina (tasa de aparición de insulina del tejido subcutáneo y degradación de insulina). El modelo descompone el sistema en subsistemas que permiten representar de manera casi independiente la tasa de aparición de la ingesta, la absorción CHO, la producción endógena de glucosa, la utilización de glucosa o la secreción de insulina. Además, este modelo tiene en cuenta hormonas contrarregulatorias como el glucagón. Aunque los subsistemas de glucosa e insulina están claramente separados, están relacionados con la glucosa e insulina plasmáticas independientes de los tejidos y la producción endógena de glucosa (Dalla Man et al., 2007; Bondia et al., 2018). Es el centro del simulador UVA-Padova, es una herramienta para la evaluación de controladores de glucosa, aceptada por la Administración de Alimentos y Medicamentos de los Estados Unidos (FDA) en el año 2010 (Kovatchev et al., 2009; Man et al., 2014) como sustituto de pruebas en animales, ha tenido varias actualizaciones: en 2014 se actualizó para tener en cuenta las

no linealidades de la acción de la insulina y la cinética del glucagón en la dinámica de la glucosa en rango hipoglucémico (Man et al., 2014), y en 2018 con el objetivo de ofrecer un escenario de prueba más realista, entre otras actualizaciones (Visentin et al., 2018). En (Schiavon et al., 2019, 2020) incorporaron modelos de insulina de acción prolongada para los análogos de glargina 100 U/ml (Gla-100), glargina 300 U/ml (Gla-300) y degludec 100 U/ml (Deg-100).

Tabla 1.4: Características de los modelos matemáticos.

Modelo	ODE y parámetros (P)	Limitaciones
Bergman	3 ODE de segundo orden que representan la BG, la concentración de insulina y 10 P	No considera el efecto contraregulatorio de hormonas como el glucagón, ni la ausencia de una condición de saturación en los compartimentos de glucosa e insulina, tampoco considera la absorción subcutánea de insulina o la ingesta de comidas ya que se concibió como un modelo para la medición
Sorensen	22 ODE de primer orden que describen la interacción de la glucosa, la insulina y el glucagón y su efecto sobre la absorción y liberación de glucosa	No puede representar la variabilidad intra-paciente (ya que todos los parámetros de este modelo son invariantes en el tiempo). Sólo contempla insulina intravenosa, no teniendo el retardo de la inyección de insulina de forma subcutánea. Presenta problemas de identificación
Hovorka	10 ODE y 15 P	Simple representación de la absorción de glucosa en el intestino y no considera la dinámica del glucagón. Presenta respuestas no fisiológicas ante la administración de dosis razonables, pero elevadas, de insulina basal, esto puede resultar en valores negativos de glucosa
Dalla Man	13 ODE y 26 P	No modelación de las variaciones diurnas de ciertos parámetros

1.3 Algoritmos de predicción para T1D

Desde la introducción de los CGM, la literatura científica ha presentado diversos enfoques para predecir los niveles de BG y eventos futuros asociados (García Jaramillo et al., 2011; De Bois et al., 2019; Woldaregay et al., 2019; Mohebbi et al., 2020; Marling and Bunescu,

2020; Martinsson et al., 2020; Tena et al., 2021; Daniels et al., 2021; Khanam and Foo, 2021; Cichosz et al., 2021; Wadghiri et al., 2022). Entre ellos, los modelos basados en datos han demostrado un buen equilibrio entre aplicabilidad y rendimiento para predecir tanto la glucemia en general como la hipoglucemia en particular. En estudios recientes se encontró que entre el 2014 y 2020, un 82.5% de los estudios se enfocaron en la predicción a corto plazo (ST). Mientras que los modelos basados en datos y los modelos híbridos representan el 80.9% y 19.1%, respectivamente. Se utilizan diversas técnicas predictivas, entre las que destacan las redes neuronales artificiales (ANN) (22.2%), aprendizaje en conjunto (27.0%), aprendizaje supervisado (20.6%), estadístico/ probabilístico (7.9%), autorregresivo (7.9%), evolutivo (6.4%), aprendizaje profundo (4.8%) y filtros adaptativos (3.2%) (Felizardo et al., 2021; Wu et al., 2022). También se han publicado estudios con regresión lineal y logística (Zhu et al., 2019). Aunque no hay una categorización estándar para el horizonte de predicción (PH), este estudio define el corto plazo cuando el PH es menor de 180 minutos, mientras que otros estudios consideran una predicción superior a 60 minutos como de largo plazo (Li, 2019; Saiti et al., 2020). La tabla 1.5 muestra el estado del arte de algunos algoritmos de predicción desde el año 2018, incluyendo solamente estudios con datos de pacientes reales con T1D que informaron los resultados de error cuadrático medio (RMSE) y el PH correspondiente. También es importante destacar los sistemas de apoyo en la toma de decisiones, que pueden incluir herramientas de registro de datos, protocolos de tratamiento, así como mejoras a su adherencia (Sim et al., 2017; Contreras and Vehi, 2018; Tyler and Jacobs, 2020).

Tabla 1.5: Estado del arte de los algoritmos de predicción.

Referencia	Algoritmo	RMSE (mg/dl) (PH)
Ali et al. (2018)	ANN	9.03 (60min)

Continúa en la siguiente página.

Referencia	Algoritmo	RMSE (mg/dl) (PH)
Hamdi et al. (2018)	Regresión de vectores de soporte (SVR) y evolución diferencial	12.95 (60 min)
Zhu et al. (2018)	Red Neuronal Convolutacional (CNN)	21.72 (30 min)
Chen et al. (2018)	Red neuronal recurrente dilatada (DRNN)	19.04 (30 min)
Li et al. (2018)	Redes neuronales de estado recurrente (ESN)	59.25 (60 min)
Griva and Basualdo (2018)	Autoregresivo con entradas exógenas (ARX)	49.1 (120 min)
Yu et al. (2018)	Fusión de filtros adaptativos	8.93 (30 min)
Midroni et al. (2018)	Impulso extremo de aumento de gradiente (XG-Boost)	16.23 (30 min)
Xie and Wang (2018)	ARX	19.59 (30 min)
Gadaleta et al. (2018)	SVR	23 (30 min)
Li and Liu (2019)	GluNet, múltiples capas de red neuronal de convolución dilatada	31.83 (60 min)
Aliberti et al. (2019)	Redes neuronales y en redes de memoria a corto plazo de largo alcance (LSTM)	41.54 (90 min)
El Idriss et al. (2019)	LSTM	12.38 (30 min)
Mirshekarian et al. (2019)	LSTM	28.19 (60min)
Chen et al. (2019)	Función de kernel optimizado por el algoritmo GWO (GWO-KELM)	13.36 (30 min)
Li (2019)	Red neuronal convolutacional recurrente (CRNN)	33.27 (60 min)
Alfian et al. (2020)	XGBoost	35.8 (60min)

Continúa en la siguiente página.

Referencia	Algoritmo	RMSE (mg/dl) (PH)
Anand et al. (2020)	Algoritmo de refuerzo adaptativo (AdaBoost)	20 (30 min)
Li et al. (2020)	Redes Neuronales de Estado Echo (ESN)	13.67 (45 min)
Xie and Wang (2020)	ARX	19.48 (30min)
Dave et al. (2021)	Bosque Aleatorio (RF)	96.21 (60 min)
Prendin et al. (2021)	Media Móvil Integrada Autoregresiva (ARIMA)	22,15 (30 min)
Zhu et al. (2021)	LSTM en el borde (edge-LSTM)	32.61 (60 min)
Zhu et al. (2022)	Redes Neuronales Recurrentes (RNN) basada en atención y evidencia	35.55 (60 min)
Mosquera-Lopez (2022)	LSTM	33,2 (60 min)
Yang et al. (2023)	Mecanismo de atención múltiple temporal para pacientes diabéticos (PBG TAM)	35.09 (60 min)
Butt et al. (2023)	LSTM	25.48 (60 min)

1.4 Análisis Composicional de Datos

En esta sección se introducen los fundamentos básicos de la teoría de los datos composicionales (CoDa), que fueron objeto para la validación e implementación de los modelos de predicción de esta tesis. La teoría del análisis composicional de los datos tiene varias décadas de haber sido establecida, y con el paso de los años sigue incrementándose el número de investigaciones cuya aportación la ha solidificado como una teoría formal (Aitchison, 1982, 1986). Históricamente estos datos han sido identificados con datos clausurados, y el simplex ha sido considerado el espacio muestral natural o soporte de este tipo de datos (S^D). Se trata de un espacio cerrado, el cual es muy diferente al espacio Euclidiano real asociado con datos sin restricciones (Aitchison,

1982).

Un dato composicional se define como un vector de D partes, $X = (x_1, x_2, \dots, x_D)$, todas las componentes positivas y cuya sumatoria es igual a una constante k definido como:

$$\begin{cases} x_i > 0, & i = 1, 2, \dots, D. \\ \sum_{i=1}^D x_i = k & k = cte. \end{cases} \quad (1.1)$$

Debido a esta definición, los datos composicionales están sujetos a la restricción de que su suma debe ser igual a 1 para proporciones o probabilidades asociadas a variables discretas; 100 para porcentajes; 10^6 para datos en ppm (partes por millón); 180 para ángulos interiores de un triángulo en análisis de formas; 24 horas del día, entre otros, y sus partes sólo pueden variar de 0 a k . Estas restricciones implican que las técnicas multivariantes habitualmente utilizadas para datos en el espacio real no son adecuadas para el análisis y modelización de los datos composicionales, ya que se basan en matrices de covarianza y de correlación de vectores de observaciones de datos y estas matrices contienen información espuria (Aitchison, 1986).

1.4.1 Principios y operaciones básicas

Los principios básicos del análisis de datos composicionales que son de especial interés en este estudio son los siguientes:

1. Invariancia de escala: lo que establece que un dato composicional solo contiene información relativa, implica que cualquier cambio en la escala de los datos originales no afecta la estructura de la composición.
2. Coherencia subcomposicional: lo que implica que los resultados obtenidos para un subconjunto de partes de una composición, es decir, una subcomposición, deben ser coherentes con los resultados obtenidos con la composición completa (Aitchison, 1982).

Operador clausura

A partir de un vector cualquiera es posible obtener una composición X de S^D solo escalando convenientemente las componentes de modo que su suma sea igual a la unidad o 100%, entonces se define el operador clausura como:

$$\mathbb{C}(x_1, x_2, \dots, x_D) = \left(\frac{x_1}{\sum_{i=1}^D x_i}, \frac{x_2}{\sum_{i=1}^D x_i}, \dots, \frac{x_D}{\sum_{i=1}^D x_i} \right). \quad (1.2)$$

Este operador se aplicará en las situaciones en las que se desee tratar los datos como composicionales y las observaciones estén almacenadas en vectores cuya suma de sus componentes no sea constante. En estas situaciones se considera que el valor absoluto de las partes del vector no aporta información por si solo, tampoco el valor de la suma de las componentes del vector, sino que la información está contenida en los cocientes o ratios entre componentes. Por tanto, es muy importante tener en cuenta que un dato composicional solo aporta información sobre las magnitudes relativas de las componentes que lo integran.

Operador de perturbación y perturbación por diferencia

La operación de perturbación es la operación que en el espacio real llamamos suma, y en el símplex se define como la multiplicación componente a componente de dos composiciones X e Y (ecuación 1.3). Posteriormente se normaliza el resultado a una suma constante con la operación de clausura (ecuación 1.2). Esta operación es una herramienta clave para comparar dos composiciones. A partir de esta se introducen de forma natural conceptos básicos para el análisis de conjuntos CoDa, como la medida de tendencia central, la variabilidad total y la distancia de Aitchison. La perturbación por diferencia a su vez sería la correspondiente a la resta en el espacio real definida en la ecuación 1.4, pero en vez de restar divide componente a componente los dos vectores y después normaliza el resultado según la ecuación 1.2 (Aitchison, 1982).

$$X \oplus Y = \mathbb{C} [x_1 * y_1, x_2 * y_2, \dots, x_D * y_D]. \quad (1.3)$$

$$X \ominus Y = \mathbb{C} \left[\frac{x_1}{y_1}, \frac{x_2}{y_2}, \dots, \frac{x_D}{y_D} \right]. \quad (1.4)$$

1.4.2 Transformaciones principales

La metodología de Aitchison se basa en la transformación de los datos composicionales al espacio real multivariante. El espacio muestral para los cocientes entre las partes es el octante positivo de \mathfrak{R}^{D-1} . Si se toman los logaritmos de los cocientes, el espacio final es todo \mathfrak{R}^{D-1} , por lo tanto, podemos aplicar cualquier técnica estadística clásica. Esta estrategia se remonta al trabajo de (McAlister, 1879) quien desarrolló los fundamentos de la ley lognormal univariante tomando el logaritmo de los datos. Tenemos diversas posibilidades de transformación de los datos, todas ellas basadas en los logaritmos de cocientes entre las partes de un dato composicional (Mateu-Figueras et al., 2003).

Al darse cuenta de la importancia del principio de invariancia de escala y que su implementación práctica requería trabajar con razones de componentes, además de que los logaritmos de proporciones son matemáticamente más manejables que las proporciones, todo esto condujo a las diferentes transformaciones.

La transformación logarítmica aditiva (*alr*) (ecuación 1.5), permite el uso de análisis multivariante estándar sin restricciones en los datos transformados (Aitchison, 1982)

$$alr(X) = \left[\ln\left(\frac{x_1}{x_D}\right), \ln\left(\frac{x_2}{x_D}\right), \dots, \ln\left(\frac{x_{D-1}}{x_D}\right) \right]. \quad (1.5)$$

En esta transformación el denominador adquiere un protagonismo especial respecto al resto, la transformación se convierte en biyectiva pero no es simétrica en las partes de X (a veces es conveniente tratar las partes simétricamente), lo que condujo a Aitchison en 1986 (Aitchison, 1986) a introducir la transformación log-cociente centrada (*clr*) (ecuación 1.6), donde $g(x)$ es la media geométrica de las D partes de X (ecuación 1.7).

$$clr(X) = \left[\ln\frac{x_1}{g(x)}, \ln\frac{x_2}{g(x)}, \dots, \ln\frac{x_D}{g(x)} \right]. \quad (1.6)$$

$$g(x) = \left(\prod_{k=1}^D x_k \right)^{\frac{1}{D}}. \quad (1.7)$$

Esta transformación es biyectiva y simétrica entre las partes. Su imagen es el hiperplano de \mathfrak{R}^D que pasa por el origen y es ortogonal al vector de unidades, es decir, la suma de las componentes del vector transformado es igual a cero.

Sea e_1, e_2, \dots, e_{D-1} una base *olr* en S^D , la función que asigna coordenadas con respecto a e_1, e_2, \dots, e_{D-1} a una composición $X \in S^D$, se denomina transformación isométrica log-ratio (*ilr*): $S^D \rightarrow \mathfrak{R}^{D-1}$ (ecuación 1.8) (Egozcue et al., 2003). La palabra isométrica en *ilr* se refiere a la preservación de la distancia. En (Martín-Fernández, 2019) se introdujo el nombre *olr* para evitar confusiones, porque la transformación *clr* también es una transformación isométrica de relación logarítmica.

$$ilr(x) = [\langle x, e_1 \rangle_a, \dots, \langle x, e_{D-1} \rangle_a]. \quad (1.8)$$

Hay varias formas de definir bases *olr* en el simplex. El criterio principal para seleccionar las composiciones que se incluirán en la base es que mejore la interpretabilidad de la representación en coordenadas. Por ejemplo, al realizar un análisis de componentes principales, se crea una base *olr* de manera que la primera coordenada *olr* (primera componente principal: PC1) represente la dirección de máxima variabilidad (Buccianti et al., 2006).

Una base ortogonal particular puede ser la vinculada a una partición binaria secuencial (SBP) de las partes del vector composicional. En (Egozcue and Pawlowsky-Glahn, 2005) definieron la SBP como una agrupación jerárquica de partes del vector de composición original, iniciando con la composición completa como un grupo y terminando con cada parte en un grupo individual. Si D es el número de partes en la composición original, entonces el número de pasos en la partición es $D - 1$, la base *olr* se calcula siguiendo la ecuación 1.9, donde p y n son el número de partes en la i -ésima fila de S (matriz de signos de la SBP) codificadas por +1 (positivas) y -1 (negativas), las cuales estarán en el numerador y denominador de la correspondiente log-ratio. Los ceros de la SBP corresponden a las partes no incluidas en la partición. La $g(x)$ es la media geométrica definida previamente (ecuación 1.7).

$$olr(x) = \sqrt{\frac{p * n}{p + n}} * \ln\left(\frac{g(x_+)^{\frac{1}{p}}}{g(x_-)^{\frac{1}{n}}}\right). \quad (1.9)$$

Por ejemplo, considere la SBP presentada en la tabla 1.6 propuesta en el trabajo (Cabrera et al., 2023) en el espacio S^5 . La primera coordenada olr en $i = 1$ se calcula como se muestra en la ecuación 1.10, donde posteriormente se calculará la media geométrica de las partes positivas y negativas (numerador y denominador, respectivamente). Siguiendo ese procedimiento se calcularán el resto de las $D - 1$ coordenadas (que para este ejemplo son 4).

Tabla 1.6: Partición Binaria Secuencial.

i	$X_{<54}$	X_{54-70}	X_{70-180}	$X_{180-250}$	$X_{>250}$	$p(+)$	$n(-)$
1	+1	+1	-1	-1	-1	2	3
2	+1	-1	0	0	0	1	1
3	0	0	-1	+1	+1	2	1
4	0	0	0	-1	+1	1	1

$$olr(i = 1) = \sqrt{\frac{2 * 3}{2 + 3}} * \ln\left(\frac{g(X_{<54}, X_{54-70})^{\frac{1}{2}}}{g(X_{70-180}, X_{180-250}, X_{>250})^{\frac{1}{3}}}\right). \quad (1.10)$$

1.4.3 Tratamiento de ceros

Cuando mencionamos el tratamiento de los ceros, nos referimos a los componentes del vector composicional que contemplan un valor cero. Debido a que las transformaciones tienen en cuenta logaritmos y cocientes, no son aplicables a datos que tengan nula alguna de sus partes, por tanto, la ocurrencia de ceros se vuelve un problema serio. En la mayoría de las situaciones prácticas se pueden observar conjuntos de datos cuyas componentes contienen valores nulos: un ejemplo en nuestro caso de estudio es el porcentaje de tiempo pasado en cada uno de los rangos de glucosa, donde al menos uno de los rangos haya sido cero debido a que el sensor no registró ninguna medición de BG del paciente en ese rango.

En el análisis estadístico de los datos composicionales se distinguen tres tipos de valores nulos o ceros: los ceros esenciales, los ceros por redondeo y los ceros de conteo. Los ceros esenciales o absolutos son ceros verdaderos que pueden aparecer en estudios con datos compo-

cionales. Estudios que comprenden variables aleatorias que pueden tomar el valor cero, para indicar la ausencia del elemento en algunos casos. Es importante no modificar estos valores para evitar alterar los datos reales originales, y a su vez la población que se esté analizando. En los ceros por redondeo la variable no es realmente cero, pero debido a la falta de herramientas o técnicas de medición adecuadas es a menudo imposible, o demasiado caro obtener algún valor computable para la variable, por lo que se redondea a cero (Martín Fernández, 2001). Los ceros de conteo aparecen en estudios de muestreo que implican recuentos. Un ejemplo de aparición y tratamiento de este tipo de ceros se desarrolló en (Cabrera et al., 2023), para validar el modelo probabilístico implementado por (Biagi et al., 2021), aquí el recuento de transiciones entre una categoría y otra pudo ser cero si no se encontraba alguna de las categorías. En este tipo de reemplazo se detecta que ningún elemento corresponde a la j -ésima categoría de recuentos, y se registra un valor cero en la j -ésima parte de una composición. Podría estar pasando desapercibida debido al tamaño limitado de la muestra, es decir, si el tiempo del período de observación o del experimento fuera mayor, entonces probablemente algún conteo entraría en esta categoría. De acuerdo con (Martín-Fernández et al., 2003) el método de imputación basado en un reemplazo multiplicativo bayesiano (BM) podría ser apropiado, la idea de esta técnica consiste en aplicar la clausura a las composiciones ($k = 1$) y, posteriormente, considerar las composiciones como vectores de estimaciones de probabilidades en un modelo de probabilidad multinomial. En consecuencia, los valores cero deben ser reemplazados por una pequeña cantidad.

El problema de la presencia de ceros en los datos composicionales ha motivado su estudio para permitir un tratamiento estadístico adecuado. Para ello, es necesario investigar si estos valores nulos presentan un patrón de distribución específico y luego, recurrir a un método de sustitución (Palarea-Albaladejo and Martín-Fernández, 2015). De acuerdo con (Rja and Rubin, 1987) se han definido tres tipos de datos faltantes:

1. Datos faltantes completamente al azar o MCAR (Missing Completely at Random) hace

- referencia a que la pérdida de datos no depende de las otras variables explicativas en el conjunto de datos. En otras palabras, cualquier dato puede perderse y el proceso de que ocurra un dato faltante depende exclusivamente del azar. Este es un caso ideal, porque el tratamiento de los datos existentes no conduce a sesgo en los parámetros estimados. Ejemplo: en un cuestionario, la omisión accidental de una respuesta.
2. Datos faltantes al azar o MAR (Missing at Random) son aquellos cuyos valores pueden deberse a diversos factores, como errores de medición, problemas de registro o simplemente a la elección aleatoria de la muestra. La probabilidad de ocurrencia de estos datos es mayor que en el caso de datos MCAR, aunque en ocasiones se tratan de manera similar. Ejemplo: en un cuestionario, donde la probabilidad de que una respuesta sea omitida depende de la respuesta a otras preguntas.
 3. Datos faltantes no al azar o NMAR (Not Missing at Random) en este caso, incluso teniendo en cuenta toda la información observada disponible, el motivo de observaciones faltantes aún depende de las propias observaciones que no se ven. Incluso después de que los datos observados se toman en cuenta, las diferencias sistemáticas permanecen entre los valores faltantes y los valores observados. En este tipo de valores existe dependencia (estocástica o determinística) entre las observaciones; es decir, un factor puede influir en la ocurrencia de otro factor. Ejemplo: una pregunta en un cuestionario ha sido omitida deliberadamente por el participante.

El preprocesamiento de los datos debe hacerse antes del análisis estadístico (cluster, regresión, MANOVA, etc). Debido a que cada tipo de datos requiere su propio procesamiento previo específico. Se han dado recomendaciones para el tratamiento de la parte faltante según el escenario. Falta de MCAR y pocos valores faltantes (el adjetivo pocos es un tema discutible y no existe una opinión única y no se ha proporcionado una solución exacta en la literatura), para este tipo de datos faltantes se ha sugerido el reemplazo multiplicativo desarrollado en (Martín-Fernández et al., 2003).

Para cuando los datos son del tipo MCAR con muchos valores faltantes o MAR, se ha sugerido el reemplazo log-ratio usando el algoritmo de Expectativa y Maximización (EM) como método más sofisticado, desarrollado en (Rja and Rubin, 1987). Este método paramétrico, supone normalidad multivariada, es un método iterativo basado en la función de verosimilitud. Se aplicará a los conjuntos CoDa en el marco de la metodología de log-ratio.

Un caso particular de NMAR son los problemas relacionados con ceros redondeados. Los datos no se pueden observar porque su valor real está por debajo del error de redondeo máximo o un límite de detección (DL): 1) cuando representamos el valor mediante una cadena finita de dígitos, el valor real se redondea a cero, 2) el valor real está por debajo del límite de detección (BDL) del dispositivo experimental y registra el valor real como un valor menor que el DL. Para resolver este problema se han planteado diferentes métodos de imputación disponibles en CoDaPack (herramienta de software (Thió-Henestrosa and Comas, 2016)) y en el paquete zCompositions de R-Studio (Palarea-Albaladejo and Martín-Fernández, 2015).

Por parte de los métodos univariados (solo usan la información en la parte para imputar), se han presentado el reemplazo multiplicativo y el reemplazo R+ normal clásico y robusto. Por parte de los métodos multivariados (solo usan información de otras partes para reemplazar el cero en una parte), se han propuesto el reemplazo log-ratio EM: clásico y robusto y algoritmo de aumento de datos.

Cuando se tienen pocos ceros por redondeo (menos del 10% de los datos), el reemplazo multiplicativo es apropiado, sin embargo, cuando es grande (más del 10%), se recomienda el algoritmo de relación logarítmica EM modificada multivariante. Este algoritmo se basa en un enfoque paramétrico porque se supone normalidad de relación logarítmica. El método consiste en modificar el algoritmo EM para incluir la información de que los valores imputados (log-ratio) deben ser inferiores al límite de detección (Palarea-Albaladejo and Martín-Fernández, 2008).

Hasta el momento, hemos discutido acerca del tratamiento de las observaciones ceros.

Sin embargo, también es importante considerar la presencia de datos NA (valores perdidos o ausentes), los cuales tienen una fuerte relación con los ceros, dado que ambos suelen estar presentes simultáneamente en conjuntos de datos. Por otro lado, existen los outliers o valores atípicos, los cuales difieren significativamente del resto de los datos, y pueden ser identificados a través de diferentes métodos estadísticos. Con frecuencia, un valor atípico multivariado se refiere a una muestra en la que se observa un valor extremo (tanto excesivamente alto como excesivamente bajo) en uno o varios componentes. Algunos valores atípicos multivariados también pueden considerarse como valores atípicos univariados, donde se encuentra un valor extremo en una única variable. Sin embargo, en algunos casos, un valor atípico multivariado se caracteriza por tener valores extremos en las asociaciones entre los diferentes componentes, lo que indica una discrepancia significativa en las relaciones entre las variables analizadas (Vives-Mestres et al., 2014*a,b*, 2016).

1.4.4 Aplicaciones con datos composicionales

En diversas áreas de la ciencia, los conjuntos CoDa han sido ampliamente utilizados en el análisis exploratorio, incluyendo la geología, geoquímica, ecología, microbiología, entre otras (Aitchison, 1982; Martin-Fernández et al., 1999; Pawlowsky-Glahn and Egozcue, 2002; Martín Fernández et al., 2015; Filzmoser and Filzmoser, 2021; Zhou et al., 2022). En la investigación de microbiomas, por ejemplo, se han utilizado modelos de regresión con datos composicionales, específicamente en la formulación general de log-contraste lineal, la optimización proximal y su aplicación en datos microbiológicos (Combettes and Müller, 2021). Además, se ha ilustrado la versatilidad de estos modelos en el análisis de datos de microbiomas. En la geología y geoquímica, las composiciones de rocas y minerales también se han considerado y se utilizan para estudiar la evolución de la corteza terrestre, la formación de yacimientos de minerales, entre otras aplicaciones (Dmitrijeva et al., 2019; Sahoo et al., 2020; Guagliardi et al., 2020; Cicchella et al., 2022). En la ecología, se ha investigado la composición

de especies en comunidades biológicas y las interacciones entre ellas, así como los cambios en la biodiversidad (Willis and Martin, 2022).

Se han identificado diversas aplicaciones en la medicina (Chastin et al., 2015; Dumuid et al., 2018; Biagi et al., 2019; Janssen et al., 2020). Relacionado a la diabetes, se han interpretado los perfiles diarios de glucosa de pacientes con T1D. Biagi y otros colaboradores presentaron una metodología para categorizar estos perfiles (Biagi et al., 2019). Posteriormente, publicaron un modelo probabilístico de transiciones entre categorías de períodos de 24 a las próximas 6 horas (Biagi et al., 2021), que luego se validó con una cohorte de pacientes reales en (Cabrera et al., 2023). Por otro lado, también se ha estudiado el tiempo dedicado a diferentes actividades durante el día, como contribuciones relativas de 24 horas (Chastin et al., 2015; Dumuid et al., 2018).

1.5 Contexto de investigación y conjuntos de datos

Este trabajo ha sido desarrollado utilizando las siguientes herramientas de software: MATLAB (MATrix LABoratory) versión R2022b, donde se implementó y validó la herramienta de simulación presentada. Posteriormente para el tratamiento y preprocesamiento de las mediciones de CGM de las cohortes de pacientes reales, análisis de las mediciones, interpolación, división en diferentes períodos, asociación de matrices de límites de detección e implementación de métricas. También se empleó la versión 2022.07.2 de R-Studio, una plataforma ampliamente utilizada en análisis estadísticos, gracias a su amplia gama de herramientas disponibles, como análisis de cluster, análisis discriminante lineal, regresión múltiple, entre otros. Para algunas visualizaciones se utilizó el programa CoDaPack versión 2.03.01, el cual es un software gratuito desarrollado en la Universidad de Girona (UdG) a partir del enfoque introducido a principios de los años ochenta por John Aitchison, y debido a que dicha metodología no era fácil de usar con paquetes estadísticos estándar (Thió-Henestrosa and Comas, 2016).

Este estudio ha sido patrocinado por el Programa de Investigadores en Formación de la

Universidad de Girona (IFUdG2019) y se ha llevado a cabo en el Laboratorio de Modelado, Identificación e Ingeniería de Control (MiceLab) de la UdG. El equipo investigador ha estado trabajando en tecnologías para la diabetes desde el año 2004 y es miembro activo del Consorcio Español de Páncreas Artificial y Tecnología de la Diabetes (eSCAPE). En el año 2018, el Consorcio se integró en el Centro de Investigación Biomédica en Red en Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), una iniciativa que agrupa a diferentes grupos de investigación en ingeniería y medicina clínica en España:

- MiceLab

El grupo de investigación MiceLab fue fundado en el año 2000 y liderado por el Dr. Josep Vehí, es un reconocido grupo de investigación financiado por la Generalitat de Catalunya. Este grupo se encuentra en el Instituto de Informática y Aplicaciones de la UdG. MiceLab es un equipo interdisciplinario que combina una sólida base teórica en ingeniería de control, inteligencia artificial, aprendizaje automático, ingeniería biomédica e informática y matemáticas aplicadas. Su enfoque se centra en la teoría de sistemas y control, modelado y control de sistemas biomédicos, incertidumbre y variabilidad intrapaciente, optimización de terapias con insulina y aplicaciones de aprendizaje automático en el campo de la diabetes. El grupo ha realizado importantes contribuciones en estas áreas de investigación, y su trabajo se ha destacado en la comunidad científica.

Los datos tratados en esta tesis se describen a continuación:

- Hospital Clínic de Barcelona

Analizamos los datos de ocho pacientes con terapia de bomba de insulina, obtenidos de un estudio piloto realizado en el Hospital Clínic de Barcelona durante aproximadamente 30 semanas, recolectados en diferentes períodos entre 2020 y 2022. Los pacientes reclutados usaban diferentes sensores, incluidos Dexcom G6, MiniMed 640G y FreeStyle Libre. Los primeros dos modelos de sensores almacenaron sus mediciones cada 5 minutos, mientras que el tercer modelo almacenó sus mediciones cada 15 minutos.

- Cohorte REPLACE-BG (NCT02258373)

Se realizó el estudio con 226 pacientes por aproximadamente 35 semanas, los datos fueron recolectados entre el 2015 y 2016. Este estudio se centró en investigar si el uso de CGM sin confirmación regular de BGM es tan seguro y efectivo como el uso de CGM con confirmación de BGM en adultos con T1D. Los pacientes portaban el sensor: Dexcom G4 Platinum Continuous Glucose Monitoring System con algoritmo modificado + Abbot Precision Xtra Blood Glucose-Ketone Meter (Aleppo et al., 2017).

1.6 Objetivos

Teniendo en cuenta el conocimiento previo acerca de la relevancia de la simulación para evaluar terapias destinadas al tratamiento de la T1D, y los desafíos inherentes a los modelos de predicción a largo plazo, se puede establecer el objetivo principal de la tesis.

El objetivo de esta tesis es proporcionar una herramienta de simulación realista con escenarios desafiantes para evaluar terapias MDI y CSII. Además, desarrollar modelos de predicción de BG a largo plazo para prever situaciones de riesgo.

Los objetivos de investigación en esta tesis se pueden dividir en los siguientes objetivos específicos:

- Implementar y validar una herramienta de simulación realística basada en el modelo compartimental de Hovorka e incorporar los modelos matemáticos de insulina de acción prolongada Gla-100, Gla-300 y Deg-100. Para este fin, se creó una cohorte de 47 VPs con alta variabilidad, en un escenario semejante a la vida cotidiana. Se realizaron simulaciones a lazo abierto para terapias MDI y CSII. Los resultados se compararon con estudios clínicos reportados en la literatura.

- Diseñar una metodología para validar el modelo probabilístico de predicción presentado previamente en (Biagi et al., 2021). Para resolver este objetivo, se utilizaron los datos de una cohorte real de pacientes con T1D. Se propuso una metodología matemática y métricas novedosas basadas en datos composicionales que pueden ser útiles en cualquier escenario que cumpla estas características.
- Implementar y validar un modelo individualizado, para predicción en amplios horizontes (2 y 4 h). Aquí, a partir de la predicción de media y del coeficiente de variación de la glucosa a cada hora del día basados en datos composicionales, se obtuvieron predicciones de estados glucémicos futuros.

1.7 Estructura de la tesis

Esta tesis está organizada de la siguiente manera: el Capítulo 2 está constituido por una copia de los artículos que permitieron la presentación de esta tesis como un compendio de publicaciones. El Capítulo 3 presenta una breve discusión sobre las principales contribuciones de los artículos que forman parte de esta tesis. Finalmente, el Capítulo 4 presenta las conclusiones y trabajos futuros. Se incluyen los apéndices A, B y C que corresponden al primer manuscrito (Estremera et al., 2022).

PREDICCIÓN PROBABILÍSTICA DE ESTADOS GLUCÉMICOS PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

Este capítulo consta de tres secciones. La Sección 2.1 presenta una herramienta de simulación que incluye escenarios desafiantes y realistas e incorporación de modelos de insulina para evaluar terapias MDI y CSII. En la Sección 2.2 se valida el modelo probabilístico de predicción de glucosa basado en datos composicionales con métricas novedosas. La Sección 2.3 implementa y valida modelos de regresión para predicción de glucosa a largo plazo. Los resultados adicionales del primer artículo presentado se incluyen en los apéndices A, B y C.

- **2.1 Simulador con escenarios realistas y desafiantes para VPs con T1D que se someten a terapia CSII y MDI.**
- **2.2 Validación del modelo probabilístico de predicción para pacientes con T1D utilizando análisis de datos composicionales.**

- **2.3 Predicción individualizada de estados glucémicos mediante el análisis de datos composicionales.**

2.1 Simulador con escenarios realistas y desafiantes para VPs con T1D que se someten a terapia CSII y MDI

En esta publicación proponemos una herramienta de simulación realista con escenarios desafiantes que incluye una cohorte de VPs y modelos de insulina de acción rápida y acción prolongada Gla-100, Gla-300 y Deg-100. La contribución del candidato para esta publicación consistió principalmente en incorporar y validar los modelos de insulina de acción prolongada al simulador, analizar los resultados y compararlos con los reportados en los ensayos clínicos, contribuir con la escritura del manuscrito, contribuir a la discusión y editar el manuscrito a lo largo de las rondas de revisión. Durante el desarrollo del trabajo, la candidata colaboró con su colega el Dr. Ernesto Estremera, Dr. Aleix Beneyto y contó con la asistencia del Dr. Josep Vehí.

Título: A simulator with realistic and challenging scenarios for virtual T1D patients undergoing CSII and MDI therapy

Autores: Ernesto Estremera, **Alvis Cabrera**, Aleix Beneyto y Josep Vehí

Revista: Journal of Biomedical Informatics

Volumen: 132, Páginas: 104-141, Published: Julio 2022

DOI: <https://doi.org/10.1016/j.jbi.2022.104141>

Índice de calidad: JCR: Q1 (3/31) en MEDICAL INFORMATICS, JIF: 8.0 en 2021

2. PREDICCIÓN PROBABILÍSTICA PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

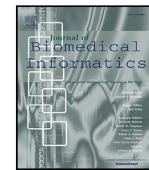
Journal of Biomedical Informatics 132 (2022) 104141



Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin



A simulator with realistic and challenging scenarios for virtual T1D patients undergoing CSII and MDI therapy

Ernesto Estremera^{a,1}, Alvis Cabrera^{a,1}, Aleix Beneyto^a, Josep Vehi^{a,b,*}

^a Department of Electrical, Electronic and Automatic Engineering, University of Girona, 17004 Girona, Spain

^b Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas, 28029 Madrid, Spain

ARTICLE INFO

Keywords:

CSII therapy
Glucose control
MDI therapy
Type 1 diabetes
Virtual patients

ABSTRACT

In silico simulations have become essential for the development of diabetes treatments. However, currently available simulators are not challenging enough and often suffer from limitations in insulin and meal absorption variability, which is unable to realistically reflect the dynamics of people with type 1 diabetes (T1D). Additionally, T1D simulators are mainly designed for the testing of continuous subcutaneous insulin infusion (CSII) therapies. In this work, a simulator is presented that includes a generated virtual patient (VP) cohort and both fast- and long-acting Glargine-100 U/ml (Gla-100), Glargine-300 U/ml (Gla-300), and Degludec-100 U/ml (Deg-100) insulin models. Therefore, in addition to CSII therapies, multiple daily injections (MDI) therapies can also be tested. The Hovorka model and its published parameter probability distributions were used to generate cohorts of VPs that represent a T1D population. Valid patients are filtered through restrictions that guarantee that they are physiologically acceptable. To obtain more realistic scenarios, basal insulin profile patterns from the literature have been used to identify variability in insulin sensitivity. A library of mixed meals identified from real data has also been included. This work presents and validates a methodology for the creation of realistic VP cohorts that include physiological variability and a simulator that includes challenging and realistic scenarios for in silico testing. A cohort of 47 VPs has been generated and in silico simulations of both CSII and MDI therapies were performed in open-loop. The simulation outcome metrics were contrasted with literature results.

1. Introduction

Diabetes mellitus is a chronic condition that has become a pandemic affecting 537 million people worldwide [1]. T1D, with a prevalence of approximately 10% of all diabetes cases, is characterized by the self-destruction of the insulin producing beta cells in the pancreas. This leads to a permanent lack of insulin that causes an abnormal glucose homeostasis state of high blood glucose levels, known as hyperglycemia, which eventually leads to chronic complications. Diabetes complications are both microvascular (retinopathy, neuropathy, nephropathy) and macrovascular (coronary artery, peripheral artery and cerebrovascular conditions), as well as other acute complications [2,3]. People with T1D rely on exogenous insulin to lower blood glucose levels back to normoglycemia (70–180 mg/dl), which has been established as a control objective [4]. To this end, either MDIs, CSII, and more recently artificial pancreas (AP) have been prescribed to them.

Although much progress has been made on the subject, many researchers are still working on more effective methods for its diagnosis and treatment. Mathematical models that contain different sets of parameters and nonlinear equations to describe the behavior of insulin and glucose dynamics in patients with T1D [5–8], are a widely used tool for creating, testing, and optimizing diagnostic and treatment methodologies. Several studies have shown that simulators containing the Hovorka [8] and Dalla Man models [7,9–11] are useful in developing MDI therapies and closed-loop control systems, also known as the AP [12–18].

The Hovorka model [8] is a compartmental model of glucose kinetics and insulin action that represents the input–output relationship between subcutaneous insulin infusion (as the input) and intravenous glucose concentration (as the output) [19]. The Dalla Man model [7] is also a compartmental model, however, this model links together

* Corresponding author at: Department of Electrical, Electronic and Automatic Engineering, University of Girona, 17004 Girona, Spain.

E-mail addresses: ernesto.estremera@udg.edu (E. Estremera), alvita.cabrera@udg.edu (A. Cabrera), aleix.beneyto@udg.edu (A. Beneyto), josep.vehi@udg.edu (J. Vehi).

¹ Contributed equally to this work.

<https://doi.org/10.1016/j.jbi.2022.104141>

Received 12 November 2021; Received in revised form 28 June 2022; Accepted 7 July 2022

Available online 12 July 2022

1532-0464/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

2.1. SIMULADOR CON ESCENARIOS REALISTAS Y DESAFIANTES PARA VPS CON T1D QUE SE SOMETEN A TERAPIA CSII Y MDI

glucose and insulin plasma concentrations with glucose fluxes (endogenous glucose production, glucose rate of appearance, glucose utilization, renal extraction) and insulin fluxes (rate of insulin appearance from the subcutaneous tissue and insulin degradation). The latest version of the Dalla Man model [20] also incorporates glucagon kinetics, secretion, and action models.

Simulation tools based and defined by mathematical models are a fundamental instrument to the investigation of diabetes treatments, as they provide a safe and economical platform to validate treatment strategies and control algorithms via in silico tests. A number of simulation environments exist and are available in the literature [9,12,21–23]. The basic characteristics that define diabetes simulators are its ability to: (1) create varied scenarios in which food intake can be defined by the time of ingestion and the amount of carbohydrates (CHO) consumed, (2) simulate physical exercise using a model that can adjust exercise schedules, frequency, and intensity, (3) simulate blood glucose control strategies that manipulate insulin and/or glucagon or suggest rescue CHO, and (4) identify the dynamic behavior of glucose considering a variety of variables and provide a series of results that allow the performance of the strategy used to be evaluated.

All these simulators use VP cohorts to evaluate open and closed-loop glycemic control strategies. For this reason, generating a physiologically adequate VP is key to effective in silico testing. Simulators can incorporate a simple set of parameters that represent an ‘average patient’ or multiple sets of parameters that represent a VP population [24].

The main limitations of these simulation tools are: (1) physiological variability of insulin absorption and meals that do not reflect real patients, (2) scenarios that do not replicate real-life characteristics which makes the current scenarios easy to control, and (3) simulators that are not designed for trials with MDI therapy.

In [25] it is stated that most of the studies in the literature have investigated meals or beverages containing glucose and, for methodological reasons, very few studies have included other simple sugars or complex CHO. They conclude that large evening meals containing complex CHO provide different challenges to the management of T1D, depending on the composition of the meals. A meal with a high glycemic load results in a pronounced, but shorter, absorption pattern, whereas a meal with a low glycemic load results in sustained and prolonged absorption and a higher overall glucose appearance. Although published simulators may incorporate some degree of variability in the rate of glucose appearance profiles, they are not able to simulate the variety of mixed meals found in real-life conditions due to limitations in the data they were built upon [26].

The consequence of these limitations is that most of today’s simulators result in scenarios that provide blood glucose dynamics that are very easy to control. Reported results in the literature using such simulators show that blood glucose control strategies are able to obtain 90% or higher time in the euglycemic range, and times in hyperglycemia or hypoglycemia are close or equal to zero [13,14,27–29] and these results are obtained regardless of the scenario used. However, it is later demonstrated that this behavior is not replicated in clinical trials [30–33].

During the past 15 years, there have been changes and advancements in long-acting insulin formulations. Models describing the pharmacokinetics of long-acting insulin [34–36] model the effect of the injected basal insulin dose. In [35], an absorption model of insulin glargine is proposed and in [37], this model is incorporated into the UVA/Padova simulator for simulations with MDI therapy. In [18], the incorporation of the insulin degludec model into the UVA/Padova simulator is described.

The objective of this work is to minimize the limitations of the simulators present in the literature, bringing in silico simulations closer to real-life conditions and creating a greater challenge for the design of controllers aimed at blood glucose control in patients with T1D. For this, a methodology to generate VP cohorts that include physiological variations and realistic scenarios incorporating basal insulin

patterns identified in the literature and adding a library of challenging mixed meals is presented. Additionally, the incorporation of insulin glargine and degludec absorption subsystems into the Hovorka model to evaluate different MDI therapies is validated.

2. Methods

2.1. Generation and validation of VP

The generation of VPs for the Hovorka model is based on public data contained in a series of publications [21,38,39]. From these publications, it is known that V_G , R_{ih} , R_{CL} , V_I , k_a , k_e , BIO , t_{max} , k_{aim} have univariate distributions, whereas the two groups of parameters $\{EGP_0, F_{01}, S_{TD}, S_{IE}, S_{IT}\}$ and $\{k_{12}, k_{a1}, k_{a2}, k_{a3}\}$ have multivariate distributions, in this case covariance matrices were calculated using the patient data presented in [39]. The probability distribution of the parameters of the Hovorka model and covariance matrices are included in Appendix B Supplementary data 1.

The first step for the creation of VPs includes randomly generating multiple sets of parameters using the Matlab commands *mvnrnd* for multivariate parameters (uses the covariance matrix), and *normrnd* or *lognrnd* (depending on the type of distribution) for univariate parameters.

Next, if a randomly sampled VP does not behave in a physiologically acceptable manner, it is discarded from the final cohort. This is determined by the following conditions: (1) all sampled parameters for each VP have to be included within the minimum and maximum allowed values [39]; (2) a glucose–insulin model must reflect an insulin-independent consumption of glucose that is always lower than the endogenous glucose production, in the Hovorka model, this is fulfilled by verifying the following relationship: $EGP_0 > F_{01}$; (3) VPs, in the absence of insulin, must have glucose levels at steady state that exceed 300 mg/dl [12]; (4) VPs without external disturbances nor variability, e.g. meals or exercise, must have an input–output relation such that for some input basal insulin infusion, the resulting output model basal glucose stays in the 90–160 mg/dl band; and (5) variations of 0.01 U/h in the basal insulin infusion rate cannot cause a change in basal glucose greater than 20 mg/dl.

The VPs that meet all the established restrictions are each assigned a 24 h circadian pattern of insulin sensitivity, a set of adjusted carbohydrate ratios (CR), correction factors (CF) and a basal insulin pattern. The basal insulin for CSII therapy is adjusted in U/h, and is calculated depending on the desired basal glucose and for MDI therapy is adjusted in mU and is dependent upon the time of day it will be administered. The CR (g/U) is generally known in clinical practice as “Insulin to Carbohydrate Ratio” and represents how many grams of CHO are covered by 1 unit of insulin, three values of CR are calculated for three different time periods, 6:00 h–12:00 h, 12:00 h–17:00 h, and 18:00 h–24:00 h, different meals from the meal library are randomly selected and simulated for each (breakfast, lunch, dinner), then we iteratively adjust the CR such that the resulting infused bolus insulin ensures minimization of hyperglycemia, trying not to cause postprandial hypoglycemia. The CF (mg/dl/U), which is defined as the amount of blood glucose that falls with 1 unit of insulin, is calculated following the test protocols detailed in [40] and consists of choosing a CF that brings glucose within ± 30 mg/dl of the target value in 5 h without falling below this limit.

2.2. Mixed meal library

The Micelab group has a mixed meal library, described in [41], that has been incorporated into the current T1D simulator to create more realistic scenarios. This library, see Appendix A Supplementary data 1, can be employed to test and optimize the design of closed-loop insulin delivery systems, insulin bolus calculators, hypoglycemia prediction algorithms, and fault detection and supervision systems for

2. PREDICCIÓN PROBABILÍSTICA PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

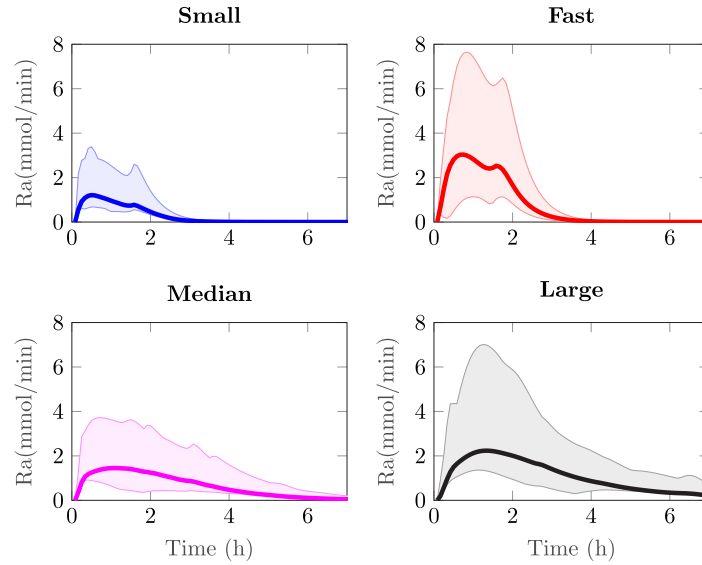


Fig. 1. Classification of meals in the library.

an AP. This library contains Ra profiles from a total of 54 different meal compositions. The mixed meals are defined by their composition (fat, protein, fiber, and energy), amount of CHO, and classified according to the Kolmogorov–Smirnov distance into large (10), medium (26), small (12), and fast (6). Fig. 1 shows the mean, maximum, and minimum of the 4 classified groups. Each meal has a time profile of 84 values taken with a sampling time of 5 min, which represents the glucose rate of appearance of 7 h after the moment of the meal. For its implementation, the term $U_g(t)$ from Eq. (A.1) of the glucose subsystem (Appendix A.1) is replaced by the parameter $Ra(t)$ according to the time profile corresponding to each meal.

2.3. Circadian variability of insulin sensitivity

Insulin sensitivity variability is a phenomenon identified as one of the main challenges in the treatment of diabetes. In 2005 [42] and 2007 [43], two clinical studies presented 24-hour basal insulin patterns identified with data from insulin pump patients. In the first study [42], an analysis is made of the characteristics of basal insulin requirements by age and gender in patients with T1D. In the second study [43], the same analysis was performed, but only for children and adolescents. In this work, we include hourly insulin sensitivity variability from randomly sampled basal insulin patterns according to the results presented in Table 1 in [42], for patients between 21–60 years of age. The procedure for generating multiple basal insulin profiles is detailed in Appendix B.1.

Fig. 2 shows the average of 100 generated basal insulin patterns, and it can be noted that the curve has similar behavior to that shown in [42]. Every single basal insulin pattern is next used to accordingly generate a unique insulin sensitivity pattern that will be given to a VP.

Insulin sensitivity is introduced into the model, in the equations in Appendix A.2, as an hourly time-varying multiplicative factor of the original sensitivity parameters of the Hovorka model (S_{ID} , S_{IE} , S_{IT}), see Eq. (1).

$$\begin{aligned} S_{ID}(t) &= \alpha(t)S_{ID} \\ S_{IE}(t) &= \alpha(t)S_{IE} \\ S_{IT}(t) &= \alpha(t)S_{IT} \end{aligned} \quad (1)$$

where $\alpha(t)$ is a piecewise constant function with 24 steps representing the factor for each hour in a day that affects the model's insulin sensitivity. To obtain $\alpha(t)$, the steady-state of a VP is solved using a generated basal profile u , this procedure is detailed in Appendix B.2.

2.4. Long-acting insulin glargine and degludec

The insulin glargine and degludec, both second-generation insulin analogs, aim to achieve stable glucose profiles with fewer peaks, while minimizing the amount of nocturnal hypoglycemia in patients with T1D. The UVA/Padova research group has developed and clinically validated a compartmental model that describes the subcutaneous absorption of insulin glargine and degludec (Appendices A.5–A.7) [18, 35].

2.4.1. Glargine model

In [35], pharmacokinetic (PK) data from 3 different clinical studies have been used to model the subcutaneous (sc) absorption of insulin glargine. In each study, euglycemic clamp protocols were conducted in patients with T1D and a validated radio immunoassay. The sc absorption model of both, Gla-100 and Gla-300 is a two-compartment structure proposed in [37] and incorporated in the UVA/Padova Simulator and validated for the Dalla Man model [37].

2.4.2. Degludec model

Second-generation insulin Deg-100 provides new basal insulin therapies for the treatment of T1D. In [18], a three-compartment PK model is presented describing subcutaneous absorption of Deg-100 based on clinical data and incorporated into the UVA/Padova simulator.

Fig. 3 shows the incorporation of both models into the Hovorka model. In both analogs, active plasma insulin (I_a) is considered as the sum of the long-acting insulin concentration (I_{la}) and the fast-acting insulin concentration (I_{fa}) (Eq. (2)). Eq. (3)–(6), shows the modifications of the insulin absorption subsystem (Appendix A.3) and the insulin action subsystem (Appendix A.2) of the Hovorka model [8]. The parameters of both long-acting models are the same as reported in [18,35] and do not vary between VPs, except for the V_I parameter, which is identified individually for each VP.

$$I_a(t) = I_{fa}(t) + I_{la}(t) \quad (2)$$

2.1. SIMULADOR CON ESCENARIOS REALISTAS Y DESAFIANTES PARA VPS CON T1D QUE SE SOMETEN A TERAPIA CSII Y MDI

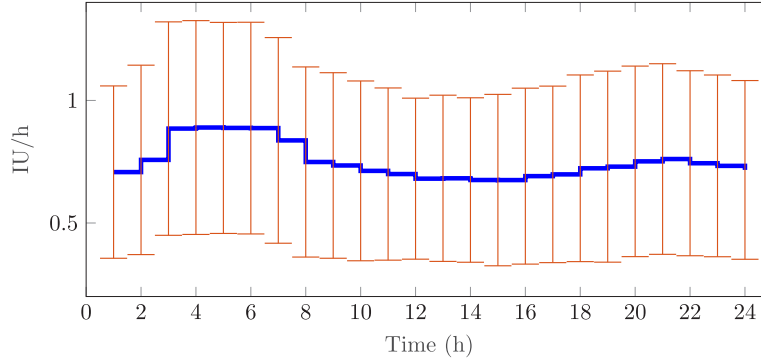


Fig. 2. Mean and SD of the generated basal insulin profiles.

$$\frac{dI_{f_d}(t)}{dt} = \frac{S_2(t)}{T_{maxI} * V_I} - k_e I_{f_d}(t) \quad (3)$$

$$\frac{dx_1(t)}{dt} = -k_{a1}x_1(t) + k_{a1}S_{IT}I_t(t) \quad (4)$$

$$\frac{dx_2(t)}{dt} = -k_{a2}x_2(t) + k_{a2}S_{ID}I_t(t) \quad (5)$$

$$\frac{dx_3(t)}{dt} = -k_{a3}x_3(t) + k_{a3}S_{IE}I_t(t) \quad (6)$$

2.5. Simulation protocol

The proposed simulator was implemented in Matlab R2021a and simulations were performed using an Intel(R) Core(TM) i7-4770 CPU @ 3.40 GHz processor with 16 GB RAM. The initial cohort size was 100 VPs for the Hovorka model, of which 47 VPs met the requirements set and were taken as valid. The Supplementary data 1. Appendix B presents a table with all parameters of the Hovorka model of the generated cohort.

For the generated cohort CSII therapy, a vector is created from 0.02 to 4 with a variation of 0.01 that represents the input of basal insulin (U/h) to the model and the one that guarantees glucose at a steady-state (basal glucose) around 160 mg/dl is selected. For MDI therapy, basal insulin was adjusted based on two defined injection times at 9:00 h or 20:00 h. The insulin bolus, given at the time of the meal, is calculated with Eq. (7) presented in [44] where CR (g/U), CF (mg/dl/U) and G_{Target} (mg/dl) are specific parameters individually adjusted for each VP. The IOB , which is defined as the amount of insulin delivered that is still active, was modeled as proposed in [45].

$$Bolus = \frac{CHO}{CR} + \frac{G - G_{Target}}{CF} - IOB \quad (7)$$

The considered scenario consists of a 60 day simulation protocol with meals of 40, 85, and 75 grams of CHO were given at 7:30 h, 13:00 h, and 19:00 h, respectively, with a variability of 30 min in the scheduled intake times and 10% in the CHO content. To simulate these meals, the amount of CHO defined in the initialization of the scenario was used to select a meal with similar characteristics from the mixed meal library. To model continuous glucose monitoring (CGM), we follow the same procedure as in [11] where starting from the calibration error of the sensor they generate measurement noise additive to the simulated blood glucose concentration.

A low frequency term of intra-patient variability is also included by using a sinusoidal pattern, on some parameters of the Hovorka model (EGP_0 , F_{01} , k_{12} , t_{maxI} , k_e) and on the long-acting insulin glargine and degludec models (m_4 , k_a). Each parameter affected by intra-patient variability becomes a time-varying parameter, and a specific $\beta(t)$ function

is generated individually. For example, EGP_0 becomes time varying by applying $EGP_0(t) = EGP_0\beta(t)$. The procedures to generate the $\beta(t)$ signal is detailed in Appendix C.

3. Results

Using the previously described simulator, generated VP cohort and scenario, we proceeded to test how adding the meal library and circadian variability affects the simulations, and compared CSII and MDI results with real reported outcomes. In silico simulations of both therapies were always performed in open-loop therapy.

3.1. Simulator complexity analysis

Table 1 shows the effects of combining the mixed meal library and circadian variability on insulin sensitivity versus the base Hovorka model. The outcomes for the base Hovorka model resulted in 87.6% of the time in range (TIR) (70–180 mg/dl) and low values of hypoglycemia (< 70 mg/dl) and hyperglycemia (> 180 mg/dl). The same table shows how this behavior is affected by incorporating both factors separately; it is important to note that these cause a decrease in TIR metrics and a notable increase in hyperglycemia values. This behavior is shown in Fig. 4. The results suggest that the inclusion of the mixed meal library and circadian variability resulted in more physiologically realistic outcomes.

3.2. Outcomes for CSII and MDI therapies

To assess the performance of both therapies, Table 2, 3, 4, and 5 show the standardized CGM metrics [46] (these values are reported as a percentage of time in the specified range), for CSII and MDI therapies. Some of the clinical trials used in this work for comparison purposes do not report the metrics as suggested in [46], that is, less severe hypoglycemia (70–54 mg/dl) and severe (< 54 mg/dl) and less severe hyperglycemia (180–250 mg/dl) and severe (> 250 mg/dl). They are presented as a single category, including less severe and severe. Therefore, the clinical outcomes presented in Tables 2, 3, and 4 span to different ranges as compared to the ‘in silico’ outcomes. The results are presented as the median (25th, 75th percentile). Fig. 5 shows the glucose outcomes per day for both therapies. We can verify the physiological plausibility of the glucose trajectories in the simulations. In addition, Supplementary data 2 shows the AGP report of two representative patients with CSII and MDI therapy.

Table 2 shows the CSII therapy results for the nominal Hovorka model and with the inclusion of the meal library and circadian variability. Comparing the results with the clinical trial it can be seen that mixed meals library and circadian variability result in more physiologically feasible VPs. Comparing the results obtained with the clinical

2. PREDICCIÓN PROBABILÍSTICA PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

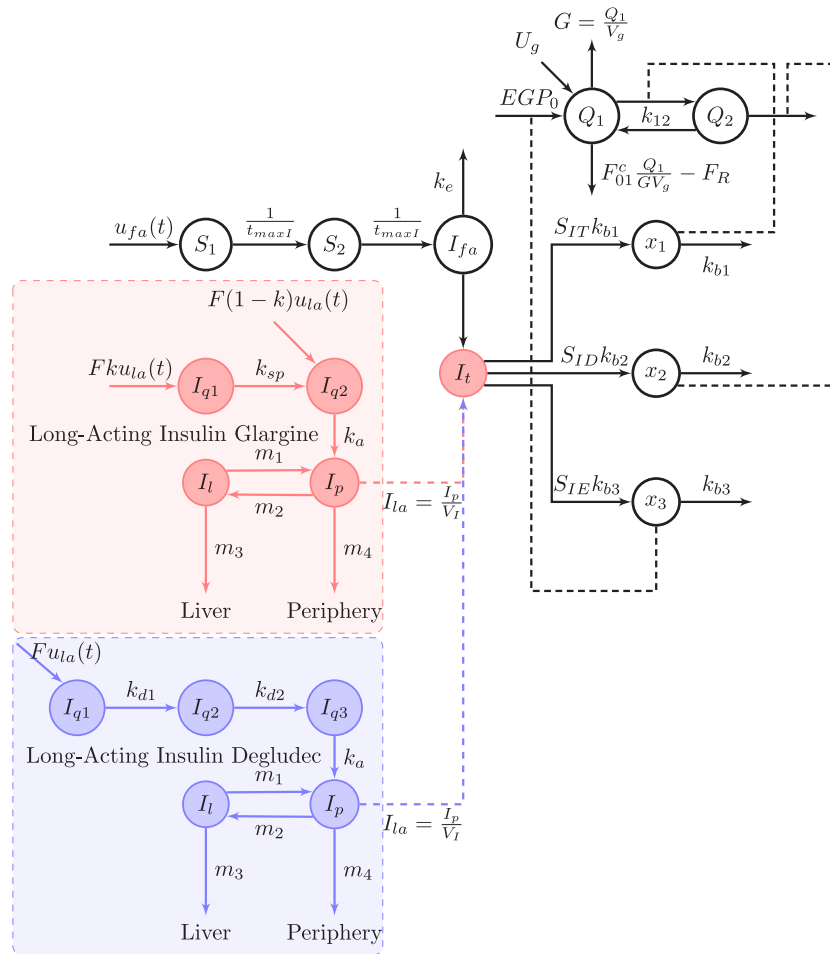


Fig. 3. Glucose-insulin compartment model with the addition of the long-acting insulin glargine (red) and degludec model (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

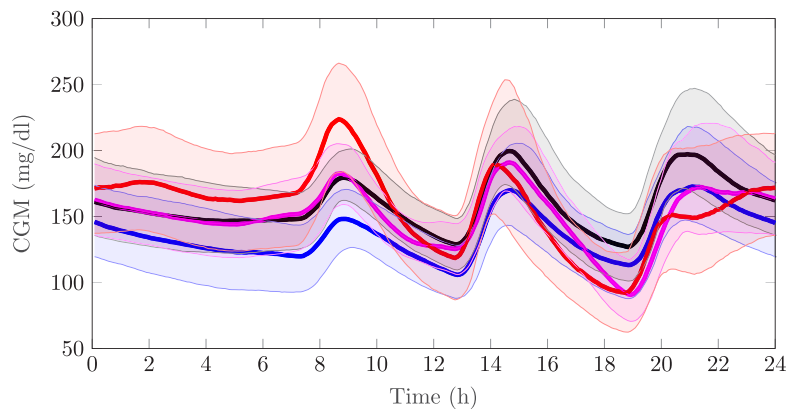


Fig. 4. Daily aggregated CGM of the cohort of Table 1 (median + IQR). The blue curve shows the Hovorka model, the black curve the Hovorka model using the meal library, the magenta curve the Hovorka model including circadian variability, and the red curve the Hovorka model with the meal library and circadian variability. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.1. SIMULADOR CON ESCENARIOS REALISTAS Y DESAFIANTES PARA VPS CON T1D QUE SE SOMETEN A TERAPIA CSII Y MDI

Table 1

Effects caused by the addition of the mixed meal library and circadian variability of insulin sensitivity in the outcomes for CSII.

Performance indicator	Nominal Hovorka model ^a	Hovorka model with mixed-meal library ^a	Hovorka model with circadian variability ^a
CGM (mg/dl)	136.4 (120.1–150.8)	158.1 (143.9–180.1)	153.3 (140–160)
%CV	22.9 (16.7–31)	20.6 (15–30)	23.9 (18.1–30.1)
% time CGM > 250 mg/dl	0 (0–0.6)	0 (0–8.7)	0.1 (0–4.7)
% time CGM 180–250 mg/dl	5.7 (2–25)	19.6 (9.6–35.1)	18.6 (9.3–28.4)
% time CGM 70–180 mg/dl	87.6 (63.2–96.8)	77.4 (47.1–87.4)	77.5 (60.3–89.8)
% time CGM 54–70 mg/dl	0.5 (0–3.2)	0 (0–0.5)	0.1 (0–2.7)
% time CGM < 54 mg/dl	0 (0–0.1)	0 (0–0)	0 (0–1)

^aValues are reported as median, interquartile range (25th, 75th percentile).

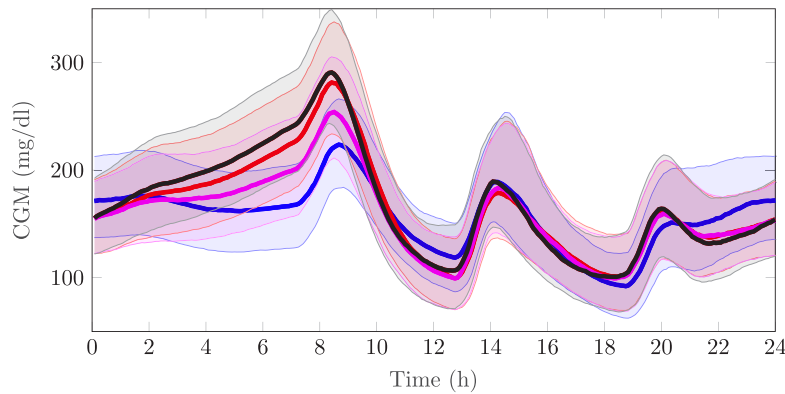


Fig. 5. Daily aggregated CGM of the generated cohort (median + IQR) for different therapies: CSII (blue), Gla-100 9 h (black), Gla-300 9 h (magenta) and Degludec 9 h (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2

Outcomes for CSII therapy and clinical trial.

Performance indicator	Nominal Hovorka model ^a	Hovorka model with variability and mixed meals ^a	Clinical trial [30] ^b
Median CGM (mg/dl)	136.4 (120.1–150.8)	159.9 (147.1–172.1)	174 (18)
%CV	22.9 (16.7–31)	35.4 (32.5–41.7)	40 (4)
% time CGM > 250	0 (0–0.6)	5.8 (3–11.9)	42 (10)
% time CGM 180–250 mg/dl	5.7 (2–25)	26 (19.5–30.4)	
% time CGM 70–180 mg/dl	87.6 (63.2–96.8)	62 (47.5–68.9)	54 (9)
% time CGM 54–70	0.5 (0–3.2)	2.9 (2.2–4.2)	
% time CGM < 54 mg/dl	0 (0–0.1)	1.6 (0.7–3)	3.9 (1.7–5.3)
Insulin			
Basal (U/day)	23 (16.9–32.5)	22.1 (15.9–31.6)	–
Basal (U/kg)	0.33 (0.23–0.44)	0.32 (0.22–0.42)	0.32 (0.10)
Bolus (U/day)	18.7 (14.4–26.2)	18.7 (14.4–26.2)	–
Bolus (U/kg)	0.29 (0.18–0.37)	0.29 (0.18–0.37)	0.39 (0.13)
Total (U/day)	45.2 (32.3–55.9)	44.2 (31.3–55)	–
Total (U/kg)	0.58 (0.47–0.81)	0.57 (0.46–0.79)	0.71 (0.19)

^aValues are reported as median, interquartile range (25th, 75th percentile).

^bValues are reported as mean, SD.

targets for CGM presented in [4], it can be noted that the TIR has a median of 62% which is less than the 70% that represents the target value. In the case of time in hypoglycemia and hyperglycemia the target value is 5% and 25%, respectively and the results obtained have a median of 4.5% and 31.8%, respectively. The glycemic variability has a median of 35.4% and the target value is $\leq 33\%$.

Tables 3 and 4 present the results for MDI therapy with insulin glargine. Comparing the results with the clinical trial it is also confirmed for this therapy that the inclusion of mixed meals and circadian variability causes a more realistic behavior in VPs. The in silico TIR for Gla-100 and Gla-300 was 54% vs. 57.1% in the morning and 65% vs.

60.3% at night, very similar to the values obtained of 53.6% and 56.2% respectively, in figure 1 of [47]. The hypoglycemic for Gla-100 6.2% and 5.4% and for Gla-300 6.1% and 6.8%. The glycemic variability for Gla-100 was greater than for Gla-300 in the morning, which coincides with the characteristics reported in clinical trials on these parameters for both analogs [37,47].

Table 5 shows the results obtained from the incorporation of insulin degludec into the Hovorka model. It was first simulated with a fixed dose of 0.4 U/kg or 0.6 U/kg as reported in [49–51] the adequacy of the 0.4 U/kg dose was confirmed as it was close to the normal daily dose of basal insulin observed. A lower glycemic variability can be observed

2. PREDICCIÓN PROBABILÍSTICA PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

Table 3
Outcomes for MDI therapy with insulin Gla-100, in silico and clinical trial [47,48].

Performance indicator	Injection time	Nominal Hovorka model ^a	Hovorka model with variability and mixed meals ^a	Clinical trial Gla-100 ^b
CGM (mg/dl)	9 h	137.5 (112.6–152.7)	166.8 (154.5–185.5)	152.1 (32.8) ^d
	20 h	117.4 (106.3–132.7)	151.3 (136.2–165.3)	169.3 ^c
%CV	9 h	25.7 (20–37.5)	40.4 (36–47)	–
	20 h	29.9 (22.8–40)	38.6 (32.9–45.3)	–
% time CGM > 250 mg/dl	9 h	0.1 (0–2.4)	12.9 (5.7–18.1)	16.1 ^c
	20 h	0 (0–1.3)	5.5 (2.9–12.5)	–
% time CGM 180–250 mg/dl	9 h	9.6 (3.1–22.2)	24.4 (18.7–30.3)	22.8 ^c
	20 h	6.9 (2–16.4)	19.7 (15.4–23.9)	–
% time CGM 70–180 mg/dl	9 h	78.8 (57.2–87.5)	54 (46.8–65.1)	53.6 ^c
	20 h	73.4 (61.1–89.8)	65 (52.8–73.9)	–
% time CGM 54–69 mg/dl	9 h	2.3 (0–9)	3.7 (2.3–4.6)	–
	20 h	5.1 (0.7–15.2)	3.8 (2.4–5.8)	8.2 ^c
% time CGM < 54 mg/dl	9h	0.2 (0–4)	2.5 (1.2–4)	–
	20 h	0.3 (0–3.7)	1.6 (0.8–3.6)	–
Insulin				
Basal (U/day)	9 h	39.9 (29.1–52.6)	30.7 (22.4–40.5)	–
	20 h	37.5 (28.8–50.9)	28.9 (22.1–39.2)	–
Basal (U/kg)	9 h	0.55 (0.41–0.7)	0.42 (0.32–0.54)	0.45 (0.19) ^d
	20 h	0.53 (0.39–0.66)	0.41 (0.3–0.51)	0.36 (0.16) ^d
Bolus (U/day)	9 h	24.3 (17.5–34.9)	24.3 (17.5–34.9)	–
	20 h	24.9 (19.3–36.4)	24.9 (19.3–36.4)	–
Total (U/day)	9 h	66.2 (47–84.1)	59 (40.5–71.8)	–
	20 h	66.4 (48.2–83.9)	58.7 (42–73.3)	–
Total (U/kg)	9 h	0.91 (0.71–1.11)	0.77 (0.61–0.94)	0.73 (0.27) ^d
	20 h	0.95 (0.72–1.12)	0.81 (0.6–1)	0.63 (0.18) ^c

^aValues are reported as median, interquartile range (25th, 75th percentile).

^bValues are reported as mean, SD.

^c[47].

^d[48].

Table 4
Outcomes for MDI therapy with insulin Gla-300, in silico and clinical trial [47,48].

Performance indicator	Injection time	Nominal Hovorka model ^a	Hovorka model with variability and mixed meals ^a	Clinical trial Gla-300 ^b
CGM (mg/dl)	9 h	128.8 (109–139.5)	161.7 (147.9–176)	158 (38) ^d
	20 h	123.2 (106.6–139.5)	154.2 (141.1–173.2)	165.0 ^c
%CV	9 h	26.1 (19–34.1)	37.9 (33.5–46.8)	–
	20 h	26.8 (20.4–37.5)	39.4 (35.2–47.5)	–
% time CGM > 250 mg/dl	9 h	0 (0–0.3)	7.8 (4–13.2)	13.6 ^c
	20 h	0 (0–0.5)	7.5 (3.6–13.2)	–
% time CGM 180–250 mg/dl	9 h	6.2 (1–15.4)	23.3 (18–30.1)	24 ^c
	20 h	5.9 (0.8–18.2)	22.4 (17.3–27)	–
% time CGM 70–180 mg/dl	9 h	82.9 (62.1–93.7)	57.1 (49.1–64.4)	56.2 ^c
	20 h	78.2 (54.3–92.8)	60.3 (47.7–69)	–
% time CGM 54–69 mg/dl	9 h	3.7 (0.1–9.3)	3.7 (2.4–5)	–
	20 h	3 (0–7.4)	4.1 (2.3–5.4)	–
% time CGM < 54 mg/dl	9 h	0.2 (0–2.8)	2.4 (1.1–5.1)	7.3 ^c
	20 h	0.1 (0–5.5)	2.7 (0.8–4.7)	–
Insulin				
Basal (U/day)	9 h	42.4 (31.7–54.4)	32.6 (24.4–41.8)	–
	20 h	34.6 (26.2–48.1)	26.6 (20.2–37)	–
Basal (U/kg)	9 h	0.59 (0.43–0.72)	0.45 (0.33–0.56)	0.49 (0.22) ^d
	20 h	0.49 (0.36–0.64)	0.37 (0.27–0.49)	0.45 (0.21) ^d
Bolus (U/day)	9 h	23.8 (17–32.4)	23.8 (17–32.4)	–
	20 h	27 (19.3–39.7)	27 (19.3–39.7)	–
Total (U/day)	9 h	71 (48.8–85.9)	62.2 (41.5–71.5)	–
	20 h	67.8 (46.2–85.3)	57.2 (40.3–73.2)	–
Total (U/kg)	9 h	0.96 (0.75–1.13)	0.8 (0.63–0.96)	0.81 (0.32)
	20 h	0.92 (0.72–1.14)	0.78 (0.61–0.98)	0.67 (0.23) ^c

^aValues are reported as median, interquartile range (25th, 75th percentile).

^bValues are reported as mean, SD.

^c[47].

^d[48].

2.1. SIMULADOR CON ESCENARIOS REALISTAS Y DESAFIANTES PARA VPS CON T1D QUE SE SOMETEN A TERAPIA CSII Y MDI

Table 5
Outcomes for MDI therapy with Deg-100.

Performance indicator	Injection time	In silico-Basal insulin of 0.4 U/kg ^a	In silico- Basal insulin of 0.6 U/kg ^a	Individually adjusted basal insulin ^a
CGM (mg/dl)	9 h	170.1 (151.4–187.9)	158.4 (142.7–169.2)	166.3 (153.6–198.1)
	20 h	165.8 (147.8–191.1)	150.3 (141–171.6)	171.4 (153–191.5)
%CV	9 h	45 (34.5–49.2)	39.8 (35.9–45.2)	41.2 (35.9–47.5)
	20 h	38.1 (33.2–45.2)	36.2 (32.9–41.7)	37.8 (34.3–42)
% time CGM > 250 mg/dl	9 h	15.4 (8.7–21.8)	8.8 (3.5–13.4)	13.4 (7.2–26.6)
	20 h	12.9 (5.7–19.8)	5.4 (2–10.2)	11.2 (6.4–26.7)
% time CGM 180–250 mg/dl	9 h	24.8 (19.5–28.2)	23.5 (17.3–30.4)	25.2 (19.3–29.1)
	20 h	26.6 (19–32.3)	20.8 (17.4–27.4)	30.7 (23–35)
% time CGM 70–180 mg/dl	9 h	48.4 (41.4–56.1)	56.4 (46.1–67.4)	52.3 (39.6–61.6)
	20 h	49.8 (41.4–65.1)	61.8 (49.4–69.5)	45.8 (38.2–54.4)
% time CGM 54–69 mg/dl	9 h	4 (2–5.9)	4.1 (3.1–5.5)	3.5 (2.2–5)
	20 h	3.1 (1.3–5.4)	4 (2.1–5.5)	3.4 (1.9–4.7)
% time CGM < 54 mg/dl	9 h	2.9 (1.1–6.9)	3.7 (1.8–5.1)	2.4 (0.8–4.9)
	20 h	1.5 (0.4–4.3)	1.9 (1–4.2)	2.6 (0.7–4.6)
Insulin				
Basal (U/kg)	9 h	0.4 (0.4–0.4)	0.6 (0.6–0.6)	0.39 (0.31–0.55)
	20 h	0.4 (0.4–0.4)	0.6 (0.6–0.6)	0.47 (0.35–0.6)
Bolus (U/kg)	9 h	0.4 (0.27–0.62)	0.34 (0.2–0.51)	0.38 (0.28–0.6)
	20 h	0.41 (0.28–0.62)	0.33 (0.21–0.53)	0.36 (0.26–0.53)
Total (U/kg)	9 h	0.8 (0.67–1.02)	0.94 (0.8–1.11)	0.85 (0.64–1.06)
	20 h	0.81 (0.68–1.02)	0.93 (0.81–1.13)	0.87 (0.68–1.06)

^aValues are reported as median, interquartile range (25th, 75th percentile).

in the nocturnal injection, 38.1 (33.2–45.2)% and 36.2 (32.9–41.7)% for 0.4 U/kg and 0.6 U/kg, respectively. In addition, a basal insulin adjustment was made and 0.39 (0.31–0.55) U/kg and 0.47 (0.35–0.6) U/kg were obtained at 9:00 h and 20:00 h, respectively, which is similar to that reported in the literature. Mean glucose, TIR, or insulin requirements are presented in [52–55]. Particularly in [52], patients who received either insulin glargine or insulin detemir twice daily were instead given insulin degludec. The median results obtained were 166.3 (153.6–198.1) mg/dl vs. 171.4 (153–191.5) mg/dl for 9:00 h vs. 20:00 h in silico and 168.6 (±23.9) mg/dl in vivo [52]. For a basal insulin of 0.39 (0.31–0.55) U/kg vs. 0.47 (0.35–0.6) U/kg in silico and 0.4 (±0.2) U/kg in vivo. The percentage of time in the range 70–180 mg/dl obtained was 52.3 (39.6–61.6)% vs. 45.8 (38.2–54.4)% in silico and 57.5 (±13.3)% in vivo (Table 3 of [52]). There were also similarities in glycemic variability of 41.2 (35.9–47.5)% vs. 37.8 (34.3–42)% in silico and 35.9 (±6.4)% in vivo (See Table 5).

4. Discussion

Treating diabetes remains a challenge today and although insulin pumps and CSII are gaining ground, MDI therapy remains the most popular treatment for T1D with slow-acting insulin formulations used to satisfy insulin daily needs [36,56]. Simulators are an indispensable tool to test different treatments of CSII and MDI and therefore, more challenging scenarios are required.

This paper presents a methodology to generate a physiologically valid VP cohort. The simulation tools referenced in this paper have a CHO absorption model to represent the action of meals and a mixed meal library that allows the simulation of a variety of realistic meals, which represents a greater challenge when testing control strategies. Regarding the variability of insulin sensitivity, it is generally represented with a sinusoidal function of the parameters that influence this phenomenon. In this paper, we add circadian patterns of insulin sensitivity identified through basal insulin patterns reported in clinical trials. The addition of the mixed meal library and circadian variability results in outcomes that more accurately reflect real-life conditions. This is shown in Tables 1, 2, 3, and 4 where it can be seen that these phenomena make the VPs more realistic compared to the nominal Hovorka model. Another advantageous feature of this work is that it

allows the simulation of both CSII and MDI therapies within a single simulation tool.

To validate the results obtained with CSII therapy, they were compared with trials reported in the literature. In [30], the authors conducted a 12-week study with patients of both genders and of average age of 21 years. The study outcomes report a mean glucose of 174 mg/dl, slightly higher when compared to the median of 154.3 mg/dl from our in silico simulation. However, the insulin requirement obtained in both cases is noticeably similar, 0.32 U/kg vs. 0.33 U/kg reported in [57]. Glucose control in [30] was slightly better for VPs, which is reflected in the differences of 54% vs. 62% in normoglycemia, 3.9% vs. 4.5% in hypoglycemia and, 42% vs. 31.8% in hyperglycemia. This may be caused by the fact that patients were free to eat any food and do any physical activity at any time in the clinical trial unlike our simulation protocol.

The incorporation of the insulin glargine and degludec into the Hovorka model allows for the simulation of MDI therapies. The results were validated by comparing them to those reported in clinical trials, [31,32,47,48] for glargine and [49–55,58–61] for degludec. The behavior of the plasma insulin concentration (Figs. 6 and 7) was similar to those reported in the literature for both analogs.

In [48], a multicenter, randomized, four-arm, parallel-group study with T1D including 275 patients for Gla-100 and 274 for Gla-300 was conducted. Morning injection of insulin glargine time was between prebreakfast and prelunch (inclusive) and evening at the diner until bedtime, while participants continued using fast acting insulin to cover meals. The insulin requirements, reported as mean (SD) for Gla-100 was 0.45 (±0.19) U/kg in the morning and 0.36 (±0.16) U/kg at night. For Gla-300 was 0.49 (±0.22) U/kg in the morning and 0.45 (±0.21) U/kg at night. These results are similar to those obtained our simulations, reported as median (25th, 75th percentile), for Gla-100 0.42 (0.32–0.54) U/kg in the morning and 0.41 (0.3–0.51) U/kg at night. For Gla-300 was 0.45 (0.33–0.56) U/kg in the morning and 0.37 (0.27–0.49) U/kg at night. The mean glucose was 152.1(±32.8) and 158.8(±38) mg/dl for Gla-100 and Gla-300, respectively, in vivo, remarkably comparable with the results obtained in silico of 166.8 (154.5–185.5) mg/dl and 161.7 (147.9–176) mg/dl for Gla-100 and Gla-300.

Degludec is an ultra-long-acting insulin analog with a flat and reproducible pharmacodynamic profile. In [52], 29 patients with a

2. PREDICCIÓN PROBABILÍSTICA PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

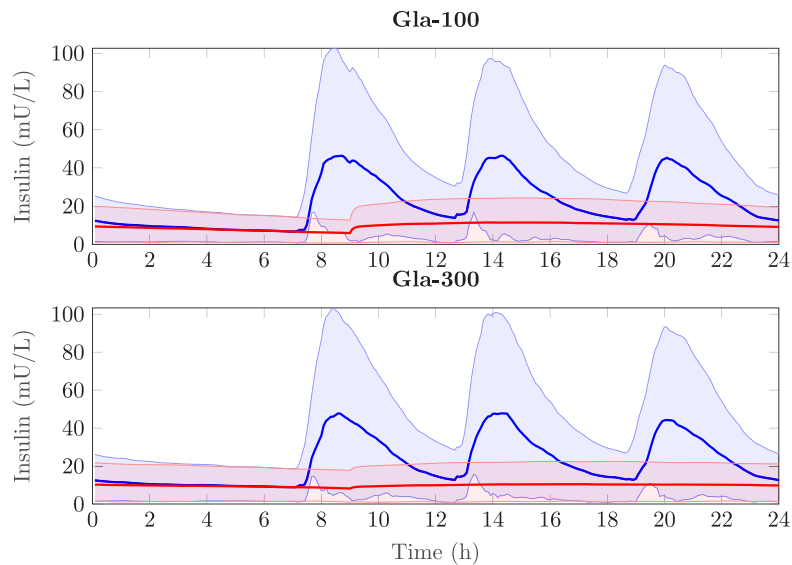


Fig. 6. Median (IQR) of plasma insulin concentration total (blue), Gla-100, and Gla-300 (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

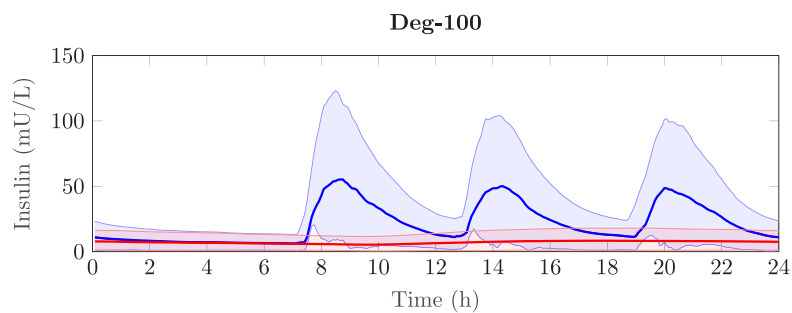


Fig. 7. Median (IQR) of plasma insulin concentration total (blue) and Deg (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

mean age of 34.8 (± 11.4) years with T1D who did not achieve 24 h coverage with glargine or detemir were chosen and the effect of changing the treatment to degludec was studied. The resulting basal insulin was 0.39 (0.31–0.55) vs. 0.47 (0.35–0.6) U/kg in our simulations and 0.3 (± 0.1) U/kg in the clinical trial after 12 weeks. The TIR obtained was 52.3 (39.6–61.6)% vs. 45.8 (38.2–54.4)% in silico and 57.5 (± 13.3)% in vivo. The times above 180 mg/dl were 38.8 (± 14.4) in vivo and 25.2 (19.3–29.1)% in the range 180–250 mg/dl vs. 13.4 (7.2–26.6)% > 250 mg/dl in silico (See Table 3 of [52]). The hypoglycemia time provided in [52] is defined by time below < 70 mg/dl and was 3.6 (± 3.7)% and for our simulations it was 5.9 (3–9.9)%.

It should be noted that there are some differences in the results of clinical trials and in silico results. This is mainly because the simulation does not include any disturbances related to free-living conditions such as physical activity, stress, or hormonal changes such as the menstrual cycle. Another factor that influences the differences found are the different eating habits of each patient, which is not taken into account in the simulator. Furthermore, the same clinical attributes of the study and the insulin titration algorithms were not implemented. However, the in silico study showed similar trends to those obtained in clinical studies, especially regarding insulin requirements.

This study presents an approach that improves upon current simulation tools, yielding results that more closely reflect real-life conditions. However, more research is still required to improve the simulators available in the literature. The main limitations of this study are given by: (1) although 24 h patterns of variability of insulin sensitivity are identified and implemented into the simulator, a model that represents it is not defined, and (2) glucagon, a hormone that is becoming more important for the treatment of T1D, was not included in the experiments.

5. Conclusion

The inclusion of the identified insulin sensitivity variability and a mixed meal library into a T1D simulator provides challenging scenarios that is better able to mimic real-life behavior, an important feature for evaluating treatments through in silico tests. The incorporation of the pharmacokinetic model of long-acting insulin glargine and degludec to the Hovorka model reproduces the main characteristics observed in the literature for these analogs, which demonstrates its validity for the testing of MDI therapies.

2.1. SIMULADOR CON ESCENARIOS REALISTAS Y DESAFIANTES PARA VPS CON T1D QUE SE SOMETEN A TERAPIA CSII Y MDI

CRedit authorship contribution statement

Ernesto Estremera: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Funding acquisition. **Alvis Cabrera:** Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Funding acquisition. **Aleix Beneyto:** Conceptualization, Methodology, Software, Investigation, Resources, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Josep Vehi:** Conceptualization, Methodology, Software, Investigation, Resources, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding and support acknowledgments

This work was partially supported by the Spanish Ministry of Science and Innovation through grant PID2019-107722RB-C22, in part by the Autonomous Government of Catalonia, Spain under Grant 2017 SGR 1551, and program for researchers in training at the University of Girona (IFUG2019) (principal investigators: MSc. Ernesto Estremera and MSc. Alvis Cabrera).

Appendix A. Model equations

A.1. Glucose subsystem of the Hovorka model

$$\frac{dQ_1}{dt} = -\left[\frac{F_{01}^c}{V_G G(t)} + x_1(t)\right] + k_{12}Q_2(t) - F_R + U_g(t) + EGP_0[1 - x_3(t)] \quad (\text{A.1})$$

$$\frac{dQ_2}{dt} = x_1(t)Q_1(t) - [k_{12} + x_2(t)]Q_2(t) \quad (\text{A.2})$$

$$G(t) = \frac{Q_2(t)}{V_G} \quad (\text{A.3})$$

$$F_{01}^c = \begin{cases} F_{01} & \text{if } G \geq 4.5 \text{ mmolL}^{-1} \\ \frac{F_{01}G}{4.5} & \text{otherwise} \end{cases} \quad (\text{A.4})$$

$$F_R = \begin{cases} 0.003(G - R_{th})V_G & \text{if } G \geq R_{th} \text{ mmolL}^{-1} \\ 0 & \text{otherwise} \end{cases} \quad (\text{A.5})$$

A.2. Insulin action subsystem of the Hovorka model

$$\frac{dx_1}{dt} = -k_{a1}x_1(t) + k_{a1}S_{IT}(t) \quad (\text{A.6})$$

$$\frac{dx_2}{dt} = -k_{a2}x_2(t) + k_{a2}S_{ID}(t) \quad (\text{A.7})$$

$$\frac{dx_3}{dt} = -k_{a3}x_3(t) + k_{a3}S_{IE}(t) \quad (\text{A.8})$$

A.3. Insulin absorption subsystem of the Hovorka model

$$\frac{dS_1}{dt} = u(t) - k_a S_1(t) \quad (\text{A.9})$$

$$\frac{dS_2}{dt} = k_a S_1(t) - k_a S_2(t) \quad (\text{A.10})$$

$$\frac{dI}{dt} = \frac{k_a S_2(t)}{V_I} - k_e I(t) \quad (\text{A.11})$$

A.4. Gut absorption subsystem of the Hovorka model

$$\frac{dG_1}{dt} = -\frac{G_1(t)}{t_{maxG}} + Bio * D(t) \quad (\text{A.12})$$

$$\frac{dG_2}{dt} = \frac{G_1(t)}{t_{maxG}} - \frac{G_2(t)}{t_{maxG}} \quad (\text{A.13})$$

$$U_g(t) = \frac{G_2(t)}{t_{maxG}} \quad (\text{A.14})$$

A.5. Glargine insulin model

$$\frac{dI_{q1}}{dt} = -k_{sp}I_{q1}(t) + k * F * u_{ia}(t) \quad (\text{A.15})$$

$$\frac{dI_{q2}}{dt} = -k_a I_{q2}(t) + k_{sp}I_{q1}(t) + (1 - k) * F * u_{ia}(t) \quad (\text{A.16})$$

$$R_{ai}(t) = k_a I_{q2}(t) \quad (\text{A.17})$$

A.6. Degludec insulin model

$$\frac{dI_{q1}}{dt} = -k_{d1}I_{q1}(t) + F * D \quad (\text{A.18})$$

$$\frac{dI_{q2}}{dt} = -k_{d2}I_{q2}(t) + k_{d1}I_{q1}(t) \quad (\text{A.19})$$

$$\frac{dI_{q3}}{dt} = -k_a I_{q3}(t) + k_{d2}I_{q2}(t) \quad (\text{A.20})$$

$$R_{ai}(t) = k_a I_{q3}(t) \quad (\text{A.21})$$

A.7. Insulin subsystem model glargine and degludec

$$\frac{dI_p}{dt} = -(m_2 + m_4)I_p(t) + m_1 I_t(t) + R_{ai}(t) \quad (\text{A.22})$$

$$\frac{dI_t}{dt} = -(m_1 + m_3)I_t(t) + m_2 I_p(t) \quad (\text{A.23})$$

$$I(t) = \frac{I_p(t)}{V_I} \quad (\text{A.24})$$

Appendix B. Procedures to generate the circadian variability of insulin sensitivity

B.1. Procedure for generation of basal insulin profile

Probability distributions are formulated how $N(\mu, \sigma^2)$ where μ represents the mean and σ is the standard deviation. For the generation of basal insulin profile the following terms are introduced.

1. The number of daily insulin basal rates $n_r \in \mathbb{N}^+$ is sampled and rounded to the nearest integer following the normal distribution $N(5.04, 3.35)$.
2. The array of insulin basal rates $\mathbf{u} = \{u_{np} \cup u_p\} \in \mathbb{R}^{n_r}$ in U/h. Each basal insulin profile contains one peak basal insulin infusion, u_p , which is the highest infusion throughout the day and an array of non-peak basal infusions, $u_{np} \in \mathbb{R}^{(n_r-1)}$. Here, the peak basal rate is sampled from $N(0.92, 0.13)$ U/h. Then, the non-peak basal rates can be computed taking into account that the average of u_{np} is at least 25% lower than u_p [42], as shown in Eq. (B.1).

$$\frac{u_{np}}{n_r} = \frac{\sum \mathbf{u} - u_p}{n_r - 1} < \frac{u_p}{1 + \frac{v}{100}} \quad (\text{B.1})$$

2. PREDICCIÓN PROBABILÍSTICA PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

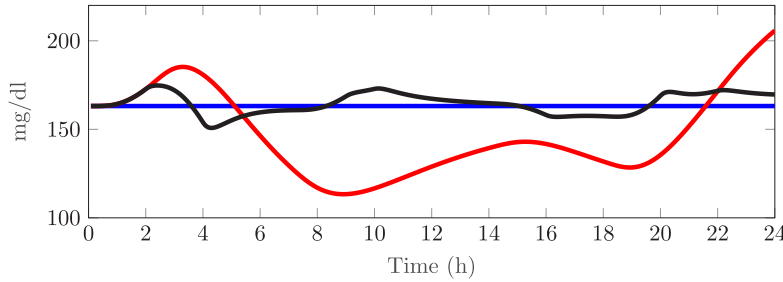


Fig. B.8. Representative blood glucose concentration $Q_1(t)/V_g$. The blue curve is the Hovorka model response with $\alpha = 1$ and using u_p , the red curve is the Hovorka model response with $\alpha = 1$ and using \bar{u}_b and the black curve is the Hovorka model response with the obtained $\alpha(t)$ and using \bar{u}_b . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Where $v \sim N(30, 2.25)\%$ is the variation percentage that ensures, at least with a 99.7% probability, that \bar{u}_{np} will be a 25% lower than u_p . After that, we randomly generate $u_{np} = \left\{ u \in (0.4u_p, 0.8u_p) \forall u \in u_{np} \mid \bar{u}_{np} = \frac{v}{100} * u_p \right\}$ with $u_{np} \in \mathbb{R}^{(n_r-1)}$.

- The array of the duration of each basal rate $\mathbf{d} = \{d_{u_{np}} \cup d_{u_p}\} \in \mathbb{R}^{n_r}$ in h . The duration of the peak basal insulin, d_{u_p} , follows $N(5.2, 11.29)$ hours. Then, we randomly generate the non-peak basal rates duration $d_{u_{np}} = \left\{ \forall d \in d_{u_{np}} \mid \sum d_{u_{np}} = 24 - d_{u_p} \right\}$ with $d_{u_{np}} \in \mathbb{R}^{(n_r-1)}$.

Once generated \mathbf{u} and \mathbf{d} each basal rate is matched to its corresponding duration. To generate the basal insulin profile each glucose rate is distributed in the 24 h. We know that u_p is distributed 80% of the times between 2:00 h–8:00 h and 20% time CGM between 20:00 h–2:00h [42]. The d_{u_p} is then centered at either 5:00 h or 24:00 h. After that, $d_{u_{np}}$ is concatenated to the ending time of the up basal rate.

B.2. Procedure for $\alpha(t)$ generation

$\alpha(t)$ is the day hourly-varying factor affecting the model's insulin sensitivity. To obtain the array α , the steady state of a VP is solved using a generated basal profile \mathbf{u} . Then, a maximization problem is iteratively solved, hour by hour, such that the daily steady state glucose remains in a 30 mg/dl band.

$$\begin{aligned} \min \quad & -\alpha \\ \text{s.t.} \quad & f(\mathbf{x}, \bar{\mathbf{u}}_b, \alpha) = 0 \\ & \frac{Q_1}{V_g} \in \left[\frac{-0.3Q_{1,basal}}{V_g}, \frac{0.3Q_{1,basal}}{V_g} \right] \\ & \bar{\mathbf{u}}_b = u_b \frac{\mathbf{u}}{u} \end{aligned} \quad (\text{B.2})$$

Where $\mathbf{x} \in \mathbb{R}^8$ is the state vector of the Hovorka model, $f(\mathbf{x}, \bar{\mathbf{u}}_b, \alpha)$ represents the nonlinear Hovorka model with the time-varying α factor, $Q_{1,basal}$ is the steady state value of Q_1 when u_b is used, u_b is the basal insulin obtained when solving for $f(\mathbf{x}, u_b) = 0$ with $Q_1/V_g = 160$ mg/dl and a constant $\alpha = 1$, \mathbf{u} is a basal pattern generated using the previous described methodology, $\bar{\mathbf{u}}$ is the average basal insulin of \mathbf{u} and $\bar{\mathbf{u}}_b$ is the basal insulin pattern scaled based on the VP original basal insulin u_b . Basically, we introduce the scaled basal insulin to the model to provoke variability and then we solve for α such that the variability in the glucose concentration state is physiologically plausible. Fig. B.8 shows the different steps involved. First, the blue trajectory shows the base steady state of the Hovorka model without variability. Then, using one of the scaled generated basal profiles we show in the red curve that the variability on glucose is significantly more complex compared to the original case. Then, we translate this effect into the α time-varying parameter such that glucose concentration is constrained to a ± 30 mg/dl band [42].

Table C.6

Parameters	P (Period)	φ (Phase)
EGP_0	61	0.1152
F_{01}	71	0.9324
k_{12}	59	0.2936
t_{maxI}	31	0.5737
k_c	37	0.8423
m_4	41	0.7567
k_6	39	0.1491

Appendix C. Procedures to generate $\beta(t)$

Intra-patient variability is introduced in model parameters by applying the following product:

$$p(t) = p_0 \beta(t) \quad (\text{C.1})$$

where $p(t)$ represents the new time-varying parameter, p_0 represents the constant nominal parameter from the VP and $\beta(t)$ is a time-varying sinusoidal function of the form (C.2).

$$\beta(t) = 1 + \frac{A}{100} \sin\left(\frac{2\pi}{P \cdot 60} t + 2\pi\varphi\right) \quad (\text{C.2})$$

where the period P , in hours, is different for each parameter and is represented by prime numbers randomly assigned between 31 and 71 to avoid periodicity between the parameters, the phase φ is randomly selected between 0 and 1 (Table C.6 shows P and φ selected for each parameter), the amplitude A was set to 20 following [62]. A total of 7 model parameters can include intra-patient circadian variability.

Appendix D. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jbi.2022.104141>.

References

- H. Sun, P. Saeedi, S. Karuranga, M. Pinkepank, K. Ogurtsova, B.B. Duncan, C. Stein, A. Basit, J.C. Chan, J.C. Mbanya, et al., IDF diabetes atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045, *Diabetes Res. Clin. Pract.* (2021) 109119.
- E.R.F. Collaboration, et al., Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies, *Lancet* 375 (9733) (2010) 2215–2222.
- J.D. Steinmetz, R.R. Bourne, P.S. Briant, S.R. Flaxman, H.R. Taylor, J.B. Jonas, A.A. Abdoli, W.A. Abrha, A. Abualhasan, E.G. Abu-Gharbieh, et al., Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the right to sight: an analysis for the global burden of disease study, *Lancet Glob. Health* 9 (2) (2021) e144–e160.
- A.D. Association, Standards of medical care in diabetes—2022 abridged for primary care providers, *Clin. Diabetes* 40 (1) (2022) 10–38.

2.1. SIMULADOR CON ESCENARIOS REALISTAS Y DESAFIANTES PARA VPS CON T1D QUE SE SOMETEN A TERAPIA CSII Y MDI

- [5] R.N. Bergman, Y.Z. Ider, C.R. Bowden, C. Cobelli, Quantitative estimation of insulin sensitivity., *Am. J. Physiol. Endocrinol. Metabol.* 236 (6) (1979) E667.
- [6] J.T. Sorensen, A Physiologic Model of Glucose Metabolism in Man and Its Use to Design and Assess Improved Insulin Therapies for Diabetes (Ph.D. thesis), Massachusetts Institute of Technology, 1985.
- [7] C. Dalla Man, R.A. Rizza, C. Cobelli, Meal simulation model of the glucose-insulin system, *IEEE Trans. Biomed. Eng.* 54 (10) (2007) 1740–1749.
- [8] R. Hovorka, V. Canonico, L.J. Chassin, U. Haueter, M. Massi-Benedetti, M.O. Federici, T.R. Pieber, H.C. Schaller, L. Schaupp, T. Vering, et al., Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes, *Physiol. Meas.* 25 (4) (2004) 905.
- [9] B.P. Kovatchev, M. Breton, C. Dalla Man, C. Cobelli, In Silico Preclinical Trials: a Proof of Concept in Closed-Loop Control of Type 1 Diabetes, SAGE Publications Sage CA, Los Angeles, CA, 2009.
- [10] R. Visentin, C. Dalla Man, C. Cobelli, One-day Bayesian cloning of type 1 diabetes subjects: toward a single-day UVA/Padova type 1 diabetes simulator, *IEEE Trans. Biomed. Eng.* 63 (11) (2016) 2416–2424.
- [11] R. Visentin, E. Campos-Náñez, M. Schiavon, D. Lv, M. Vettoretti, M. Breton, B.P. Kovatchev, C. Dalla Man, C. Cobelli, The UVA/Padova type 1 diabetes simulator goes from single meal to single day, *J. Diabetes Sci. Technol.* 12 (2) (2018) 273–281.
- [12] N. Resalat, J. El Youssef, N. Tyler, J. Castle, P.G. Jacobs, A statistical virtual patient population for the glucoregulatory system in type 1 diabetes with integrated exercise model, *PLoS One* 14 (7) (2019) e0217301.
- [13] A. Bertachi, L. Biagi, A. Beneyto, J. Vehí, Dynamic rule-based algorithm to tune insulin-on-board constraints for a hybrid artificial pancreas system, *J. Healthc. Eng.* 2020 (2020).
- [14] A. Beneyto, A. Bertachi, J. Bondia, J. Vehí, A new blood glucose control scheme for unannounced exercise in type 1 diabetic subjects, *IEEE Trans. Control Syst. Technol.* 28 (2) (2018) 593–600.
- [15] A. Bertachi, C.M. Ramkissoon, A. Beneyto, J. Vehí, Exercise-induced hypoglycemia in type 1 diabetes: in-silico comparison between announced and unannounced strategies in closed-loop control, *IFAC-PapersOnLine* 52 (1) (2019) 1000–1005.
- [16] O. Orozco-López, A. Rodríguez-Herrero, C.E. Castañeda, G. García-Sáez, M.E. Hernando, Method to generate a large cohort in-silico for type 1 diabetes, *Comput. Methods Programs Biomed.* 193 (2020) 105523.
- [17] C.M. Ramkissoon, A. Bertachi, A. Beneyto, J. Bondia, J. Vehí, Detection and control of unannounced exercise in the artificial pancreas without additional physiological signals, *IEEE J. Biomed. Health Inf.* 24 (1) (2019) 259–267.
- [18] M. Schiavon, R. Visentin, C. Giegerich, J. Sieber, C. Dalla Man, C. Cobelli, T. Klabunde, In silico head-to-head comparison of insulin glargine 300 U/mL and insulin degludec 100 U/mL in type 1 diabetes, *Diabetes Technol. Ther.* 22 (8) (2020) 553–561.
- [19] N.F.M. Yusof, A.M. Som, An overview of recent development on artificial pancreas.
- [20] C.D. Man, F. Micheletto, D. Lv, M. Breton, B. Kovatchev, C. Cobelli, The UVA/PADOVA type 1 diabetes simulator: new features, *J. Diabetes Sci. Technol.* 8 (1) (2014) 26–34.
- [21] M.E. Wilinska, L.J. Chassin, C.L. Acerini, J.M. Allen, D.B. Dunger, R. Hovorka, Simulation environment to evaluate closed-loop insulin delivery systems in type 1 diabetes, *J. Diabetes Sci. Technol.* 4 (1) (2010) 132–144.
- [22] M. Rashid, S. Samadi, M. Sevil, I. Hajizadeh, P. Kolodziej, N. Hobbs, Z. Maloney, R. Brandt, J. Feng, M. Park, et al., Simulation software for assessment of non-linear and adaptive multivariable control algorithms: glucose-insulin dynamics in type 1 diabetes, *Comput. Chem. Eng.* 130 (2019) 106565.
- [23] M.R. Smaoui, R. Rabasa-Lhoret, A. Haidar, Development platform for artificial pancreas algorithms, *PLoS One* 15 (12) (2020) e0243139.
- [24] M.E. Wilinska, R. Hovorka, Simulation models for in silico testing of closed-loop glucose controllers in type 1 diabetes, *Drug Discov. Today Dis. Models* 5 (4) (2008) 289–298.
- [25] D. Elleri, J. Allen, J. Harris, K. Kumareswaran, M. Nodale, L. Leelarathna, C. Acerini, A. Haidar, M. Wilinska, N. Jackson, et al., Absorption patterns of meals containing complex carbohydrates in type 1 diabetes, *Diabetologia* 56 (5) (2013) 1108–1117.
- [26] N. Camerlingo, M. Vettoretti, S. Del Favero, A. Facchinetti, G. Sparacino, Mathematical models of meal amount and timing variability with implementation in the type-1 diabetes patient decision simulator, *J. Diabetes Sci. Technol.* 15 (2) (2021) 346–359.
- [27] C. Toffanin, R. Visentin, M. Messori, F. Di Palma, L. Magni, C. Cobelli, Toward a run-to-run adaptive artificial pancreas: In silico results, *IEEE Trans. Biomed. Eng.* 65 (3) (2017) 479–488.
- [28] P. Colmegna, F. Garelli, H. De Battista, R. Sánchez-Peña, Automatic regulatory control in type 1 diabetes without carbohydrate counting, *Control Eng. Pract.* 74 (2018) 22–32.
- [29] M. Messori, G.P. Incredonza, C. Cobelli, L. Magni, Individualized model predictive control for the artificial pancreas: In silico evaluation of closed-loop glucose control, *IEEE Control Syst. Mag.* 38 (1) (2018) 86–104.
- [30] M. Tauschmann, H. Thabit, L. Bally, J.M. Allen, S. Hartnell, M.E. Wilinska, Y. Ruan, J. Sibayan, C. Kollman, P. Cheng, et al., Closed-loop insulin delivery in suboptimally controlled type 1 diabetes: a multicentre, 12-week randomised trial, *Lancet* 392 (10155) (2018) 1321–1329.
- [31] M.S. Ruiz-de Adana, M.-E. Dominguez-Lopez, I. Gonzalez-Molero, A. Machado, V. Martin, I. Cardona, M. de-la Higuera, M.-J. Tapia, F. Soriguer, M.T. Anarte, et al., Comparison between a multiple daily insulin injection regimen (basal once-daily glargine plus mealtime lispro) and continuous subcutaneous insulin infusion (lispro) using continuous glucose monitoring in metabolically optimized type 1 diabetes patients: a randomized open-labelled parallel study, *Medicina Clínica* 146 (6) (2016) 239–246.
- [32] G.B. Bolli, D. Kerr, R. Thomas, E. Torlone, A. Sola-Gazagnes, E. Vitacolonna, J.L. Selam, P.D. Home, Comparison of a multiple daily insulin injection regimen (basal once-daily glargine plus mealtime lispro) and continuous subcutaneous insulin infusion (lispro) in type 1 diabetes: a randomized open parallel multicenter study, *Diabetes Care* 32 (7) (2009) 1170–1176.
- [33] M. Somali, S.A. Paschou, Z. Mouslech, Insulin pumps use in Greece: Efficacy and safety data from 140 patients with type 1 diabetes mellitus, *Diabetes Res. Clin. Pract.* 160 (2020) 108026.
- [34] M. Berger, D. Rodbard, Computer simulation of plasma insulin and glucose dynamics after subcutaneous insulin injection, *Diabetes Care* 12 (10) (1989) 725–736.
- [35] M. Schiavon, R. Visentin, C. Giegerich, T. Klabunde, C. Cobelli, C. Dalla Man, Modeling subcutaneous absorption of long-acting insulin glargine in type 1 diabetes, *IEEE Trans. Biomed. Eng.* 67 (2) (2019) 624–631.
- [36] A. El Fathi, R.E. Kearney, E. Palisaitis, B. Boulet, A. Haidar, A model-based insulin dose optimization algorithm for people with type 1 diabetes on multiple daily injections therapy, *IEEE Trans. Biomed. Eng.* (2020).
- [37] R. Visentin, M. Schiavon, C. Giegerich, T. Klabunde, C. Dalla Man, C. Cobelli, Incorporating long-acting insulin glargine into the UVA/Padova type 1 diabetes simulator for in silico testing of MDI therapies, *IEEE Trans. Biomed. Eng.* 66 (10) (2019) 2889–2896.
- [38] R. Hovorka, F. Shojae-Moradie, P.V. Carroll, L.J. Chassin, L.J. Gowrie, N.C. Jackson, R.S. Tudor, A.M. Umpleby, R.H. Jones, Partitioning glucose distribution/transport, disposal, and endogenous production during IVGTT, *Am. J. Physiol. Endocrinol. Metabol.* 282 (5) (2002) E992–E1007.
- [39] L.J. Chassin, In Silico Testing of Glucose Controllers: Methodology and Sample Application (Ph.D. thesis), City University London, 2005.
- [40] J. Walsh, R. Roberts, Pumping Insulin: Everything You Need for Success on a Smart Insulin Pump, Vol. 4, Torrey Pines Press San Diego, CA, 2006.
- [41] F.M.L. Vargas, Design and Implementation of a Closed-Loop Blood Glucose Control System in Patients with Type 1 Diabetes (Ph.D. thesis), Universitat de Girona, 2013.
- [42] G. Scheiner, B.A. Boyer, Characteristics of basal insulin requirements by age and gender in type-1 diabetes patients using insulin pump therapy, *Diabetes Res. Clin. Pract.* 69 (1) (2005) 14–21.
- [43] P.M. Holterhus, R. Odendahl, S. Oesingmann, R. Lepler, V. Wagner, O. Hiort, R. Holl, G.D. Initiative, G.P.C.W. Group, et al., Classification of distinct baseline insulin infusion patterns in children and adolescents with type 1 diabetes on continuous subcutaneous insulin infusion therapy, *Diabetes Care* 30 (3) (2007) 568–573.
- [44] P. Herrero, J. Bondia, O. Adewuyi, P. Pesl, M. El-Sharkawy, M. Reddy, C. Toumazou, N. Oliver, P. Georgiou, Enhancing automatic closed-loop glucose control in type 1 diabetes with an adaptive meal bolus calculator—in silico evaluation under intra-day variability, *Comput. Methods Programs Biomed.* 146 (2017) 125–131.
- [45] R. Hu, C. Li, An improved PID algorithm based on insulin-on-board estimate for blood glucose control with type 1 diabetes, *Comput. Math. Methods Med.* 2015 (2015).
- [46] A.D. Association, et al., 6. Glycemic Targets: standards of medical care in diabetes—2020, *Diabetes Care* 43 (Supplement 1) (2020) S66–S76.
- [47] R.M. Bergenstal, T.S. Bailey, D. Rodbard, M. Ziemien, H. Guo, I. Muehlen-Bartmer, A.J. Ahmann, Comparison of insulin glargine 300 units/mL and 100 units/mL in adults with type 1 diabetes: continuous glucose monitoring profiles and variability using morning or evening injections, *Diabetes Care* 40 (4) (2017) 554–560.
- [48] P.D. Home, R.M. Bergenstal, G.B. Bolli, M. Ziemien, M. Rojeski, M. Espinasse, M.C. Riddle, New insulin glargine 300 units/mL versus glargine 100 units/mL in people with type 1 diabetes: a randomized, phase 3a, open-label clinical trial (EDITION 4), *Diabetes Care* 38 (12) (2015) 2217–2225.
- [49] T. Heise, M. Norskov, L. Nosek, K. Kaplan, S. Famulla, H.L. Haahr, Insulin degludec: Lower day-to-day and within-day variability in pharmacodynamic response compared with insulin glargine 300 U/mL in type 1 diabetes, *Diabetes Obes. Metab.* 19 (7) (2017) 1032–1039.
- [50] T. Heise, K. Kaplan, H.L. Haahr, Day-to-day and within-day variability in glucose-lowering effect between insulin degludec and insulin glargine (100 U/mL and 300 U/mL): a comparison across studies, *J. Diabetes Sci. Technol.* 12 (2) (2018) 356–363.

2. PREDICCIÓN PROBABILÍSTICA PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

E. Estremera et al.

Journal of Biomedical Informatics 132 (2022) 104141

- [51] T. Bailey, J. Pettus, R. Roussel, W. Schmider, M. Maroccia, N. Nassr, O. Klein, G. Bolli, R. Dahmen, Morning administration of 0.4 U/kg/day insulin glargine 300 U/mL provides less fluctuating 24-hour pharmacodynamics and more even pharmacokinetic profiles compared with insulin degludec 100 U/mL in type 1 diabetes, *Diabetes Metab.* 44 (1) (2018) 15–21.
- [52] S. Galasso, A. Facchinetti, B. Bonora, V. Mariano, F. Boscardi, E. Cipponeri, A. Maran, A. Avogaro, G. Fadini, D. Bruttomesso, Switching from twice-daily glargine or detemir to once-daily degludec improves glucose control in type 1 diabetes. An observational study, *Nutr. Metab. Cardiovasc. Dis.* 26 (12) (2016) 1112–1119.
- [53] D.C. Henao-Carrillo, O.M. Muñoz, A.M. Gómez, M. Rondón, C. Colón, L. Chica, C. Rubio, F. León-Vargas, M.A. Calvachi, A.M. Perea, Reduction of glycemic variability with degludec insulin in patients with unstable diabetes, *J. Clin. Transl. Endocrinol.* 12 (2018) 8–12.
- [54] R. Iga, H. Uchino, K. Kanazawa, S. Usui, M. Miyagi, N. Kumashiro, H. Yoshino, Y. Ando, T. Hirose, Glycemic variability in type 1 diabetes compared with degludec and glargine on the morning injection: an open-label randomized controlled trial, *Diabetes Ther.* 8 (4) (2017) 783–792.
- [55] S. Iwasaki, J. Kozawa, T. Kimura, K. Fukui, H. Iwahashi, A. Imagawa, I. Shimomura, Insulin degludec is associated with less frequent and milder hypoglycemia in insulin-deficient patients with type 1 diabetes compared with insulin glargine or detemir, *Diabetol. Int.* 8 (2) (2017) 228–236.
- [56] J.B. McGill, A. Ahmann, Continuous glucose monitoring with multiple daily insulin treatment: outcome studies, *Diabetes Technol. Ther.* 19 (S3) (2017) S–3.
- [57] M.I. Maiorino, G. Bellastella, O. Casciano, P. Cirillo, V. Simeon, P. Chiodini, M. Petrizzo, M. Gicchino, O. Romano, P. Caruso, et al., The effects of subcutaneous insulin infusion versus multiple insulin injections on glucose variability in young adults with type 1 diabetes: the 2-year follow-up of the observational METRO study, *Diabetes Technol. Ther.* 20 (2) (2018) 117–126.
- [58] A.J. Garber, A.B. King, S. Del Prato, S. Sreenan, M.K. Balci, M. Muñoz-Torres, J. Rosenstock, L.A. Endahl, A.M.O. Francisco, P. Hollander, et al., Insulin degludec, an ultra-longacting basal insulin, versus insulin glargine in basal-bolus treatment with mealtime insulin aspart in type 2 diabetes (BEGIN basal-bolus type 2): a phase 3, randomised, open-label, treat-to-target non-inferiority trial, *Lancet* 379 (9825) (2012) 1498–1507.
- [59] C. Mathieu, P. Hollander, B. Miranda-Palma, J. Cooper, E. Franek, D. Russell-Jones, J. Larsen, S.C. Tamer, S.C. Bain, N.-B.F.T.T. Investigators, Efficacy and safety of insulin degludec in a flexible dosing regimen vs insulin glargine in patients with type 1 diabetes (BEGIN: Flex T1): a 26-week randomized, treat-to-target trial with a 26-week extension, *J. Clin. Endocrinol. Metab.* 98 (3) (2013) 1154–1162.
- [60] F.C. San Laureano, M.T. Fernández-Ladreda, A.I.J. Millán, C.G. Calzado, M.d.C.A. Ortega, Insulin doses requirements in patients with type 1 diabetes using glargine U300 or degludec in routine clinical practice, *J. Invest. Med.* 69 (5) (2021) 983–988.
- [61] J. Rosenstock, A. Cheng, R. Ritzel, Z. Bosnyak, C. Devisme, A.M. Cali, J. Sieber, P. Stella, X. Wang, J.P. Frías, et al., More similarities than differences testing insulin glargine 300 units/ml versus insulin degludec 100 units/ml in insulin-naive type 2 diabetes: the randomized head-to-head BRIGHT trial, *Diabetes Care* 41 (10) (2018) 2147–2154.
- [62] F. León-Vargas, F. Garelli, H. De Battista, J. Vehí, Postprandial blood glucose control using a hybrid adaptive PD controller with insulin-on-board limitation, *Biomed. Signal Process. Control* 8 (6) (2013) 724–732.

2.2 Validación del modelo probabilístico de predicción para pacientes con T1D utilizando análisis de datos composicionales

En esta publicación, proponemos una metodología basada en datos composicionales para validar el modelo de transición probabilística presentado previamente (Biagi et al., 2021). Desarrollamos una metodología y una nueva métrica de exactitud que puede emplearse en cualquier enfoque cuyos datos sean composicionales, adaptamos las métricas de error comunes para calcular la precisión del modelo. Consideramos mediciones de CGM de una cohorte real de ocho pacientes adultos con T1D. La contribución del candidato para esta publicación consistió en el procesamiento de los datos obtenidos del ensayo clínico, creación de la metodología, redacción del manuscrito, contribución a la discusión y la edición del manuscrito a lo largo de las rondas de revisión. Durante el desarrollo del trabajo, la candidata trabajó con sus colegas Dr. Lyvia Biagi, Dr. Ernesto Estremera, Dr. Aleix Beneyto. También fueron asistidos por el Dr. Iván Contreras, Dra. Marga Giménez, Dr. Ignacio Conget, Dr. Jorge Bondia, Dr. Josep Vehí y Dr. Josep Antoni Martín-Fernández, quien brindó orientación para el análisis de la metodología de datos composicionales y contribuyó en la interpretación de los resultados.

Título: Validation of a Probabilistic Prediction Model for Patients with Type 1 Diabetes Using Compositional Data Analysis

Autores: **Alvis Cabrera**, Lyvia Biagi, Aleix Beneyto, Ernesto Estremera, Iván Contreras, Marga Giménez, Ignacio Conget, Jorge Bondia, Josep Antoni Martín-Fernández y Josep Vehí

Revista: Mathematics











Volúmen: 11, Número: 5, Páginas: 12-41, Publicado: Marzo 2023

DOI: <https://doi.org/10.3390/math11051241>

Índice de calidad: JCR: Q1 (21/333) en MATHEMATICS, JIF: 2.592 en 2021

Article

Validation of a Probabilistic Prediction Model for Patients with Type 1 Diabetes Using Compositional Data Analysis

Alvis Cabrera ¹, Lyvia Biagi ², Aleix Beneyto ¹, Ernesto Estremera ¹, Iván Contreras ¹, Marga Giménez ^{3,4}, Ignacio Conget ^{3,4}, Jorge Bondia ^{4,5}, Josep Antoni Martín-Fernández ⁶ and Josep Vehí ^{1,4,*}

- ¹ Department of Electrical, Electronic and Automatic Engineering, University of Girona, 17003 Girona, Spain
 - ² Campus Guarapuava, Federal University of Technology–Paraná (UTFPR), Guarapuava 85053-525, Brazil
 - ³ Diabetes Unit, Endocrinology and Nutrition Department, Hospital Clínic de Barcelona, 08036 Barcelona, Spain
 - ⁴ Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Instituto de Salud Carlos III, 28029 Madrid, Spain
 - ⁵ Instituto Universitario de Automática e Informática Industrial, Universitat Politècnica de València, 46022 València, Spain
 - ⁶ Department of Computer Science, Applied Mathematics and Statistics, University of Girona, 17003 Girona, Spain
- * Correspondence: josep.vehi@udg.edu

Abstract: Glycemia assessment in people with type 1 diabetes (T1D) has focused on the time spent in different glucose ranges. As this time reflects the relative contributions to the finite duration of a day, it should be treated as compositional data (CoDa) that can be applied to T1D data. Previous works presented a tool for the individual categorization of days and proposed a probabilistic transition model between categories, although validation has hitherto not been presented. In this study, we consider data from eight real adult patients with T1D obtained from continuous glucose monitoring (CGM) sensors and introduce a methodology based on compositional methods to validate the previously presented probability transition model. We conducted 5-fold cross-validation, with both the training and validation data being CoDa vectors, which requires developing new performance metrics. We design new accuracy and precision measures based on statistical error calculations. The results show that the precision for the entire model is higher than 95% in all patients. The use of a probabilistic transition model can help doctors and patients in diabetes treatment management and decision-making. Although the proposed method was tested with CoDa applied to T1D data obtained from CGM, the newly developed accuracy and precision measures apply to any other data or validation based on CoDa.

Keywords: compositional data; continuous glucose monitoring; prediction model; time in range; type 1 diabetes

MSC: 62H99



Citation: Cabrera, A.; Biagi, L.; Beneyto, A.; Estremera, E.; Contreras, I.; Giménez, M.; Conget, I.; Bondia, J.; Martín-Fernández, J.A.; Vehí, J. Validation of a Probabilistic Prediction Model for Patients with Type 1 Diabetes Using Compositional Data Analysis. *Mathematics* **2023**, *11*, 1241. <https://doi.org/10.3390/math11051241>

Academic Editor: Vasile Preda

Received: 18 January 2023

Revised: 1 March 2023

Accepted: 2 March 2023

Published: 4 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Diabetes mellitus is considered one of the chronic diseases, significantly impacting the quality of life of the world population and constituting a real health problem. It belongs to the group of diseases that cause physical disability because of its various multi-organ complications, and has undoubtedly led to an increase in morbidity and mortality in recent years [1]. Individuals with type 1 diabetes (T1D) rely on external insulin to regulate blood glucose (BG) levels, which can be delivered through multiple injections of insulin or continuous subcutaneous insulin infusion. The different characteristics of patients with T1D render it difficult for clinicians to adjust insulin doses to the patient's activities appropriately [2]. The integration of patient measurements into a decision support system

2.2. VALIDACIÓN DEL MODELO PROBABILÍSTICO DE PREDICCIÓN PARA PACIENTES CON T1D UTILIZANDO ANÁLISIS DE DATOS COMPOSICIONALES

could help clinicians in consultations or even guide the patient when using devices for insulin or carbohydrate administration [3]. Different authors presented an exhaustive review of decision support systems based on artificial and computational intelligence to manage T1D [4–7], where the preceding information was continuous glucose monitoring (CGM).

The standardized clinical levels are defined as the percentage in each of the following glucose ranges: hypoglycemia level 1: $54 \leq BG < 70$ mg/dL, hypoglycemia level 2: $BG < 54$ mg/dL, hyperglycemia level 1: $180 < BG \leq 250$ mg/dL, hyperglycemia level 2: $BG > 250$ mg/dL and time in range (TIR): 70–180 mg/dL [8]. Several studies have already used different approaches to treat the times in each glucose range of people with T1D [9–12]. Given that these percentage of times in each of the ranges are codependent and only provides relative information, log-ratio (logarithm of a ratio) techniques of compositional data (CoDa) are appropriate to deal with this type of data [13].

The CoDa analysis has been studied and developed for several decades, and the number of investigations continues to increase over the years [9,14–17], with applications in medicine [10–12,18]. Biagi and colleagues [12] present a methodology based on CoDa analysis to categorize the daily glucose profiles of patients with T1D. The CoDa analysis involves positive component vectors describing the contribution of several parts to a whole. For example, the time spent in different activities during a day are 24 h relative contributions, and thus, are CoDa [10,11]. Similarly, the time spent in each glucose range is CoDa [12].

Several statistical procedures exist for the validation of probabilistic models. According to Mayer et al. [19], the empirical validation of comparing model predictions with real-world observations must be performed using appropriate statistical methods. In this work, we aim to complement the analysis of the results and validate the probabilistic transition model presented in Biagi et al. [13]. First, a CoDa approach is used to categorize glucose data of 24 h and 6 h duration, then a 5-fold cross-validation method is applied. The main focus is to propose an accuracy metric based on CoDa, calculate the errors associated with CoDa, and evaluate the model's accuracy. We employ glucose data from eight real patients for the validation of the model. The methodology allows the probabilistic prediction of the glucose profile category for the next 6 h and can be used to help clinicians provide individualized adjustments in their patient therapies.

2. Materials and Methods

2.1. Data Set

We analyzed data from eight patients with insulin pump therapy obtained from a pilot study performed at the Hospital Clinic of Barcelona. We use data of 30 weeks, approximately, recollected in different periods between 2020 and 2022. The demographic characteristics of patients are presented in Table 1. All patients provided written informed consent to participate in this study.

Table 1. Demographic characteristics of the cohort.

Variables	Mean ± SD
Age (years)	36.3 ± 10.9
Weight (kg)	70.5 ± 6.4
Height (cm)	167.3 ± 5.5
HbA1c (%)	6.9 ± 0.9
Time with T1D (years)	25.2 ± 12.7
Time with pump (years)	13 ± 7.2
Sex	4 (M) 4 (F)

The recruited patients wore different sensors, including the Dexcom G6, the MiniMed 640G, and the FreeStyle Libre. The first two sensor models stored their measurements every 5 min, while the third model stored its measurements every 15 min. First, CGM

2. PREDICCIÓN PROBABILÍSTICA PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

measurements were preprocessed. Then, 24 h periods starting at 0h, 6h, 12h, and 18h and their corresponding subsequent period of 6 h were obtained (Figure 1). The measurements in each period are divided into the five previously mentioned glucose ranges.

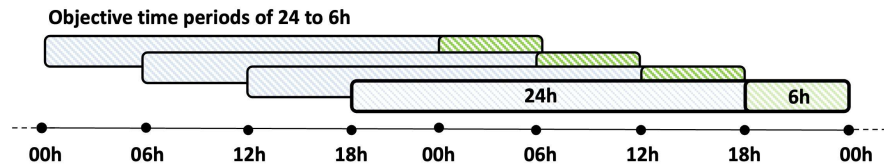


Figure 1. Distribution of the 24 h and 6 h periods.

The patients’ glucose profiles have unrecorded measurements. Data were linearly interpolated when the gaps of missing data were not greater than two consecutive hours. The missing data that was interpolated is actual missing data. Surely if smaller periods were interpolated or not interpolated at all the quality of the data would improve. In this procedure, the days that had more than 2 h of missing data in the same period of 6 h were eliminated. Finally, a day was considered valid if each of its four 6 h periods had at least 75% of the data. Table 2 shows the total number of days analyzed, sensors, and valid periods of 24 h and consecutive 6 h periods for each patient.

Table 2. Characteristics of the sensors and measurements of the eight patients.

Patient	Valid days	Sensor	Periods 24 h to 6 h			
			0 h–0 h 0 h–6 h	6 h–6 h 6 h–12 h	12 h–12 h 12 h–18 h	18 h–18 h 18 h–0 h
1	90	Dexcom G6	85	83	83	84
2	90	MiniMed 640G	54	51	54	53
3	226	MiniMed 640G	57	44	62	56
4	90	Dexcom G6	74	81	81	75
5	134	Dexcom G6	83	82	85	81
6	232	Dexcom G6	76	68	72	73
7	115	MiniMed 640G	80	79	71	81
8	556	FreeStyle Libre	229	231	232	227

2.2. CoDa

A composition is a vector $X = (x_1, x_2, \dots, x_D)$, with D number of parts whose components are all strictly positive and of constant sum (which can be the unit, 100% (as is our case study), sum of the hours of the day (24 h) or some other constant sum defined by the researcher), according to the “scale invariance” property the chosen value for k it is irrelevant for the analysis and is only useful for the interpretation of the results:

$$\begin{cases} x_i > 0, & i = 1, 2, \dots, D. \\ \sum_{i=1}^D x_i = k & k = cte. \end{cases} \quad (1)$$

Historically, CoDa have been identified with closed data, with the simplex being the natural sample space for this data type, while the real Euclidean space is associated with unrestricted data. The basic principles of CoDa analysis that are of special interest in this study are as follows:

(1) scale invariance, which states that CoDa only contains relative information, implying that any change in the scale of the original data do not affect the structure of the composition;

2.2. VALIDACIÓN DEL MODELO PROBABILÍSTICO DE PREDICCIÓN PARA PACIENTES CON T1D UTILIZANDO ANÁLISIS DE DATOS COMPOSICIONALES

(2) subcompositional coherence, which implies that the results obtained for a subset of parts of a composition, that is, a subcomposition, must be coherent with the results obtained with the complete composition; and

(3) permutation invariance, which indicates that the results do not depend on the order in which the parts appear in the composition [14].

In this work, the compositional vector X is defined as a composition where each D-part corresponds to the percentage of time in each of the glucose ranges:

- Hypoglycemia level 2 ($X_{<54}$),
- Hypoglycemia level 1 (X_{54-70}),
- Target BG (X_{70-180}),
- Hyperglycemia level 1 ($X_{180-250}$), and
- Hyperglycemia level 2 ($X_{>250}$)

and is treated as the 5-part composition:

$$X = (X_{<54}, X_{54-70}, X_{70-180}, X_{180-250}, X_{>250}). \tag{2}$$

whose constant sum is 100%. Some of the parts of the composition would be zero if no measurements were found in some of the glucose ranges. For example, a composition (0, 0, 100, 0, 0) would mean 100% TIR, another example could be (0, 10, 80, 10, 0) where it would have 10% in hypo and hyper level 1 and 80% in TIR. As CoDa analysis is based on log-ratios of parts, treating zeros appropriately and analyzing incidence patterns is necessary [20]. In the analyzed measurements, three types of patterns of zeros were identified: non-consecutive zeros, two consecutive zeros, and three consecutive zeros.

The detection limits (dl) matrix used in zero imputation was obtained considering 5- and 15-min fractions, as in Biagi and colleagues [12], depending on the position of the zero in the compositions. For the sensors that save the measurements every 5 min, one day has $1440/5 = 288$ measurement recordings, then the dl is calculated as $dl = 5/1440 = 1/288 = 0.0035$; following the same procedure for the sensors that store the samples every 15 min, $1440/15 = 96$ measurements, then the $dl = 15/1440 = 1/96 = 0.0104$. Measured values below the thresholds defined in the dl matrix cannot be distinguished from a blank signal with a specified confidence level. We considered that the further the zero is from the non-zero value, the smaller this value must be in the dl matrix, as presented in Table A5 of Appendix A. In none of the periods, the glucose range was found to be always zero for any patient. Therefore, we consider these zeros to be rounded zeros of continuous data, not essential ones.

The replacement of zeros in the vector of times at each of the glucose ranges is performed following Biagi and colleagues [12] using the robust expectation-maximization (lrEM) [20,21]. This model-based function imputes left-censored data (e.g., values below the dl , rounded zeros) by representing CoDa coordinates incorporating the relative covariance structure information. When the matrix of zero patterns has a whole column of zeros, the multiplicative replacement method (multRepl) [20,22] was considered. This method preserves the covariance structure and is consistent with the properties of CoDa; it consists of multiplicatively imputing the null values with a small preset value. The modification in the values that are not zero is multiplicative; it is consistent with the basic operations in the simplex and the structure of the compositions [23,24]. Figure 2 describes the methodology for this work, first for proper preprocessing of BG measurements, then for the validation of the model, and finally, the update of the probabilistic prediction model if it has already been previously validated.

2. PREDICCIÓN PROBABILÍSTICA PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

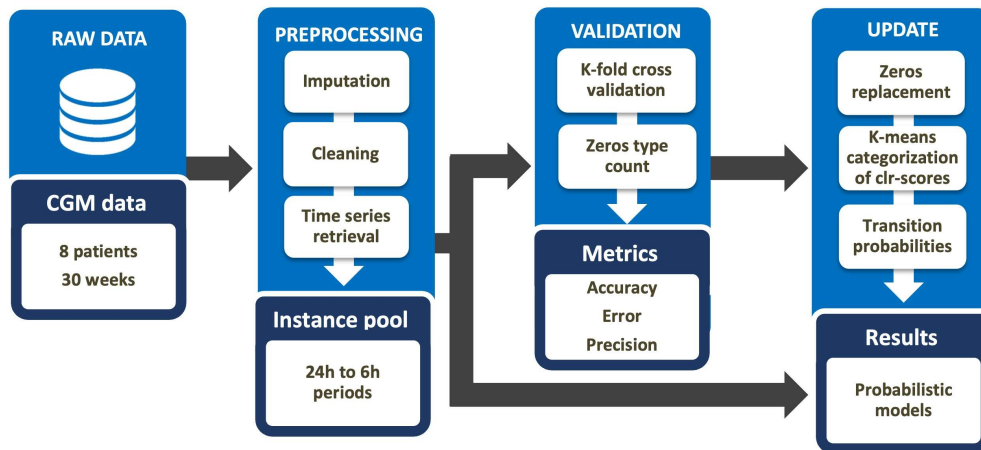


Figure 2. Methodology for data analysis, validation, and update of the probabilistic transition model.

2.2.1. Log-Ratio Coordinates

CoDa can be translated into real space via clr-scores and olr-coordinates, in which, traditional statistical methods can be applied [14], and are calculated using Equations (3) and (4) [25], where g is the geometric mean (Equation (5)), r and s are the number of parts in the i -th row of S (sign matrix of the sequential binary partition (SBP)) coded by +1 (positive) and -1 (negative), which will be in the numerator and denominator of the corresponding log-ratio, respectively. Egozcue and Pawłowsky-Glahn [26] defined the SBP as a hierarchical grouping of parts of the original compositional vector, starting with the complete composition as a group and ending with each part in a single group. If D is the number of parts in the original composition, then the number of steps in the partition is $D - 1$. In this study, we considered the SBP presented in Table 3, which was established according to Biagi and colleagues [12].

$$clr(x) = [clr_1, clr_2, \dots, clr_D] = \left[\ln \frac{x_1}{g(x)}, \ln \frac{x_2}{g(x)}, \dots, \ln \frac{x_D}{g(x)} \right]. \quad (3)$$

$$olr(x) = \sqrt{\frac{r * s}{r + s}} * \ln \left(\frac{g(x_+)^{\frac{1}{r}}}{g(x_-)^{\frac{1}{s}}} \right). \quad (4)$$

Table 3. Sequential binary partition.

i	$X_{<54}$	X_{54-70}	X_{70-180}	$X_{180-250}$	$X_{>250}$	$r(+)$	$s(-)$
1	+1	+1	-1	-1	-1	2	3
2	+1	-1	0	0	0	1	1
3	0	0	-1	+1	+1	2	1
4	0	0	0	-1	+1	1	1

2.2.2. Compositional Measurements

The geometric mean is a representative measure of the center of the CoDa set and identifies the components that better discriminate in the composition. Let $X = (x_1, x_2, \dots, x_n)$ be a compositional data set of S^D . The compositional geometric mean (g_k) of the set X is defined as:

$$g(X) = \mathcal{G}(g_1, g_2, \dots, g_D) = \left(\frac{g_1}{\sum g_k}, \frac{g_2}{\sum g_k}, \dots, \frac{g_D}{\sum g_k} \right), g_k = \left(\prod_{i=1}^n x_{ik} \right)^{\frac{1}{n}}. \quad (5)$$

2.2. VALIDACIÓN DEL MODELO PROBABILÍSTICO DE PREDICCIÓN PARA PACIENTES CON T1D UTILIZANDO ANÁLISIS DE DATOS COMPOSICIONALES

where g_k represents the geometric mean represents the geometric mean of the k -th component of the data. The variation matrix shows the pairwise log-ratio variance for all parts of the composition. It allows the analysis of the data dispersion. This matrix is defined as $T = [T_{ij}]$, and in its extended form, is equal to:

$$T = [T_{ij}] = \begin{pmatrix} \text{var} \left[\ln \frac{x_1}{x_1} \right] & \cdots & \text{var} \left[\ln \frac{x_1}{x_D} \right] \\ \vdots & \ddots & \vdots \\ \text{var} \left[\ln \frac{x_D}{x_1} \right] & \cdots & \text{var} \left[\ln \frac{x_D}{x_D} \right] \end{pmatrix}. \tag{6}$$

where $T_{ij} = \text{var}[\ln(x_i/x_j)]$ represents the expected variance of the log-ratio of parts i and j . This matrix is based on the contribution of variance for each pairwise log-ratio. However, the variation array is usually preferred in practice. This array is based on the variation matrix where the upper diagonal of the array contains the log-ratio variances and the lower diagonal contains the log-ratio means. That is, the way we show the results in Table A2 of the Appendix A, the ij -th component of the upper diagonal is $\text{var}[\ln(X_i/X_j)]$ and the ij -th component of the lower diagonal is $E[\ln(X_i/X_j)]$, where $i, j = 1, 2, \dots, D$ [14].

In real space, the most widely used measure of dispersion is the trace of the covariance matrix associated with the ensemble. However, the interpretability of the direct covariance matrix of a CoDa set is lacking. As this measure is not compatible with CoDa, Pawlowsky-Glahn and Egozcue [16] defined a measure of variability $\text{totvar}(X)$ equal to the trace of the covariance matrix of the clr-transformed data set:

$$\text{totvar}(X) = \sum_{k=1}^D \text{var}[\text{clr}_k(x)]. \tag{7}$$

An example of specific results of these measures for patient 1 (P1) can be observed in Tables A1 and A2 of the Appendix A.

2.3. Probabilistic Model of Transition

The probabilistic transition model proposed by Biagi et al. [13] was implemented for 3, 4, and 5 clusters, where the categories from the previous 24 h period to the next 6 h period are counted. The procedure is as follows: suppose the categories are defined as A, B, C, D, and E. First find all the A categories, examine the period after these days and complete a matrix with the counts of each column, as described in Table 4. Then, with the closure operator defined in Equation (8), the transition probabilities are calculated at different times of the day. Table A3 of the Appendix A shows a particular example of the model for P1.

$$\mathbb{C}(X_1, X_2, \dots, X_D) = \left(\frac{X_1}{\sum_{i=1}^D X_i}, \frac{X_2}{\sum_{i=1}^D X_i}, \dots, \frac{X_D}{\sum_{i=1}^D X_i} \right). \tag{8}$$

Table 4. Methodology of the probabilistic transition model.

		A	B	6 h C	D	E
24 h	A	AA	AB	AC	AD	AE
	B	BA	BB	BC	BD	BE
	C	CA	CB	CC	CD	CE
	D	DA	DB	DC	DD	DE
	E	EA	EB	EC	ED	EE

To validate this probabilistic transition model, clusters 3, 4, and 5 mentioned previously are analyzed. We consider 5-fold cross-validation, randomly selecting 75% of the data for training and the remaining 25% for validation. We employ linear discriminant analysis to assign groups to the validation data, following the methodology of Biagi et al. [13]. Both

the training and validation data of the model are CoDa vectors, whose constant sum is 100%. Therefore, a metric is needed to compare the training and validation results to obtain the accuracy of the model.

2.4. Accuracy Metric

We propose the calculation of an accuracy metric based on CoDa. The first step is to create the Training (T) and Validation (V) matrices (Equation (9)), which contain the transition probabilities of the categories from one 24 h period to the next 6 h, where D is the dimension of the vector. For this, the categories are counted, and the zero type counts are substituted.

In Martín-Fernández et al. [27], count-type vectors are defined as categorical data in which the counts represent the number of elements located in each of several categories. This type of zeros is related to a sampling problem because the components may not be observed given the limited size of the sample. The count-zero multiplicative replacement is implemented in the R package “zCompositions” [20], following what was established in Martín-Fernández et al. [27], where the multiplicative replacement by rounded zero defined in Martín-Fernández and colleagues [24] was adapted for the case of counting zeros. Although this method satisfies the condition that the imputed zero value does not depend on the D parts of the composition, it is recommended only when the number of zeros in the data matrix is insignificant.

$$T = \begin{pmatrix} T_{11} & \cdots & T_{1D} \\ \vdots & \ddots & \vdots \\ T_{D1} & \cdots & T_{DD} \end{pmatrix} V = \begin{pmatrix} V_{11} & \cdots & V_{1D} \\ \vdots & \ddots & \vdots \\ V_{D1} & \cdots & V_{DD} \end{pmatrix}. \tag{9}$$

The accuracy is a difference measure between the training data (expected), and the validation data (observed) for each k model created. The higher the accuracy, the more similar the probability vectors between the transitions from one period to another; therefore, it would also suggest the most appropriate number of groups for each patient. From the analysis of the distances and the norms of the T and V vectors that are detailed below, the accuracy metric is defined as Equation (10).

$$Accuracy = 100 - \frac{\|\vec{V} \ominus \vec{T}\|_a}{\|\vec{V}\|_a + \|\vec{T}\|_a} * 100. \tag{10}$$

To implement this Equation (10), mathematical operators defined for CoDa were considered. The T or V matrix vectors where no transitions were found from one period to another were treated as null or empty (\emptyset). Then, the difference perturbation operator introduced by Martín-Fernández and their colleagues [9] is applied as:

$$X \ominus Y = \mathbb{C} \left[\frac{X_1}{Y_1}, \frac{X_2}{Y_2}, \dots, \frac{X_D}{Y_D} \right]. \tag{11}$$

Applying Equation (11) to the previously defined matrices T and V leads to Equation (12). The perturbation difference operation is analogous to subtraction in Euclidean space. Therefore, the process is performed row-wise.

$$V \ominus T = \begin{pmatrix} V_{11} & \cdots & V_{1D} \\ \vdots & \ddots & \vdots \\ V_{D1} & \cdots & V_{DD} \end{pmatrix} \ominus \begin{pmatrix} T_{11} & \cdots & T_{1D} \\ \vdots & \ddots & \vdots \\ T_{D1} & \cdots & T_{DD} \end{pmatrix} = \mathbb{C} \begin{pmatrix} \frac{V_{11}}{T_{11}} & \cdots & \frac{V_{1D}}{T_{1D}} \\ \vdots & \ddots & \vdots \\ \frac{V_{D1}}{T_{D1}} & \cdots & \frac{V_{DD}}{T_{DD}} \end{pmatrix}. \tag{12}$$

The difference perturbation operation also includes the closure operator (\mathbb{C}), as defined in Equation (8). This is a technique to simplify the use of closed-form compositions, that is, positive vectors whose parts add up to a constant positive k (in our case, $k = 100\%$,

2.2. VALIDACIÓN DEL MODELO PROBABILÍSTICO DE PREDICCIÓN PARA PACIENTES CON T1D UTILIZANDO ANÁLISIS DE DATOS COMPOSICIONALES

percentage type data). In this context, the triangular inequality theorem of Euclidean geometry is applied, which has been generalized to normed vector spaces, obtaining:

$$\|\vec{V} \ominus \vec{T}\|_a \leq \|\vec{V}\|_a + \|\vec{T}\|_a. \tag{13}$$

where the Aitchison norm $\|\cdot\|_a$ of a composition can be calculated as the Euclidean norm $\|\cdot\|$ of the clr-scores [25].

$$\|\vec{V} \ominus \vec{T}\|_a = \|\text{clr}(\vec{V} \ominus \vec{T})\| = \sqrt{\sum_{i=1}^D \text{clr}_i(\vec{V} \ominus \vec{T})^2}. \tag{14}$$

$$\|\vec{V}\|_a = \|\text{clr}(\vec{V})\| = \sqrt{\sum_{i=1}^D \text{clr}_i(\vec{V})^2}. \tag{15}$$

$$\|\vec{T}\|_a = \|\text{clr}(\vec{T})\| = \sqrt{\sum_{i=1}^D \text{clr}_i(\vec{T})^2}. \tag{16}$$

Rearranging Equation (13), we obtain:

$$0 \leq \frac{\|\vec{V} \ominus \vec{T}\|_a}{\|\vec{V}\|_a + \|\vec{T}\|_a} \leq 1. \tag{17}$$

The Aitchison distance [28] between two compositions is known as $d_a(X, Y)$ (Equation (18)), which is the norm of the difference perturbation operation of these compositions; therefore, in the numerator of Equation (17), the Aitchison distance of the composition created between the components of the training vector and those of the validation vector is calculated.

$$d_a(X, Y) = \|\vec{X} \ominus \vec{Y}\|_a. \tag{18}$$

2.5. Precision Metric of the Transition Model

In this work, we also propose the adaptation of known statistical errors in terms of CoDa. This is achieved through the transformation of basic Euclidean operations of the mathematical equations to their corresponding operation in the simplex, as follows:

$$MAE = e_a = \frac{1}{total} \sum_{k=1}^{total} \|\vec{V}_D \ominus \vec{T}_D\|_a. \tag{19}$$

$$MRE = e_r = \frac{e_a}{\|\vec{T}_D\|_a}. \tag{20}$$

$$RMSE = e_c = \sqrt{\frac{\sum_{k=1}^{total} \|\vec{V}_D \ominus \vec{T}_D\|_a^2}{total}}. \tag{21}$$

where:

e_a : mean absolute error (MAE)

e_r : mean relative error (MRE)

e_c : root mean square error (RMSE)

D is the number of parts of the composition

V_D : is the observed composition (that is, the data validation)

T_D : is the expected composition (the training data)

$total$: is the number of corresponding vectors for each 5-fold

2. PREDICCIÓN PROBABILÍSTICA PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

$k=3, 4$ or 5 : corresponds to the number of clusters being analyzed
 The precision of the model is calculated according to:

$$Precision = \begin{cases} 100 - e_a \\ 100 - e_r \\ 100 - e_c \end{cases} \quad (22)$$

3. Results

The median accuracy after 5-fold validation for 3, 4, and 5 clusters are presented in Table 5 for the cohort. Accuracy greater than 50% is considered exceptionally good, indicating that at least half of the parts of the composition coincided satisfactorily (i.e., the data of V with the data of T). Clarifying that as the clusters increase, the T vector and the V vector also increase, is necessary. Therefore, more probabilities of dissimilarity exist among the parts of the vectors, which causes the result of the accuracy measurement to decrease. As evident from Table 5, the accuracy results for 3 clusters are larger than those for 5 clusters. The probabilistic model has been successfully validated because the validation data confirms what was predicted by the training data. The Appendix illustrates an example of the application of this proposed metric for P1.

Table 5. Accuracy results after 5-fold cross-validation for the cohort.

Cluster		At 0 h	At 6 h	At 12 h	At 18 h
3	A	56.4 (43.4–69.7)	67.8 (54.6–76.9)	57.9 (37.6–73.4)	58.1 (39.1–72.1)
	B	56.6 (35.2–65.6)	67.3 (55.0–73.2)	50.5 (43.2–72.4)	57.9 (40.5–66.9)
	C	48.8 (25.4–64.3)	63.3 (48.1–71.0)	58.2 (48.2–63.9)	48.3 (31.5–57.7)
4	A	56.3 (39.1–69.6)	68.6 (49.1–75.1)	59.7 (38.3–72.7)	55.9 (40.2–66.2)
	B	65.0 (44.9–76.8)	62.5 (50.7–68.1)	60.9 (45.2–68.1)	46.9(36.6–62.2)
	C	37.5 (26.5–54.6)	57.3 (45.9–65.2)	50.5 (40.2–58.5)	45.6 (28.7–57.1)
	D	35.9 (17.5–51.6)	40.8 (27.3–50.2)	43.2 (31.9–54.5)	40.5 (28.1–51.5)
5	A	55.3 (34.8–64.3)	48.4 (37.9–59.8)	46.6 (31.0–62.6)	48.8 (28.8–61.1)
	B	56.9 (41.2–70.2)	50.4 (38.8–62.5)	38.8 (23.1–53.8)	43.8 (21.5–53.3)
	C	40.1 (29.7–54.2)	57.6 (44.2–66.9)	59.2 (36.9–62.8)	36.6 (24.0–48.4)
	D	38.1 (24.4–48.1)	45.7 (33.0–52.0)	42.2 (31.1–52.8)	32.6 (18.0–44.7)
	E	34.1 (24.3–48.9)	47.8 (30.1–57.5)	38.3 (33.8–46.8)	39.6 (29.9–46.5)

The result is the median (interquartile range (25th–75th)) of the accuracy for the validation of the probabilistic transition model for the cohort.

The box plot of Figure 3 shows the summary of the precision based on the errors of the validation of the probabilistic transition model for clusters 3, 4, and 5 of each one of the patients and the cohort in general. The top of the box (third quartile) illustrates that 75% of the values are less than or equal to this (98%), the bottom of the box (first quartile) shows that 25% of the values are less than or equal to this value (94%). The median (second quartile) divides the distribution into two equal parts. For cluster 3, only P1 has a symmetric precision distribution (mean, median, and mode coincide), although for the remaining patients, as well as the cohort, the median has a negative asymmetry or is skewed to the left (the longest part below the median). Thus, the data is concentrated in the upper part of the distribution and the mean is less than the median. This same behavior is evident for all the clusters, indicating that in all cases, the lower part of the boxes is larger (the data is more dispersed). Notably, the respective clusters of each patient exhibit similar behavior. However, cluster 5 of patients 3 and 4 stands out—their minimum precision was close to 94% and this was due to the variability of the patients (43% and 39.6%, respectively). In all cases, the dimensions of the boxes determined by the distance of the interquartile range grew as the clusters increased, which evidences the dispersion of the data around the median, and in turn, reflects how it coincides with the results of the accuracy (the greater the number of clusters, the lower the number of samples per group).

2.2. VALIDACIÓN DEL MODELO PROBABILÍSTICO DE PREDICCIÓN PARA PACIENTES CON T1D UTILIZANDO ANÁLISIS DE DATOS COMPOSICIONALES

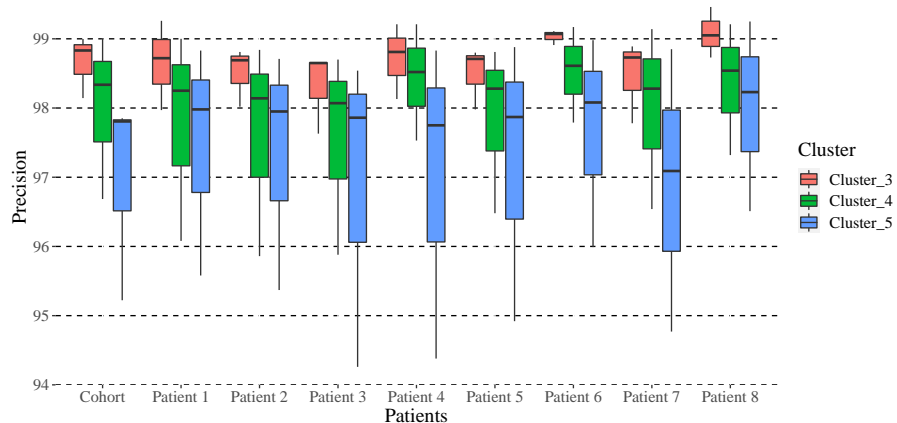


Figure 3. Precision of the probabilistic transition model for each patient and the cohort.

These results are relevant for the research, application, and validation of CoDa methodologies in any branch of science. The metrics obtained allow the comparison of dissimilar compositions, the disregard of parts of the composition that are not significant, the validation of models, and the identification of patterns. However, it is important to highlight that the researcher is the one who proposes, based on their experience in the field, the parts of the composition and the constant to be added according to the characteristics of CoDa, so they must also interpret the results obtained. Furthermore, as previously mentioned, the more parts the composition vector has, the greater the possibility that the accuracy metric will be lower. Therefore, the percentage obtained by this metric will be associated with the interpretation given by the researcher.

4. Discussion

Data-driven decision support systems have always helped physicians and patients [4,5]. This work is the continuation of previous studies to characterize glucose profiles of times in different ranges applying CoDa. Biagi and colleagues [12] created a methodology for the categorization of glucose profiles of six T1D patients who were monitored for eight weeks. Then, the k-means algorithm was applied to the clr-scores, obtaining different groups. Subsequently, Biagi et al. [13] obtained a discriminant model to determine the category of 24 h periods, achieving an average of more than 94% correct classification. Furthermore, the authors proposed a probabilistic transition model to predict the future 6 h period.

These two investigations set the path to the CoDa analysis tool applied to the glucose profiles of patients with T1D. However, a validation of the probabilistic transition model obtained in Biagi et al. [13] was not presented, which is considered a limitation of the study by the authors. This model would serve as a decision support tool to manage T1D for patients and physicians. In this study, the limitations of [12] are minimized by comprehensively analyzing the compositional statistics of these data, which provides valuable information on the behavior of glucose, traceability, and improvement of the patient's glucose profiles. Thus, the characteristics of the classificatory groups are identified not only qualitatively but also quantitatively, as stated in [12]. In addition, other methodologies were tested to determine the detection limits matrix of the data set; however, placing a lower dl on zero parts farther from the non-zero parts had less effect on variability.

Finally, the probabilistic prediction model is validated to ensure its reliability. Notably, during the validation of the model, the greater the number of groups, the fewer the observations per group, which suggests lower percentages for each part because of the distribution of the category counts. Hence, the relevance of accuracy. A new accuracy metric based on the difference in compositional vectors was proposed in this work, which allowed the validation of the proposed prediction model. The validation of the model

2. PREDICCIÓN PROBABILÍSTICA PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

provides an idea of what the correct amount of cluster should be for each patient to achieve the highest accuracy in the prediction.

In some cases, the measure of the accuracy of some transition probabilities was not high. This happens fundamentally because of the few days, which leads to sufficient transitions not being always counted in some cases, creating a model where the probability percentages are dispersed. Similarly, we consider that the update of the model and incorporation of more recent data improves the reliability of prediction, including predicting for a shorter period. This would help clinicians to assess patient outcomes and to customize their insulin dosing profile.

This methodology is designed for an individualized probabilistic transition model and not for the cohort. As mentioned in [13], in this study, we did not intend to present the prediction of BG values or trends but a probability of the behavior of glucose in the following 6 h as a decision support tool for the management of T1D. Furthermore, considering that glucose sensors lack of accuracy, patients's insulin sensitivity and the dynamic of insulin response in the body are also estimations, there is not a method that could guarantee the verification of the effects of insulin in the glucose drop. In that way, the prediction of the behavior of the patient through the category obtained from composition of times in different glucose ranges in future periods could decrease the effects of intrinsic inaccuracies of devices used for diabetes management that can jeopardize patient care. The intent was that, as the model suggests the patient, it will be updated and adjusted considering the habits and characteristics of individual patients over time.

5. Conclusions

In this work, a novel methodology was presented to validate the probabilistic transition model presented in [13]. New measures of accuracy and precision based on CoDa were proposed. Glucose measurements from eight T1D patients were processed. Obtaining satisfactory average results with accuracy and precision greater than 50% and 95% for the entire cohort, respectively, suggests the reliability of the model. This methodology can be extended to CoDa analysis for other studies that need to be validated or for comparisons between compositions where the components represent parts of a whole. The novelty of this work stems from the absence of this type of measure in the extant literature.

Author Contributions: A.C. and L.B. wrote the manuscript, contributed to discussion, and reviewed/edited the manuscript; conceptualization, A.C., J.V. and J.A.M.-F.; methodology, A.C.; software, A.C., L.B., E.E., A.B. and I.C. (Iván Contreras); validation, A.C., E.E., A.B., I.C. (Iván Contreras) and J.B.; investigation, A.C.; writing—review and editing, A.C., E.E., L.B., A.B., I.C. (Iván Contreras), J.A.M.-F., M.G., I.C. (Ignacio Conget) and J.V.; funding acquisition, A.C., E.E., J.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by grants PID2019-107722RB-C22 and PID2019-107722RB-C12 funded by MCIN/AEI/10.13039/501100011033, in part by the Autonomous Government of Catalonia under Grant 2017 SGR 1551, in part by the Spanish Ministry of Universities, and by the European Union through Next GenerationEU (Margarita Salas), and by the program for researchers in training at the University of Girona (IFUdG2019).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank all the participants who dedicated their time and effort to complete this study.

Conflicts of Interest: The authors declare no conflict of interest.

2.2. VALIDACIÓN DEL MODELO PROBABILÍSTICO DE PREDICCIÓN PARA PACIENTES CON T1D UTILIZANDO ANÁLISIS DE DATOS COMPOSICIONALES

Abbreviations

The following abbreviations are used in this manuscript:

T1D	Type 1 diabetes
TIR	Time in range
BG	Blood glucose
MDI	Multiple daily injections
CSII	Continuous subcutaneous insulin infusion
CGM	Continuous glucose monitoring
CoDa	Compositional Data
clr	Centered log-ratio
ilr	Isometric log-ratio
SBP	Sequential binary partition
BGV	Blood glucose variation
CHO	Carbohydrate
SD	Standard deviation
SEM	Standard error of the mean
MAE	Mean absolute error
MRE	Mean relative error
RMSE	Root mean square error
P1	Patient 1

Appendix A. Specific Results for P1

Appendix A.1. Compositional Statistics of the Data

Once the 24 h and 6 h consecutive data have been categorized with the k-means method, the different groups are obtained (in this example, 4 groups were set). The compositional geometric mean is calculated for each of these groups (Equation (5)). This vector provides a quantitative interpretation of each of the groups (Table A1). The 24 h and 6 h periods were qualitatively characterized in terms of the relative time spent in the different glucose ranges, according to the log-ratio approach. Although groups of different patients may present comparable characteristics in terms of the relative interpretation of the time spent in different glucose ranges, the results must be interpreted individually and in a relative sense and not in an absolute manner. Both the 24 h and 6 h periods were classified taking into account the standardized metrics [8] where the following glucose targets are pursued: <54 mg/dL (<1%), 54–70 mg/dL (<4%), 70–180 mg/dL (>70%), 180–250 mg/dL (<20%) and >250 mg/dL (<5%). If we analyze the example shown in (Table A1) we can see that group 1 has an average of 2.61% in hypo level 1, 87.38% in TIR and 10% in hyper level 1. In view of these data, we assign classification A, which we will qualitatively define as periods with a moderate percentage of hypo and hyper level 1. Following this logic, for this example, the groups are classified as follows:

The 24 h periods were classified as:

- A—periods with percentages in hyperglycemia and moderate hypoglycemia of level 1.
- B—periods with high percentages of level 1 and level 2 hyperglycemia.
- C—periods with high percentages of time in range with slight occurrences of level 1 hypoglycemia.
- D—periods with percentages of level 1 hyperglycemia.

The 6 h periods were classified as:

- A—periods with percentages in level 1 hyperglycemia and hypoglycemia.
- B—periods with very high percentages of level 1 and level 2 hyperglycemia.
- C—periods with high percentages of time in range.
- D—periods with percentages of level 1 hyperglycemia.

2. PREDICCIÓN PROBABILÍSTICA PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

Table A1. Compositional center of each of the parts by a group of P1.

Compositional center of each group of 24 h periods for group 4				
	G_A(179)	G_B(96)	G_C(23)	G_D(37)
<54 [mg/dL]	0.00	0.00	0.00	0.00
54–70 [mg/dL]	2.61	0.99	0.22	0.00
70–180 [mg/dL]	87.38	76.18	99.77	86.47
180–250 [mg/dL]	10	18.79	0.00	13.52
>250 [mg/dL]	0.00	4.02	0.00	0.00
Compositional center of each group of the next period of 6 h for group 4				
	G_A(67)	G_B(33)	G_C(119)	G_D(116)
<54 [mg/dL]	0.00	0.00	0.00	0.00
54–70 [mg/dL]	5.65	0.00	0.00	0.00
70–180 [mg/dL]	80.00	51.23	99.95	83.28
180–250 [mg/dL]	14.34	32.54	0.00	16.72
>250 [mg/dL]	0.00	16.21	0.00	0.00
Median and percentile (25th,75th) of each of the parts by group				
	G_A(246)	G_B(129)	G_C(142)	G_D(153)
<54 [mg/dL]	0.00(0.00–0.69)	0.00(0.00–0.69)	0.00(0.00–0.00)	0.00(0.00–0.00)
54–70 [mg/dL]	3.47(1.39–5.56)	1.39(0.00–3.47)	0.00(0.00–5.21)	0.00(0.00–0.00)
70–180 [mg/dL]	83.33(74.31–89.58)	68.06(55.21–77.43)	100(94.44–100)	83.33(72.22–91.67)
180–250 [mg/dL]	12.15(6.60–19.79)	20.83(15.28–26.74)	0.00(0.00–0.00)	16.67(8.33–27.78)
>250 [mg/dL]	0.00(0.00–0.00)	5.90(2.43–13.89)	0.00(0.00–0.00)	0.00(0.00–0.00)

Group (Number of observations per group).

Table A2. Variation array of P1.

$\frac{X_i}{X_j}$	Variances $\ln(\frac{X_i}{X_j})$					clr variances
	<54	54–70	70–180	180–250	>250	
<54		14.5736	18.7269	34.8205	16.6793	7.1033
54–70	3.7926		26.8440	41.9360	29.9660	12.8072
70–180	11.0020	7.2094		15.6085	19.6058	6.3003
180–250	7.3020	3.5094	−3.7001		27.6585	14.1479
>250	0.3664	−3.4262	−10.6357	−6.9356		8.9251
			Mean $\ln(\frac{X_i}{X_j})$			TV: 49.28

Table A2 shows the variation array of the data for P1 (according to Equation (7)). The greatest compositional variability is associated with the parts (180–250) with (54–70) (41.9), (180–250) with (<54) (34.8), and (>250) with (54–70) (29.9) (bold font). According to [29], the components with the greatest variability turn out to be adequate to obtain a subcomposition of three parts and illustrate the data dispersion. The total variance (TV) for this data set was 49.28, and (54–70) with 12.8, (180–250) with 14.14, and (>250) with 8.9 being the parts with the largest contribution (see column clr variances in Table A2).

Appendix A.2. Compositional Biplot for P1 Categorized for 4 Clusters

Akin to classical statistical analysis, CoDa analysis requires data visualization tools; one tool is the compositional biplot [30]. The biplot is a dimensional reduction technique used to represent data with three or more variables. This technique aims to approximate the elements of a matrix from vectors called markers associated with the rows and columns [31].

In this work, the rows correspond to the days and are displayed as points in the compositional clr-biplot. The columns correspond to the times in each of the glucose ranges, represented as rays. The quality of the representation depends on the percentage of variance that is retained with the two axes that are represented. It is constructed by obtaining a singular value decomposition of the covariance matrix using the clr transformation. The interpretation is based on the links between the rays: each ray represents a clr variable,

2.2. VALIDACIÓN DEL MODELO PROBABILÍSTICO DE PREDICCIÓN PARA PACIENTES CON T1D UTILIZANDO ANÁLISIS DE DATOS COMPOSICIONALES

and its length is associated with the variance explained in the projection. The directions of the rays indicate those observations with a greater domain of the compositional part [29]. Figure A1 illustrates the distribution of days when they have been categorized into four clusters from different perspectives and the retained variance.

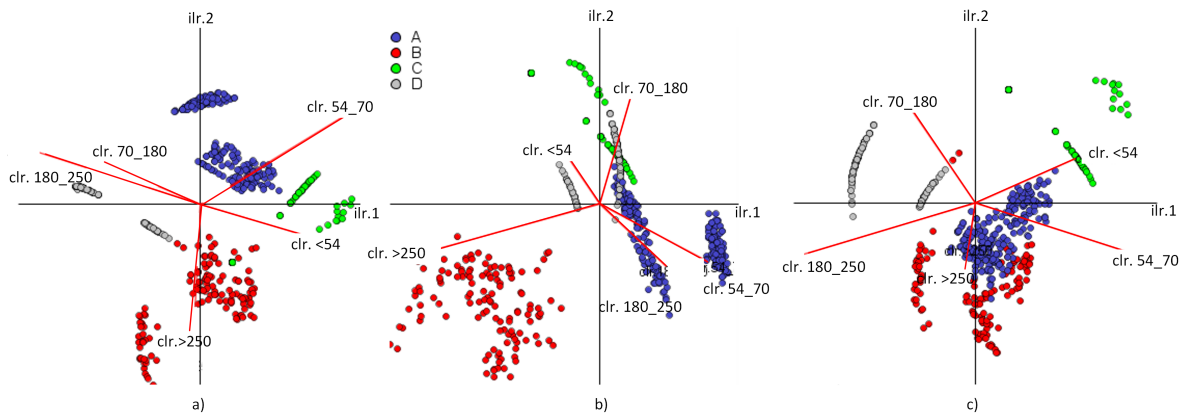


Figure A1. Compositional biplot in space for P1: (a) XY plane; (b) YZ plane, and (c) XZ plane. The three axes of the biplot retain 51%, 77%, and 90% of the TV, respectively.

Appendix A.3. Probabilistic of Transition Model

Table A3 shows the probabilities of transition at different times of the day (0 h, 6 h, 12 h, and 18 h) for P1. Let us consider the patient at 18 h when they have been categorized with four clusters. First, we analyze the glucose composition of the previous 24 h period using Table A1 (Compositional center of each group of 24 h periods) and verify that this period is categorized as type D (86.47% in normoglycemia and 13.52% in hyperglycemia, no hypos observed). Then the probability that the category of the next 6 h period (from 18 h to 0 h) is of type D 75% can be known. Table A1 (Compositional center of each group of 6 h periods) demonstrates how group D was characterized by having 83.28% time in normoglycemia and 16.72% in hyperglycemia. In other words, P1 is expected to continue in normoglycemia with a tendency to hyperglycemic excursions for the next 6 h.

Table A3. Probabilistic transition model for clusters 3, 4, and 5 for P1.

	At 0 h					At 6 h					At 12 h					At 18 h				
	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E
A	23.0	46.1	30.7			31.1	31.1	37.7			40.4	23.4	36.1			50.9	21.5	27.4		
B	28.0	48.0	24.0			20.6	31.0	48.2			30.7	26.9	42.3			46.1	30.7	23.0		
C	25.0	12.5	62.5			22.2	33.3	44.4			50.0	40.0	10.0			14.2	14.2	71.4		
A	10.6	17.0	44.6	27.6		13.9	2.3	46.5	37.2		29.5	4.5	29.5	36.3		37.7	6.6	28.8	26.6	
B	4.3	21.7	52.1	21.7		16.0	20.0	12.0	52.0		17.3	13.0	26.0	43.4		36.0	12.0	32.0	20.0	
C	0.0	0.0	42.8	57.1		0.0	0.0	60.0	40.0		0.0	20.0	60.0	20.0		16.6	0.0	50.0	33.3	
D	12.5	0.0	25.0	62.5		10.0	10.0	40.0	40.0		36.3	9.0	36.3	18.1		12.5	0.0	12.5	75.0	
A	5.0	20.0	15.0	25.0	35.0	0.0	0.0	7.1	35.7	57.1	0.0	0.0	42.1	21.0	36.8	16.6	0.0	16.6	33.3	33.3
B	4.3	21.7	0.0	52.1	21.7	0.0	20.8	12.5	12.5	54.1	13.6	13.6	4.5	22.7	45.4	8.0	12.0	28.0	32.0	20.0
C	0.0	14.8	3.7	59.2	22.2	3.3	3.3	16.6	50.0	26.6	0.0	7.6	19.2	38.4	34.6	14.8	11.1	25.9	25.9	22.2
D	0.0	0.0	0.0	42.8	57.1	0.0	0.0	0.0	60.0	40.0	0.0	20.0	0.0	60.0	20.0	16.6	0.0	0.0	50.0	33.3
E	0.0	0.0	12.5	25.0	62.5	10.0	10.0	0.0	40.0	40.0	0.0	9.0	36.3	36.3	18.1	0.0	0.0	12.5	12.5	75.0

Appendix A.4. Analysis of the Result of the Accuracy Metric for P1

The median accuracy and interquartile range after 5-fold validation for 3, 4, and 5 clusters for P1 is shown in Table A4. It was observed that considering 3 clusters in

2. PREDICCIÓN PROBABILÍSTICA PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

the evaluation resulted in higher accuracy compared to the evaluation with 5 clusters. Furthermore, it was found that the median accuracy for this particular patient significantly exceeded the cohort median, approaching 90% in the 75th percentile.

Table A4. Accuracy results after 5-fold cross-validation for P1.

Cluster		At 0 h	At 6 h	At 12 h	At 18 h
3	A	67.2 (66.3–77.2)	83.9 (55.6–84.6)	80.1 (78.2–83.8)	80.4 (72.2–83.3)
	B	54.6 (2.1–68.7)	75.1 (66.9–76.7)	61.7 (52.6–73.2)	69.1 (39.3–71.0)
	C	55.6 (44.8–69.9)	86.1 (62.8–86.7)	40.3 (19.4–48.5)	58.6 (52.6–65.6)
4	A	59.8 (45.7–72.0)	83.7 (24.3–88.6)	74.3 (39.7–78.9)	69.8 (67.5–75.3)
	B	56.7 (44.9–63.1)	64.9 (58.4–67.8)	46.1 (28.2–47.9)	45.3 (33.7–63.0)
	C	47.7 (42.8–49.4)	64.6 (60.2–66.4)	35.8 (18.6–46.6)	43.9 (31.8–56.7)
	D	45.7 (32.2–66.1)	38.8 (24.1–54.8)	20.6 (13.8–37.6)	54.6 (39.3–64.7)
5	A	44.0 (29.2–55.2)	44.8 (36.3–61.9)	25.4 (18.2–60.5)	43.8 (11.1–57.4)
	B	60.9 (47.1–65.8)	51.3 (42.4–62.1)	33.9 (6.6–44.5)	48.3 (31.9–60.5)
	C	50.7 (39.8–58.6)	46.2 (30.8–71.3)	58.8 (39.3–62.5)	48.8 (33.5–58.2)
	D	36.2 (30.7–36.7)	60.6 (34.6–64.7)	30.8 (21.0–34.6)	35.5 (22.7–53.9)
	E	20.5 (16.4–40.8)	28.3 (8.6–46.9)	13.0 (10.2–23.5)	48.8 (37.9–54.1)

The result is the median (interquartile range (25th–75th)) of the accuracy for the validation of the probabilistic transition model for P1.

Below is an example of the accuracy metric for P1, for k-fold = 2, when the transitions at 0 h to the next period of 0–6 h are counted. After categorizing with k-means the clr-scores corresponding to the times in the range. The categories from the 24 h to the next 6 h period of the *T* and *V* data are counted, which in this case, were *T* = [19 16 7] and for *V* = [6 2 6]. Then, these vectors are verified as not having zeros; in case they do, they are replaced as explained in Section 2.4. Subsequently, Equation (8) is applied, whose constant sum is 100% and can be treated as a CoDa, and *T* = [45.23 14.28 40.47] and *V* = [42.85 14.28 42.85] are obtained.

Calculating the numerator for Equation (10): Applying the difference operator (Equation (11)) and then the closure operator (Equation (8)):

$$\vec{V} \ominus \vec{T} = \mathbb{C} \left[\frac{V_1}{T_1}, \frac{V_2}{T_2}, \frac{V_3}{T_3} \right] = \left[\frac{42.85}{45.23}, \frac{14.28}{14.28}, \frac{42.85}{40.47} \right] = \mathbb{C}[0.9474, 1, 1.0588] = [0.3151, 0.3326, 0.3522]. \tag{A1}$$

Then, the clr-scores are calculated according to Equation (3) and the denominator of the clr-scores (the geometric mean) is calculated according to Equation (5).

$$clr(\vec{V} \ominus \vec{T}) = clr(0.3151, 0.3326, 0.3522) = \left[\ln \frac{0.3151}{0.3330}, \ln \frac{0.3326}{0.3330}, \ln \frac{0.3522}{0.3330} \right] = [-0.0551, -0.0010, 0.0561]. \tag{A2}$$

Applying Equation (14), the norm of the clr-scores is calculated as:

$$\|\vec{V} \ominus \vec{T}\|_a = \sqrt{(-0.0551)^2 + (-0.0010)^2 + (0.0561)^2} = 0.0786. \tag{A3}$$

Calculating the denominator: According to Equation (3), the clr-scores are first calculated, where the geometric mean $g(\vec{V}) = 27.70$ and $g(\vec{T}) = 29.67$ according to Equation (5). Subsequently, the Aitchison norm of the vectors *V* and *T* is calculated (Equations (15) and (16)).

$$clr(\vec{V}) = \left[\ln \frac{42.85}{29.70}, \ln \frac{14.28}{29.70}, \ln \frac{42.85}{29.70} \right] = [0.3663, -0.7326, 0.3663]. \tag{A4}$$

$$clr(\vec{T}) = \left[\ln \frac{45.23}{29.67}, \ln \frac{14.28}{29.67}, \ln \frac{40.47}{29.67} \right] = [0.4214, -0.7315, 0.3102]. \tag{A5}$$

2.2. VALIDACIÓN DEL MODELO PROBABILÍSTICO DE PREDICCIÓN PARA PACIENTES CON T1D UTILIZANDO ANÁLISIS DE DATOS COMPOSICIONALES

$$\|\vec{V}\|_a = \|clr(\vec{V})\| = \sqrt{(0.3663)^2 + (-0.7326)^2 + (0.3663)^2} = 0.8972. \quad (A6)$$

$$\|\vec{T}\|_a = \|clr(\vec{T})\| = \sqrt{(0.4214)^2 + (-0.7314)^2 + (0.3102)^2} = 0.8994. \quad (A7)$$

Finally substituting in Equation (10):

$$Accuracy = 100 - \frac{\|\vec{V} \ominus \vec{T}\|_a}{\|\vec{V}\|_a + \|\vec{T}\|_a} * 100 = 100 - \frac{0.0786}{0.8972 + 0.8994} * 100 = 95.62\%. \quad (A8)$$

Figure A2 shows the previously discussed example of accuracy in compositional biplot, where: A_T , B_T , and C_T correspond to the training data and A_V , B_V , and C_V to the validation data when the data have only been categorized for three clusters. How related the T and V data are in each cluster is evident. In addition, the length of all the pairwise links between the rays suggests no redundant information. That is, amalgamating parts is not recommended. The variance retained by the two first axes in the biplot is 79%, implying that it has high quality, thus suggesting caution in the interpretations.

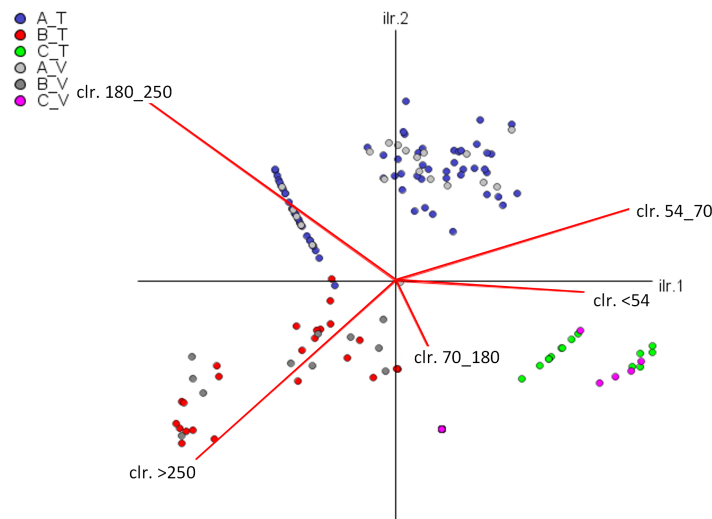


Figure A2. Example of visualization of training and validation data for P1.

Appendix A.5. Detection Limits

Table A5 shows the detection limits for different patterns of zeros.

2. PREDICCIÓN PROBABILÍSTICA PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

Table A5. Detection limits for different patterns of zeros.

For measurements recorded every 5 min. 1440 min/5 min= 288 measurements dl= 1/288=0.0035				
Consecutive zeros	Position 1	Position 2	Position 3	Position 4
1	dl =0.0035			
2	dl/3=0.0012	2dl/3=0.0023		
3	dl/9=0.00038	2dl/9=0.00077	2dl/3=0.0023	
4	dl/27=0.00012	2dl/27=0.00026	2dl/9=0.00077	2dl/3=0.0023
For measurements recorded every 15 min. 1440 min/15 min= 96 measurements dl= 1/96=0.010				
1	dl=0.010			
2	dl/3=0.0033	2dl/3=0.0067		
3	dl/9=0.0011	2dl/9=0.0022	2dl/3=0.0067	
4	dl/27=0.00037	2dl/27=0.00074	2dl/9=0.0022	2dl/3=0.0067

References

- da Silva, J.A.; de Souza, E.C.F. Diagnosis of diabetes mellitus and living with a chronic condition: Participatory study. *Bmc Public Health* **2018**, *18*, 1-8.
- Jeitler, K.; Horvath, K.; Berghold, A. Continuous subcutaneous insulin infusion versus multiple daily insulin injections in patients with diabetes mellitus: Systematic review and meta-analysis. *Diabetologia* **2008**, *51*, 941–951.
- Livvi, S.; Wei, L.; Kim, B.K.H.; Wee, T.T. Development of a clinical decision support system for diabetes care: A pilot study. *PLoS ONE* **2017**, *12*, e0173021.
- Tyler, N.S.; Jacobs, P.G. Artificial intelligence in decision support systems for type 1 diabetes. *Sensors* **2020**, *20*, 3214.
- Ivan, C.; Josep, V. Artificial intelligence for diabetes management and decision support: Literature review. *J. Med. Internet Res.* **2018**, *20*, e10775.
- Revita, N.; Tadej, B. Insulin dose optimization using an automated artificial intelligence-based decision support system in youths with type 1 diabetes. *Nat. Med.* **2020**, *26*, 9.
- Ellahham, S. Artificial intelligence: The future for diabetes care. *Am. J. Med.* **2020**, *133*, 8.
- American Diabetes Association. Care in Diabetes-2022. *Diabetes Care* **2022**, *17*, 254–255.
- Fernández, M.; Antoni, J.; Estadella, D. On the interpretation of differences between groups for compositional data. *Sort Stat. Oper. Res. Trans.* **2015**, *39*, 231-252.
- Dumuid, D.; Stanford, T.E.; Martín-Fernández, J.-A. Compositional data analysis for physical activity, sedentary time and sleep research. *Stat. Methods Med. Res.* **2018**, *27*, 3726-3738.
- Chastin, S.; Palarea-Albaladejo, J.; Dontje, M.L. Combined effects of time spent in physical activity, sedentary behaviors and sleep on obesity and cardio-metabolic health markers: A novel compositional data analysis approach. *PLoS ONE* **2015**, *10*, e0139984.
- Biagi, L.; Bertachi, A.; Giménez, M. Individual categorisation of glucose profiles using compositional data analysis. *Stat. Methods Med. Res.* **2019**, *28*, 3550-3567.
- Biagi, L.; Bertachi, A. Probabilistic Model of Transition between Categories of Glucose Profiles in Patients with Type 1 Diabetes Using a Compositional Data Analysis Approach. *Sensors* **2021**, *21*, 3593.
- Aitchison, J. The statistical analysis of compositional data. *J. R. Stat. Soc. Ser. (Methodol.)* **1982**, *44*, 139–160.
- Martín-Fernández, J.A.; Bren, M.; Barceló-Vidal, C.; Pawłowsky-Glahn, V. A measure of difference for compositional data based on measures of divergence. *Proc. Iamg* **1999**, *99*, 211–216.
- Pawłowsky-Glahn, V.; Egozcue, J.J. BLU estimators and compositional data. *Math. Geol.* **2002**, *34*, 259–274.
- Filzmoser, P.; Hron, K.; Martín-Fernández, J.A.; Palarea-Albaladejo, J. *Advances in Compositional Data Analysis: Festschrift in Honour of Vera Pawłowsky-Glahn*; Springer International Publishing: Berlin/Heidelberg, Germany, 2021.
- Janssen, I.; Clarke, A.E. A systematic review of compositional data analysis studies examining associations between sleep, sedentary behaviour, and physical activity with health outcomes in adults. *Appl. Physiol. Nutr. Metab.* **2020**, *45*, 10.
- Mayer, D.G.; Stuart, M.A.; Swain, A.J. Regression of real-world data on model output: An appropriate overall test of validity. *Agric. Syst.* **1994**, *45*, 93–104.
- Palarea-Albaladejo, J.; Martín, F.; Josep, A. zCompositions-R package for multivariate imputation of left-censored data under a compositional approach. *Chemom. Intell. Lab. Syst.* **2015**, *143*, 85–96.
- Quinn, T.P.; Erb, I. A field guide for the compositional analysis of any-omics data. *GigaScience* **2019**, *8*, 9.
- Lubbe, S.; Filzmoser, P. Comparison of zero replacement strategies for compositional data with large numbers of zeros 2021. *Chemom. Intell. Lab. Syst.* **2021**, *210*, 104248
- Egozcue, J.J.; Pawłowsky-Glahn, V. Compositional data: The sample space and its structure. *Test* **2019**, *28*, 3.
- Martín-Fernández, J.A.; Barceló-Vidal, C.; Pawłowsky-Glahn, V. Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Math. Geol.* **2003**, *35*, 253–278.

2.2. VALIDACIÓN DEL MODELO PROBABILÍSTICO DE PREDICCIÓN PARA PACIENTES CON T1D UTILIZANDO ANÁLISIS DE DATOS COMPOSICIONALES

25. Egozcue, J.J.; Pawlowsky-Glahn, V. *Simplicial Geometry for Compositional Data*; Geological Society, London, Special Publications: London, UK, 2006; Volume 264, pp. 145–159.
26. Egozcue, J.J.; Pawlowsky-Glahn, V. Groups of parts and their balances in compositional data analysis. *Math. Geol.* **2005**, *37*, 795–828.
27. Martín-Fernández, J.A.; Hron, K.; Templ, M.; Filzmoser, P.; Palarea-Albaladejo, J. Bayesian-multiplicative treatment of count zeros in compositional data sets. *Stat. Model.* **2015**, *15*, 134–158.
28. Aitchison, J. Principal component analysis of compositional data. *Biometrika* **1983**, *70*, 57–65.
29. Thió-Henestrosa, S.; Comas, M. *CoDaPack v2 User's Guide*; University of Girona, Department of Computer Science and Applied Mathematics: Girona, Spain, 2016.
30. Aitchison, J.; Greenacre, M. Biplots of compositional data. *J. R. Stat. Soc. Ser. (Applied Stat.)* **2002**, *51*, 375–392.
31. Pinzón, L.M. *Biplot Consenso Para análisis de Tablas Múltiples*; Ediciones Universidad de Salamanca: Salamanca, Spain 2012; p. 300.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

2.3 Predicción individualizada de estado glucémico mediante el análisis de datos composicionales

En esta publicación, proponemos un modelo individualizado de regresión lineal múltiple basado en datos composicionales, para predecir la media y el CV de la BG en personas con T1D en PH amplios (2 y 4 horas). A partir de estas predicciones, se calculan los valores mínimos y máximos estimados de BG. Se ha utilizado una base de datos de 226 pacientes reales con T1D (Replace BG (NCT02258373)). La contribución de la candidata para esta publicación consistió en el procesamiento de los datos, creación de la metodología e implementación de métricas, redacción del manuscrito, contribución a la discusión y la edición del manuscrito. Durante el desarrollo del trabajo, la candidata trabajó con sus colegas Dr. Ernesto Estremera, Dr. Aleix Beneyto y Dr. Lyvia Biagi, asistidos también por el Dr. Iván Contreras, Dr. Josep Vehí y Dr. Josep Antoni Martín-Fernández, quien brindó orientación para el análisis de la metodología de datos composicionales.

Título: Individualized prediction of blood glucose outcomes using Compositional Data analysis

Autores: **Alvis Cabrera**, Ernesto Estremera, Aleix Beneyto, Lyvia Biagi, Iván Contreras, Josep Antoni Martín-Fernández y Josep Vehí

Revista: International Journal of Applied Mathematics and Computer Science

Estado: Enviado

Índice de calidad: JCR: Q1 (66/267) en MATHEMATICS, APPLIED, JIF: 2.157 en 2021



INDIVIDUALIZED PREDICTION OF BLOOD GLUCOSE OUTCOMES USING COMPOSITIONAL DATA ANALYSIS

ALVIS CABRERA^a, ERNESTO ESTREMER A^a, ALEIX BENEYTO^a, LYVIA BIAGI^b, IVÁN CONTRERAS^a, JOSEP ANTONI MARTÍN-FERNÁNDEZ^c, JOSEP VEHI^{a,d,*}

^a Department of Electrical, Electronic and Automatic Engineering, University of Girona, 17003 Girona, Spain
e-mail: alvita.cabrera@udg.edu, ernesto.estremera@udg.edu, aleix.beneyto@udg.edu, ivan.contreras@udg.edu, josep.vehi@udg.edu

^b Campus Guarapuava, Federal University of Technology–Paraná (UTFPR), Guarapuava 85053-525, Brazil
e-mail: lyviar@utfpr.edu.br

^c Department of Computer Science, Applied Mathematics and Statistics, University of Girona, 17003 Girona, Spain
e-mail: josepantoni.martin@udg.edu

^d Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM) Instituto de Salud Carlos III, 28029 Madrid, Spain

This paper presents an individualized multiple linear regression model based on compositional data, where we predict the mean and coefficient of variation of blood glucose in individuals with type 1 diabetes for long-term (2 and 4 hours). From these predictions, we estimate the minimum and maximum glucose values to provide future glycemic status. The proposed methodology has been validated using a dataset of 226 real adult patients with type 1 diabetes (Replace BG (NCT02258373)). The obtained results show a median balanced accuracy and sensitivity of over 90% and 80%, respectively. A “traffic light with symbols” information system has been implemented and validated to update patients on their glycemic status and associated risks for the next few hours.

Keywords: compositional data, decision support system, diabetes type 1, blood glucose prediction.

1. Introduction

Type 1 diabetes (T1D) is a metabolic disorder that causes abnormal regulation of blood glucose (BG), which can lead to short and long-term health complications, and even death if not adequately controlled (Silva *et al.*, 2018). Prediction models can learn personalized glucose and insulin dynamics based on sensor measurements and daily activity of each individual. Notwithstanding the widespread use of machine learning techniques for glucose prediction (Contreras and Vehi, 2018; Mohebbi *et al.*, 2020; Martinsson *et al.*, 2020; Daniels *et al.*, 2021; Tena *et al.*, 2021; Cichosz *et al.*, 2021; Wadghiri *et al.*, 2022), a dearth of up-to-date literature reviews exists on the subject of modeling strategies applied to personalized BG prediction, as pointed out in (Woldaregay *et al.*, 2019). Currently, glucose prediction models exhibit

significant discrepancies with reality due to factors such as sensor noise and delays. As a result, long-term glucose prediction remains poor and continuous to be a very challenging task despite the increase of data availability (Sun *et al.*, 2021).

Chronic hyperglycemia is the main risk factor for the development of complications in diabetes mellitus, however, it is believed that large or frequent glucose fluctuations may contribute independently to these complications. Glycemic variability (GV) refers to this fluctuation of glucose levels, describing variations throughout the day, including hypoglycemic episodes and postprandial increases, as well as variations in glucose levels at different times of the day and at the same time on different days (Henao-Carrillo *et al.*, 2018; Kovatchev, 2019).

Glycemic control can be assessed by continuous glucose monitoring (CGM) using time in range (TIR),

*Josep Vehi

2.3. PREDICCIÓN INDIVIDUALIZADA DE ESTADO GLUCÉMICO MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

serving as a surrogate for glycated hemoglobin (HbA1c) for use in clinical management (ElSayed *et al.*, 2022). Compositional data (CoDa) are data that transmit information about the parts of a whole expressed in proportions or percentages, as is the case of the vector of daily times in each of the glucose ranges: time below range (TBR) (< 70 mg/dl), TIR (70 – 180 mg/dl) and time above range (TAR) (> 180 mg/dl) (ElSayed *et al.*, 2023), where all the components are positive and of constant sum. Research has been published in which the time profiles in different glucose ranges are treated as a composition (Biagi *et al.*, 2019; Biagi *et al.*, 2021; Cabrera *et al.*, 2023). Several studies have developed models for prediction in the field of diabetes, such as the relationship between HbA1c and glucose values, adaptive adjustment of bolus calculator parameters, and glucose prediction (Vigersky and McMahon, 2019; Vettoretti *et al.*, 2020; Noaro *et al.*, 2020). In the literature, regression models for the prediction of diabetes have been previously reported (Khanam and Foo, 2021). In (Makroum *et al.*, 2022), a total of 89 studies published between 2011 and 2021 were included.

Although regression analysis is a widely used statistical technique, there is limited literature available when it comes to CoDa (Aitchison, 1982; Egozcue *et al.*, 2003; Egozcue *et al.*, 2012; Van den Boogaart and Tolosana-Delgado, 2013; Pawlowsky-Glahn *et al.*, 2015; Thió i Fernández de Henestrosa and Martín Fernández, 2015; Fišerová *et al.*, 2016). No research has been found that specifically examines the application of CoDa to individualized regression models for diabetes. None of them were related to glucose prediction, mean or coefficient of variation (CV). Although short-term prediction reviews have been found, there are not many publications with relevant metrics for long-term glycemic state predictions (Ståhl and Johansson, 2009; Mhaskar *et al.*, 2017; Rodríguez-Rodríguez *et al.*, 2019; Katsarou *et al.*, 2022).

This study presents personalized multiple regression models for each hour of the day for predicting BG for long-term. The models incorporate a CoDa type regressor (TBR, TIR, TAR) in addition to other scalar variables that were useful in distinguishing when compositional variables were similar. The dependent variables in the models are the mean and CV of glucose measurements for the next 2 and 4 hours.

2. Materials and methods

2.1. Dataset. The REPLACE-BG dataset, publicly available (NCT02258373) (Aleppo *et al.*, 2017), was employed and consists of 226 adult subjects with T1D who underwent CGM for 26 weeks. The study was conducted between May 2015 and March 2016 in adult participants with T1D of more than 1-year duration

and with HbA1c of 9.0% (75 mmol/mol) or less. All participants used the Dexcom G4 Platinum CGM system with a modified algorithm and the Abbott Precision Xtra BG and ketone meter.

Data Preprocessing The CGM measurements of the patients' glucose profiles contain unrecorded measurements, thus the data were linearly interpolated when the missing data gap did not exceed 30 consecutive minutes (6 measurements). Subsequently, the measurements were organized for the 2h before and 2h and 4h after each hour of the day (00h, 01h, 02h, ..., 23h) (Fig. 1). With these measurements already divided into groups of 2h and 4h, the times in the different glucose ranges for the 2h prior to the prediction were calculated, which were treated as three-part CoDa (< 70, 70 – 180, > 180) whose sum is constant at 100%.

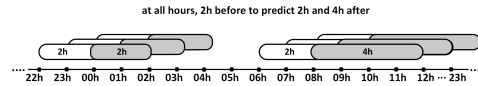


Fig. 1. The distribution of 2h periods prior to the prediction (white) and the following 2h and 4h periods (gray). The 2h period preceding the prediction was treated as a three-part CoDa.

2.2. CoDa. A compositional vector of D parts, whose sample space is the simplex S^D , is defined as a vector in which the only relevant information is contained in the relationships between its components (Eq. (1)). One way to simplify the use of compositions is to represent them in closed form, that is, as positive vectors, whose parts add up to a positive constant k (in our case 100%). From any vector it is possible to obtain a composition X of S^D by conveniently scaling the components so that their sum is equal to constant k . In other words, applying the closure operator defined by equation 2 (Aitchison, 1982).

$$\begin{cases} x_i > 0, & i = 1, 2, \dots, D, \\ \sum_{i=1}^D x_i = k & k = cte, \end{cases} \quad (1)$$

$$\begin{aligned} \mathbb{C}(X_1, X_2, \dots, X_D) \\ = \left(\frac{X_1}{\sum_{i=1}^D X_i}, \frac{X_2}{\sum_{i=1}^D X_i}, \dots, \frac{X_D}{\sum_{i=1}^D X_i} \right). \end{aligned} \quad (2)$$

The importance of the scale invariance principle has been demonstrated, where the value of k is not relevant, and it has been observed that its practical implementation requires working with component ratios. Therefore, the analysis of logarithmic ratios was implemented for composition problems. Logarithms of ratios are mathematically more manageable than ratios, which has

led to the use of log-ratio functions for obtaining the components (Aitchison, 1982).

The centered logratio function (*clr*) (Eq. (3)) is symmetric, where $g(x)$ is the geometric mean (Eq. (4)).

$$\begin{aligned} clr(x) &= [clr_1x, clr_2x, \dots, clr_Dx] \\ &= \left[\ln \frac{x_1}{g(x)}, \ln \frac{x_2}{g(x)}, \dots, \ln \frac{x_D}{g(x)} \right]. \end{aligned} \quad (3)$$

$$g(x) = \left(\prod_{i=1}^D x_i \right)^{\frac{1}{D}}. \quad (4)$$

Let e_1, e_2, \dots, e_{D-1} be an *olr*-basis in S^D , the function that assigns coordinates with respect to e_1, e_2, \dots, e_{D-1} to a composition $x \in S^D$ is called the isometric transformation logratio *ilr*: $S^D \rightarrow \mathbb{R}^{D-1}$ (Eq. (5)) (Egozcue *et al.*, 2003). The *olr* base associated with a Sequential Binary Partition (SBP) can be defined in several ways. The word isometric in *ilr* refers to the preservation of distance. In (Martín-Fernández, 2019) the name *olr* was introduced to avoid confusion because the *clr* transformation is also an isometric log-ratio transformation.

$$ilr(x) = [\langle x, e_1 \rangle_a, \dots, \langle x, e_{D-1} \rangle_a]. \quad (5)$$

2.3. Regression model with CoDa. In general, there are three types of linear regression models (LRM) that involve CoDa (Aitchison, 1982; Hron *et al.*, 2012; Navarro-Lopez *et al.*, 2022). Type I (multivariate model) has a composition as the response and one or more non-compositional (scalar) variables as explanatory (Aitchison, 1982). Type II has composition as explanatory and a non-compositional response; if the response is univariate, it is a multiple LRM. Finally, type III has both composition as explanatory and composition as response, becoming a multivariate multiple LRM. For each type, the regression model can be constructed using the Euclidean structure of the simplex or the *olr* coordinates or transformed *clr* scores. However, since there are infinite possibilities to construct *olr* coordinates (Egozcue *et al.*, 2003), it is important to focus on those that allow interpretation of the model and the corresponding regression coefficients.

2.3.1. LRM with compositional predictor and scalar response. Multiple linear regression (MLR) models are a statistical technique widely used to predict a response variable (y) from one or more explanatory variables (x). In the context of an MLR model, the compositional vector x (belonging to the simplex composition space, S^D) is used as the explanatory variable of the model to predict

the response variable y . In this type of models, no statistical assumptions are made about the composition of x , but only about the residuals u of the response variable y that is being predicted. It is assumed that the residuals are normally distributed and have constant variance. Residual diagnostics are performed in the same way as in a standard MLR and a single equation model is fitted, whose coefficient of determination (\mathbb{R}^2) is directly interpretable (Coenders and Pawlowsky-Glahn, 2020).

Steps for the creation of the model based on CoDa

1. An *olr* base is selected in S^D using an SBP (Table 1) (Coenders and Pawlowsky-Glahn, 2020).

Table 1. Sequential binary partition.

i	$X_{<70}$	X_{70-180}	$X_{>180}$	$p(+)$	$n(-)$
1	+1	+1	-1	2	1
2	+1	-1	0	1	1

2. The predictor is represented in *olr*-coordinates (Eq. (6)). The compositions are by definition multivariate and therefore must be mapped in some way, linear or non-linear, to a single number. To compute such a regression model, the principle of working in coordinates is used to transform the model into a multiple regression problem. The *olr*-coordinate x^* , of a composition x with respect to a base linked to a SBP is calculated as Eq. (7).

$$x_j^* = olr(x) \in \mathbb{R}^{D-1}. \quad (6)$$

$$x_j^* = \sqrt{\frac{p_j * n_i}{p_j + n_i}} \ln \frac{(x_{k_1} \dots x_{k_{p_j}})^{1/p_j}}{(x_{l_1} \dots x_{l_{n_j}})^{1/n_j}}. \quad (7)$$

where k_1, \dots, k_{p_j} are the labels of the parts in the numerator (encoded by +1 in the i -th row of SBP), l_1, \dots, l_{n_j} are the labels of the parts in the denominator (encoded by -1 in the same row) and $j : 1, \dots, D - 1$.

3. The ordinary regression model is solved to obtain the coefficients β^* with Eq. (8) for $i = 1, \dots, n$.

$$y_i = \beta_0 + \beta_1^* x_{i1}^* + \dots + \beta_{D-1}^* x_{iD-1}^* + u_i. \quad (8)$$

The *ilr* transformation has been used as it satisfies the requirement that the analysis has to be permutation invariant. On the other hand, the *clr* transformation is not easy to interpret with compositional explanatory variables because it produces numerical problems with singular matrices in the tests (Van den Boogaart and Tolosana-Delgado, 2013).

2.3. PREDICCIÓN INDIVIDUALIZADA DE ESTADO GLUCÉMICO MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

2.3.2. Data preprocessing. The compositional input could contain zeros if some of the parts of the CoDa vector were zero, therefore a pre-treatment was done since CoDa is based on log-ratios of parts. The detection matrix (dl) used in the imputation of the zeros was interpreted as in (Cabrera *et al.*, 2023) taking into account the consecutive zeros. In this case, where we are only analyzing three parts, there could only be 2 consecutive zeros, the dl value will then be calculated by dividing 5 min (sensor measurement interval) by 120 min, which is the time analyzed from the previous 2h, $dl = 0.04166$. We consider that the further zero is from the non-zero value, the smaller this value should be in the dl matrix, as presented in Table 2. To make the replacement, we used the multRepl (Multiplicative simple replacement) function implemented in the package “zCompositions” of R, this method provides a compositional counterpart to the common simple substitution by a fixed fraction of the censoring threshold. The remaining components are multiplicatively adjusted to preserve the relative multivariate structure of the data (Martín-Fernández *et al.*, 2003; Palarea-Albaladejo and Martín-Fernández, 2015; Martín-Fernández *et al.*, 2015). The scalar variables have also been transformed (function ln) beforehand to estimate the ordinary regression models (Step 3). This decision is shared by all LRMs because it is an option due to the nature of the covariate (sample space, distribution, etc.). In addition, an outlier analysis could be performed at this step (Coenders and Pawlowsky-Glahn, 2020).

Table 2. Detection limit matrix for 2 and 4h.

Consecutive zeros	Position 1	Position 2
For 2h, $5/120 = 0.04166$		
1	dl= 0.04166	
2	dl/3= 0.01388	2dl/3=0.0277
For 4h, $5/240 = 0.02083$		
1	dl= 0.02083	
2	dl/3= 0.00694	2dl/3=0.01388

In this study, 24 multiple LRMs were implemented for each hour of the day. With compositional input based on the time vector in each of the glucose ranges that we have treated as CoDa from 2h before and scalar outputs of the mean and the CV of glucose 2h and 4h later. The R-Studio program was used to develop and validate the models. A validation was carried out with 80% of the data for training and the remaining 20% for validation. Although it is an individualized model, the results are presented for the entire cohort.

The study methodology is described in Fig. 2, which includes the analysis of BG measurements, data processing and implementation of the multiple regression model, whose inputs are the $olr(x)$,

$olr2(x)$ corresponding to the CoDa vector (TBR, TIR, TAR), transformed scalar variables (mean, CV, minimum and maximum) of the 2h before and the outputs are the transform of the mean and CV of 2 and 4 hours after, model validation and finally the application of “traffic light with symbols” as a decision support system (DSS).

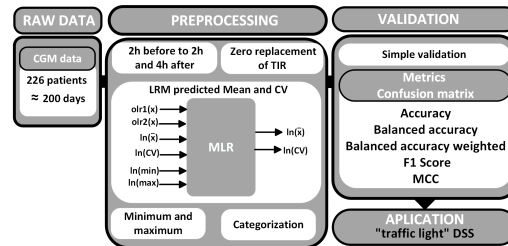


Fig. 2. Methodology for data analysis, validation, and application.

2.4. Confusion matrix. Metrics for multi-class classification. Classification tasks in machine learning that involve more than two classes are known as “multi-class classification”. Performance indicators are very useful when the objective is to evaluate and compare different classification models or machine learning techniques (Rácz *et al.*, 2019; Grandini *et al.*, 2020; Tharwat, 2021). The confusion matrix is a cross tab (Table 3) that records the number of occurrences between two raters, the actual classification (rows), and the predicted classification (columns). The classes are listed in the same order in the rows as in the columns, therefore, the correctly classified elements are located on the main diagonal of the matrix and correspond to the number of times the prediction was accurate.

		Predicted data		
		class	A	B
Real data	A	AA	AB	AC
	B	BA	BB	BC
	C	CA	CB	CC

Table 3. Confusion matrix.

2.4.1. Accuracy, balanced accuracy and balanced accuracy weighted. The accuracy metric (Eq. (9)) considers the sum of the true positive (TP) and true negative (TN) elements in the numerator, which are the elements correctly classified by the model (main diagonal), and in the denominator the sum of all the confusion matrix entries that have been incorrectly classified by the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

Balanced accuracy (BA) is an evaluation measure commonly used to assess the performance of a classification model, especially in cases where the training data is unbalanced. It is defined as the average of the recall obtained in each true class, where recall is the proportion of true positives correctly classified over the total number of samples in that class. It takes into account the imbalance in the training data and provides a fairer evaluation of the model performance by considering all classes equally. Then Eq. (10) describes the BA from Table 3.

$$BA = \frac{\frac{AA}{\sum row_1} + \frac{BB}{\sum row_2} + \frac{CC}{\sum row_3}}{total_{class}} \quad (10)$$

The BA weighted (BAW) (Eq. (11)) takes advantage of the BA formula by multiplying each recall by the weight of its class, i.e., the frequency of the class in the entire dataset. This metric allows us to keep separate algorithm performances across the different classes, so that we can track which class is causing poor performance. At the same time, it keeps track of the importance of each class thanks to the frequency. This ensures obtaining a reliable value of the overall performance of the dataset: we can interpret this metric as the probability of correctly predicting a given unit (Grandini *et al.*, 2020).

$$BAW = \frac{\frac{AA}{\sum row_1} * AA + \frac{BB}{\sum row_2} * BB + \frac{CC}{\sum row_3} * CC}{AA + BB + CC} \quad (11)$$

2.4.2. F1-Score for multi-class classification. The F1-Score evaluates the performance of a classification model based on the confusion matrix and combines measures of accuracy and recall under the concept of harmonic mean. In cases of multiple classes, F1-Score should involve all classes. To achieve this, a multiclass measure of accuracy and recall needs to be incorporated into the harmonic mean. Such metrics can have two different specifications, leading to two different metrics: Micro F1-Score and Macro F1-Score.

To calculate **Macro F1-Score**, we first need to calculate Macro-Precision and Macro-Recall. Macro-Precision is computed by taking the average of the precision (Eq. (12)) for each predicted class, while Macro-Recall is computed by taking the average of the recall (Eq. (13)) (also known as sensitivity) for each real class. The Macro approach considers all classes equally and does not distinguish between classes with high or low population. To perform the calculations, we use the confusion matrix and focus on one class at a time,

labeling the tiles accordingly. TP represents the units correctly classified for the given class (i.e., the main diagonal of each case), while False Positive (FP) and False Negative (FN) represent misclassified elements in the column and row, respectively. True Negative (TN) represents the rest of the classifications.

$$Precision_{class} = \frac{TP_{class}}{TP_{class} + FP_{class}} \quad (12)$$

$$Recall_{class} = \frac{TP_{class}}{TP_{class} + FN_{class}} \quad (13)$$

The Macro Average Precision (MaAP) and Macro Average Recall (MaAR) are simply calculated as the arithmetic mean of the metrics for individual classes such as Eq. (14,15). Then, Macro F1-Score is the harmonic mean of MaAP and MaAR (Eq. (16)).

$$MaAP = \frac{\sum_{class=1}^{class_{max}} Precision_{class}}{class_{max}} \quad (14)$$

$$MaAR = \frac{\sum_{class=1}^{class_{max}} Recall_{class}}{class_{max}} \quad (15)$$

$$Macro\ F1\text{-Score} = 2 * \frac{MAP * MAR}{MAP + MAR} \quad (16)$$

To get **Micro F1-Score**, we need to calculate Micro Average Precision (MiAP) (Eq. (17)) and Micro Average Recall (MiAR) (Eq. (18)) before. The idea is to consider all units together, without taking into account possible differences between classes.

$$MiAP = \frac{\sum_{class=1}^{class_{max}} TP_{class}}{total\ per\ column} \quad (17)$$

$$MiAR = \frac{\sum_{class=1}^{class_{max}} TP_{class}}{total\ per\ row} \quad (18)$$

Both formulas are based on the TP in the numerator and in the denominator the total sum both by rows and by columns coincides, therefore, clearing in the F1-Score equation we are left:

$$Micro\ F1\text{-Score} = \frac{\sum_{class=1}^{class_{max}} TP_{class}}{Total} \quad (19)$$

2.4.3. Matthews Correlation Coefficient. The Matthews Correlation Coefficient (MCC) metric undergoes some changes from binary classification. It is defined in terms of a confusion matrix C (Eq. (20)).

$$MCC = \frac{\sum_{i=1}^K \sum_{j=1}^K n_{ii} n_{jj} - \sum_{i=1}^K n_{ii} \sum_{j=1}^K n_{ij}}{\sqrt{(\sum_{i=1}^K n_{ii} - \sum_{i=1}^K \sum_{j=1}^K n_{ij}) (\sum_{i=1}^K n_{ii} - \sum_{j=1}^K n_{ij})}} \quad (20)$$

2.3. PREDICCIÓN INDIVIDUALIZADA DE ESTADO GLUCÉMICO MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES



A. Cabrera *et al.*

where K is the number of classes, n_{ii} is the number of samples correctly classified in class i , and n_{ij} is the number of samples that were classified as i but belong to class j . The numerator of the formula represents the covariance between the predictions and the true labels, while the denominator is a normalization to bring the result in the range $[-1, 1]$.

3. Results

Below are the detailed results for glucose mean and CV prediction as well as metrics for the classification and DSS.

3.1. Overall LRM test results. Compared to univariate linear regression, it is not possible to display the strength of the relationship between multiple composition variables (orthogonal basis of different time in ranges of glucose) and a dependent variable (mean, CV) in a single XY scatter plot because X has several potentially influential components (Van den Boogaart and Tolosana-Delgado, 2013).

To test the normality assumption of the residuals, the Shapiro-Wilk test was used, which showed a p-value > 0.05 , suggesting that we cannot reject the null hypothesis that the data come from a normally distributed population.

Non-Constant Variance Score Test and Breusch-Pagan test were performed to verify the homoscedasticity assumption, that is, "all errors have the same variance". The results showed a p-value > 0.05 , suggesting that the homoscedasticity assumption is met. Additionally, the independence assumption of the errors was checked using the Durbin Watson test, and no evidence of violation of this assumption was found (p-value > 0.05).

3.2. Validation of the multivariable LRM of mean and CV prediction. The results are presented in terms of root mean squared error (RMSE) and mean absolute error (MAE) to estimate performance and evaluate the model fit for the entire cohort at different times of the day. Figure 3 shows the results for the mean and CV prediction model for the next 2h and 4h. We analyzed both errors since the MAE error is more robust and does not give much importance to outliers, unlike the RMSE, which gives more importance to outliers by squaring the absolute value of the difference. As expected, the RMSE error is higher than the MAE error.

The results show that for the CV prediction, both the RMSE and MAE errors for all models were higher when predicting the next 4h than when predicting the next 2h. However, this did not happen with the mean glucose prediction, which remained more uniform.

It is very useful to identify glycemic trends at different times of the day, quantify glycemic variability,

and stratify the risk of hypoglycemia based on the hours. In the early morning hours (01:00 to 08:00 h), the RMSE and MAE errors were lower for the Mean model compared to the rest of the hours. Similarly, for the CV model, the RMSE error during the hours from 00:00 to 07:00 h was lower than the rest of the hours, and the MAE error was lower from 23:00 to 07:00 h. This shows that our model is capable of predicting early morning hours with higher reliability (lower errors). This factor is significant both the risk of experiencing nocturnal hypoglycemia and the dawn phenomenon, which typically happens between 04:00 - 08:00 h in the morning.

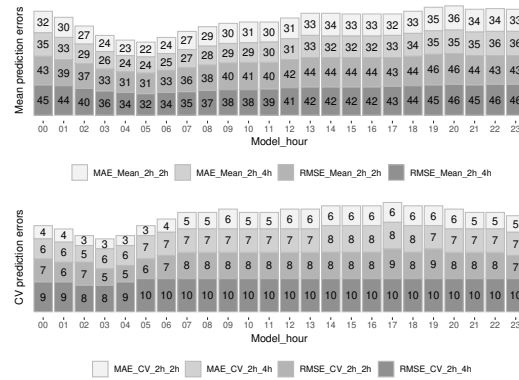


Fig. 3. RMSE and MAE results from the mean and CV prediction model for the next 2h and 4h.

Also, the distributions between the real and predicted means and CV were compared to detect if there were differences between them. The Kolmogorov-Smirnov statistic was used. The main advantage of this statistic is that it is sensitive to differences in both the location and shape of the cumulative distribution function. The results showed a p-value > 0.05 in all time periods, suggesting that we cannot reject the null hypothesis that the analyzed data follows the same distribution.

3.3. Prediction of minimum and maximum glucose. CV is calculated according to Eq. (21), it is a measure of variability relative to the mean (Gulhar *et al.*, 2012), solving for Eq. (22) (glucose standard deviation (STD)) is obtained. Knowing the mean and CV previously predicted by the multiple LRM for the next 2h and 4h. Under the assumption of normality, it can be said that the minimum and maximum glucose values are in the range of $\pm 3STD(99.7\% \sim 100\%)$, where \bar{x} is the mean glucose.

$$CV = \frac{STD}{\bar{x}} * 100 \quad (21)$$

$$STD = \frac{CV * \bar{x}}{100} \quad (22)$$

3.4. Application, example of the “traffic light with symbols” proposed for a specific patient. “Traffic light” systems for clinical information and clinical support are well known (Evans *et al.*, 2021; Garden *et al.*, 2021). A methodology was implemented to categorize the results of the predicted glucose levels at 2 and 4 hours for each hour of the day, taking into account the mean, minimum, and maximum glucose levels, following the standards defined in (ElSayed *et al.*, 2022). To define the categories, specific limits were evaluated: for three categories, the following limits were used: < 70 mg/dl, 70-180 mg/dl and, > 180 mg/dl. For 5 categories, more restrictive limits were used: < 54 mg/dl, 54-70 mg/dl, 70-180 mg/dl, 180-250 mg/dl and > 250 mg/dl. An example of application is presented below for patient 1 from one day with high variability.

3.4.1. Patient 1, day 3 characterized by high variability. Table 4 presents an example of the proposed “traffic light with symbols” system for patient 1. We have analyzed day 3 as it is a day with high glucose variability (36.53%), severe hyperglycemia both during the day and at night, and also the presence of hypoglycemia. Column 4 shows the description for each of the previously mentioned class. Analyzing the predictions of the states for 3 class (column 2 of Table 4), it can be seen that from 00:00 to 18:00 h, every hour in that interval, the model predicted that the patient would be there for the next 2h in hyperglycemia (> 180 mg/dl), the actual states validate that the model was correct every time. During the night period, from 22:00 h of the previous day to 8:00 h, this patient experienced a glucose variability of 6.5%, with a minimum reading of 269 mg/dl and a maximum of 371 mg/dl, indicating severe hyperglycemia.

From 19:00 to 20:00 h, he was in the target glucose range (70-180 mg/dl), a situation that the model also correctly predicted. However, from 21:00 to 23:00 h, the patient was in hypoglycemia, a situation predicted by the model.

Still considering the prediction of 2h, by analyzing the results for 5 class, from 00:00 to 17:00 h, the model predicted severe hyperglycemia, being more specific than when it was analyzed for 3 class. It was found that the minimum glucose was 244 mg/dl and the maximum was 329 mg/dl, and the CV for 2h was between 2% and 8%. However, at 18:00 h the model predicted risk of hyperglycemia, here we verified that the patient had a minimum of 70 mg/dl and a maximum of 321 mg/dl with a CV of glucose for the next 2h of 40%, and vector time in range was 0% below 70 mg/dl, and 50% for both TIR and hyperglycemia above 180 mg/dl, that is, half of the

next 2h was spent time in normoglycemia and the rest in hyperglycemia.

Hence, at 19:00 and 20:00 h, the patient will behave in range time. At 21:00 and 22:00 h, the model predicted risk of hypoglycemia; however, the validation corroborated that it was accurate for 21:00 h, but for 22:00 h, the real state reported severe hypoglycemia, the time vector in range glucose reported 66% of time below 70 mg/dl, 33.3% in TIR and 0% above 180 mg/dl. For 23:00 h both the model and reality reported severe hypoglycemia. In practice, as we have shown in this example, it is expected that the patient will have the 24 models for each hour of the day, and the prediction model will update him on his future status for the next 2h.

Figure 4 displays the BG measurements for Patient 1 for day 3. This day showed severe hyperglycemia for over 50% of the time, with the first minimum peak at 70 mg/dl occurring at 20:00 h, increasing glucose levels and remaining in range until 22:00 h before dropping to hypoglycemia level 1 with few normoglycemic measurements.

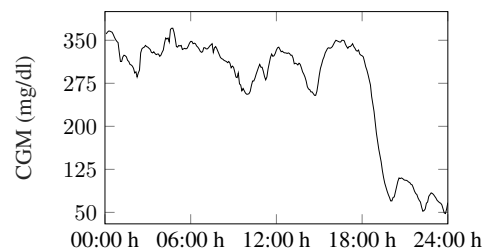


Fig. 4. BG of patient 1, day 3.

3.5. Results of the metrics for multi-class classification. Once the actual and predicted data from the validation data were classified, the confusion matrix was created for each of the 24 models and each of the 226 patients. Although this is an individualized model, the metrics results are shown for the entire cohort. Figure 5 shows the results for accuracy, BA, BAW, sensitivity, Macro and Micro F1-Score, each of the results will be discussed below.

3.5.1. Accuracy Results. The accuracy returns a general measure of how correctly the model predicts for all samples, the results for the entire cohort are shown in the boxplot in Fig. 5 (first graph on the left).

The diagrams show the results of the predictions of the 24 models (M_00, M_01, ..., M_23) corresponding to each hour of the day. The prediction of 2h and 4h with 3 and 5 classes. This type of graph allows us to identify outliers and compare distributions, as well as knowing

2.3. PREDICCIÓN INDIVIDUALIZADA DE ESTADO GLUCÉMICO MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

Hour	3 class		5 class		Characteristics of the states
	Predicted state	Real state	Predicted state	Real state	
00:00	●	●	⊗	⊗	<p>3 class</p> <p>Categorization for 3 classes</p> <p>⊗ Period characterized by risk of hyperglycemia</p> <p>⊙ Period characterized by tendency to normoglycemia</p> <p>⊗ Period characterized by risk of hypoglycemia</p> <p>5 class</p> <p>Categorization for 5 classes</p> <p>⊗ Period characterized by risk of hyperglycemia level 2</p> <p>⊙ Period characterized by risk of hyperglycemia level 1</p> <p>⊙ Period characterized by tendency to normoglycemia</p> <p>⊙ Period characterized by risk of hypoglycemia level 1</p> <p>⊗ Period characterized by risk of hypoglycemia level 2</p>
01:00	●	●	⊗	⊗	
02:00	●	●	⊗	⊗	
03:00	●	●	⊗	⊗	
04:00	●	●	⊗	⊗	
05:00	●	●	⊗	⊗	
06:00	●	●	⊗	⊗	
07:00	●	●	⊗	⊗	
08:00	●	●	⊗	⊗	
09:00	●	●	⊗	⊗	
10:00	●	●	⊗	⊗	
11:00	●	●	⊗	⊗	
12:00	●	●	⊗	⊗	
13:00	●	●	⊗	⊗	
14:00	●	●	⊗	⊗	
15:00	●	●	⊗	⊗	
16:00	●	●	⊗	⊗	
17:00	●	●	⊗	⊗	
18:00	●	●	●	●	
19:00	⊙	⊙	⊙	⊙	
20:00	⊙	⊙	⊙	⊙	
21:00	⊗	⊗	●	●	
22:00	⊗	⊗	●	⊗	
23:00	⊗	⊗	⊗	⊗	

Table 4. Example of “traffic light with symbols” for patient 1, day 3 with 3 and 5 class to predict the next 2h

in a comfortable and fast way how 50% of the central values are distributed. The dimensions of the boxes are determined by the distance of the 25th – 75th interquartile ranges. At all times, these distances were greater when the prediction horizon (PH) was longer (4h) and they increased for the 5-class categorization.

For the prediction of 2h, 3 and 5 classes, it can be evidenced that the median is located in the center of the box, then the distribution is symmetric and both the mean, median and mode coincide. Except for 2h 3 class (M_04, M_07, M_11, M_14, M_20) and for 2h, 5 class (M_07, M_08). For the prediction of 4h 3 classes for schedules

M_00, M_06 to M_18, negative asymmetry was shown since the longest part is the lower part of the median, therefore the data were concentrated in the upper part of the distribution, here the mean is usually less than the median, this shows dispersion in the data, not a greater amount. For the prediction of 2h and 4h for 3 classes at all times, an accuracy greater than 85% was reached at all times of the day with a 75th quartile close to 100%. For 5 classes, the 4h forecast presented better performance, although the data were more dispersed, with a 75th close to 90% for all times.

2. PREDICCIÓN PROBABILÍSTICA PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

Individualized prediction of blood glucose outcomes using Compositional Data analysis

9 amcs

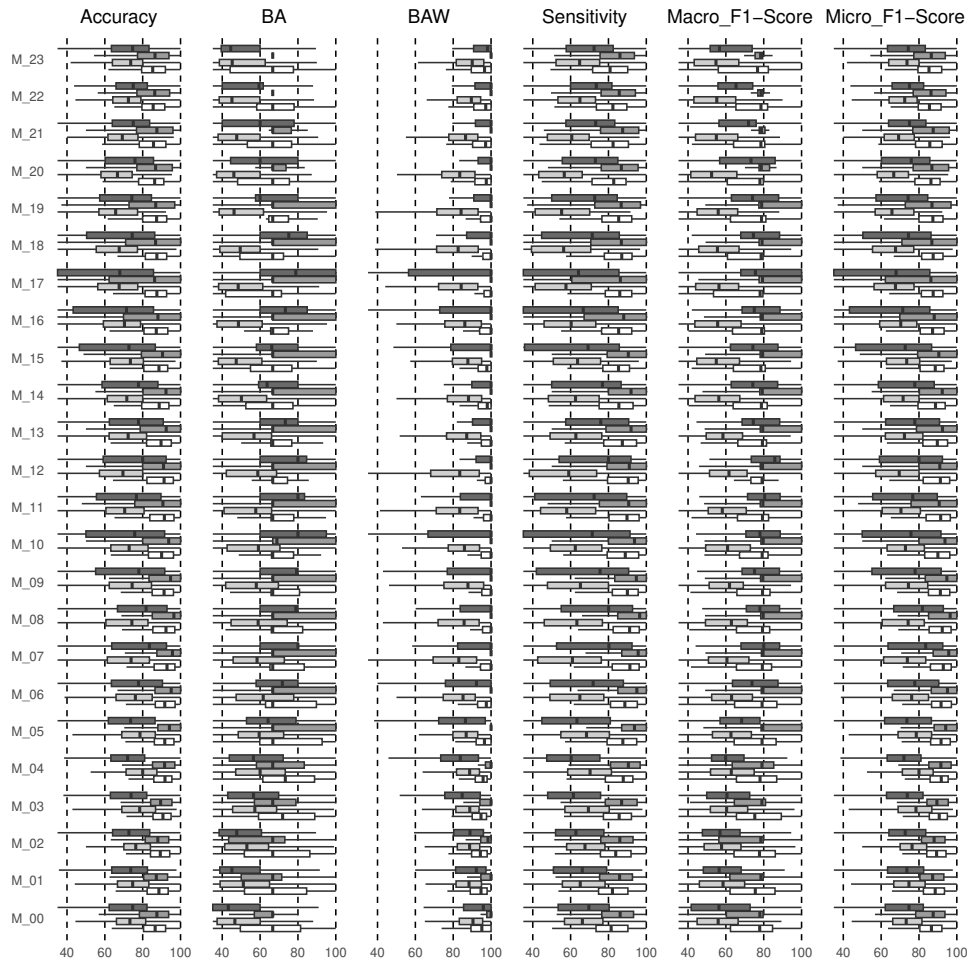


Fig. 5. Results for multi-class classification, 2h_3 classes, 2h_5classes, 4h_3classes and 4h_5classes (from the bottom up, from white to the darkest gray).

3.5.2. Balanced Accuracy and Balanced Accuracy Weighted Results. Figure 5 show the results of the BA and BAW respectively (second and third graph from left to right). The results of the BA for 2h, 3 classes for schedules M_00 to M_05 and M_15 behave symmetrically, however the model for schedules M_13, M_17, M_20, M_21 and M_23 show negative asymmetry. For the 4h forecast, except for the hours M_00 to M_02, there was positive asymmetry. And for M_22 and M_23 all the results were concentrated in the median. At all times, the 75th quartile was above 70%. For the prediction to 5 classes, symmetry was observed only for 2h in M_03 and M_04, M_07 to M_10, for the rest, generally positive asymmetry. Here the 75th quartile was above 60%, however it improved for

the 4h forecast, exceeding 80%. The results for 3 classes are satisfactory, although no symmetric distribution was observed in the results for any model. In all cases the median was greater than 90% and the 75th quartile close to 100%. For the prediction with 5 classes, the results were observed to be more dispersed, especially in hours from M_08 to M_10, M_16 and M_17. Symmetry was not observed.

3.5.3. Sensitivity Results. The results show that for the prediction with 3 classes the median was above 80% in all cases, with the 75th quartile close to 100%. However, the cohort data were more dispersed when 5 classes were evaluated, finding the median close to 75% for all hours

2.3. PREDICCIÓN INDIVIDUALIZADA DE ESTADO GLUCÉMICO MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

and with a greater dispersion in daytime hours from M.05 to M.20 (Fig. 5 (fourth graph from left to right)).

3.5.4. F1-Score. In this study for the prediction with 3 classes, the results of the median for the entire model for the prediction at both 2h and 4h was higher than 80% and a 75th quartile close to 100%, thus same in the hours from M.05 to M.19 indicating that the algorithm performs well in all classes. However, for 5 classes the median for 2h in all cases was above 60% but for 4h in some cases above 70% (M.06 to M.21).

Micro-average considers all units together, without taking into account possible differences between classes just like accuracy. Both measures give more importance to large classes, because they only consider all units together. In our case, all classes are important, we should not underestimate the small ones. In addition, at some times the large classes for our model are usually TIR, which, although they provide information, do not suggest any corrective action. Even so, the results showed a median higher than 8% and 75% for when there are 3 classes and 5 classes respectively. Very scattered results were not observed in any case, although there was a difference between the prediction with 3 and 5 classes.

3.5.5. Mattheus Correlation Coefficient for multi-class classification. Among the advantages of this metric we can see that MCC includes all the entries of the confusion matrix both in the numerator and in the denominator (Vehí *et al.*, 2020; Parcerisas *et al.*, 2022). Our results (Fig. 6) show that for the prediction with 3 classes, especially for 4h, the median for the hours from M.05 to M.23 was 1, indicating a perfect prediction. However, for 5 classes, such a median was only obtained for the models from M.07 to M.22 for 4h. The rest of the hours, the median was close to 0.5 (greater than 0.5 is considered good). Some isolated cases where it was close to 0, which corresponds to a random prediction of the model and some very isolated samples were below zero, which indicated a totally incorrect prediction. For 3 classes it could be considered as an accurate model, however for some schedules of 5 classes it indicates that the model is not better than a random prediction.

4. Discussion

DSS have proven to be useful tools for patients and physicians (Contreras and Vehí, 2018; Evans *et al.*, 2021; Garden *et al.*, 2021). Although glucose profiles have been treated as CoDa vectors in previous studies (Biagi *et al.*, 2019; Biagi *et al.*, 2021; Cabrera *et al.*, 2023), there is no application in this branch of mathematics that is focused on predicting the mean and the CV as an information system or DSS tool for patients with T1D at specific hours of the day oriented to wide PH (2h and

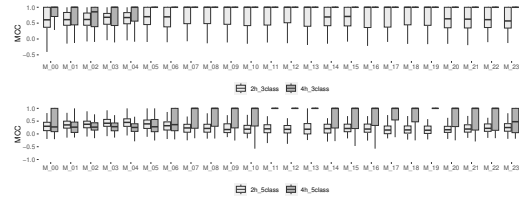


Fig. 6. MCC results for multiclass classification, 2- and 4h prediction.

4h). In this work, CoDa variables and transformed scalars have been used to predict the mean and CV of glucose in patients with T1D. In addition, the different times of the day of the patients have been categorized to provide an idea of the behavior of glucose in the next 2h and 4h. The results have been validated using a sample of 226 adult patients from a real cohort.

Although no study was found that predicted the mean and CV for patients with T1D at a PH of 120 and 240 minutes, studies have been reported that have focused on glucose prediction for horizons from 15 to 120 minutes (Mohebbi *et al.*, 2020; Martinsson *et al.*, 2020; Cichosz *et al.*, 2021; Daniels *et al.*, 2021; Tena *et al.*, 2021; Wadghiri *et al.*, 2022). As expected, the longer the forecast horizon, the greater the error. For a PH of 120 min, errors generally higher than 45 mg/dl have been reported ((De Bois *et al.*, 2019; Daniels *et al.*, 2021; Tena *et al.*, 2021; Wadghiri *et al.*, 2022)).

The results show that the MAE mean prediction error is between 23 – 36 mg/dl for all times, both when predicting at 2h and 4h. The CV is between 4-7% for the 2h prediction and between 6-8% for the 4h prediction. The RMSE and MAE prediction error of the mean and CV at all times of the day was higher for the 4h forecast horizon in the entire cohort, but the early morning times presented a lower error. It was confirmed that the CV at this time was lower than during the daytime hours.

Previous studies have used some of these metrics based on the confusion matrix to evaluate the performance of different methodologies (Vehí *et al.*, 2020; Parcerisas *et al.*, 2022). In (Vehí *et al.*, 2020) population outcomes for the mid-term continuous prediction module to predict hypoglycaemia and population outcomes for the nocturnal hypoglycaemic events predictor module are reported. An average mean of accuracy of 86.1% and 80.1%, respectively. A mean sensitivity of 48.5% and 44%, respectively was reported. Here a mean MCC of 0.51 with a minimum of -0.18 and a maximum of 0.86 for the mid-term continuous prediction module to predict hypoglycaemia. In (Parcerisas *et al.*, 2022) the results of the cohort (10 real patients) of the model using support vector machines including physical activity measures and

excluding physical activity measures were reported. A median sensitivity of 71% and 70% respectively, was reported. The median F1-scores for each individual patient were reported ranging from 37% (patient 12) to 80% (patient 45) and there were no significant changes in this metric when excluding physical activity measures was assessed. For the MCC, medians of 0.2 (patient 12) to 0.67 (patient 56) were reported.

The DSS provided interesting results in different metrics, such as accuracy, BA, BAW, sensitivity, F1-Score and MCC. They were higher than 90% for the entire cohort for 3 classes, but for 5 classes they decreased, obtaining results above 80%. Therefore, the system will be more reliable and accurate when 3 class are used according to some metrics. It should be noted that the results for the 4h prediction for both 3 and 5 classes were more dispersed, this evidences the variability of the cohort, even so, but they showed satisfactory results. The results presented in this paper refer to the entire cohort, yet, it is a personalized model, as some patients achieved superior results compared to others. Therefore, the outcomes are displayed in median and interquartile range format. Furthermore, a model is suggested for every hour of the day due to the impact of interday variability on hourly ranges, which must be taken into account. The authors foresee that this model should be updated and adjusted considering the habits and characteristics of individual patients over time.

Acknowledgment

This research was partially supported by grants PID2019-107722RB-C22 and PID2019-107722RB-C12 funded by MCIN/AEI/10.13039/501100011033, in part by the Autonomous Government of Catalonia under Grant 2017 SGR 1551, in part by the Spanish Ministry of Universities, and by the European Union through Next GenerationEU (Margarita Salas), and by the program for researchers in training at the University of Girona (IFUdG2019).

References

- Aitchison, J. (1982). The statistical analysis of compositional data, *Journal of the Royal Statistical Society: Series B (Methodological)* **44**(2): 139–160.
- Aleppo, G., Ruedy, K. J., Riddlesworth, T. D., Kruger, D. F., Peters, A. L., Hirsch, I., Bergenstal, R. M., Toschi, E., Ahmann, A. J., Shah, V. N. et al. (2017). Replace-bg: a randomized trial comparing continuous glucose monitoring with and without routine blood glucose monitoring in adults with well-controlled type 1 diabetes, *Diabetes care* **40**(4): 538–545.
- Biagi, L., Bertachi, A., Giménez, M., Conget, I., Bondia, J., Martín-Fernández, J. A. and Vehí, J. (2019). Individual categorisation of glucose profiles using compositional data analysis, *Statistical methods in medical research* **28**(12): 3550–3567.
- Biagi, L., Bertachi, A., Giménez, M., Conget, I., Bondia, J., Martín-Fernández, J. A. and Vehí, J. (2021). Probabilistic model of transition between categories of glucose profiles in patients with type 1 diabetes using a compositional data analysis approach, *Sensors* **21**(11): 3593.
- Cabrera, A., Biagi, L., Beneyto, A., Estremera, E., Contreras, I., Giménez, M., Conget, I., Bondia, J., Martín-Fernández, J. A. and Vehí, J. (2023). Validation of a probabilistic prediction model for patients with type 1 diabetes using compositional data analysis, *Mathematics* **11**(5): 1241.
- Cichosz, S. L., Kronborg, T., Jensen, M. H. and Hejlesen, O. (2021). Penalty weighted glucose prediction models could lead to better clinically usage, *Computers in Biology and Medicine* **138**: 104865.
- Coenders, G. and Pawlowsky-Glahn, V. (2020). On interpretations of tests and effect sizes in regression models with a compositional predictor, *SORT-Statistics and Operations Research Transactions* pp. 201–220.
- Contreras, I. and Vehí, J. (2018). Artificial intelligence for diabetes management and decision support: literature review, *Journal of medical Internet research* **20**(5): e10775.
- Daniels, J., Herrero, P. and Georgiou, P. (2021). A multitask learning approach to personalized blood glucose prediction, *IEEE Journal of Biomedical and Health Informatics* **26**(1): 436–445.
- De Bois, M., Ammi, M. and El Yacoubi, M. A. (2019). Model fusion to enhance the clinical acceptability of long-term glucose predictions, *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, IEEE, pp. 258–264.
- Egozcue, J. J., Daunis-I-Estadella, J., Pawlowsky-Glahn, V., Hron, K. and Filzmoser, P. (2012). *Simplicial regression. The normal model*, na.
- Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G. and Barcelo-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis, *Mathematical geology* **35**(3): 279–300.
- ElSayed, N. A., Aleppo, G., Aroda, V. R., Bannuru, R. R., Brown, F. M., Bruemmer, D., Collins, B. S., Cusi, K., Das, S. R., Gibbons, C. H. et al. (2023). Introduction and methodology: Standards of care in diabetes 2023.
- ElSayed, N. A., Aleppo, G., Aroda, V. R., Bannuru, R. R., Brown, F. M., Bruemmer, D., Collins, B. S., Hilliard, M. E., Isaacs, D., Johnson, E. L., Kahan, S., Khunti, K., Leon, J., Lyons, S. K., Perry, M. L., Prahalad, P., Pratley, R. E., Seley, J. J., Stanton, R. C. and Gabbay, Robert A., o. b. o. t. A. D. A. (2022). 6. glycemic targets: Standards of care in diabetes 2023, *Diabetes Care* **46**(Supplement1): S97–S110.
- Evans, M., Morgan, A. R., Patel, D., Dhatriya, K., Greenwood, S., Newland-Jones, P., Hicks, D., Yousef, Z., Moore, J., Kelly, B. et al. (2021). Risk prediction of the diabetes missing million: identifying individuals at high risk of

2.3. PREDICCIÓN INDIVIDUALIZADA DE ESTADO GLUCÉMICO MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

- diabetes and related complications, *Diabetes Therapy* **12**: 87–105.
- Fišerová, E., Donevska, S., Hron, K., Bábek, O. and Vaňkátová, K. (2016). Practical aspects of log-ratio coordinate representations in regression with compositional response, *Measurement Science Review* **16**(5): 235–43.
- Garden, G. L., Frier, B. M., Hine, J. L., Hutchison, E. J., Mitchell, S. J., Shaw, K. M., Heller, S. R., Koehler, G., Hofmann, V., Gaffney, T. P. et al. (2021). Blood glucose monitoring by insulin-treated pilots of commercial and private aircraft: An analysis of out-of-range values, *Diabetes, Obesity and Metabolism* **23**(10): 2303–2310.
- Grandini, M., Bagli, E. and Visani, G. (2020). Metrics for multi-class classification: an overview, *arXiv preprint arXiv:2008.05756*.
- Gulhar, M., Kibria, B. G., Albatineh, A. N. and Ahmed, N. U. (2012). A comparison of some confidence intervals for estimating the population coefficient of variation: a simulation study, *SORT-Statistics and Operations Research Transactions* pp. 45–68.
- Henaó-Carrillo, D. C., Muñoz, O. M., Gómez, A. M., Rondón, M., Colón, C., Chica, L., Rubio, C., León-Vargas, F., Calvachi, M. A. and Perea, A. M. (2018). Reduction of glycemic variability with degludec insulin in patients with unstable diabetes, *Journal of Clinical & Translational Endocrinology* **12**: 8–12.
- Hron, K., Filzmoser, P. and Thompson, K. (2012). Linear regression with compositional explanatory variables, *Journal of applied statistics* **39**(5): 1115–1128.
- Katsarou, D. N., Georga, E. I., Christou, M., Tigas, S., Papaloukas, C. and Fotiadis, D. I. (2022). Short term glucose prediction in patients with type 1 diabetes mellitus, *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, pp. 329–332.
- Khanam, J. J. and Foo, S. Y. (2021). A comparison of machine learning algorithms for diabetes prediction, *ICT Express* **7**(4): 432–439.
- Kovatchev, B. (2019). Glycemic variability: risk factors, assessment, and control, *Journal of diabetes science and technology* **13**(4): 627–635.
- Makroum, M. A., Adda, M., Bouzouane, A. and Ibrahim, H. (2022). Machine learning and smart devices for diabetes management: Systematic review, *Sensors* **22**(5): 1843.
- Martín-Fernández, J. A. (2019). Comments on: Compositional data: The sample space and its structure, *Test* **28**(3): 653–657.
- Martín-Fernández, J. A., Barceló-Vidal, C. and Pawlowsky-Glahn, V. (2003). Dealing with zeros and missing values in compositional data sets using nonparametric imputation, *Mathematical Geology* **35**: 253–278.
- Martín-Fernández, J.-A., Hron, K., Templ, M., Filzmoser, P. and Palarea-Albaladejo, J. (2015). Bayesian-multiplicative treatment of count zeros in compositional data sets, *Statistical Modelling* **15**(2): 134–158.
- Martinsson, J., Schliep, A., Eliasson, B. and Mogren, O. (2020). Blood glucose prediction with variance estimation using recurrent neural networks, *Journal of Healthcare Informatics Research* **4**: 1–18.
- Mhaskar, H. N., Pereverzyev, S. V. and Van der Walt, M. D. (2017). A deep learning approach to diabetic blood glucose prediction, *Frontiers in Applied Mathematics and Statistics* **3**: 14.
- Mohebbi, A., Johansen, A. R., Hansen, N., Christensen, P. E., Tarp, J. M., Jensen, M. L., Bengtsson, H. and Mørup, M. (2020). Short term blood glucose prediction based on continuous glucose monitoring data, *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, pp. 5140–5145.
- Navarro-Lopez, C., Linares-Mustaros, S. and Mulet-Forteza, C. (2022). The statistical analysis of compositional data by John Aitchison (1986): A bibliometric overview, *SAGE Open* **12**(2): 21582440221093366.
- Noaro, G., Cappon, G., Vettoretti, M., Sparacino, G., Del Favero, S. and Facchinetti, A. (2020). Machine-learning based model to improve insulin bolus calculation in type 1 diabetes therapy, *IEEE Transactions on Biomedical Engineering* **68**(1): 247–255.
- Palarea-Albaladejo, J. and Martín-Fernández, J. A. (2015). zcompositions r package for multivariate imputation of left-censored data under a compositional approach, *Chemometrics and Intelligent Laboratory Systems* **143**: 85–96.
- Parcerisas, A., Contreras, I., Delecourt, A., Bertachi, A., Beneyto, A., Conget, I., Viñals, C., Giménez, M. and Vehi, J. (2022). A machine learning approach to minimize nocturnal hypoglycemic events in type 1 diabetic patients under multiple doses of insulin, *Sensors* **22**(4): 1665.
- Pawlowsky-Glahn, V., Egozcue, J. J. and Tolosana-Delgado, R. (2015). *Modeling and analysis of compositional data*, John Wiley & Sons.
- Rácz, A., Bajusz, D. and Héberger, K. (2019). Multi-level comparison of machine learning classifiers and their performance metrics, *Molecules* **24**(15): 2811.
- Rodríguez-Rodríguez, I., Chatzigiannakis, I., Rodríguez, J.-V., Maranghi, M., Gentili, M. and Zamora-Izquierdo, M.-Á. (2019). Utility of big data in predicting short-term blood glucose levels in type 1 diabetes mellitus through machine learning techniques, *Sensors* **19**(20): 4482.
- Silva, J. A. d., Souza, E. C. F. d., Echazú Böschemeier, A. G., Costa, C. C. M. d., Bezerra, H. S. and Feitosa, E. E. L. C. (2018). Diagnosis of diabetes mellitus and living with a chronic condition: participatory study, *BMC Public Health* **18**: 1–8.
- Ståhl, F. and Johansson, R. (2009). Diabetes mellitus modeling and short-term prediction based on blood glucose measurements, *Mathematical biosciences* **217**(2): 101–117.
- Sun, X., Rashid, M., Askari, M. R. and Cinar, A. (2021). Latent variables model based mpc for people with type 1 diabetes, *IFAC-PapersOnLine* **54**(3): 294–299.

2. PREDICCIÓN PROBABILÍSTICA PARA PACIENTES CON DIABETES TIPO 1 MEDIANTE EL ANÁLISIS DE DATOS COMPOSICIONALES

Tena, F., Garnica, O., Lanchares, J. and Hidalgo, J. I. (2021). Ensemble models of cutting-edge deep neural networks for blood glucose prediction in patients with diabetes, *Sensors* **21**(21): 7090.

Tharwat, A. (2021). Classification assessment methods, *Applied computing and informatics* **17**(1): 168–192.

Thió i Fernández de Henestrosa, S. and Martín Fernández, J. A. (2015). *Proceedings of the 6th International Workshop on Compositional Data Analysis: Girona, 1-7 de juny de 2015*, Universitat de Girona. Departament d'Informàtica, Matemàtica Aplicada.

Van den Boogaart, K. G. and Tolosana-Delgado, R. (2013). *Analyzing compositional data with R*, Vol. 122, Springer.

Vehí, J., Contreras, I., Oviedo, S., Biagi, L. and Bertachi, A. (2020). Prediction and prevention of hypoglycaemic events in type-1 diabetic patients using machine learning, *Health informatics journal* **26**(1): 703–718.

Vettoretti, M., Cappon, G., Facchinetti, A. and Sparacino, G. (2020). Advanced diabetes management using artificial intelligence and continuous glucose monitoring sensors, *Sensors* **20**(14): 3870.

Vigersky, R. A. and McMahon, C. (2019). The relationship of hemoglobin a1c to time-in-range in patients with diabetes, *Diabetes technology & therapeutics* **21**(2): 81–85.

Wadghiri, M., Idri, A., El Idrissi, T. and Hakkoum, H. (2022). Ensemble blood glucose prediction in diabetes mellitus: A review, *Computers in Biology and Medicine* p. 105674.

Woldaregay, A. Z., Årsand, E., Walderhaug, S., Albers, D., Mamykina, L., Botsis, T. and Hartvigsen, G. (2019). Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes, *Artificial intelligence in medicine* **98**: 109–134.



Alvis Cabrera. Is a Ph.D. researcher at the University of Girona (UdG) in the Modeling, Identification & Control Engineering Laboratory since december 2019. Her research work focuses on applying mathematical and software tools for glucose control and prediction in patients with type 1 diabetes. She is working on implementing compositional data for glycemic state prediction. Her research interests include applied mathematics, data analysis and processing, and simulators for evaluating therapies.



Ernesto Estremera. Is a Ph.D. researcher at the UdG in the Modeling, Identification & Control Engineering Laboratory since December 2019. His research work is focused on designing and implementing modern control strategies. He is working on the design of controllers to improve glycemic control in the treatment of diabetes. His research interests include control systems engineering, type 1 diabetes simulators for in silico simulations, and biomedical engineering.

Aleix Beneyto. Received the Ph.D. degree in technology from the UdG in 2020. He is currently research director at the Modeling, Identification & Control Engineering Laboratory of the Institute of Informatics

and Applications at the UdG. He has been actively involved in the development of projects related to diabetes technology for the last 8 years, focused on the research and development of artificial pancreas technology. His research interests include the development of control algorithms for the artificial pancreas, including fault tolerant control strategies, control with carbohydrates, robust control, artificial intelligence applied to diabetes management for patient assessment, control, and prediction.



Lyvia Biagi. Holds a Bachelor's degree in Electrical Engineering from the Federal University of Technology-Paraná (UTFPR)-Cornélio Procópio, Paraná, Brazil (2011), a Master's of Science degree in Electrical Engineering, from the same university (2014). In 2019, she completed her Ph.D. in Technology, at the UdG, Spain. Currently she is an assistant professor at UTFPR-Guarapuava, Paraná, Brazil, where she has been a faculty member since April 2014. Her research interests are digital signal processing, pattern recognition techniques, intelligent systems, artificial pancreas, diabetes management, and compositional data analysis.



Ivan Contreras. Is a Ph.D. researcher at the UdG. He is focused on system optimization, data analytics, and predictive modeling in economics and healthcare. He has worked five years as a research assistant in the IE Business School, and in the Complutense University of Madrid to optimize investment decisions in the stock market and forex. Since 2014, Ivan is member of the MiceLAB group, where he is applying his expertise in the development of machine-learning techniques and its application for diabetes management.



Josep Martín-Fernández. Holds a degree in mathematics (Universitat Autònoma de Barcelona) and a PhD in statistics by the Universitat Politècnica de Catalunya. Since 1990, he is working in the Computer Science, Applied Mathematics and Statistics Department of the UdG. His area of interest is the statistical analysis of compositional data (CoDa). Currently he is the principal investigator of the CoDa-research group at the UdG. Since June 2015 he is the treasurer of the CoDa Association, which aims to bring together scientists interested in developing methods for CoDa modeling and their application.



Josep Vehi. is full Professor at the UdG and leads the "Laboratory of Modelling, Identification, and Control Engineering" (MiceLab). Since 2018 belongs to the Spanish centre of excellence CIBERDEM: "Networked Research Centre on Diabetes and associated Metabolic Diseases". His research interests include control systems with application to bioengineering, applications of artificial intelligence and machine learning to biomedicine, modelling and simulation of biomedical and biological processes and human-in-the-loop control systems design. He is a co-author of more than 150 peer-reviewed international journal papers and 12 books and chapters of books.

Received:

Accepted:

DISCUSIÓN

Los resultados presentados en esta tesis abarcan el análisis de mediciones de CGM en una cohorte de 47 VPs, un conjunto de datos de ocho pacientes con terapia CSII obtenidos de un estudio piloto realizado en el Hospital Clínic de Barcelona, y una base de datos disponible públicamente de un estudio compuesto por 226 pacientes con T1D. El objetivo principal ha sido desarrollar una herramienta de simulación que incluye escenarios desafiantes y realistas, donde hemos generado VPs, a los cuales se les ha incluido variabilidad en la sensibilidad a la insulina. Además, hemos incorporado modelos de insulina de acción prolongada (Gla-100, Gla-300 y Deg-100) para evaluar terapias tanto CSII como MDI. El segundo objetivo fue validar un modelo probabilístico de transición entre categorías de perfiles de glucosa, para lo cual se creó una metodología y métricas novedosas basadas en análisis de datos composicionales. Finalmente, siguiendo esta línea de analizar los tiempos en cada uno de los rangos de glucosa como datos composicionales, se propuso un modelo de regresión lineal múltiple para predecir la media y el CV de glucosa a largo plazo (2 y 4 horas). Las siguientes secciones de este capítulo resumen el trabajo completado, discuten los beneficios, comparan resultados particulares y las posibles limitaciones de los enfoques propuestos.

3.1 Análisis de la incorporación de las insulinas de acción prolongada al modelo de Hovorka

Las simulaciones se han convertido en una herramienta esencial para evaluar los tratamientos de la T1D, así como para validar estrategias y algoritmos de control. Estas simulaciones permiten realizar experimentos virtuales en entornos controlados, lo que brinda la posibilidad de probar diferentes enfoques terapéuticos y analizar su eficacia en la regulación de los niveles de BG. Al simular la respuesta del organismo a distintas intervenciones, se pueden obtener predicciones sobre el comportamiento metabólico y los efectos de las terapias propuestas. Esto no solo ahorra tiempo y costos asociados a los ensayos clínicos tradicionales, sino que también facilita la optimización de los enfoques de tratamiento antes de su implementación en pacientes reales. Además, las simulaciones proporcionan una plataforma para evaluar y comparar diferentes estrategias de control, permitiendo identificar aquellas que logren un mejor control glucémico y minimicen los riesgos asociados a la T1D. Sin embargo, los simuladores disponibles actualmente no son lo suficientemente desafiantes y a menudo tienen limitaciones en la variabilidad de la absorción de insulina y comida (Kovatchev et al., 2009; Visentin et al., 2018; Resalat et al., 2019; Rashid et al., 2019; Smaoui et al., 2020). Además, los simuladores de T1D están diseñados principalmente para probar terapias de CSII. La motivación principal de este trabajo fue crear una herramienta de simulación con pacientes realistas y escenarios desafiantes, donde pudiéramos probar terapias tanto MDI como CSII y que tuviera incorporado los modelos de insulina de acción prolongada.

En (Estremera et al., 2022) presentamos una herramienta de simulación que incluyó modelos de análogos de insulina de acción prolongada de Gla-100, Gla-300 y Deg-100. Por lo tanto, además de las terapias CSII, también se pueden probar terapias de MDI. Se utilizó el modelo de Hovorka y las distribuciones de probabilidad de sus parámetros publicados para generar cohortes de VPs que representan una población con T1D. Para obtener escenarios más realistas,

3.1. ANÁLISIS DE LA INCORPORACIÓN DE LAS INSULINAS DE ACCIÓN PROLONGADA AL MODELO DE HOVORKA

se utilizaron patrones de insulina basal publicados de pacientes reales para identificar patrones de variabilidad en la sensibilidad a la insulina e incorporarlos a los VPs. También se incluyó una biblioteca de comidas mixtas identificadas a partir de datos reales. Finalmente, se generó una cohorte de 47 VPs y se realizaron simulaciones de terapias CSII y MDI a lazo abierto. Las métricas de los resultados de la simulación se validaron comparándolos con resultados informados en ensayos clínicos en las mismas condiciones (Bolli et al., 2009; Home et al., 2015; Ruiz-de Adana et al., 2016; Galasso et al., 2016; Bergenstal et al., 2017; Heise et al., 2018; Bailey et al., 2018). En resumen, ese trabajo presentó y validó una metodología para la creación de cohortes realistas de VP que incluyen variabilidad fisiológica y una herramienta de simulación que incluye escenarios desafiantes y realistas.

En (Bergenstal et al., 2017) se realizó un estudio multicéntrico que incluyó 275 pacientes para Gla-100 y 274 para Gla-300. La inyección matutina de insulina glargina se realizó entre el desayuno y el almuerzo (incluidos), y la inyección de la noche entre la cena y hasta la hora de acostarse, mientras que los participantes continuaron usando insulina de acción rápida para cubrir las comidas. Los requerimientos de insulina informados como media \pm desviación estándar, para Gla-100 fueron de 0,45 (\pm 0,19) U/kg por la mañana y 0,36 (\pm 0,16) U/kg por la noche. Para Gla-300 fue de 0,49 (\pm 0,22) U/kg por la mañana y 0,45 (\pm 0,21) U/kg por la noche. Estos resultados son similares a los obtenidos en nuestras simulaciones, informados como mediana (percentil 25-75), para Gla-100 obtuvimos 0,42 (0,32–0,54) U/kg por la mañana y 0,41 (0,3–0,51) U/kg por la noche. Para Gla-300 fue de 0,45 (0,33–0,56) U/kg por la mañana y 0,37 (0,27–0,49) U/kg por la noche. La glucosa media fue de 152,1 (\pm 32,8) y 158,8 (\pm 38) mg/dl para Gla-100 y Gla-300, respectivamente, in vivo, valores comparables con los resultados obtenidos en simulación de 166,8 (154,5–185,5) mg/dl y 161,7 (147,9–176) mg/dl para Gla-100 y Gla-300, respectivamente.

En (Galasso et al., 2016) se eligieron 29 pacientes con una edad media de 34,8 (\pm 11,4) años con T1D que no conseguían una cobertura de 24 h con glargina o detemir y se estudió el

efecto de cambiar el tratamiento a degludec. La insulina basal resultante fue de 0,3 ($\pm 0,1$) U/kg después de 12 semanas obteniendo un TIR de 57,5 ($\pm 13,3$)%. En nuestras simulaciones se obtuvo una insulina basal de 0,39 (0,31–0,55) U/kg frente a 0,47 (0,35–0,6) U/kg suministradas a las 9 y 20 horas, respectivamente, lo que proporcionó unos TIR de 52,3 (39,6–61,6)% y 45,8 (38,2–54,4)%, respectivamente. En el ensayo clínico se reportó 38,8 ($\pm 14,4$) por encima de 180 mg/dl y en nuestras simulaciones en el rango 180–250 mg/dl reportamos 25,2 (19,3–29,1)% y por encima de 250 mg/dl 13,4 (7,2–26,6)%, para un total de 38,6% por encima del rango. El ensayo reportó un tiempo en hipoglucemia (< 70 mg/dl) de 3,6 ($\pm 3,7$)% y para nuestras simulaciones fue del 5,9 (3–9,9)%.

Aunque existieron diferencias en nuestros resultados con los reportados en los ensayos clínicos, esto se debió principalmente a que la simulación no incluye ningún trastorno relacionado con las condiciones de vida libre, como la actividad física, el estrés o los cambios hormonales como el ciclo menstrual. Otro factor que influye en las diferencias encontradas son los diferentes hábitos alimenticios de cada paciente, que no se tienen en cuenta en nuestra herramienta de simulación. Además, no se implementaron los mismos atributos clínicos del estudio y los algoritmos de ajustes dosis de insulina. Sin embargo, el estudio mostró tendencias similares a las obtenidas en los estudios clínicos, especialmente en cuanto a los requerimientos de insulina. Debemos mencionar que entre las limitaciones de este estudio se encuentra que aunque se identifican e implementan patrones de variabilidad de sensibilidad a la insulina de 24 horas en la herramienta de simulación, no se define un modelo que lo represente. Además, no se incluyó en los experimentos la hormona del glucagón.

3.2 Metodología y métricas para la validación del modelo probabilístico de transición

Los resultados de las métricas propuestas basadas en datos composicionales para validar el modelo probabilístico de transición los consideramos correctos a pesar de obtener una exactitud de 50% para toda la cohorte (Cabrera et al., 2023), lo que indica que la mitad de las partes de la composición coincidieron satisfactoriamente (es decir, el vector composicional de validación y el de entrenamiento).

Varias causas han conducido a este resultado, en algunos casos, no se contaron suficientes transiciones, lo que llevó a un modelo en el que las probabilidades estaban dispersas. De ahí que hemos planteado que la incorporación de más datos permitiría actualizar el modelo y obtener una mayor fiabilidad en la predicción. Los resultados que se muestran en la tabla 5 de (Cabrera et al., 2023) corresponden a la precisión del modelo para toda la cohorte. Sin embargo, en la tabla A.4 del Apéndice de ese manuscrito se muestran los resultados del paciente 1 que en muchos casos superan el 70% de precisión. La validación del modelo mostró una mejora en los resultados a medida que se disponía de un mayor volumen de datos. Esta observación plantea una posible limitación en nuestra métrica de evaluación, dado que incluye los ceros de conteo y, en situaciones con una menor cantidad de datos, existe una mayor probabilidad de que se presenten estos ceros. Como consecuencia, la métrica puede disminuir, lo cual podría no estar necesariamente relacionado con la confiabilidad del modelo en sí. Se puede ejemplificar esto comparando los resultados del paciente 2, del cuál solo se contaba con 54 combinaciones de períodos y el paciente 8 con más de 200 combinaciones (tablas 3.1 y 3.2, respectivamente), donde se puede observar que el resultado en porcentaje de la métrica aumenta con la disposición de mayor cantidad de datos.

Es importante recalcar que este es un modelo individualizado. Entonces, desde una perspectiva clínica, este modelo podría usarse para dar una probabilidad individualizada del

Tabla 3.1: Resultados de exactitud (métrica propuesta) después de una validación cruzada de 5 veces para el Paciente 2 (54 combinaciones de períodos).

Cluster	A las 0h	A las 6h	A las 12h	A las 18h	
3	A	45.2 (23.7-49.2)	77.2 (73.0-81.2)	74.2 (61.2-82.3)	21.4 (2-62.1)
	B	56.4 (42.0-69.1)	41.9 (7.5-53.4)	38.1 (14.5-65.5)	53.0 (47.4-62.6)
	C	22.4 (16.4-37.5)	no transiciones	82.3 (82.3-82.3)	35.9 (23.6-42.2)
4	A	44.3 (32.2-78.9)	72.1 (65.3-76.5)	52.2 (39.8-66.1)	26.6 (17.9-45.0)
	B	65.4 (37.1-83.0)	68.4 (42.6-79.8)	63.2 (49.2-70.9)	31.7 (22.0-35.5)
	C	10.0 (5.9-13.0)	no transiciones	60.4 (60.4-60.4)	no transiciones
	D	21.0 (18.9-42.4)	37.7 (22.1-56.7)	15.6 (5.6-39.3)	43.6 (43.6-43.6)
5	A	41.7 (15.5-58.8)	53.2 (40.7-66.0)	54.2 (34.3-62.9)	50.5 (32.3-66.5)
	B	60.1 (40.8-78.0)	36.3 (26.5-55.4)	45.1 (17.1-54.1)	36.0 (2-46.2)
	C	no transiciones	no transiciones	no transiciones	3.3 (1.9-22.9)
	D	7.2 (2-30.1)	34.9 (24.1-35.9)	40.2 (27.3-49.9)	31.0 (12.2-47.2)
	E	23.5 (18.4-25.1)	45.5 (12.8-52.0)	15.6 (13.2-27.9)	42.1 (42.1-42.1)

comportamiento de la glucosa en las próximas 6 horas. Por ejemplo, el médico o paciente analiza su composición de glucosa en el período de 24 h anterior, y comprueba que existe una alta probabilidad de que el siguiente período de 6 horas continuará con un control glucémico insatisfactorio, podría tomar una acción correctiva (es decir, ajustar la insulina basal, tomar algunos CHO o bolos correctores), según la categoría del futuro período de 6 horas. Aunque este modelo no presenta la predicción de valores o tendencias de BG, sino una probabilidad de comportamiento de la glucosa en las siguientes 6 horas. Consideramos que es un PH muy amplio en comparación con los modelos encontrados en la literatura que muestran hasta 120 min de predicción.

Los sensores de glucosa tienen errores de precisión, la sensibilidad a la insulina de los pacientes y la dinámica de respuesta de la insulina en el cuerpo también son estimaciones, no existe un método que pueda garantizar la verificación de los efectos de la insulina en la caída de glucosa. De esa forma, la predicción del comportamiento del paciente podría disminuir los efectos de las imprecisiones intrínsecas de los dispositivos. Este modelo podría ayudar a los médicos a evaluar los resultados de los pacientes y personalizar su régimen de dosificación de insulina. Se prevé que el modelo se actualice y ajuste según los hábitos y características

3.3. PROPUESTA DE PREDICCIÓN DE ESTADOS GLUCÉMICOS EN DÍAS CON ALTA VARIABILIDAD

Tabla 3.2: Resultados de exactitud (métrica propuesta) después de una validación cruzada de 5 veces para el Paciente 8 (más de 200 combinaciones de períodos).

Cluster		A las 0h	A las 6h	A las 12h	A las 18h
3	A	72.7(54.0-76.1)	71.7(61.1-82.9)	79.6(65.4-81.0)	80.71(77.7-86.5)
	B	88.9(76.1-92.9)	84.2(74.2-86.1)	71.7(69.2-79.5)	73.3(62.4-90.5)
	C	60.9(42.5-75.6)	70.4(47.8-87.5)	78.4(74.2-84.9)	73.8(48.9-75.7)
4	A	65.4(29.4-69.6)	73.1(53.0-82.3)	66.0(59.1-76.1)	74.3(71-81.3)
	B	80.2(61.5-86.7)	78.2(56.5-80.5)	70.0(62.1-72.4)	56.8(52.0-72.3)
	C	30.8(26.8-43.7)	70.1(48.2-74.9)	68.4(50.2-73.4)	50.6(39.3-65.7)
	D	62.5(27.6-68.8)	67.8(52.2-69.2)	69.3(55.6-71.5)	55.1(38.1-62.5)
5	A	77.8(66.6-79.8)	35.1(18.7-49.9)	68.3(50.0-71.2)	59.3(19.0-69.2)
	B	59.7(55.3-78.3)	75.2(58.1-77.8)	71.8(46.7-76.1)	74.4(63.2-77.0)
	C	78.4(71.7-80.8)	75.8(65.4-77.4)	77.3(64.3-82.9)	59.3(46.0-81.2)
	D	41.8(36.9-52.3)	54.1(44.4-66.6)	70.9(57.2-73.2)	54.2(51.0-65.0)
	E	75.0(59.2-79.5)	65.8(59.3-71.5)	65.1(62.4-74.3)	56.7(35.9-62.3)

individuales de los pacientes a lo largo del tiempo.

3.3 Propuesta de predicción de estados glucémicos en días con alta variabilidad

Esta sección presenta un ejemplo de la metodología propuesta en el último manuscrito de esta tesis, donde se propone la predicción a largo plazo mediante modelos de regresión lineal múltiple basados en datos composicionales.

La figura 3.1 muestra las curvas de CGM del paciente 1 para dos días caracterizados con alta variabilidad, el día 3 (azul) y 49 (rojo). El día 3 mostró hiperglucemia severa por más de la mitad del día, descendiendo a 70 mg/dl a las 20:00 h, permaneció en rango objetivo por al menos 2 horas, para luego caer en hipoglucemia. Sin embargo, el día 49, se encontró en hiperglucemia desde las 00:00 h hasta las 08:00 h, seguido de TIR hasta las 14:00 h. A partir de entonces, los niveles de BG disminuyeron, alcanzando la hipoglucemia hasta estabilizar los niveles de glucosa a las 20:00 h. Esta figura ilustra la significativa variabilidad diaria en la glucemia del paciente.

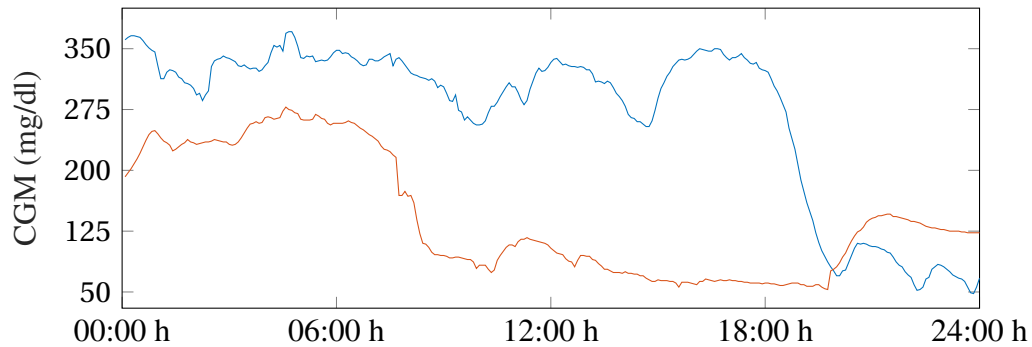


Figura 3.1: BG del paciente 1, día 3 (azul) y día 49 (rojo).

La tabla 3.3 presenta un ejemplo del sistema “traffic light” propuesto para el paciente 1. Analizamos el día 3 caracterizado por alta variabilidad con $CV = 36.53\%$, hiperglucemia severa tanto durante el día como la noche, y también la presencia de hipoglucemia. La columna 4 muestra la descripción para cada una de las clases (3 ó 5 clases, que hemos diferenciado también por colores).

Analizando las predicciones de los estados para 3 clases (columna 2 de la tabla 3.3), se puede observar que desde las 00:00 h hasta las 18:00 h, cada hora en ese intervalo, el modelo predijo que el paciente estaría durante las próximas 2 horas en riesgo de hiperglucemia (>180 mg/dl). Comprobamos que durante el período nocturno, desde las 22:00 h del día anterior hasta las 8:00 h del día analizado, este paciente tuvo un CV de 6.5% , con una BG mínima de 269 mg/dl y una máxima de 371 mg/dl. Desde las 19:00 h a las 20:00 h, se mantuvo en TIR y desde las 21:00 h a las 23:00 h, se categorizaron esas horas por riesgo de hipoglucemia, predicho correctamente.

Al analizar los resultados para 5 clases (columna 3 de la tabla 3.3), desde las 00:00 h hasta las 17:00 h, el modelo predijo riesgo de hiperglucemia severa (nivel 2), siendo una categoría más específica que cuando se analizó para 3 clases. La glucosa mínima fue de 244 mg/dl y la máxima de 329 mg/dl, con un CV entre el 2% y el 8% . Sin embargo, a las 18:00 h, el modelo predijo un riesgo de hiperglucemia nivel 1, aquí verificamos que el paciente tenía un mínimo

3.3. PROPUESTA DE PREDICCIÓN DE ESTADOS GLUCÉMICOS EN DÍAS CON ALTA VARIABILIDAD

de 70 mg/dl y un máximo de 321 mg/dl con un CV para las próximas 2 horas del 40%. A las 19:00 h y 20:00 h, el paciente estuvo en TIR, a las 21:00 h y 22:00 h, el modelo predijo un riesgo de hipoglucemia; sin embargo, la validación corroboró que fue precisa para las 21:00 h, pero para las 22:00 h, el estado real informó de una hipoglucemia severa. A las 23:00 h, tanto el modelo como la realidad informaron de una hipoglucemia severa.

La tabla 3.4 muestra otro ejemplo del mismo paciente para el día 49, caracterizado por alta variabilidad de glucosa con $CV = 53.15\%$, con un mínimo de 53 mg/dl y un máximo de 278 mg/dl. Cuando se examinó la predicción del estado con tres clases, se encontró que el modelo predijo con precisión el riesgo de hiperglucemia durante el período de 00:00 h a 06:00 h, pero cometió un error a las 07:00 h. Vale la pena destacar que el período anterior de 2 horas se caracterizó por una alta variabilidad ($CV = 33,27\%$) en un lapso de tiempo relativamente corto. Hemos validado nuestra propuesta para emitir una caracterización del tipo "Alta variabilidad" cuando el período previo se caracterizó por un CV superior al 36% (que es la métrica estandarizada), pero también consideramos que un CV de 33,27% en un PH de 2 horas es alta. De 08:00 h a 13:00 h se comportó en TIR, de 14:00 h a 19:00 h se predijo riesgo de hipoglucemia correctamente, sin embargo, el modelo predijo hipoglucemia a las 20:00 h de forma errónea. A partir de las 21:00 h se comportaron con tendencia a la normoglucemia de forma acertada.

Para la categorización de 5 clases, se presentaron resultados correctos de 00:00 h a 06:00 h. A las 07:00 h, el modelo predijo el riesgo de hiperglucemia y lo real fue normoglucemia. Como se explicó anteriormente, esto se debe a la alta variabilidad en las 2 horas previas. A las 14:00 horas, el modelo predijo la normoglucemia y el paciente estaba en riesgo de hipoglucemia. A las 20:00 h, el modelo predijo el riesgo de hipoglucemia y los datos reales mostraron normoglucemia. En este momento, el coeficiente de variabilidad anterior fue del 35,9%.

Incluso en días con alta variabilidad, nuestro modelo puede lograr una efectividad de

predicción alta para las próximas 2 horas. Sin embargo, cuando se obtiene un coeficiente de variabilidad previo superior al 30%, el modelo tiende a cometer errores de predicción. Cabe señalar que los errores pueden estar dados por nuestra metodología de categorización y no a la predicción del modelo de media y CV. Estos ejemplos confirman la importancia de tener modelos de predicción a largo plazo confiables, sobretodo para horarios nocturnos, lo que puede ser difícil de manejar. Aunque el modelo podría dar información más específica cuando se analizan 5 clases, el porcentaje de predicciones exitosas es mayor cuando solo se consideran tres clases, como era de esperar. Conocer su estado glucémico con tal precisión puede ser de mucha ayuda al paciente a la hora de tomar decisiones.

En este trabajo se ha evidenciado la importancia de los modelos de predicción individualizados, los cuales desempeñan un papel crucial en el manejo de la T1D al proporcionar estimaciones personalizadas y precisas de los niveles de BG. Estos modelos integran características y variables específicas de cada paciente, como perfiles metabólicos, patrones de alimentación, actividad física y administración de insulina. Esto permite desarrollar estrategias terapéuticas personalizadas y adaptadas a las necesidades y condiciones particulares de cada individuo. Mediante su uso, los profesionales de la salud pueden optimizar la terapia de insulina al obtener una visión más precisa de las posibles fluctuaciones de los niveles de BG a lo largo del tiempo. Esto ayuda a prevenir episodios de hiperglucemia o hipoglucemia y facilita el ajuste efectivo de las dosis de insulina para lograr un control glucémico óptimo. Además, hemos validado la propuesta de predicción a cada hora del día, que permite analizar el comportamiento de la glucosa de forma específica, distinguiendo horarios con más o menos variabilidad.

3.3. PROPUESTA DE PREDICCIÓN DE ESTADOS GLUCÉMICOS EN DÍAS CON ALTA VARIABILIDAD

Hora	3 clases		5 clases		Características de los estados
	Predicción	Real	Predicción	Real	
00:00	●	●	●	●	<p>3 clases</p> <p>Categorización para 3 clases</p> <ul style="list-style-type: none"> Período caracterizado por riesgo de hiperglucemia Período caracterizado por tendencia a la normoglucemia Período caracterizado por riesgo de hipoglucemia <p>5 clases</p> <p>Categorización para 5 clases</p> <ul style="list-style-type: none"> Período caracterizados por riesgo de hiperglucemia nivel 2 Período categorizados por riesgo de hiperglucemia nivel 1 Período caracterizados por tendencia a la normoglucemia Período caracterizados por riesgo de hipoglucemia nivel 1 Período caracterizados por riesgo de hipoglucemia nivel 2
01:00	●	●	●	●	
02:00	●	●	●	●	
03:00	●	●	●	●	
04:00	●	●	●	●	
05:00	●	●	●	●	
06:00	●	●	●	●	
07:00	●	●	●	●	
08:00	●	●	●	●	
09:00	●	●	●	●	
10:00	●	●	●	●	
11:00	●	●	●	●	
12:00	●	●	●	●	
13:00	●	●	●	●	
14:00	●	●	●	●	
15:00	●	●	●	●	
16:00	●	●	●	●	
17:00	●	●	●	●	
18:00	●	●	●	●	
19:00	●	●	●	●	
20:00	●	●	●	●	
21:00	●	●	●	●	
22:00	●	●	●	●	
23:00	●	●	●	●	

Tabla 3.3: Ejemplo de predicción glucémica para las próximas 2h para el paciente 1, día 3 con 3 y 5 clases

3. DISCUSIÓN

Hora	3 clases		5 clases		Características de los estados
	Predicción	Real	Predicción	Real	
00:00	●	●	●	●	<p>3 clases</p> <p>Categorización para 3 clases</p> <ul style="list-style-type: none"> ● Período caracterizado por riesgo de hiperglucemia ● Período caracterizado por tendencia a la normoglucemia ● Período caracterizado por riesgo de hipoglucemia
01:00	●	●	●	●	
02:00	●	●	●	●	
03:00	●	●	●	●	
04:00	●	●	●	●	
05:00	●	●	●	●	
06:00	●	●	●	●	
07:00	●	●	●	●	<p>5 clases</p> <p>Categorización para 5 clases</p> <ul style="list-style-type: none"> ● Período caracterizados por riesgo de hiperglucemia nivel 2 ● Período categorizados por riesgo de hiperglucemia nivel 1 ● Período caracterizados por tendencia a la normoglucemia ● Período caracterizados por riesgo de hipoglucemia nivel 1 ● Período caracterizados por riesgo de hipoglucemia nivel 2
08:00	●	●	●	●	
09:00	●	●	●	●	
10:00	●	●	●	●	
11:00	●	●	●	●	
12:00	●	●	●	●	
13:00	●	●	●	●	
14:00	●	●	●	●	
15:00	●	●	●	●	
16:00	●	●	●	●	
17:00	●	●	●	●	
18:00	●	●	●	●	
19:00	●	●	●	●	
20:00	●	●	●	●	
21:00	●	●	●	●	
22:00	●	●	●	●	
23:00	●	●	●	●	

Tabla 3.4: Ejemplo de predicción glucémica para las próximas 2h para el paciente 1, día 49 con 3 y 5 clases

CONCLUSIONES Y TRABAJO FUTURO

Esta tesis ha contribuido en el desarrollo de una herramienta de simulación para evaluar terapias MDI y CSII para T1D, que se ha validado comparando los resultados de simulación con los publicados de ensayos clínicos reportados. Además, ha propuesto una metodología y métricas novedosas para la validación del modelo de predicción probabilística de la glucosa. También ha contribuido y validado un nuevo enfoque para la predicción de glucosa a largo plazo.

4.1 Contribuciones

El trabajo desarrollado en esta tesis lleva a exponer como contribuciones más destacadas:

- **Desarrollo de una herramienta de simulación realista que incluye simulaciones de insulinas de larga duración.** Hemos propuesto una herramienta de simulación basada en el modelo de Hovorka que incluye escenarios desafiantes y realistas, donde se pueden generar cohortes de VPs. Hemos incorporado modelos de insulina de acción rápida y prolongada (Gla-100, Gla-300 y Deg-100) para evaluar terapias MDI y CSII. Obtuvimos

escenarios más realistas, utilizando patrones de perfil de insulina basal de la literatura para identificar la variabilidad en la sensibilidad a la insulina. También se ha incluido una biblioteca de comidas mixtas identificadas a partir de datos reales.

- **Validación de la incorporación de insulinas de acción prolongada al modelo de Hovorka.** Se han incorporado los modelos de las insulinas de acción prolongada Gla-100, Gla y Deg-100 en la herramienta de simulación. Para su validación se han simulado con VPs realistas comparando los resultados obtenidos con ensayos clínicos realizados en condiciones similares.
- **Validación del modelo probabilístico de transiciones entre categorías.** Se ha evaluado la precisión del modelo probabilístico de transición entre las categorías de perfiles de glucosa basado en un enfoque de análisis de datos composicionales, mediante su validación con una cohorte de 8 pacientes reales. Se ha utilizado el método de k-medias para categorizar los períodos de 24 horas previos y próximas 6 horas.
- **Métricas novedosas con enfoque de conjuntos CoDa.** Se han desarrollado y validado nuevas métricas de exactitud y se han adaptado las métricas de precisión comunes al enfoque composicional, lo que permite utilizar estas métricas en cualquier trabajo con conjuntos de datos de entrenamiento y validación que cumplan los principios composicionales.
- **Regresión múltiple para predicción de media y coeficientes de variación.** Se ha implementado una secuencia de modelos de regresión lineal múltiple, con entradas composicionales y entradas escalares, para predecir la media y el CV de la glucosa a cada hora del día. Este método complementa la predicción del estado glucémico, lo que podría ayudar a los pacientes a tomar medidas de corrección ante situaciones adversas en horizontes de predicción amplios.

- **Predicción del coeficiente de variación.** Hasta donde sabemos no se encontró ningún estudio que predijese el CV de la glucosa a cada hora del día. La variabilidad de la glucosa puede variar significativamente en diferentes pacientes, e incluso en el mismo paciente en horarios y días de la semana particulares. En este trabajo hemos predicho el CV de la glucosa a cada hora del día para las próximas 2 y 4 horas.
- **Modelos individualizados y enfoque de modelos múltiples.** Aunque los resultados de estos modelos se presentan como una cohorte de pacientes, han sido interpretados y aplicados como modelos individualizados. Nuestra metodología propone 24 modelos, uno para cada hora del día, de forma individualizada. La implementación de múltiples modelos es más compleja, no obstante, es capaz de caracterizar la dinámica de cada paciente de forma más sólida.

4.2 Trabajos futuros

Los resultados expuestos en esta tesis son prometedores y ofrecen una herramienta de análisis adicional para los pacientes y médicos. Sin embargo, teniendo en cuenta que todos los análisis realizados en este trabajo se llevaron a cabo de manera retrospectiva, se necesitan pruebas para implementarlos en tiempo real. Aunque ya existen herramientas para sincronizar Matlab y R-Studio, se espera el desarrollo de un sistema que ponga en marcha estas metodologías para aplicarlas en tiempo real en un dispositivo móvil de fácil acceso a los pacientes y médicos.

Además, a pesar de haber validado el modelo probabilístico de transición con datos de pacientes reales, sería interesante evaluar el comportamiento para mayor cantidad de días de un mismo paciente y verificar la métrica propuesta. Este modelo de predicción se podría implementar en tiempo real e intentar monitorizar las predicciones y acciones de los pacientes de forma individualizada. Además, considerando una técnica similar de categorización de perfiles de glucosa, utilizando el enfoque log-ratio para el análisis del tiempo pasado en

diferentes rangos de glucosa, se podrían encontrar similitudes para categorizar los diferentes días de la semana, incluso diferenciarlos de los fines de semana y predecir el comportamiento en estas situaciones.

La predicción a largo plazo presentado en este trabajo podría utilizarse como un sistema de apoyo a la toma de decisiones o incorporarse como un umbral en un controlador para terapias MDI o CSII. Proporcionar información personalizada adicional ayudaría a abordar la variabilidad de la glucosa y optimizar un algoritmo de control de BG de forma automática.

Teniendo en cuenta que los ensayos clínicos realizados en pacientes con T1D recopilan una gran cantidad de mediciones de glucosa en tiempo real y variables categóricas y binarias, es posible evaluar algunas de estas variables mediante el uso de análisis de conjuntos CoDa. A continuación se enumeran algunos ejemplos de variables que podrían ser tratadas mediante este enfoque analítico:

- **Tiempos en rango:** hasta donde sabemos los únicos datos que se han tratado como composicionales para pacientes con T1D son los diferentes tiempos en rango de glucosa. Sin embargo, no se han utilizado para relacionarlos con otras variables importantes, como por ejemplo: 1) identificar patrones y tendencias en los niveles de glucemia y su relación con factores como la dieta, la actividad física y el uso de medicamentos; 2) evaluar la efectividad de diferentes estrategias de tratamiento en el control de la glucemia; 3) identificar pacientes que podrían estar en riesgo de desarrollar complicaciones relacionadas con la diabetes, como la hipoglucemia o hiperglucemia; 4) identificar pacientes que podrían beneficiarse de cambios en la dosis de insulina o en el horario de las comidas; y 5) ayudar a los pacientes a alcanzar y mantener niveles de glucemia saludables (aumentando el TIR). Algunas herramientas que podrían implementarse son los análisis de serie de tiempo que se utilizan para identificar patrones y tendencias en los niveles de glucemia a lo largo del tiempo.
- **Dietas:** Los pacientes con diabetes a menudo tienen que controlar su ingesta de CHO y

calorías para mantener el control metabólico. Un vector composicional podría tratarse como [CHO, grasas, proteínas] en la dieta. Esto podría ayudar a los pacientes a alcanzar y mantener un peso saludable, así como identificar alimentos particulares que estén afectando en sus dietas de acuerdo a sus partes de la composición. Identificar pacientes que podrían estar en riesgo de desarrollar complicaciones relacionadas con la diabetes, como la enfermedad cardiovascular, y brindarles intervenciones tempranas. Además en este aspecto también podrían evaluarse el metabolismo de los macronutrientes (CHO, proteínas y ácidos grasos). Las herramientas reconocidas que podrían evaluarse en este aspecto incluyen el análisis de componentes principales para identificar patrones y tendencias en la ingesta de nutrientes y su relación con los niveles de glucemia. Podría profundizarse en análisis estadístico multivariado, análisis de cluster, métodos matemáticos de optimización, discriminante lineal y MANOVA.

- **Medicación:** Aunque este aspecto está más relacionado con pacientes con T2D. Algunos pacientes con T1D tienen otras patologías y son frecuentemente medicados. Los registros de las dosis y el tipo de medicamentos tomado podría tratarse como un conjunto CoDa y visualizar cómo podrían afectar a los niveles de CGM y cómo se relacionan para encontrar la dosis más apropiada e incluso el horario del día más indicado para ingerirlos, predecir posibilidades de eventos de hipoglucemia o hiperglucemia, incluso ayudar a los médicos a comprender y mejorar la adherencia de los pacientes a sus regímenes de medicación.
- **Actividad física:** Los registros de frecuencia, duración y tipo de actividad física realizada por los pacientes con T1D podría tratarse como conjunto CoDa. También, la proporción de tiempo dedicado a diferentes tipos de actividad física (actividad moderada o rigurosa) en relación con el tiempo dedicado a actividad sedentaria. Esta información se puede recoger a través de acelerómetros, dispositivos de seguimiento de actividad, o mediante registros diarios de actividad física. Estos estudios podrían sugerir la actividad más

apropiada según el perfil del paciente y su influencia en las mediciones de CGM.

- **HbA1c:** Los resultados de las pruebas de HbA1c se pueden tratar como conjunto CoDa si se comparan entre diferentes pacientes o en diferentes momentos en el tiempo, así como su proporción en relación al total de hemoglobina en la sangre. Podría ser utilizada como un indicador del control metabólico a largo plazo. La herramientas de aprendizaje automático podría ser de utilidad para predecir los resultados de las pruebas de laboratorio, identificar patrones o pacientes con riesgos. La cohorte utilizada por (Cabrera et al., 2023) registra estas mediciones.
- **Medidas de calidad de vida:** Las medidas del índice de masa corporal (IMC) o el índice de resistencia a la insulina (HOMA-IR) se pueden tratar como conjunto CoDa al analizarlos por encima o por debajo de ciertos umbrales predeterminados. Esta proporción puede ser utilizada para evaluar el sobrepeso y si está relacionada con la resistencia a la insulina.
- **Relación insulina y CHO:** Se podría tratar como conjunto CoDa la proporción de dosis de insulina por cada gramo de CHO ingeridos, o la proporción de dosis de insulina necesaria para controlar los niveles de glucemia en relación a la ingesta de CHO. Podrían ser utilizadas para evaluar la eficacia de diferentes regímenes de insulina y para identificar pacientes con riesgo de hipoglucemia.
- **Otras aplicaciones:** La frecuencia cardíaca, calidad del sueño, enfermedades particulares y recuento de pasos son variables que podrían tratarse como conjunto CoDa para la detección de complicaciones, ayudando a mantener un equilibrio saludable durante la enfermedad. Además, se podría predecir a partir de modelos la probabilidad de que algunos pacientes desarrollen patrones similares asociados a estas patologías.



LIBRERÍA DE COMIDAS MIXTAS

En este apéndice se presenta la librería de comidas mixtas utilizada en (Estremera et al., 2022). La biblioteca incluye vectores de RA para un total de 54 composiciones diferentes de comidas divididas en 17 grupos. Estas comidas se definen por su composición nutricional (grasas, proteínas, fibras y energía), cantidad de CHO y se clasifican en cuatro categorías según la distancia de Kolmogorov- Smirnov: grandes (10), medianas (26), pequeñas (12) y rápidas (6). Cada perfil de comida está compuesto por 84 valores, obtenidos con una frecuencia de muestreo de 5 minutos, que representan la tasa de aparición de BG 7 horas después del momento de la ingesta (Estremera et al., 2022).

A. LIBRERÍA DE COMIDAS MIXTAS

Grupo 1							
Alimentos	Clasificación	CHO (g)	Grasa (g)	Proteína (g)	Fibra (g)	Energía (kCal)	Tiempo (min)
1. Leche, arroz, pera, galletas de salvado, queso bajo en grasa, aceite	Mediana	52.00	10.50	14.50	2.90	362.09	300
2. Leche, pan, queso bajo en grasa, mantequilla, aceite	Mediana	52.50	10.50	14.50	2.60	362.33	300
3. IG bajo (250ml de agua): Cebada perlada hervida	Mediana	50.00	1.30	9.20	5.10	243.15	240
4. Pan blanco, huevos, margarina y jugo de naranja	Mediana	50.00	13.00	13.00	2.00	360.90	300
5. Suplemento nutricional en polvo	Mediana	50.00	12.00	14.00	0.00	366.87	300
6. Pasta y salsa de tomate + aceite	Mediana	50.00	-	-	-	509.94	240

Grupo 2							
Alimentos	Clasificación	CHO (g)	Grasa (g)	Proteína (g)	Fibra (g)	Energía (kCal)	Tiempo (min)
7. Papa al horno, gelatina, pechuga de pavo	Mediana	45.00	18.00	15.00	-	-	180
8. Arroz hervido maíz, pechuga de pavo	Mediana	50.00	18.00	15.00	-	-	180
9. Pasta y salsa de tomate	Mediana	50.00	-	-	-	240.04	240

Grupo 3							
Alimentos	Clasificación	CHO (g)	Grasa (g)	Proteína (g)	Fibra (g)	Energía (kCal)	Tiempo (min)
10. Pasta + bajo contenido en aceite de girasol	Grandes	75.00	3.20	12.80	-	331.00	430
11. Pasta + bajo contenido medio en aceite de girasol	Grandes	75.00	45.50	14.50	-	952.76	420
12. Desayuno estándar	Mediana	70.30	70.30	15.60	-	952.76	300
Grupo 4							
Alimentos	Clasificación	CHO (g)	Grasa (g)	Proteína (g)	Fibra (g)	Energía (kCal)	Tiempo (min)
13. Frijoles, pan integral, salami, queso	Mediana	50.00	-	-	-	750.00	300
14. Pasta y salsa de tomate + aceite + psyllium	Mediana	50.00	-	-	-	509.94	240
15. Desayuno estándar	Mediana	50.00	80.00	14.40	-	955.86	300
Grupo 5							
Alimentos	Clasificación	CHO (g)	Grasa (g)	Proteína (g)	Fibra (g)	Energía (kCal)	Tiempo (min)
16. Arroz con leche, azúcar y canela	Mediana	50.00	-	-	-	270.00	300
17. Avena templada	Pequeña	25.00	1.80	4.60	13.70	181.76	120
18. Chips de proteína de soja ligeramente salados	Pequeña	25.00	3.80	13.50	0.00	192.99	120

A. LIBRERÍA DE COMIDAS MIXTAS

Grupo 6							
Alimentos	Clasificación	CHO (g)	Grasa (g)	Proteína (g)	Fibra (g)	Energía (kCal)	Tiempo (min)
19. Tortilla de queso, pan y margarina	Pequeña	38.00	24.00	24.00	-	466.94	120
20. Pan de Trigo Blanco	Mediana	50.00	3.00	8.40	3.10	266.79	180
21. Pan Integral	Mediana	50.00	2.60	7.40	12.80	258.91	180
22. Pan de Centeno con B-glucano	Mediana	50.00	2.40	10.50	17.10	270.85	180
23. Pasta Integral	Mediana	50.00	4.70	12.10	5.60	298.56	180

Grupo 7							
Alimentos	Clasificación	CHO (g)	Grasa (g)	Proteína (g)	Fibra (g)	Energía (kCal)	Tiempo (min)
24. Tempe de cebada	Pequeña	25.00	1.80	4.60	5.70	150.47	120
25. Tortilla de huevo, pan, espinacas, pimiento rojo, tomate grillado	Pequeña	20.20	-	17.50	-	236.94	120
26. Tortilla de huevo, miel, pan, espinacas, pimiento rojo	Pequeña	15.50	10.30	17.10	-	223.08	120

Grupo 8							
Alimentos	Clasificación	CHO (g)	Grasa (g)	Proteína (g)	Fibra (g)	Energía (kCal)	Tiempo (min)
27. Barra de chocolate con frambuesa	Pequeña	10.00	12.50	45.00	2.00	526.66	120
28. Batido de chocolate con sucralosa	Pequeña	10.00	5.00	40.00	4.00	260.82	120

Grupo 9							
Alimentos	Clasificación	CHO (g)	Grasa (g)	Proteína (g)	Fibra (g)	Energía (kCal)	Tiempo (min)
29. Leche, arroz blanco, pera, salvado, aceite	Grandes	111.00	22.60	31.00	6.40	776.01	300
30. Comida rica en CHO	Grandes	93.00	30.00	146.00	22.00	2913.92	360
31. Galleta, yogur sin grasa	Grandes	94.00	17.00	29.00	-	673.78	360
32. Copos de trigo, queso fresco, leche	Grandes	91.00	25.00	28.00	-	685.73	360

Grupo 10							
Alimentos	Clasificación	CHO (g)	Grasa (g)	Proteína (g)	Fibra (g)	Energía (kCal)	Tiempo (min)
33. comida rica en grasas	Mediana	27.00	60.00	135.00	18.00	2842.27	360
34. Desayuno cereales de avena, leche, mermelada de fresa y zumo de naranja	Mediana	69.00	12.80	13.10	-	441.86	300
35. Barra de chocolate, leche, helado, crema espesa	Grandes	80.00	80.00	18.00	-	1111.11	360
36. Cebada laminada al 100%	Mediana	75.00	1.70	7.20	9.20	361.85	240
37. Cebada laminada al 100%	Mediana	75.00	1.10	5.40	3.20	348.72	240

A. LIBRERÍA DE COMIDAS MIXTAS

Grupo 11							
Alimentos	Clasificación	CHO (g)	Grasa (g)	Proteína (g)	Fibra (g)	Energía (kCal)	Tiempo (min)
38. Pasta + psyllium bajo en grasa	Mediana	52.00	-	-	1.70	240.00	240
39. IG alto : (285ml de agua) Puré de patata	Mediana	50.00	4.50	5.50	6.60	244.34	240

Grupo 12							
Alimentos	Clasificación	CHO (g)	Grasa (g)	Proteína (g)	Fibra (g)	Energía (kCal)	Tiempo (min)
40. Barra de chocolate con mantequilla de maní	Pequeña	25.00	4.80	12.90	0.80	194.18	120
41. Batido de chocolate Daydream – fructosa	Pequeña	25.00	1.80	14.70	1.50	177.22	120
42. Espaguetis de soja	Pequeña	25.00	1.10	10.60	0.80	151.91	120
43. Muffin, mantequilla de maní	Pequeña	27.30	15.40	4.50	-	265.12	120

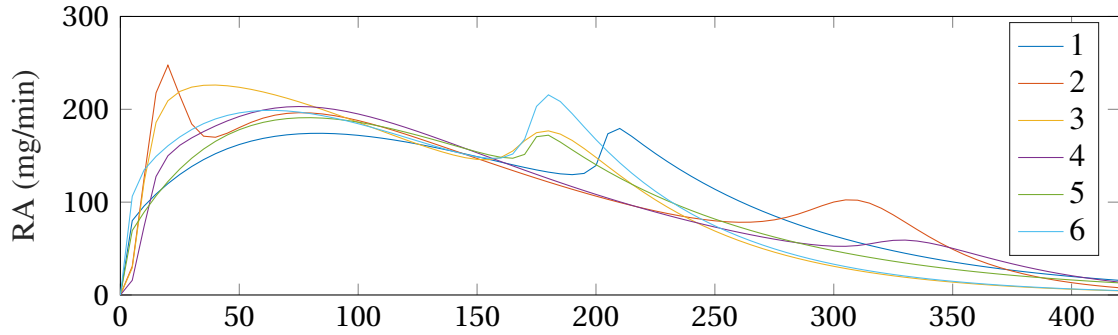
Grupo 13							
Alimentos	Clasificación	CHO (g)	Grasa (g)	Proteína (g)	Fibra (g)	Energía (kCal)	Tiempo (min)
44. Avena, le- che pan y margarina	Rápida	62.00	17.00	12.00	-	452.61	120
45. Cebada con tomate y queso	Rápida	68.00	8.00	12.00	-	394.10	120

Grupo 14							
Alimentos	Clasificación	CHO (g)	Grasa (g)	Proteína (g)	Fibra (g)	Energía (kCal)	Tiempo (min)
46. Cereales con fibra y frutas, leche, melón, piña	Rápida	50.00	2.60	10.20	-	265.12	120
47. Cereal alto en fibra, leche, fresas, toronja	Rápida	42.00	2.50	11.00	-	234.07	120
48. Sirope, leche desnatada, helado sin grasa	Grandes	80.00	0.80	18.00	-	399.35	360

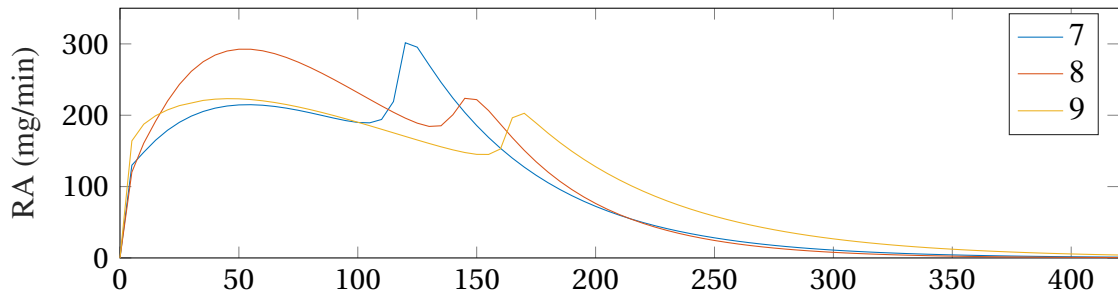
Grupo 15							
Alimentos	Clasificación	CHO (g)	Grasa (g)	Proteína (g)	Fibra (g)	Energía (kCal)	Tiempo (min)
49. Arroz blanco	Mediana	75.00	0.90	4.60	0.50	342.74	240
50. Cebada laminada al 50%	Mediana	75.00	1.30	5.90	4.90	352.78	240

Grupo 16							
Alimentos	Clasificación	CHO (g)	Grasa (g)	Proteína (g)	Fibra (g)	Energía (kCal)	Tiempo (min)
51. Espaguetis con tomate, queso y lentejas	Rápida	87.00	9.00	25.00	-	527.85	120
52. Copos de maíz, leche, pan y margarina	Rápida	104.00	13.00	18.00	-	609.06	120

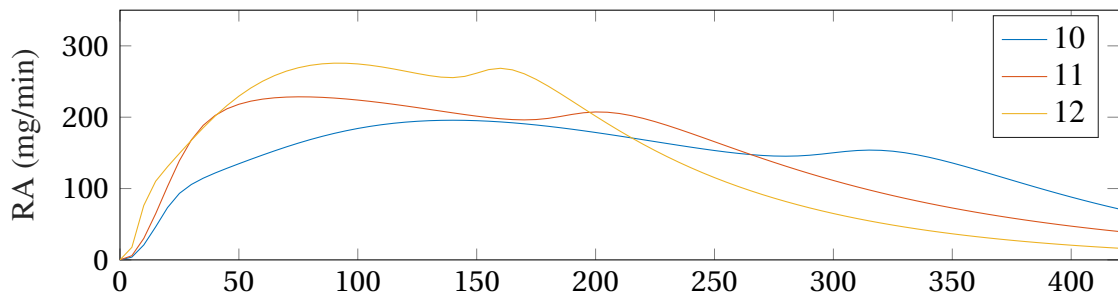
Curvas de RA del grupo 1



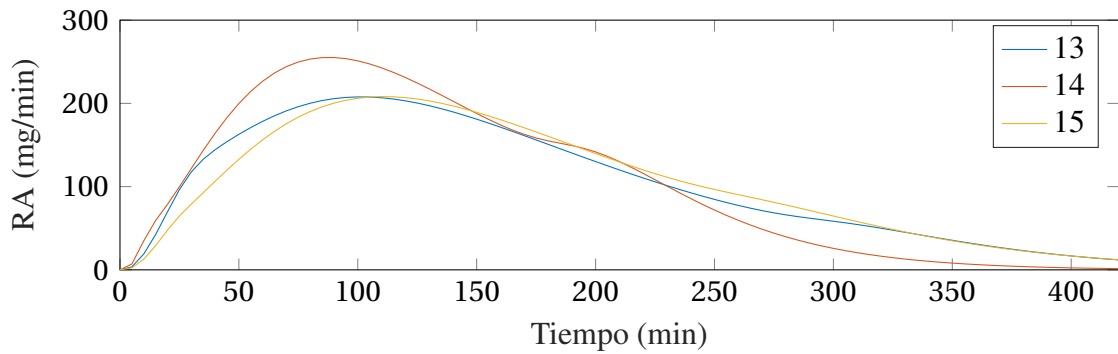
Curvas de RA del grupo 2



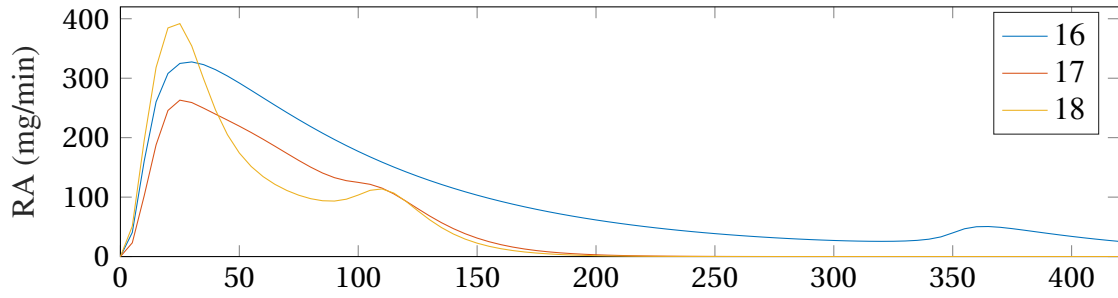
Curvas de RA del grupo 3



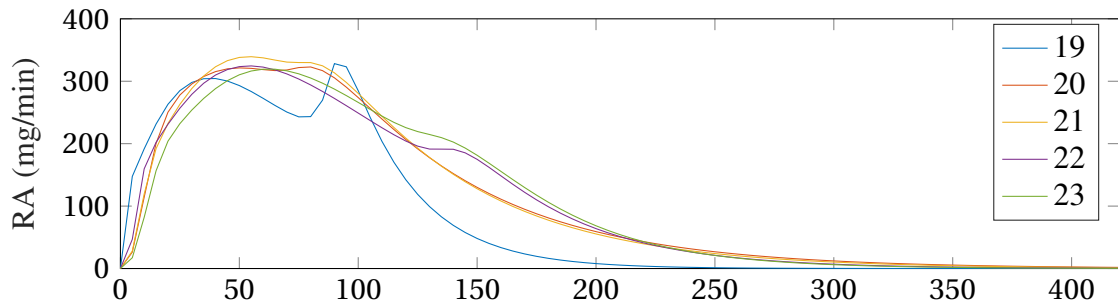
Curvas de RA del grupo 4



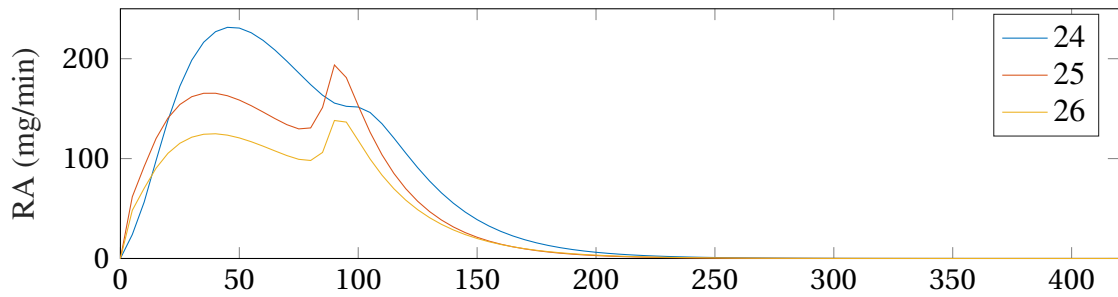
Curvas de RA del grupo 5



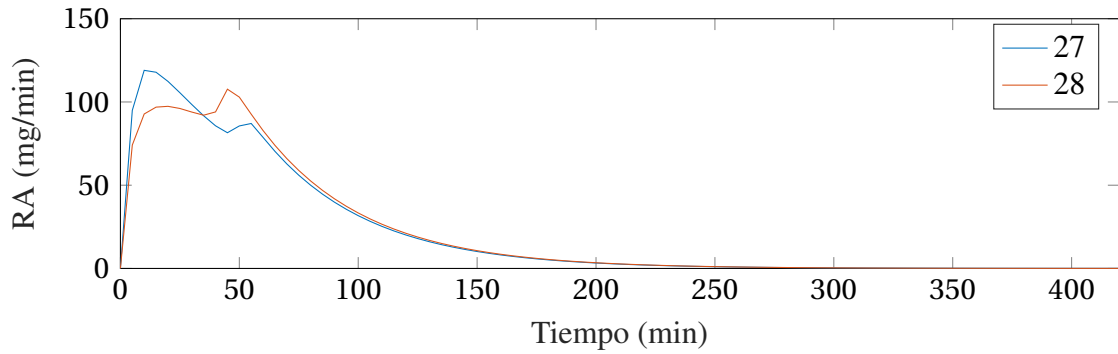
Curvas de RA del grupo 6



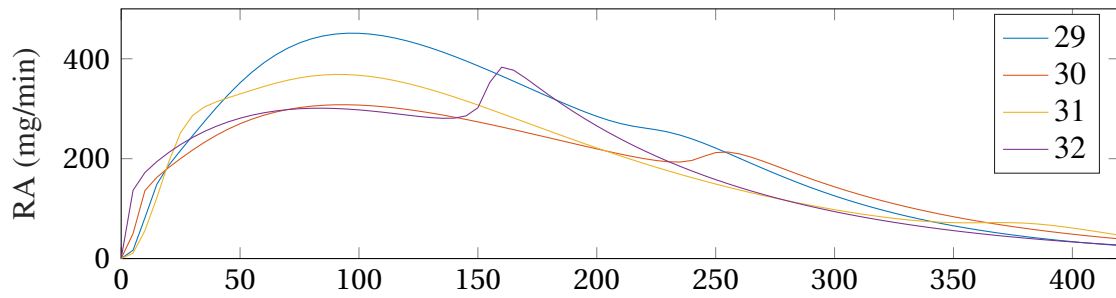
Curvas de RA del grupo 7



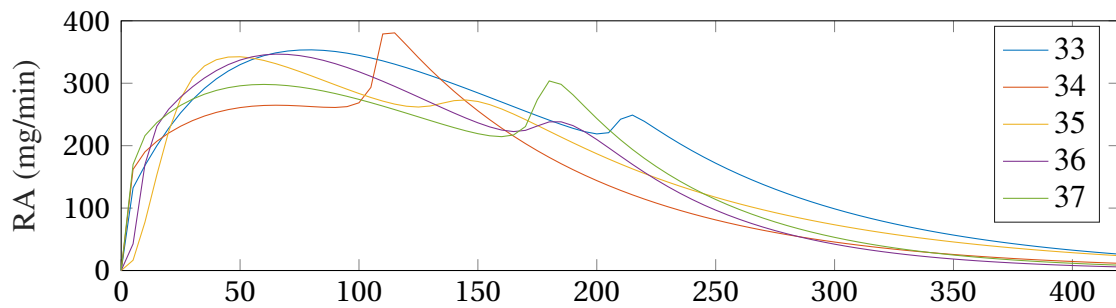
Curvas de RA del grupo 8



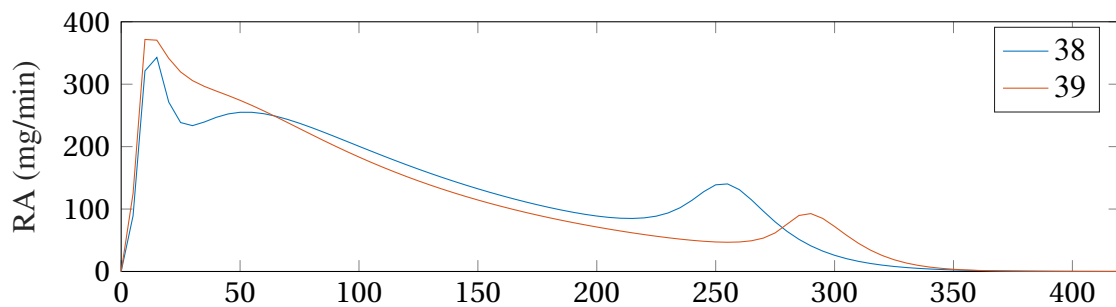
Curvas de RA del grupo 9



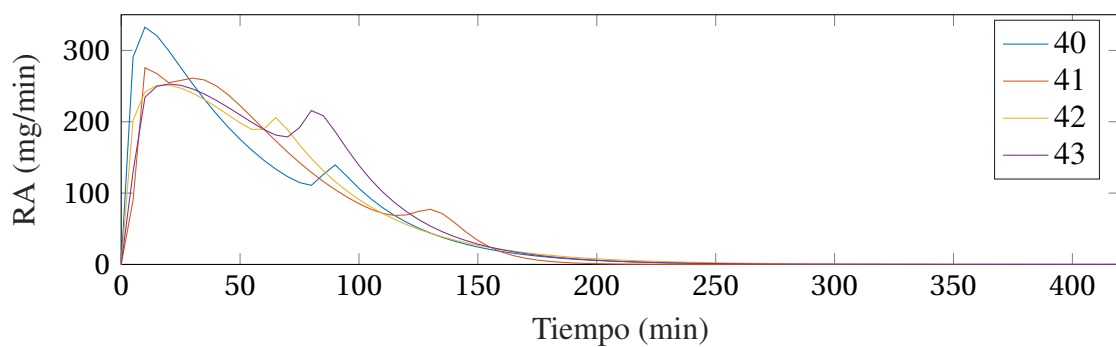
Curvas de RA del grupo 10



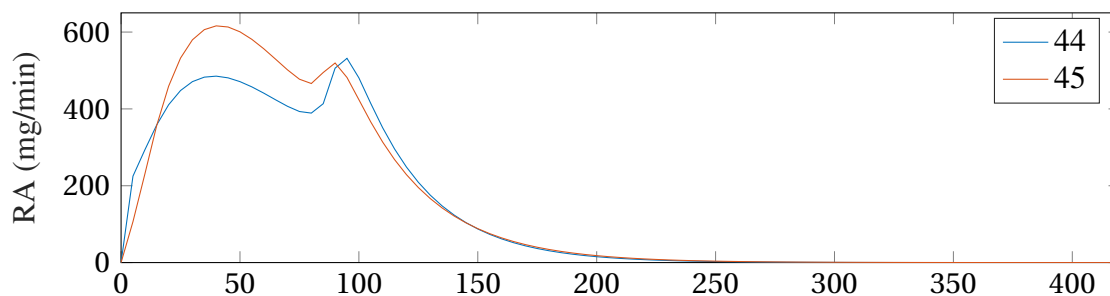
Curvas de RA del grupo 11



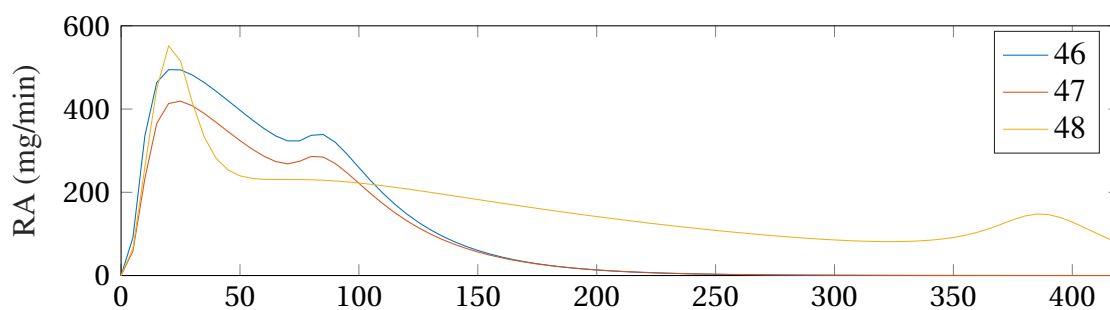
Curvas de RA del grupo 12



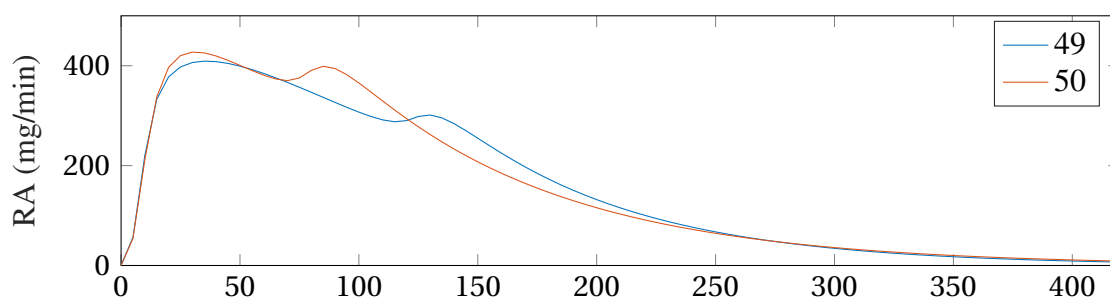
Curvas de RA del grupo 13



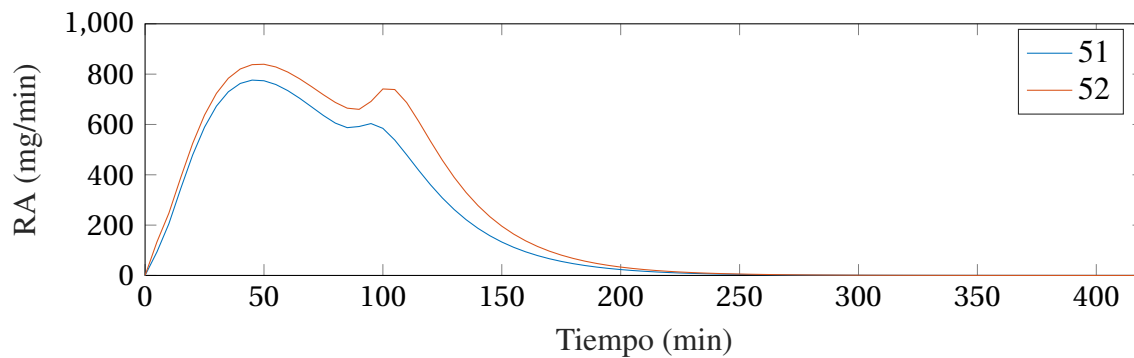
Curvas de RA del grupo 14



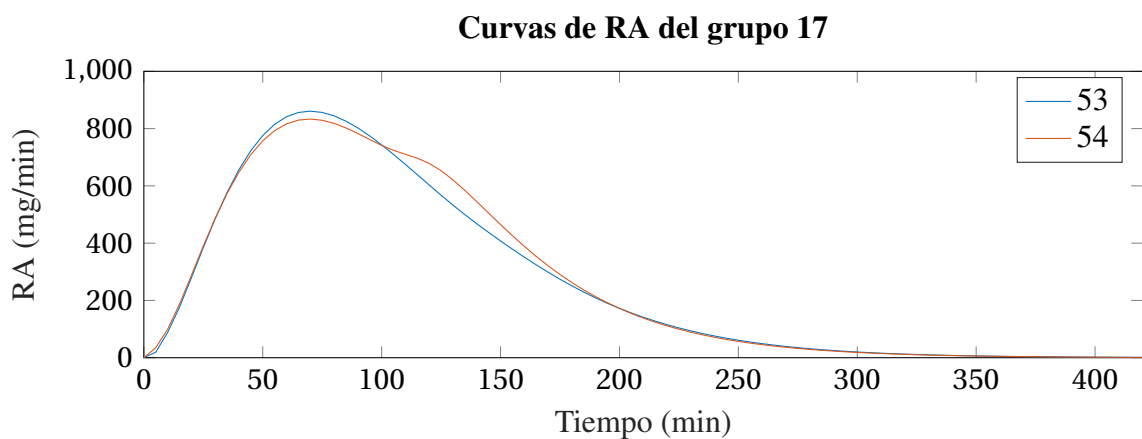
Curvas de RA del grupo 15



Curvas de RA del grupo 16



Grupo 17							
Alimentos	Clasificación	CHO (g)	Grasa (g)	Proteína (g)	Fibra (g)	Energía (kCal)	Tiempo (min)
53. Galletas de trigo, leche baja en grasa, pan integral, frutas, aves, vegetales	Grandes	123.00	29.00	17.00	-	673.78	180
54. Galletas de trigo, leche baja en grasa, pan integral, frutas, aves, vegetales	Grandes	126.00	28.00	25.00	-	685.73	180



PARÁMETROS Y DISTRIBUCIONES PARA GENERAR LA COHORTE VIRTUAL

En este apéndice se presentan los parámetros del modelo de Hovorka y sus distribuciones de probabilidad según (Hovorka et al., 2002, 2004; Chassin, 2005; Wilinska and Hovorka, 2008; Wilinska et al., 2010). Luego se muestra la cohorte virtual generada por la metodología expuesta en (Estremera et al., 2022).

Tabla B.1: Distribución de probabilidad y límites de los parámetros univariados del modelo de Hovorka.

Parámetros	Unidades	Distribución de probabilidad	Límites
K_a	min^{-1}	$K_a \sim N(0.018, 0.27 * 10^{-4})$	[0.005-0.060]
V_i	$l kg^{-1}$	$V_i \sim N(0.12, 0.12 * 10^{-3})$	[0.080-0.180]
k_e	min^{-1}	$k_e \sim N(0.147, 0.0011)$	[0.050-0.300]
T_{maxI}	min	$1/K_a$	[16.67-200]
V_g	$l kg^{-1}$	$Exp(V_g) \sim N(-1.869, 0.051)$	[0.09-0.25]
R_{cl}	min^{-1}	Cte=0.003	[0.003-0.03]
R_{th}	$mmoll^{-1}$	$R_{th} \sim N(9.52, 1.69)$	[7.5-15]
$k_{a_{int}}$	min^{-1}	$Exp(k_{a_{int}}) \sim N(-2.412, 1.37)$	-
t_{maxG}	min	$Exp(1/t_{maxG}) \sim N(-3.694, 0.04)$	[25-100]

$$M[F_{01}, S_{IT}, S_{ID}, S_{IE}, EGP_0] = [-4.336, -5.185, -6.53, -7.687, -3.732] \quad (B.1)$$

$$Cov[F_{01}, S_{IT}, S_{ID}, S_{IE}, EGP_0] = \begin{bmatrix} 0.2210 & 0.0529 & -0.2830 & -0.0240 & 0.2420 \\ 0.0529 & 0.1550 & -0.1772 & 0.2211 & 0.1578 \\ -0.2830 & -0.0177 & 0.8300 & -0.0717 & -0.2687 \\ -0.0240 & 0.2211 & -0.0717 & 0.4100 & 0.1476 \\ 0.2420 & 0.1578 & -0.2687 & 0.1476 & 0.4463 \end{bmatrix} \quad (B.2)$$

$$M[K_{a1}, K_{a2}, K_{a3}, K_{12}] = [-2.852, -5.773, -3.090, -3.846] \quad (B.3)$$

$$Cov[K_{a1}, K_{a2}, K_{a3}, K_{12}] = \begin{bmatrix} 0.1641 & -0.0560 & 0.0829 & 0.0499 \\ -0.0560 & 0.6732 & -0.0672 & 0.0908 \\ 0.0829 & -0.0672 & 0.8043 & 0.0067 \\ 0.0499 & 0.0908 & 0.0067 & 0.2817 \end{bmatrix} \quad (B.4)$$

Tabla B.2: Parámetros de los pacientes 1 hasta el 24 (parte 1).

	TmaxI	BW	Vi	ke	ka1	Sit	ka2	Sid	ka3	Sie
P1	50.7054	73.0570	0.1080	0.1184	0.0068	0.0014	0.0693	0.0005	0.0611	0.0466
P2	83.2205	67.4902	0.1314	0.1606	0.0016	0.0028	0.0315	0.0008	0.0253	0.0284
P3	36.7312	62.5456	0.1323	0.1595	0.0011	0.0008	0.0389	0.0017	0.0168	0.0251
P4	50.4517	84.4897	0.0939	0.1168	0.0021	0.0010	0.0182	0.0003	0.0144	0.0332
P5	40.8755	89.0595	0.1242	0.1887	0.0095	0.0016	0.0211	0.0013	0.0273	0.0133
P6	134.8036	88.7000	0.0958	0.1257	0.0017	0.0025	0.0162	0.0002	0.0262	0.0236
P7	45.8047	100.7474	0.1222	0.1356	0.0016	0.0024	0.0453	0.0002	0.0194	0.0447
P8	51.7814	72.5089	0.1037	0.1607	0.0045	0.0004	0.0148	0.0004	0.0220	0.0227
P9	58.0923	82.1498	0.1170	0.1144	0.0040	0.0018	0.0071	0.0009	0.0425	0.0296
P10	53.4906	51.9898	0.1240	0.1003	0.0026	0.0021	0.2765	0.0003	0.0274	0.0210
P11	60.7265	75.5530	0.1240	0.2193	0.0020	0.0015	0.0769	0.0001	0.0259	0.0160
P12	64.6443	101.6634	0.1117	0.1352	0.0010	0.0007	0.0373	0.0007	0.0361	0.0413
P13	54.0750	77.4315	0.1072	0.1615	0.0041	0.0006	0.0696	0.0017	0.0204	0.0141
P14	104.6895	57.9530	0.1317	0.0642	0.0129	0.0013	0.0216	0.0003	0.0119	0.0362
P15	64.8806	41.3623	0.1255	0.1689	0.0126	0.0040	0.1018	0.0003	0.0443	0.0227
P16	40.4574	75.6511	0.1127	0.1665	0.0013	0.0013	0.1179	0.0008	0.0099	0.0295
P17	61.4001	71.4204	0.1358	0.1808	0.0090	0.0007	0.0959	0.0008	0.0129	0.0260
P18	54.9328	64.1663	0.1296	0.1751	0.0032	0.0015	0.0433	0.0006	0.0284	0.0236
P19	34.1652	50.9014	0.1201	0.1505	0.0023	0.0023	0.0261	0.0001	0.0218	0.0352
P20	53.3345	47.1362	0.1408	0.1572	0.0129	0.0022	0.0597	0.0004	0.0176	0.0395
P21	44.9178	78.2702	0.1208	0.1944	0.0020	0.0019	0.1720	0.0001	0.0157	0.0168
P22	55.4636	48.0221	0.1146	0.1788	0.0046	0.0007	0.1028	0.0005	0.0198	0.0197
P23	92.0104	91.0572	0.1349	0.1751	0.0021	0.0009	0.0200	0.0005	0.0181	0.0185
P24	86.4673	89.6619	0.1379	0.1681	0.0080	0.0009	0.0227	0.0021	0.0118	0.0247

Tabla B.3: Parámetros de los pacientes 1 hasta el 24 (parte 2).

	k12	EGP0	Vg	F01	TmaxG	Bio	R_cl	R_th	Ka_int
P1	0.0523	0.0180	0.1581	0.0046	46.8619	0.8719	0.0030	10.0060	0.0095
P2	0.0587	0.0157	0.1755	0.0063	26.4019	1.0346	0.0030	7.7654	0.1409
P3	0.1009	0.0101	0.1614	0.0038	65.6238	1.0101	0.0030	9.8251	0.0395
P4	0.1065	0.0145	0.1581	0.0034	52.8370	0.8676	0.0030	9.1559	0.0583
P5	0.0768	0.0098	0.1999	0.0048	32.6598	1.0704	0.0030	12.1906	0.0843
P6	0.0343	0.0143	0.1832	0.0094	36.4230	1.1279	0.0030	10.5153	0.0296
P7	0.0495	0.0211	0.1221	0.0075	50.0021	1.1170	0.0030	9.6852	0.5800
P8	0.0830	0.0114	0.1497	0.0026	36.2175	0.9185	0.0030	11.5465	0.0350
P9	0.0531	0.0139	0.1899	0.0066	47.9656	0.8542	0.0030	10.4157	0.0443
P10	0.0655	0.0149	0.1356	0.0108	41.9200	1.0354	0.0030	11.9734	0.3575
P11	0.0570	0.0141	0.1311	0.0090	31.4096	0.8964	0.0030	9.5435	0.2154
P12	0.0798	0.0139	0.1457	0.0037	42.7982	0.7045	0.0030	9.5653	0.1091
P13	0.0583	0.0078	0.1420	0.0029	42.4002	0.8507	0.0030	8.0434	0.2015
P14	0.0514	0.0166	0.1888	0.0049	47.3528	0.7004	0.0030	10.5554	0.1825
P15	0.0544	0.0152	0.1587	0.0096	35.6043	0.9984	0.0030	10.4769	0.0583
P16	0.1052	0.0143	0.1313	0.0045	47.4410	0.8990	0.0030	9.0004	0.1508
P17	0.0631	0.0108	0.1211	0.0031	36.7603	1.0283	0.0030	8.0738	0.4532
P18	0.0344	0.0130	0.1416	0.0050	28.8100	1.1568	0.0030	10.2301	0.2985
P19	0.0276	0.0176	0.1775	0.0072	51.0802	1.0163	0.0030	7.5800	0.1765
P20	0.1063	0.0168	0.1365	0.0056	48.1697	0.7124	0.0030	12.0533	0.4121
P21	0.0892	0.0131	0.1998	0.0087	54.9185	0.7998	0.0030	11.2415	0.0933
P22	0.0527	0.0089	0.1759	0.0038	32.7802	0.9404	0.0030	8.1058	0.0200
P23	0.0452	0.0117	0.1069	0.0060	52.6429	0.8226	0.0030	9.4561	0.2799
P24	0.0335	0.0094	0.1076	0.0031	40.0346	1.1691	0.0030	8.5044	0.0562

Tabla B.4: Parámetros de los pacientes 25 hasta el 47 (parte 1).

TmaxI	BW	Vi	ke	ka1	Sit	ka2	Sid	ka3	Sie
P25	142.4826	87.7430	0.1149	0.1300	0.0060	0.0036	0.0488	0.0001	0.0282
P26	45.6655	110.9315	0.1148	0.1108	0.0022	0.0026	0.0664	0.0007	0.0162
P27	35.7436	73.7382	0.1255	0.1460	0.0011	0.0017	0.1052	0.0021	0.0168
P28	56.9384	71.6295	0.1226	0.1266	0.0024	0.0022	0.0141	0.0002	0.0145
P29	63.6907	104.9821	0.1117	0.1182	0.0023	0.0016	0.0130	0.0002	0.0220
P30	60.5539	82.9378	0.1193	0.1308	0.0025	0.0006	0.0083	0.0008	0.0485
P31	54.3101	53.8385	0.1194	0.1723	0.0042	0.0012	0.0865	0.0003	0.0124
P32	62.7992	86.9580	0.1330	0.1197	0.0015	0.0015	0.0342	0.0005	0.0145
P33	101.8119	73.5168	0.1045	0.0948	0.0046	0.0008	0.1084	0.0003	0.0407
P34	41.9235	85.4476	0.1221	0.1605	0.0065	0.0020	0.0102	0.0006	0.0213
P35	64.1941	77.8899	0.1029	0.1530	0.0018	0.0014	0.0284	0.0012	0.0400
P36	56.1583	44.1437	0.1214	0.1531	0.0073	0.0012	0.1484	0.0005	0.0253
P37	39.6967	68.7992	0.1443	0.1280	0.0030	0.0009	0.0415	0.0007	0.0417
P38	50.5858	62.9689	0.1359	0.2018	0.0043	0.0022	0.0319	0.0000	0.0241
P39	76.7056	38.7647	0.1208	0.1191	0.0050	0.0018	0.0264	0.0005	0.0184
P40	72.4680	70.3096	0.1164	0.1818	0.0051	0.0008	0.0869	0.0016	0.0094
P41	47.2757	64.4521	0.1186	0.1977	0.0017	0.0031	0.0764	0.0001	0.0175
P42	45.5005	64.6709	0.1331	0.1452	0.0023	0.0023	0.0952	0.0009	0.0110
P43	42.3443	96.3243	0.1288	0.1253	0.0026	0.0015	0.1229	0.0005	0.0182
P44	48.8809	87.9729	0.1079	0.0900	0.0025	0.0009	0.0411	0.0003	0.0096
P45	43.4992	74.8541	0.1199	0.1491	0.0014	0.0024	0.0641	0.0020	0.0202
P46	88.5559	47.6192	0.1230	0.1965	0.0032	0.0072	0.0566	0.0001	0.0618
P47	65.2773	90.5243	0.1115	0.0998	0.0010	0.0007	0.1569	0.0009	0.0182

Tabla B.5: Parámetros de los pacientes 25 hasta el 47 (parte 2).

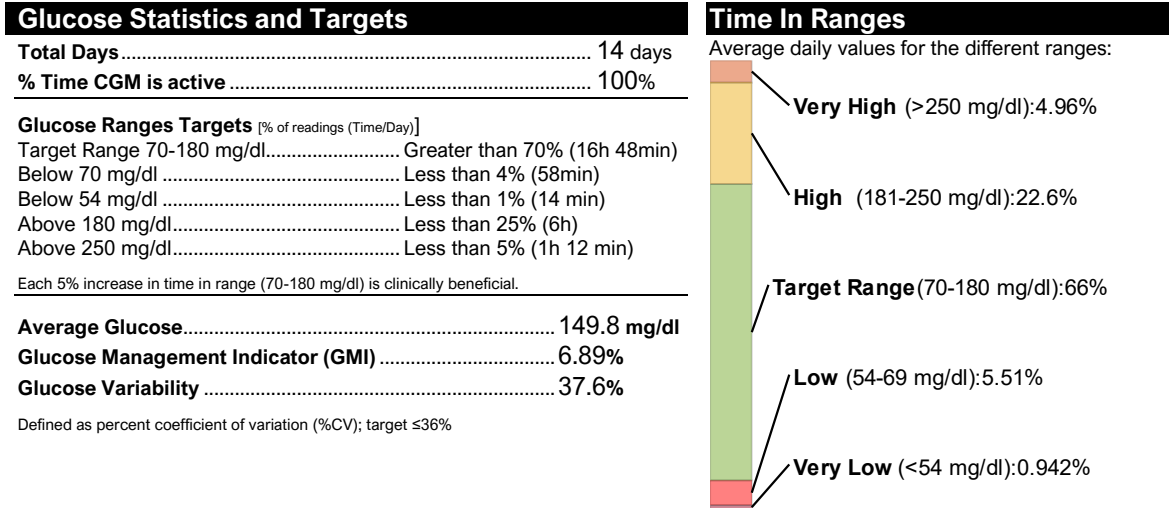
	k12	EGP0	Vg	F01	TmaxG	Bio	R_cl	R_th	Ka_int
P25	0.0498	0.0153	0.1773	0.0089	48.0804	1.1305	0.0030	7.6909	0.0850
P26	0.0308	0.0164	0.1769	0.0062	33.7650	1.1054	0.0030	8.5370	0.4276
P27	0.0862	0.0103	0.1797	0.0042	38.1493	0.9398	0.0030	7.7346	0.2741
P28	0.0395	0.0189	0.1982	0.0103	32.4643	1.0259	0.0030	9.9416	0.0198
P29	0.0559	0.0168	0.1820	0.0086	26.8392	0.7461	0.0030	11.4027	0.1751
P30	0.0513	0.0129	0.1461	0.0029	34.9771	1.0232	0.0030	7.8673	0.0921
P31	0.0458	0.0114	0.2189	0.0054	42.7419	0.8182	0.0030	8.1037	0.0356
P32	0.0396	0.0094	0.1518	0.0057	35.4251	1.0804	0.0030	10.5278	0.2611
P33	0.0676	0.0109	0.1869	0.0044	40.4510	1.1085	0.0030	9.2409	0.0466
P34	0.0762	0.0154	0.1311	0.0057	44.3240	0.7923	0.0030	7.9835	0.0554
P35	0.0463	0.0126	0.1920	0.0039	42.0581	1.1591	0.0030	8.0807	0.0104
P36	0.0554	0.0106	0.2147	0.0062	45.7015	0.7319	0.0030	10.0767	0.0640
P37	0.0760	0.0092	0.1137	0.0062	43.9258	1.1901	0.0030	9.1627	0.0317
P38	0.0217	0.0218	0.1451	0.0104	38.3927	0.8198	0.0030	11.0835	0.1938
P39	0.0418	0.0109	0.0940	0.0062	34.2850	1.1076	0.0030	7.6188	0.0787
P40	0.0908	0.0094	0.0997	0.0045	36.5737	1.0366	0.0030	9.0093	0.0676
P41	0.0789	0.0185	0.1558	0.0091	37.9304	1.0143	0.0030	8.6419	0.0334
P42	0.0594	0.0126	0.2195	0.0073	33.8559	0.9709	0.0030	10.0794	0.1745
P43	0.0462	0.0089	0.1122	0.0068	41.0820	0.9163	0.0030	10.8852	0.0144
P44	0.0385	0.0115	0.1440	0.0032	33.6673	0.9625	0.0030	9.0313	0.2395
P45	0.0470	0.0127	0.2287	0.0058	46.3173	0.8313	0.0030	9.8339	0.5548
P46	0.1049	0.0228	0.1065	0.0128	31.3947	0.8543	0.0030	9.3438	0.0034
P47	0.1466	0.0107	0.1585	0.0039	58.3057	1.1782	0.0030	10.3234	0.1662

REPORTE DE PERFIL DE GLUCOSA AMBULATORIO (AGP)

En este apéndice se presenta el archivo Suplementario 2 correspondiente al primer documento del Capítulo 2 (Estremera et al., 2022), muestra el informe AGP de dos pacientes representativos con terapia CSII y MDI.

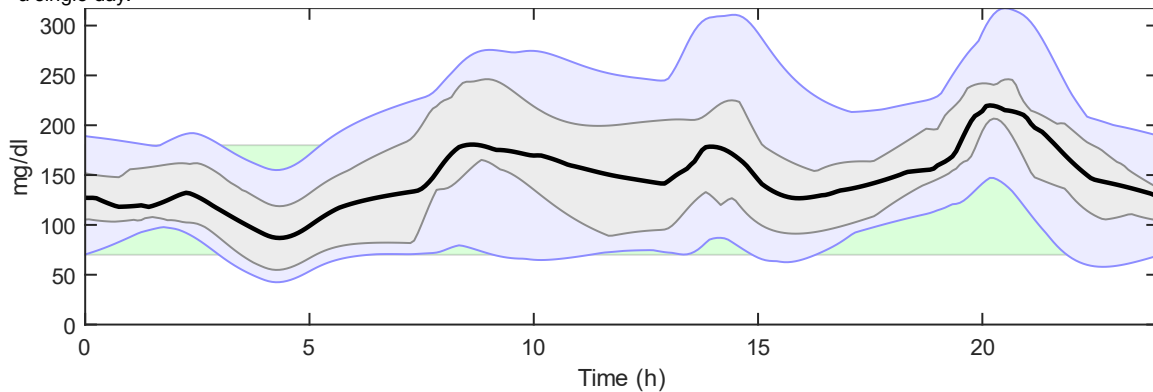
AGP Report

Name: CSII Subject7
Date: 18-Jun-2021

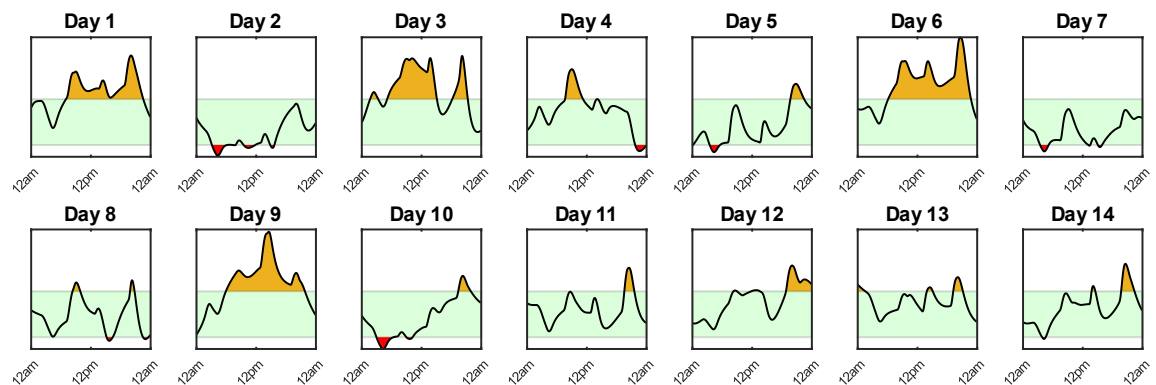


Ambulatory Glucose Profile (AGP)

AGP is a summary of glucose values from the report period, with median (50%) and other percentiles shown as if occurring in a single day.



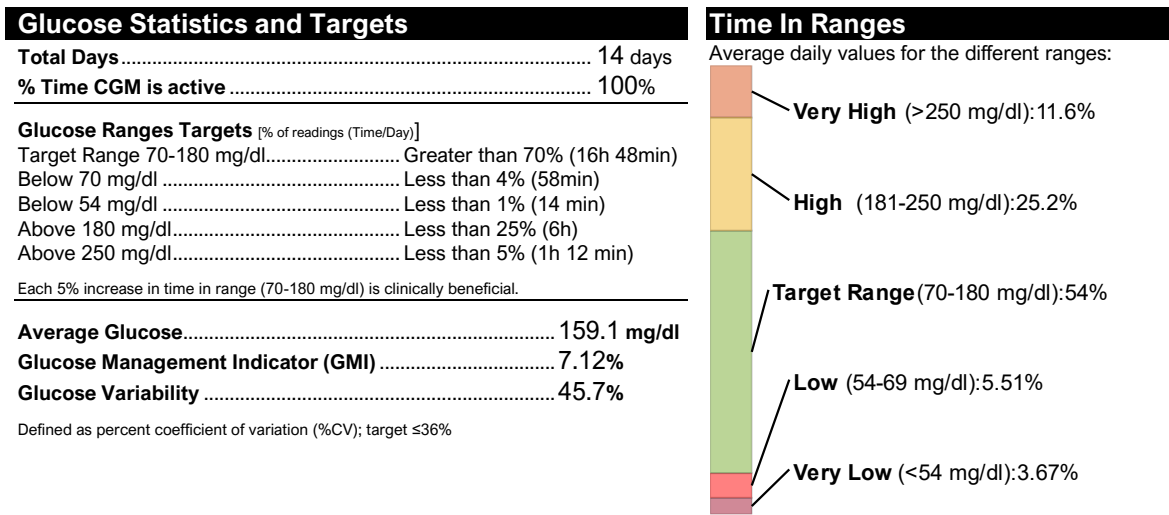
Daily Glucose Profiles



Target area of 70-180 mg/dl is green colored, hyperglycemia is colored in orange and hypoglycemia in red.

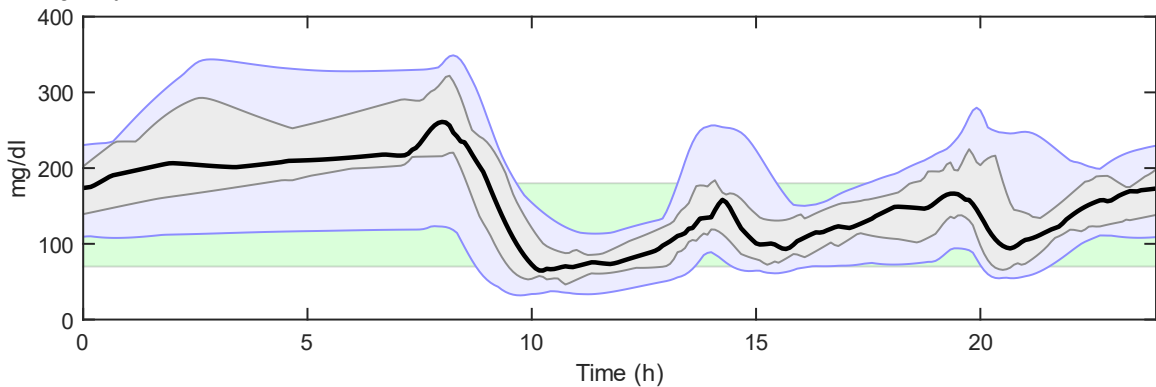
AGP Report

Name: Gla-300_9H Subject37
Date: 23-Jun-2021

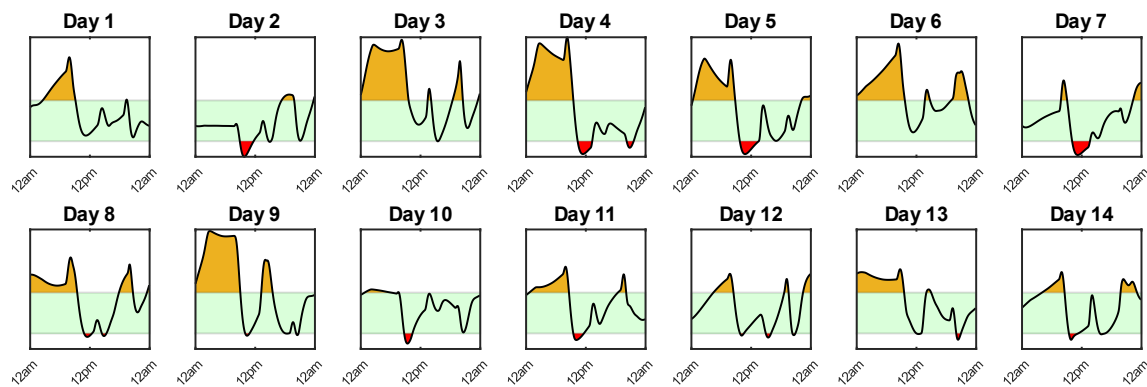


Ambulatory Glucose Profile (AGP)

AGP is a summary of glucose values from the report period, with median (50%) and other percentiles shown as if occurring in a single day.



Daily Glucose Profiles



Target area of 70-180 mg/dl is green colored, hyperglycemia is colored in orange and hypoglycemia in red.

REFERENCIAS BIBLIOGRÁFICAS

- Ackerman, E., Rosevear, J. W. and McGuckin, W. F. (1964), 'A mathematical model of the glucose-tolerance test', *Physics in medicine & Biology* **9**(2), 203.
- Adolfsson, P., Hartvig, N. V., Kaas, A., Møller, J. B. and Hellman, J. (2020), 'Increased time in range and fewer missed bolus injections after introduction of a smart connected insulin pen', *Diabetes Technology & Therapeutics* **22**(10), 709–718.
- Advani, A. (2020), 'Positioning time in range in diabetes management', *Diabetologia* **63**(2), 242–252.
- Aitchison, J. (1982), 'The statistical analysis of compositional data', *Journal of the Royal Statistical Society: Series B (Methodological)* **44**(2), 139–160.
- Aitchison, J. (1986), 'Monographs on statistics and applied probability'.
- Aleppo, G., Ruedy, K. J., Riddlesworth, T. D., Kruger, D. F., Peters, A. L., Hirsch, I., Bergenstal, R. M., Toschi, E., Ahmann, A. J., Shah, V. N. et al. (2017), 'Replace-bg: a randomized trial comparing continuous glucose monitoring with and without routine blood glucose monitoring in adults with well-controlled type 1 diabetes', *Diabetes care* **40**(4), 538–545.
- Alfian, G., Syafrudin, M., Rhee, J., Anshari, M., Mustakim, M. and Fahrurrozi, I. (2020), Blood glucose prediction model for type 1 diabetes based on extreme gradient boosting, in 'IOP Conference Series: Materials Science and Engineering', Vol. 803, IOP Publishing, p. 012012.

- Ali, J. B., Hamdi, T., Fnaiech, N., Di Costanzo, V., Fnaiech, F. and Ginoux, J.-M. (2018), ‘Continuous blood glucose level prediction of type 1 diabetes based on artificial neural network’, *Biocybernetics and Biomedical Engineering* **38**(4), 828–840.
- Aliberti, A., Pupillo, I., Terna, S., Macii, E., Di Cataldo, S., Patti, E. and Acquaviva, A. (2019), ‘A multi-patient data-driven approach to blood glucose prediction’, *IEEE Access* **7**, 69311–69325.
- Alsofiani, W. A., Alessa, B. H., Alsabaan, F., Althemery, A. U., Ghith, A. M. and Alfaifi, A. A. (2022), ‘Effects of ultra-long-acting insulin compared to long-acting insulin on diabetic ketoacidosis incidence in type 1 diabetes mellitus patients’, *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy* pp. 733–739.
- Anand, P. K., Shin, D. R. and Memon, M. L. (2020), ‘Adaptive boosting based personalized glucose monitoring system (pgms) for non-invasive blood glucose prediction with improved accuracy’, *Diagnostics* **10**(5), 285.
- Assessment, G. (2022), ‘6. glycemetic targets: standards of medical care in diabetes—2022’, *Diabetes Care* **45**, S83.
- Atkinson, M. A., Bluestone, J. A., Eisenbarth, G. S., Hebrok, M., Herold, K. C., Accili, D., Pietropaolo, M., Arvan, P. R., Von Herrath, M., Markel, D. S. et al. (2011), ‘How does type 1 diabetes develop? the notion of homicide or β -cell suicide revisited’, *Diabetes* **60**(5), 1370–1379.
- Avgerinos, I., Papanastasiou, G., Karagiannis, T., Michailidis, T., Liakos, A., Mainou, M., Matthews, D. R., Tsapas, A. and Bekiari, E. (2021), ‘Ultra-rapid-acting insulins for adults with diabetes: A systematic review and meta-analysis’, *Diabetes, Obesity and Metabolism* **23**(10), 2395–2401.

- Ayano-Takahara, S., Ikeda, K., Fujimoto, S., Hamasaki, A., Harashima, S.-i., Toyoda, K., Fujita, Y., Nagashima, K., Tanaka, D. and Inagaki, N. (2015), 'Glycemic variability is associated with quality of life and treatment satisfaction in patients with type 1 diabetes', *Diabetes Care* **38**(1), e1–e2.
- Babiker, A., Alammari, N., Aljuraishi, A., Alharbi, R., Alqarni, H., Masuadi, E. and Alfaraidi, H. (2022), 'The effectiveness of insulin pump therapy versus multiple daily injections in children with type 1 diabetes mellitus in a specialized center in riyadh', *Clinical Medicine Insights: Endocrinology and Diabetes* **15**, 11795514221128495.
- Bailey, T., Pettus, J., Roussel, R., Schmider, W., Maroccia, M., Nassr, N., Klein, O., Bolli, G. and Dahmen, R. (2018), 'Morning administration of 0.4 u/kg/day insulin glargine 300 u/ml provides less fluctuating 24-hour pharmacodynamics and more even pharmacokinetic profiles compared with insulin degludec 100 u/ml in type 1 diabetes', *Diabetes & metabolism* **44**(1), 15–21.
- Benjamin, E. M. (2002), 'Self-monitoring of blood glucose: the basics'.
- Bergental, R. M., Bailey, T. S., Rodbard, D., Ziemien, M., Guo, H., Muehlen-Bartmer, I. and Ahmann, A. J. (2017), 'Comparison of insulin glargine 300 units/ml and 100 units/ml in adults with type 1 diabetes: continuous glucose monitoring profiles and variability using morning or evening injections', *Diabetes care* **40**(4), 554–560.
- Bergman, R. N. (2021), 'Origins and history of the minimal model of glucose regulation', *Frontiers in Endocrinology* **11**, 583016.
- Bergman, R. N. and Cobelli, C. (1980), Minimal modeling, partition analysis, and the estimation of insulin sensitivity, in 'Federation proceedings', Vol. 39, pp. 110–115.

- Bergman, R. N., Ider, Y. Z., Bowden, C. R. and Cobelli, C. (1979), ‘Quantitative estimation of insulin sensitivity.’, *American Journal of Physiology-Endocrinology And Metabolism* **236**(6), E667.
- Bergman, R. N. and Urquhart, J. (1971), The pilot gland approach to the study of insulin secretory dynamics, *in* ‘Proceedings of the 1970 Laurentian Hormone Conference’, Elsevier, pp. 583–605.
- Biagi, L., Bertachi, A., Giménez, M., Conget, I., Bondia, J., Martín-Fernández, J. A. and Vehí, J. (2019), ‘Individual categorisation of glucose profiles using compositional data analysis’, *Statistical methods in medical research* **28**(12), 3550–3567.
- Biagi, L., Bertachi, A., Giménez, M., Conget, I., Bondia, J., Martín-Fernández, J. A. and Vehí, J. (2021), ‘Probabilistic model of transition between categories of glucose profiles in patients with type 1 diabetes using a compositional data analysis approach’, *Sensors* **21**(11), 3593.
- Bjerg, L., Hulman, A., Carstensen, B., Charles, M., Witte, D. R. and Jørgensen, M. E. (2019), ‘Effect of duration and burden of microvascular complications on mortality rate in type 1 diabetes: an observational clinical cohort study’, *Diabetologia* **62**, 633–643.
- Bolli, G. B., Kerr, D., Thomas, R., Torlone, E., Sola-Gazagnes, A., Vitacolonna, E., Selam, J. L. and Home, P. D. (2009), ‘Comparison of a multiple daily insulin injection regimen (basal once-daily glargine plus mealtime lispro) and continuous subcutaneous insulin infusion (lispro) in type 1 diabetes: a randomized open parallel multicenter study’, *Diabetes care* **32**(7), 1170–1176.
- Bondia, J., Romero-Vivo, S., Ricarte, B. and Diez, J. L. (2018), ‘Insulin estimation and prediction: A review of the estimation and prediction of subcutaneous insulin pharmacokinetics in closed-loop glucose control’, *IEEE Control Systems Magazine* **38**(1), 47–66.

- Brink, S. J. (2022), ‘Insulin past, present, and future: 100 years from leonard thompson’, *Diabetology* **3**(1), 117–158.
- Buccianti, A., Mateu-Figueras, G. and Pawlowsky-Glahn, V. (2006), *Compositional data analysis in the geosciences: from theory to practice*, Geological Society of London.
- Butt, H., Khosa, I. and Iftikhar, M. A. (2023), ‘Feature transformation for efficient blood glucose prediction in type 1 diabetes mellitus patients’, *Diagnostics* **13**(3), 340.
- Cabrera, A., Biagi, L., Beneyto, A., Estremera, E., Contreras, I., Giménez, M., Conget, I., Bondía, J., Martín-Fernández, J. A. and Vehí, J. (2023), ‘Validation of a probabilistic prediction model for patients with type 1 diabetes using compositional data analysis’, *Mathematics* **11**(5), 1241.
- Care, D. (2022), ‘Diabetes: Standards of medical care in diabetes—2022’, *Diabetes Care* **45**(1), S113–S124.
- Chassin, L. J. (2005), *In silico testing of glucose controllers: methodology and sample application*, PhD thesis, City University London.
- Chastin, S. F., Palarea-Albaladejo, J., Dontje, M. L. and Skelton, D. A. (2015), ‘Combined effects of time spent in physical activity, sedentary behaviors and sleep on obesity and cardio-metabolic health markers: a novel compositional data analysis approach’, *PloS one* **10**(10), e0139984.
- Chen, J., Li, K., Herrero, P., Zhu, T. and Georgiou, P. (2018), Dilated recurrent neural network for short-time prediction of glucose concentration., *in* ‘KHD@ IJCAI’, pp. 69–73.
- Chen, X., Tuo, J. and Wang, Y. (2019), A prediction method for blood glucose based on grey wolf optimization evolving kernel extreme learning machine, *in* ‘2019 Chinese Control Conference (CCC)’, IEEE, pp. 3000–3005.

- Cicchella, D., Ambrosino, M., Gramazio, A., Coraggio, F., Musto, M. A., Caputi, A., Avagliano, D. and Albanese, S. (2022), 'Using multivariate compositional data analysis (coda) and clustering to establish geochemical backgrounds in stream sediments of an onshore oil deposits area. the agri river basin (italy) case study', *Journal of Geochemical Exploration* **238**, 107012.
- Cichosz, S. L., Kronborg, T., Jensen, M. H. and Hejlesen, O. (2021), 'Penalty weighted glucose prediction models could lead to better clinically usage', *Computers in Biology and Medicine* **138**, 104865.
- Cobry, E., McFann, K., Messer, L., Gage, V., VanderWel, B., Horton, L. and Chase, H. P. (2010), 'Timing of meal insulin boluses to achieve optimal postprandial glycemic control in patients with type 1 diabetes', *Diabetes technology & therapeutics* **12**(3), 173–177.
- Collaboration, E. R. F. et al. (2010), 'Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies', *The lancet* **375**(9733), 2215–2222.
- Combettes, P. L. and Müller, C. L. (2021), 'Regression models for compositional data: General log-contrast formulations, proximal optimization, and microbiome data applications', *Statistics in Biosciences* **13**(2), 217–242.
- Contreras, I. and Vehi, J. (2018), 'Artificial intelligence for diabetes management and decision support: literature review', *Journal of medical Internet research* **20**(5), e10775.
- Dalla Man, C., Rizza, R. A. and Cobelli, C. (2007), 'Meal simulation model of the glucose-insulin system', *IEEE Transactions on biomedical engineering* **54**(10), 1740–1749.
- Daniels, J., Herrero, P. and Georgiou, P. (2021), 'A multitask learning approach to personalized blood glucose prediction', *IEEE Journal of Biomedical and Health Informatics* **26**(1), 436–445.

- Dave, D., DeSalvo, D. J., Haridas, B., McKay, S., Shenoy, A., Koh, C. J., Lawley, M. and Erraguntla, M. (2021), 'Feature-based machine learning model for real-time hypoglycemia prediction', *Journal of Diabetes Science and Technology* **15**(4), 842–855.
- De Bois, M., Ammi, M. and El Yacoubi, M. A. (2019), Model fusion to enhance the clinical acceptability of long-term glucose predictions, in '2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)', IEEE, pp. 258–264.
- DeVries, J. H. (2013), 'Glucose variability: where it is important and how to measure it', *Diabetes* **62**(5), 1405.
- Didyuk, O., Econom, N., Guardia, A., Livingston, K. and Klueh, U. (2021), 'Continuous glucose monitoring devices: past, present, and future focus on the history and evolution of technological innovation', *Journal of diabetes science and technology* **15**(3), 676–683.
- Dmitrijeva, M., Ehrig, K. J., Ciobanu, C. L., Cook, N. J., Verdugo-Ihl, M. R. and Metcalfe, A. V. (2019), 'Defining iocg signatures through compositional data analysis: A case study of lithochemical zoning from the olympic dam deposit, south australia', *Ore Geology Reviews* **105**, 86–101.
- Dowling, E. and Maidment, D. W. (2023), 'Assessing depressive symptoms and diabetes distress in type 1 diabetic adults: A comparison of continuous subcutaneous insulin infusion (csii) and multiple daily injections (mdi) users', *Journal of Health Psychology* **28**(2), 149–161.
- Dumuid, D., Stanford, T. E., Martin-Fernández, J.-A., Pedišić, Ž., Maher, C. A., Lewis, L. K., Hron, K., Katzmarzyk, P. T., Chaput, J.-P., Fogelholm, M. et al. (2018), 'Compositional data analysis for physical activity, sedentary time and sleep research', *Statistical methods in medical research* **27**(12), 3726–3738.
- Egozcue, J. J. and Pawlowsky-Glahn, V. (2005), 'Groups of parts and their balances in compositional data analysis', *Mathematical Geology* **37**, 795–828.

- Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G. and Barcelo-Vidal, C. (2003), 'Isometric logratio transformations for compositional data analysis', *Mathematical geology* **35**(3), 279–300.
- El Idriss, T., Idri, A., Abnane, I. and Bakkoury, Z. (2019), Predicting blood glucose using an lstm neural network, in '2019 Federated Conference on Computer Science and Information Systems (FedCSIS)', IEEE, pp. 35–41.
- Estremera, E., Cabrera, A., Beneyto, A. and Vehi, J. (2022), 'A simulator with realistic and challenging scenarios for virtual t1d patients undergoing csii and mdi therapy', *Journal of Biomedical Informatics* **132**, 104141.
- Felizardo, V., Garcia, N. M., Pombo, N. and Megdiche, I. (2021), 'Data-based algorithms and models using diabetics real data for blood glucose and hypoglycaemia prediction—a systematic literature review', *Artificial Intelligence in Medicine* **118**, 102120.
- Filzmoser, P. and Filzmoser (2021), *Advances in Compositional Data Analysis*, Springer.
- Freckmann, G. (2020), 'Basics and use of continuous glucose monitoring (cgm) in diabetes therapy', *Journal of Laboratory Medicine* **44**(2), 71–79.
- Gadaleta, M., Facchinetti, A., Grisan, E. and Rossi, M. (2018), 'Prediction of adverse glycemic events from continuous glucose monitoring signal', *IEEE journal of biomedical and health informatics* **23**(2), 650–659.
- Galasso, S., Facchinetti, A., Bonora, B., Mariano, V., Boscari, F., Cipponeri, E., Maran, A., Avogaro, A., Fadini, G. and Bruttomesso, D. (2016), 'Switching from twice-daily glargine or detemir to once-daily degludec improves glucose control in type 1 diabetes. an observational study', *Nutrition, Metabolism and Cardiovascular Diseases* **26**(12), 1112–1119.

- Gao, R. and Chilibeck, P. D. (2020), 'Nutritional interventions during bed rest and spaceflight: prevention of muscle mass and strength loss, bone resorption, glucose intolerance, and cardiovascular problems', *Nutrition Research* **82**, 11–24.
- García Jaramillo, M. A. et al. (2011), 'Prediction of postprandial blood glucose under inpatient variability and uncertainty and its use in the design of insulin dosing strategies for type 1 diabetic patients'.
- Griva, L. O. and Basualdo, M. S. (2018), Evaluating clinical accuracy of models for predicting glycemic behavior for diabetes care, in '2018 Argentine conference on automatic control (AADECA)', IEEE, pp. 1–6.
- Group, N. D. D. (1979), 'Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance', *diabetes* **28**(12), 1039–1057.
- Guagliardi, I., Zuzolo, D., Albanese, S., Lima, A., Cerino, P., Pizzolante, A., Thiombane, M., De Vivo, B. and Cicchella, D. (2020), 'Uranium, thorium and potassium insights on campania region (italy) soils: Sources patterns based on compositional data analysis and fractal model', *Journal of geochemical exploration* **212**, 106508.
- Hamdi, T., Ali, J. B., Di Costanzo, V., Fnaiech, F., Moreau, E. and Ginoux, J.-M. (2018), 'Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm', *Biocybernetics and Biomedical Engineering* **38**(2), 362–372.
- Heise, T., Kaplan, K. and Haahr, H. L. (2018), 'Day-to-day and within-day variability in glucose-lowering effect between insulin degludec and insulin glargine (100 u/ml and 300 u/ml): a comparison across studies', *Journal of Diabetes Science and Technology* **12**(2), 356–363.

- Hernández-Ordoñez, M. and Campos-Delgado, D. U. (2008), ‘An extension to the compartmental model of type 1 diabetic patients to reproduce exercise periods with glycogen depletion and replenishment’, *Journal of biomechanics* **41**(4), 744–752.
- Herrero, P., Bondia, J., Adewuyi, O., Pesl, P., El-Sharkawy, M., Reddy, M., Toumazou, C., Oliver, N. and Georgiou, P. (2017), ‘Enhancing automatic closed-loop glucose control in type 1 diabetes with an adaptive meal bolus calculator—in silico evaluation under intra-day variability’, *Computer methods and programs in biomedicine* **146**, 125–131.
- Herrero, P., Pesl, P., Bondia, J., Reddy, M., Oliver, N., Georgiou, P. and Toumazou, C. (2015), ‘Method for automatic adjustment of an insulin bolus calculator: In silico robustness evaluation under intra-day variability’, *Computer methods and programs in biomedicine* **119**(1), 1–8.
- Hipszer, B. R. (2008), *A mathematical model of glucose metabolism in hospitalized patients with diabetes and stress hyperglycemia*, Vol. 70.
- Hirsch, I. B., Juneja, R., Beals, J. M., Antalis, C. J. and Wright Jr, E. E. (2020), ‘The evolution of insulin and how it informs therapy and treatment choices’, *Endocrine reviews* **41**(5), 733–755.
- Holt, R. I., Cockram, C., Flyvbjerg, A. and Goldstein, B. J. (2017), *Textbook of diabetes*, John Wiley & Sons.
- Home, P. D., Bergenstal, R. M., Bolli, G. B., Ziemien, M., Rojas, M., Espinasse, M. and Riddle, M. C. (2015), ‘New insulin glargine 300 units/ml versus glargine 100 units/ml in people with type 1 diabetes: a randomized, phase 3a, open-label clinical trial (edition 4)’, *Diabetes care* **38**(12), 2217–2225.
- Hovorka, R. (2006), ‘Continuous glucose monitoring and closed-loop systems’, *Diabetic medicine* **23**(1), 1–12.

- Hovorka, R. (2011), ‘Closed-loop insulin delivery: from bench to clinical practice’, *Nature Reviews Endocrinology* **7**(7), 385–395.
- Hovorka, R., Canonico, V., Chassin, L. J., Haueter, U., Massi-Benedetti, M., Federici, M. O., Pieber, T. R., Schaller, H. C., Schaupp, L., Vering, T. et al. (2004), ‘Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes’, *Physiological measurement* **25**(4), 905.
- Hovorka, R., Shojaee-Moradie, F., Carroll, P. V., Chassin, L. J., Gowrie, I. J., Jackson, N. C., Tudor, R. S., Umpleby, A. M. and Jones, R. H. (2002), ‘Partitioning glucose distribution/transport, disposal, and endogenous production during ivgtt’, *American Journal of Physiology-Endocrinology and Metabolism* **282**(5), E992–E1007.
- Janssen, I., Clarke, A. E., Carson, V., Chaput, J.-P., Giangregorio, L. M., Kho, M. E., Poitras, V. J., Ross, R., Saunders, T. J., Ross-White, A. et al. (2020), ‘A systematic review of compositional data analysis studies examining associations between sleep, sedentary behaviour, and physical activity with health outcomes in adults’, *Applied physiology, nutrition, and metabolism* **45**(10), S248–S257.
- Khanam, J. J. and Foo, S. Y. (2021), ‘A comparison of machine learning algorithms for diabetes prediction’, *ICT Express* **7**(4), 432–439.
- Kovatchev, B. (2019), ‘Glycemic variability: risk factors, assessment, and control’, *Journal of diabetes science and technology* **13**(4), 627–635.
- Kovatchev, B. P., Breton, M., Dalla Man, C. and Cobelli, C. (2009), ‘In silico preclinical trials: a proof of concept in closed-loop control of type 1 diabetes’.
- Lachin, J. M., Genuth, S., Nathan, D. M., Zinman, B., Rutledge, B. N. and Group, D. R. (2008), ‘Effect of glycemic exposure on the risk of microvascular complications in the diabetes control and complications trial—revisited’, *Diabetes* **57**(4), 995–1001.

- Leal Moncada, Y. T. et al. (2013), 'Enfoques metodológicos para mejorar la precisión y fiabilidad de los sistemas de monitorización continua de glucosa en pacientes con diabetes tipo 1 y en pacientes críticos mediante técnicas lineales y no lineales'.
- Lee, T.-Y., Kuo, S., Yang, C.-Y. and Ou, H.-T. (2020), 'Cost-effectiveness of long-acting insulin analogues vs intermediate/long-acting human insulin for type 1 diabetes: A population-based cohort followed over 10 years', *British journal of clinical pharmacology* **86**(5), 852–860.
- Li, K. e. a. (2019), 'Convolutional recurrent neural networks for glucose prediction', *IEEE journal of biomedical and health informatics* **24**(2), 603–613.
- Li, K. and Liu, C. e. a. (2019), 'Glunet: A deep learning framework for accurate glucose forecasting', *IEEE journal of biomedical and health informatics* **24**(2), 414–423.
- Li, N., Tuo, J. and Wang, Y. (2018), Chaotic time series analysis approach for prediction blood glucose concentration based on echo state networks, in '2018 Chinese control and decision conference (CCDC)', IEEE, pp. 2017–2022.
- Li, N., Tuo, J., Wang, Y. and Wang, M. (2020), 'Prediction of blood glucose concentration for type 1 diabetes based on echo state networks embedded with incremental learning', *Neurocomputing* **378**, 248–259.
- Li, T., Jia, Y., Wang, S., Wang, A., Gao, L., Yang, C. and Zou, H. (2019), 'Retinal microvascular abnormalities in children with type 1 diabetes mellitus without visual impairment or diabetic retinopathy', *Investigative ophthalmology & visual science* **60**(4), 990–998.
- Lucidi, P., Porcellati, F., Marinelli Andreoli, A., Carriero, I., Candeloro, P., Cioli, P., Bolli, G. B. and Fanelli, C. G. (2015), 'Pharmacokinetics and pharmacodynamics of nph insulin in type 1 diabetes: the importance of appropriate resuspension before subcutaneous injection', *Diabetes Care* **38**(12), 2204–2210.

- Makroum, M. A., Adda, M., Bouzouane, A. and Ibrahim, H. (2022), ‘Machine learning and smart devices for diabetes management: Systematic review’, *Sensors* **22**(5), 1843.
- Man, C. D., Micheletto, F., Lv, D., Breton, M., Kovatchev, B. and Cobelli, C. (2014), ‘The uva/padova type 1 diabetes simulator: new features’, *Journal of diabetes science and technology* **8**(1), 26–34.
- Marling, C. and Bunescu, R. (2020), The ohiot1dm dataset for blood glucose level prediction: Update 2020, in ‘CEUR workshop proceedings’, Vol. 2675, NIH Public Access, p. 71.
- Martín Fernández, J. A. (2001), *Medidas de diferencia y clasificación automática no paramétrica de datos composicionales*, Universitat Politècnica de Catalunya.
- Martín-Fernández, J. A. (2019), ‘Comments on: Compositional data: The sample space and its structure’, *Test* **28**(3), 653–657.
- Martín-Fernández, J. A., Barceló-Vidal, C. and Pawlowsky-Glahn, V. (2003), ‘Dealing with zeros and missing values in compositional data sets using nonparametric imputation’, *Mathematical Geology* **35**, 253–278.
- Martin-Fernández, J. A., Bren, M., Barceló-Vidal, C. and Pawlowsky-Glahn, V. (1999), A measure of difference for compositional data based on measures of divergence, in ‘Proceedings of IAMG’, Vol. 99, pp. 211–216.
- Martín Fernández, J. A., Daunis-i Estadella, P. and Mateu i Figueras, G. (2015), ‘On the interpretation of differences between groups for compositional data’, *SORT: statistics and operations research transactions, 2015, vol. 39, núm. 2, p. 231-252* .
- Martinsson, J., Schliep, A., Eliasson, B. and Mogren, O. (2020), ‘Blood glucose prediction with variance estimation using recurrent neural networks’, *Journal of Healthcare Informatics Research* **4**, 1–18.

- Mateu-Figueras, G., Martín-Fernández, J., Pawlowsky-Glahn, V. and Barceló-Vidal, C. (2003), El problema del análisis estadístico de datos composicionales, *in* '27 Congreso Nacional de Estadística e Investigación Operativa, Lleida, abril del', Vol. 8.
- Matschinsky, F. M. (1996), 'A lesson in metabolic regulation inspired by the glucokinase glucose sensor paradigm', *Diabetes* **45**(2), 223–241.
- McAlister, D. (1879), 'Xiii. the law of the geometric mean', *Proceedings of the Royal Society of London* **29**(196-199), 367–376.
- Midroni, C., Leimbigler, P. J., Baruah, G., Kolla, M., Whitehead, A. J. and Fossat, Y. (2018), 'Predicting glycemia in type 1 diabetes patients: experiments with xgboost', *heart* **60**(90), 120.
- Mirshekarian, S., Shen, H., Bunescu, R. and Marling, C. (2019), Lstms and neural attention models for blood glucose prediction: Comparative experiments on real and synthetic data, *in* '2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)', IEEE, pp. 706–712.
- Misra, S. and Mathieu, C. (2020), 'Are newer insulin analogues better for people with type 1 diabetes?', *Diabetic Medicine* **37**(4), 522–531.
- Mohebbi, A., Johansen, A. R., Hansen, N., Christensen, P. E., Tarp, J. M., Jensen, M. L., Bengtsson, H. and Mørup, M. (2020), Short term blood glucose prediction based on continuous glucose monitoring data, *in* '2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)', IEEE, pp. 5140–5145.
- Moscoso-Vasquez, M., Colmegna, P. and Sanchez-Pena, R. S. (2016), Intra-patient dynamic variations in type 1 diabetes: A review, *in* '2016 IEEE Conference on Control Applications (CCA)', IEEE, pp. 416–421.
- Mosquera-Lopez, C. e. a. (2022), 'Incorporating glucose variability into glucose forecasting accuracy assessment using the new glucose variability impact index and the prediction

- consistency index: An lstm case example', *Journal of Diabetes Science and Technology* **16**(1), 7–18.
- Mughal, I. S., Patanè, L., Xibilia, M. G. and Caponetto, R. (2023), 'Variable structure-based controllers applied to the modified hovorka model for type 1 diabetes', *International Journal of Dynamics and Control* pp. 1–17.
- Noaro, G., Cappon, G., Vettoretti, M., Sparacino, G., Del Favero, S. and Facchinetti, A. (2020), 'Machine-learning based model to improve insulin bolus calculation in type 1 diabetes therapy', *IEEE Transactions on Biomedical Engineering* **68**(1), 247–255.
- Organization, W. H. et al. (1980), 'World health organization expert committee on diabetes mellitus: second report', *World Health Organ Tech Rep Ser* **646**, 1–180.
- Organization, W. H. et al. (1985), *Diabetes Mellitus: Report of a WHO Study Group [meeting held in Geneva from 11 to 16 February 1985]*, World Health Organization.
- Organization, W. H. et al. (1999a), Definition, diagnosis and classification of diabetes mellitus and its complications: report of a who consultation. part 1, diagnosis and classification of diabetes mellitus, Technical report, World health organization.
- Organization, W. H. et al. (1999b), *The world health report: 1999: making a difference*, World Health Organization.
- Pala, L., Dicembrini, I. and Mannucci, E. (2019), 'Continuous subcutaneous insulin infusion vs modern multiple injection regimens in type 1 diabetes: an updated meta-analysis of randomized clinical trials', *Acta diabetologica* **56**, 973–980.
- Palarea-Albaladejo, J. and Martín-Fernández, J.-A. (2008), 'A modified em algorithm for replacing rounded zeros in compositional data sets', *Computers & Geosciences* **34**(8), 902–917.

- Palarea-Albaladejo, J. and Martín-Fernández, J. A. (2015), ‘zcompositions r package for multivariate imputation of left-censored data under a compositional approach’, *Chemometrics and Intelligent Laboratory Systems* **143**, 85–96.
- Palladino, R., Tabak, A. G., Khunti, K., Valabhji, J., Majeed, A., Millett, C. and Vamos, E. P. (2020), ‘Association between pre-diabetes and microvascular and macrovascular disease in newly diagnosed type 2 diabetes’, *BMJ Open Diabetes Research and Care* **8**(1), e001061.
- Pawlowsky-Glahn, V. and Egozcue, J. J. (2002), ‘Blu estimators and compositional data’, *Mathematical Geology* **34**, 259–274.
- Pickup, J. C., Freeman, S. C. and Sutton, A. J. (2011), ‘Glycaemic control in type 1 diabetes during real time continuous glucose monitoring compared with self monitoring of blood glucose: meta-analysis of randomised controlled trials using individual patient data’, *Bmj* **343**.
- Plank, J., Blaha, J., Cordingley, J., Wilinska, M. E., Chassin, L. J., Morgan, C., Squire, S., Haluzik, M., Kremen, J., Svacina, S. et al. (2006), ‘Multicentric, randomized, controlled trial to evaluate blood glucose control by the model predictive control algorithm versus routine glucose management protocols in intensive care unit patients’, *Diabetes care* **29**(2), 271–276.
- Prabhakar, O. et al. (2021), ‘Inflammatory biomarkers as a part of diagnosis in diabetic peripheral neuropathy’, *Journal of Diabetes and Metabolic Disorders* **20**(1), 869.
- Prendin, F., Del Favero, S., Vettoretti, M., Sparacino, G. and Facchinetti, A. (2021), ‘Forecasting of glucose levels and hypoglycemic events: head-to-head comparison of linear and nonlinear data-driven algorithms based on continuous glucose monitoring data only’, *Sensors* **21**(5), 1647.
- Rashid, M., Samadi, S., Sevil, M., Hajizadeh, I., Kolodziej, P., Hobbs, N., Maloney, Z., Brandt, R., Feng, J., Park, M. et al. (2019), ‘Simulation software for assessment of nonlinear and

- adaptive multivariable control algorithms: glucose–insulin dynamics in type 1 diabetes’, *Computers & chemical engineering* **130**, 106565.
- Raskin, P., Klaff, L., Bergenstal, R., Halle, J.-P., Donley, D. and Mecca, T. (2000), ‘A 16-week comparison of the novel insulin analog insulin glargine (hoe 901) and nph human insulin used with insulin lispro in patients with type 1 diabetes.’, *Diabetes care* **23**(11), 1666–1671.
- Resalat, N., El Youssef, J., Tyler, N., Castle, J. and Jacobs, P. G. (2019), ‘A statistical virtual patient population for the glucoregulatory system in type 1 diabetes with integrated exercise model’, *PloS one* **14**(7), e0217301.
- Rja, L. and Rubin, D. B. (1987), ‘Statistical analysis with missing data’, *New York* .
- Rodríguez-Rodríguez, I., Chatzigiannakis, I., Rodríguez, J.-V., Maranghi, M., Gentili, M. and Zamora-Izquierdo, M.-Á. (2019), ‘Utility of big data in predicting short-term blood glucose levels in type 1 diabetes mellitus through machine learning techniques’, *Sensors* **19**(20), 4482.
- Rosenstock, J., Park, G., Zimmerman, J. and Group, U. I. G. H. . T. . D. I. (2000), ‘Basal insulin glargine (hoe 901) versus nph insulin in patients with type 1 diabetes on multiple daily insulin regimens. us insulin glargine (hoe 901) type 1 diabetes investigator group.’, *Diabetes care* **23**(8), 1137–1142.
- Ross, L. J. and Neville, K. A. (2019), ‘Continuous subcutaneous insulin infusion versus multiple daily injections for type 1 diabetes’, *Journal of paediatrics and child health* **55**(6), 718–722.
- Ruiz-de Adana, M. S., Dominguez-Lopez, M.-E., Gonzalez-Molero, I., Machado, A., Martin, V., Cardona, I., de-la Higuera, M., Tapia, M.-J., Soriguer, F., Anarte, M. T. et al. (2016), ‘Comparison between a multiple daily insulin injection regimen (basal once-daily glargine plus mealtime lispro) and continuous subcutaneous insulin infusion (lispro) using continuous glucose monitoring in metabolically optimized type 1 diabetes patients: a randomized open-labelled parallel study’, *Medicina clínica* **146**(6), 239–246.

- Sahoo, P. K., Dall’Agnol, R., Salomão, G. N., Junior, J. d. S. F., Silva, M. S., e Souza Filho, P. W. M., da Costa, M. L., Angélica, R. S., Medeiros Filho, C. A., da Costa, M. F. et al. (2020), ‘Regional-scale mapping for determining geochemical background values in soils of the itacaiúnas river basin, brazil: The use of compositional data analysis (coda)’, *Geoderma* **376**, 114504.
- Saiti, K., Macaš, M., Lhotská, L., Štechová, K. and Pithová, P. (2020), ‘Ensemble methods in combination with compartment models for blood glucose level prediction in type 1 diabetes mellitus’, *Computer Methods and Programs in Biomedicine* **196**, 105628.
- Sartore, G., Chillelli, N. C., Burlina, S., Stefano, P. D., Piarulli, F., Fedele, D., Mosca, A. and Lapolla, A. (2012), ‘The importance of hba1c and glucose variability in patients with type 1 and type 2 diabetes: outcome of continuous glucose monitoring (cgm)’, *Acta diabetologica* **49**, 153–160.
- Schiavon, M., Dalla Man, C. and Cobelli, C. (2017), ‘Modeling subcutaneous absorption of fast-acting insulin in type 1 diabetes’, *IEEE Transactions on Biomedical Engineering* **65**(9), 2079–2086.
- Schiavon, M., Visentin, R., Giegerich, C., Klabunde, T., Cobelli, C. and Dalla Man, C. (2019), ‘Modeling subcutaneous absorption of long-acting insulin glargine in type 1 diabetes’, *IEEE Transactions on Biomedical Engineering* **67**(2), 624–631.
- Schiavon, M., Visentin, R., Giegerich, C., Sieber, J., Dalla Man, C., Cobelli, C. and Klabunde, T. (2020), ‘In silico head-to-head comparison of insulin glargine 300 u/ml and insulin degludec 100 u/ml in type 1 diabetes’, *Diabetes technology & therapeutics* **22**(8), 553–561.
- Scott, E. S., McGrath, R. T., Januszewski, A. S., Calandro, D., Hardikar, A. A., O’Neal, D. N., Fulcher, G. and Jenkins, A. J. (2019), ‘Hba1c variability in adults with type 1 diabetes on continuous subcutaneous insulin infusion (csii) therapy compared to multiple daily injection (mdi) treatment’, *BMJ open* **9**(12), e033059.

- Seyed Ahmadi, S., Westman, K., Pivodic, A., Ólafsdóttir, A. F., Dahlqvist, S., Hirsch, I. B., Hellman, J., Ekelund, M., Heise, T., Polonsky, W. et al. (2020), 'The association between hba1c and time in hypoglycemia during cgm and self-monitoring of blood glucose in people with type 1 diabetes and multiple daily insulin injections: a randomized clinical trial (gold-4)', *Diabetes Care* **43**(9), 2017–2024.
- Sharma, A. K., Taneja, G., Kumar, A., Sahu, M., Sharma, G., Kumar, A., Sardana, S. and Deep, A. (2019), 'Insulin analogs: Glimpse on contemporary facts and future prospective', *Life sciences* **219**, 90–99.
- Sim, L. L. W., Ban, K. H. K., Tan, T. W., Sethi, S. K. and Loh, T. P. (2017), 'Development of a clinical decision support system for diabetes care: A pilot study', *PloS one* **12**(2), e0173021.
- Simon, B., Treat, V., Marco, C., Rosenberg, D., Joseph, J., Hipszer, B., Li, Y., Chervoneva, I., Padron-Massara, L. and Jabbour, S. (2008), 'A comparison of glycaemic variability in csii vs. mdi treated type 1 diabetic patients using cgms', *International journal of clinical practice* **62**(12), 1858–1863.
- Smaoui, M. R., Rabasa-Lhoret, R. and Haidar, A. (2020), 'Development platform for artificial pancreas algorithms', *Plos one* **15**(12), e0243139.
- Som, A. M., binti Mohd Yusof, N. F. M., Ali, S. A. and Fuzil, N. S. (2019), 'Meal disturbance effect on blood glucose control for type 1 diabetes using improved hovorka equations', *Key Engineering Materials* **797**, 158–167.
- Sorensen, J. T. (1985), A physiologic model of glucose metabolism in man and its use to design and assess improved insulin therapies for diabetes, PhD thesis, Massachusetts Institute of Technology.

- Stamati, A., Karagiannis, T., Tsapas, A. and Christoforidis, A. (2022), ‘Efficacy and safety of ultra-rapid insulin analogues in insulin pumps in patients with type 1 diabetes mellitus: A systematic review and meta-analysis’, *Diabetes Research and Clinical Practice* p. 110144.
- Steinmetz, J. D., Bourne, R. R., Briant, P. S., Flaxman, S. R., Taylor, H. R., Jonas, J. B., Abdoli, A. A., Abrha, W. A., Abualhasan, A., Abu-Gharbieh, E. G. et al. (2021), ‘Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to vision 2020: the right to sight: an analysis for the global burden of disease study’, *The Lancet Global Health* **9**(2), e144–e160.
- Sun, H., Saeedi, P., Karuranga, S., Pinkepank, M., Ogurtsova, K., Duncan, B. B., Stein, C., Basit, A., Chan, J. C., Mbanya, J. C. et al. (2022), ‘Idf diabetes atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045’, *Diabetes research and clinical practice* **183**, 109119.
- Tena, F., Garnica, O., Lanchares, J. and Hidalgo, J. I. (2021), ‘Ensemble models of cutting-edge deep neural networks for blood glucose prediction in patients with diabetes’, *Sensors* **21**(21), 7090.
- Thió-Henestrosa, S. and Comas, M. (2016), ‘Codapack v2 user’s guide’.
- Toschi, E., Slyne, C., Sifre, K., O’Donnell, R., Greenberg, J., Atakov-Castillo, A., Carl, S. and Munshi, M. (2020), ‘The relationship between cgm-derived metrics, a1c, and risk of hypoglycemia in older adults with type 1 diabetes’, *Diabetes Care* **43**(10), 2349–2354.
- Tyler, N. S. and Jacobs, P. G. (2020), ‘Artificial intelligence in decision support systems for type 1 diabetes’, *Sensors* **20**(11), 3214.
- Vallejo-Mora, M. d. R., Carreira-Soler, M., Linares-Parrado, F., Oliveira, G., Rojo-Martínez, G., Domínguez-López, M., Ruiz-de Adana-Navas, M. S. and González-Romero, M. S. (2017), ‘The calculating boluses on multiple daily injections (cbmdi) study: A randomized controlled

- trial on the effect on metabolic control of adding a bolus calculator to multiple daily injections in people with type 1 diabetes’, *Journal of diabetes* **9**(1), 24–33.
- Vettoretti, M., Cappon, G., Facchinetti, A. and Sparacino, G. (2020), ‘Advanced diabetes management using artificial intelligence and continuous glucose monitoring sensors’, *Sensors* **20**(14), 3870.
- Vigersky, R. A. and McMahon, C. (2019), ‘The relationship of hemoglobin a1c to time-in-range in patients with diabetes’, *Diabetes technology & therapeutics* **21**(2), 81–85.
- Visentin, R., Campos-Náñez, E., Schiavon, M., Lv, D., Vettoretti, M., Breton, M., Kovatchev, B. P., Dalla Man, C. and Cobelli, C. (2018), ‘The uva/padova type 1 diabetes simulator goes from single meal to single day’, *Journal of diabetes science and technology* **12**(2), 273–281.
- Visentin, R., Schiavon, M., Giegerich, C., Klabunde, T., Dalla Man, C. and Cobelli, C. (2019), ‘Incorporating long-acting insulin glargine into the uva/padova type 1 diabetes simulator for in silico testing of mdi therapies’, *IEEE Transactions on Biomedical Engineering* **66**(10), 2889–2896.
- Vives-Mestres, M., Daunis-I-Estadella, J. and Martín-Fernández, J.-A. (2014a), ‘Individual t 2 control chart for compositional data’, *Journal of Quality Technology* **46**(2), 127–139.
- Vives-Mestres, M., Daunis-i Estadella, J. and Martín-Fernández, J.-A. (2014b), ‘Out-of-control signals in three-part compositional t2 control chart’, *Quality and Reliability Engineering International* **30**(3), 337–346.
- Vives-Mestres, M., Daunis-i Estadella, J. and Martín-Fernández, J.-A. (2016), ‘Signal interpretation in hotelling’s t 2 control chart for compositional data’, *IIE Transactions* **48**(7), 661–672.
- Wadghiri, M., Idri, A., El Idrissi, T. and Hakkoum, H. (2022), ‘Ensemble blood glucose prediction in diabetes mellitus: A review’, *Computers in Biology and Medicine* p. 105674.

- Walsh, J., Roberts, R. and Bailey, T. (2011), 'Guidelines for optimal bolus calculator settings in adults', *Journal of diabetes science and technology* **5**(1), 129–135.
- Wilinska, M. E., Chassin, L. J., Acerini, C. L., Allen, J. M., Dunger, D. B. and Hovorka, R. (2010), 'Simulation environment to evaluate closed-loop insulin delivery systems in type 1 diabetes', *Journal of diabetes science and technology* **4**(1), 132–144.
- Wilinska, M. E. and Hovorka, R. (2008), 'Simulation models for in silico testing of closed-loop glucose controllers in type 1 diabetes', *Drug Discovery Today: Disease Models* **5**(4), 289–298.
- William, J., McCluskey, J. and Gleeson, N. (2022), 'Rt-cgm in conjunction with csii vs mdi in optimizing glycaemic control in t1dm: Systemic review and meta-analysis', *Endocrinology, Diabetes & Metabolism* **5**(2), e00324.
- Willis, A. D. and Martin, B. D. (2022), 'Estimating diversity in networked ecological communities', *Biostatistics* **23**(1), 207–222.
- Woldaregay, A. Z., Årsand, E., Walderhaug, S., Albers, D., Mamykina, L., Botsis, T. and Hartvigsen, G. (2019), 'Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes', *Artificial intelligence in medicine* **98**, 109–134.
- Wong, E. Y. and Kroon, L. (2021), 'Ultra-rapid-acting insulins: How fast is really needed?', *Clinical Diabetes* **39**(4), 415–423.
- Wright Jr, E. E., Morgan, K., Fu, D. K., Wilkins, N. and Guffey, W. J. (2020), 'Time in range: How to measure it, how to report it, and its practical application in clinical decision-making', *Clinical Diabetes: A Publication of the American Diabetes Association* **38**(5), 439.

- Wu, Y., Zhang, Q., Hu, Y., Sun-Woo, K., Zhang, X., Zhu, H., Li, S. et al. (2022), ‘Novel binary logistic regression model based on feature transformation of xgboost for type 2 diabetes mellitus prediction in healthcare systems’, *Future Generation Computer Systems* **129**, 1–12.
- Xie, J. and Wang, Q. (2018), Benchmark machine learning approaches with classical time series approaches on the blood glucose level prediction challenge., in ‘KHD@ IJCAI’, pp. 97–102.
- Xie, J. and Wang, Q. (2020), ‘Benchmarking machine learning algorithms on blood glucose prediction for type i diabetes in comparison with classical time-series models’, *IEEE Transactions on Biomedical Engineering* **67**(11), 3101–3124.
- Yang, G., Liu, S., Li, Y. and He, L. (2023), ‘Short-term prediction method of blood glucose based on temporal multi-head attention mechanism for diabetic patients’, *Biomedical Signal Processing and Control* **82**, 104552.
- Yoo, J. H. and Kim, J. H. (2020), ‘Time in range from continuous glucose monitoring: a novel metric for glycemic control’, *Diabetes & metabolism journal* **44**(6), 828–839.
- Yu, X., Turksoy, K., Rashid, M., Feng, J., Hobbs, N., Hajizadeh, I., Samadi, S., Sevil, M., Lazaro, C., Maloney, Z. et al. (2018), ‘Model-fusion-based online glucose concentration predictions in people with type 1 diabetes’, *Control engineering practice* **71**, 129–141.
- Zhou, H., He, K., Chen, J. and Zhang, X. (2022), ‘Linda: linear models for differential abundance analysis of microbiome compositional data’, *Genome biology* **23**(1), 1–23.
- Zhu, C., Idemudia, C. U. and Feng, W. (2019), ‘Improved logistic regression model for diabetes prediction by integrating pca and k-means techniques’, *Informatics in Medicine Unlocked* **17**, 100179.
- Zhu, T., Kuang, L., Daniels, J., Herrero, P., Li, K. and Georgiou, P. (2022), ‘Iomt-enabled real-time blood glucose prediction with deep learning and edge computing’, *IEEE Internet of Things Journal* .

REFERENCIAS BIBLIOGRÁFICAS

Zhu, T., Kuang, L., Li, K., Zeng, J., Herrero, P. and Georgiou, P. (2021), Blood glucose prediction in type 1 diabetes using deep learning on the edge, *in* '2021 IEEE International Symposium on Circuits and Systems (ISCAS)', IEEE, pp. 1–5.

Zhu, T., Li, K., Herrero, P., Chen, J. and Georgiou, P. (2018), A deep learning algorithm for personalized blood glucose prediction., *in* 'KHD@ IJCAI', pp. 64–78.

