



Eslam Montaser was born in Dakahlia, Egypt in 1985. He received a B.Sc in Statistics and Computer science in 2006 from the Department of Mathematics, Mansoura University, Egypt. He also received the M.Sc degree in Statistics and Computer science in 2007 from the Department of Mathematics, Mansoura University, Egypt. He also received the M.Sc degree in Mathematical engineering in 2013 from the Department of Mathematics, Universidad Carlos III de Madrid, Spain. He is pursuing a Ph.D. in Control Engineering, Robotics, and Industrial Informatics at the Universitat Politècnica de València, Spain.

He worked in the Tecnología UCM research group in 2014 at the Universidad Complutense de Madrid, Spain. He also worked in the PMMCRG research group under the supervision of Prof. Ali Cinar at the Illinois Institute of Technology, Chicago, IL, the United States of America for doctoral stay in 2017. Currently, he working in the artificial pancreas and technologies for diabetes (Tecnodabetes) research group under the supervision of Prof. Jorge Bondia at the Universitat Politècnica de València, Spain.

His research interests include stochastic modeling in type 1 diabetes, statistical pattern recognition, time series analysis, machine learning, biostatistics, big data analysis, biomedical applications, and predictive modeling of physiological systems.



January 2020

Stochastic Seasonal Models for Glucose Prediction in Type 1 Diabetes

Eslam Montaser



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

Ph.D. Dissertation

# Stochastic Seasonal Models for Glucose Prediction in Type 1 Diabetes

Author:  
Eslam Montaser  
Supervisors:  
Prof. Jorge Bondia  
Dr. José Luis Díez

January 2020



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

*PhD in Automation, Robotics and Computer Science for Industry*  
DOCTORADO EN AUTOMÁTICA, ROBÓTICA E INFORMÁTICA INDUSTRIAL

PHD DISSERTATION

---

# Stochastic Seasonal Models for Glucose Prediction in Type 1 Diabetes

---

*Author:* D. Eslam Montaser Roushdi Ali

*Supervisors:* Prof. Dr. Jorge Bondia Company

Dr. José Luis Díez Ruano

*Department of Systems Engineering and Control*  
DEPARTAMENTO DE INGENIERÍA DE SISTEMAS Y AUTOMÁTICA

**January 2020**



This work has been supported by the Spanish Ministry of Economy and Competitiveness (MINECO) under the FPI grant BES-2014-069253 and projects DPI2013-46982-C2-1-R and DPI2016-78831-C2-1-R. Moreover, with relation to this grant, a short stay was done at the end of 2017 at the Illinois Institute of Technology, Chicago, United States of America, under the supervision of Prof. Ali Cinar, for four months from 01/09/2017 to 29/12/2017.





*To my late father, my inspiration and the one who gave me every  
opportunity to realize my dreams.*

*To my family and my wife and all those who supported me through the  
University years.*



## Acknowledgements

Alhamdulillah, first and foremost, I would like to praise and thank Allah SWT, for His greatness, mercifulness and for giving me the strength to accomplish this thesis.

I wish to express my sincere thanks to all the people who directly or indirectly contributed and supported the development of the thesis. It is very difficult to mention them all, but I want everyone to know that I am deeply thankful.

I would like to express my sincere gratitude to my supervisors Prof. Dr. Jorge Bondia and Dr. José Luis Díez. I thank Jorge for the continuous support and encouragement of my research during the last four years, for his great patience, motivation, guidance, confidence, and immense knowledge in many fields. His right vision for solving the complex problems during my doctoral study. I could not have imagined having a better supervisor for my dissertation. To José Luis, thanks for your academic support, valuable contributions, guidance, encouragement, endless help, and creative and comprehensive advice. I have learned a lot from them, especially the spirit of teamwork in order to achieve the work's objectives.

Particular thanks go to all the members of our research group, especially, Juanfer, Iván, Vanessa, and Clara, for their friendship, support, and kind endless help. I also thank Dr. Paolo Rossetti, for his clinical suggestions concerning my research, encouragement, and generous advice. As well, I thank Frank and Yadira for their friendship, kind help and co-operation throughout my study period.

I would also like to thank all the members of Prof. Cinar's research group, especially, Prof. Ali Cinar, for his hospitality, cooperation, and support during my short stay at the Illinois Institute of Technology (IIT). As well, Dr. Mudassir Rashid, for his suggestions, and cooperation during my stay in Chicago.

Last but not least, I would like to thank my family for supporting and encouraging me throughout writing this thesis and my life in general. My deepest recognition goes to my beloved parents, especially to my late father who helped me in any imaginable way to achieve my goals and dreams. They have been an inexhaustible source of love and inspiration all my life. As well, to my mother, thanks for your support, encouragement, sacrifice, and love. My special thanks go to my wife, for her great patience, understanding, supporting, and encouragement, without which it would have been difficult to complete this thesis. As well, I thank my beloved son, who puts a smile on my face every day. This smile was a high support and powerful for me to complete this thesis, hoping that the effort of these years may offer him a more plentiful life in the coming years.

Finally, I would like to extend my deepest gratitude to my family, friends, and all the people who care about me and about whom I care.

Thank you all!

*Eslam Montaser*  
*Valencia, Spain*  
*January 2020*



## Abstract

Diabetes is a significant global health problem, one of the most serious non-communicable diseases after cardiovascular diseases, cancer and chronic respiratory diseases. Diabetes prevalence has been steadily increasing over the past decades, especially in low- and middle-income countries. It is estimated that 425 million people worldwide had diabetes in 2017, and by 2045 this number may rise to 629 million. About 10% of people with diabetes suffer from type 1 diabetes, characterized by autoimmune destruction of the  $\beta$ -cells in the pancreas, responsible for the secretion of the hormone insulin. Without insulin, plasma glucose rises to deleterious levels, provoking long-term vascular complications. Until a cure is found, the management of diabetes relies on technological developments for insulin replacement therapies. With the advent of continuous glucose monitors, technology has been evolving towards automated systems. Coined as “artificial pancreas”, closed-loop glucose control devices are nowadays a game-changer in diabetes management. Research in the last decades has been intense, yielding a first commercial system in late 2017 and many more are in the pipeline of the main medical devices industry. However, as a first-generation device, many issues still remain open and new technological advancements will lead to system improvements for better glycemic control outputs and reduced patient’s burden, improving significantly the quality of life of people with type 1 diabetes.

At the core of any artificial pancreas system is glucose prediction, the topic addressed in this thesis. The ability to predict glucose along a given prediction horizon, and estimation of future glucose trends, is the most important feature of any artificial pancreas system, in order to be able to take preventive actions to entirely avoid risk to the patient. Glucose prediction can appear as part of the control algorithm itself, such as in systems based on model predictive control (MPC) techniques, or as part of a monitoring system to avoid hypoglycemic episodes. However, predicting glucose is a very challenging problem due to the large inter- and intra-subject variability that patients suffer, whose sources are only partially understood. These limits models forecasting performance, imposing relatively short prediction horizons, despite the modeling technique used (physiological, data-driven or hybrid approaches). The starting hypothesis of this thesis is that the complexity of glucose dynamics requires the ability to characterize clusters of behaviors in the patient’s historical data naturally yielding to the concept of local modeling. Besides, the similarity of responses in a cluster can be further exploited to introduce the classical concept of seasonality into glucose prediction. As a result, seasonal local models are at the core of this thesis. Several clinical databases including mixed meals and exercise are used to demonstrate the feasibility and superiority of the performance of this approach.





## Resumen

La diabetes es un importante problema de salud mundial, siendo una de las enfermedades no transmisibles más graves después de las enfermedades cardiovasculares, el cáncer y las enfermedades respiratorias crónicas. La prevalencia de la diabetes ha aumentado constantemente en las últimas décadas, especialmente en países de ingresos bajos y medios. Se estima que 425 millones de personas en todo el mundo tenían diabetes en 2017, y para 2045 este número puede aumentar a 629 millones. Alrededor del 10% de las personas con diabetes padecen diabetes tipo 1, caracterizada por una destrucción autoinmune de las células  $\beta$  en el páncreas, responsables de la secreción de la hormona insulina. Sin insulina, la glucosa plasmática aumenta a niveles nocivos, provocando complicaciones vasculares a largo plazo. Hasta que se encuentre una cura, el manejo de la diabetes depende de los avances tecnológicos para terapias de reemplazo de insulina. Con la llegada de los monitores continuos de glucosa, la tecnología ha evolucionado hacia sistemas automatizados. Acuñados como “páncreas artificial”, los dispositivos de control de glucosa en lazo cerrado suponen hoy en día un cambio de juego en el manejo de la diabetes. La investigación en las últimas décadas ha sido intensa, dando lugar al primer sistema comercial a fines de 2017, y muchos más están siendo desarrollados por las principales industrias de dispositivos médicos. Sin embargo, como dispositivo de primera generación, muchos problemas aún permanecen abiertos y nuevos avances tecnológicos conducirán a mejoras del sistema para obtener mejores resultados de control glucémico y reducir la carga del paciente, mejorando significativamente la calidad de vida de las personas con diabetes tipo 1.

En el centro de cualquier sistema de páncreas artificial se encuentra la predicción de glucosa, tema abordado en esta tesis. La capacidad de predecir la glucosa a lo largo de un horizonte de predicción dado, y la estimación de las tendencias futuras de glucosa, es la característica más importante de cualquier sistema de páncreas artificial, para poder tomar medidas preventivas que eviten por completo el riesgo para el paciente. La predicción de glucosa puede aparecer como parte del algoritmo de control en sí, como en sistemas basados en técnicas de control predictivo basado en modelo (MPC), o como parte de un sistema de supervisión para evitar episodios de hipoglucemia. Sin embargo, predecir la glucosa es un problema muy desafiante debido a la gran variabilidad inter e intra-sujeto que sufren los pacientes, cuyas fuentes solo se entienden parcialmente. Esto limita las prestaciones predictivas de los modelos, imponiendo horizontes de predicción relativamente cortos, independientemente de la técnica de modelado utilizada (modelos fisiológicos, basados en datos o híbridos). La hipótesis de partida de esta tesis es que la complejidad de la dinámica de la glucosa requiere la capacidad de caracterizar grupos de comportamientos en los datos

históricos del paciente que llevan naturalmente al concepto de modelado local. Además, la similitud de las respuestas en un grupo puede aprovecharse aún más para introducir el concepto clásico de estacionalidad en la predicción de glucosa. Como resultado, los modelos locales estacionales están en el centro de esta tesis. Se utilizan varias bases de datos clínicas que incluyen comidas mixtas y ejercicio para demostrar la viabilidad y superioridad de las prestaciones de este enfoque.



## Resum

La diabetis és un important problema de salut mundial, sent una de les malalties no transmissibles més greus després de les malalties cardiovasculars, el càncer i les malalties respiratòries cròniques. La prevalença de la diabetis ha augmentat constantment en les últimes dècades, especialment en països d'ingressos baixos i mitjans. S'estima que 425 milions de persones a tot el món tenien diabetis en 2017, i per 2045 aquest nombre pot augmentar a 629 milions. Al voltant del 10% de les persones amb diabetis pateixen diabetis tipus 1, caracteritzada per una destrucció autoimmunitària de les cèl·lules  $\beta$  en el pàncrees, responsables de la secreció de l'hormona insulina. Sense insulina, la glucosa plasmàtica augmenta a nivells nocius, provocant complicacions vasculars a llarg termini. Fins que es trobi una cura, el maneig de la diabetis depèn dels avenços tecnològics per a teràpies de reemplaçament d'insulina. Amb l'arribada dels monitors continus de glucosa, la tecnologia ha evolucionat cap a sistemes automatitzats. Encunyats com "pàncrees artificial", els dispositius de control de glucosa en llaç tancat suposen avui dia un canvi de joc en el maneig de la diabetis. La investigació en les últimes dècades ha estat intensa, donant lloc al primer sistema comercial a finals de 2017, i molts més estan sent desenvolupats per les principals indústries de dispositius mèdics. No obstant això, com a dispositiu de primera generació, molts problemes encara romanen oberts i nous avenços tecnològics conduiran a millores del sistema per obtenir millors resultats de control glucèmic i reduir la càrrega del pacient, millorant significativament la qualitat de vida de les persones amb diabetis tipus 1.

Al centre de qualsevol sistema de pàncrees artificial es troba la predicció de glucosa, tema abordat en aquesta tesi. La capacitat de predir la glucosa al llarg d'un horitzó de predicció donat, i l'estimació de les tendències futures de glucosa, és la característica més important de qualsevol sistema de pàncrees artificial, per poder prendre mesures preventives que evitin completament el risc per el pacient. La predicció de glucosa pot aparèixer com a part de l'algoritme de control en si, com en sistemes basats en tècniques de control predictiu basat en model (MPC), o com a part d'un sistema de supervisió per evitar episodis d'hipoglucèmia. No obstant això, predir la glucosa és un problema molt desafiant degut a la gran variabilitat inter i intra-subjecte que pateixen els pacients, les fonts només s'entenen parcialment. Això limita les prestacions predictives dels models, imposant horitzons de predicció relativament curts, independentment de la tècnica de modelatge utilitzada (models fisiològics, basats en dades o híbrids). La hipòtesi de partida d'aquesta tesi és que la complexitat de la dinàmica de la glucosa requereix la capacitat de caracteritzar grups de comportaments en les dades històriques del pacient que porten naturalment al concepte de modelatge local. A més, la similitud de les respostes en un grup pot aprofitar-se encara més per introduir

el concepte clàssic d'estacionalitat en la predicció de glucosa. Com a resultat, els models locals estacionals estan al centre d'aquesta tesi. S'utilitzen diverses bases de dades clíniques que inclouen menjars mixtes i exercici per demostrar la viabilitat i superioritat de les prestacions d'aquest enfocament.



# TABLE OF CONTENTS

	Page
<b>I Thesis scope and objectives</b>	<b>1</b>
<b>1 Diabetes and the artificial pancreas</b>	<b>2</b>
1.1 Glucose regulation . . . . .	2
1.1.1 Diabetes overview . . . . .	4
1.1.2 Types of diabetes . . . . .	5
1.1.3 Diabetes complications . . . . .	7
1.1.4 Insulin therapy in type 1 diabetes . . . . .	8
1.1.5 Challenges to glycemic control . . . . .	10
1.2 Closed-loop glucose control in type 1 diabetes . . . . .	13
1.2.1 The artificial pancreas . . . . .	15
1.2.2 State of development . . . . .	18
<b>2 Thesis objectives</b>	<b>20</b>
2.1 Glucose prediction in the artificial pancreas . . . . .	20
2.2 General objective . . . . .	21
2.3 Specific objectives . . . . .	21
<b>II State of the art</b>	<b>24</b>
<b>3 Glucose prediction in type 1 diabetes</b>	<b>25</b>
3.1 Challenges of blood glucose prediction . . . . .	25
3.1.1 Inter- and intra-individual variability . . . . .	26
3.1.2 Postprandial response . . . . .	28
3.1.3 Exercise response . . . . .	29
3.2 Categories of prediction models . . . . .	31
3.3 Physiological models . . . . .	34
3.3.1 Minimal model based approaches . . . . .	34
3.3.2 Medium complexity models . . . . .	37
3.3.3 Maximal models . . . . .	39
3.4 Data-driven models . . . . .	41

3.4.1	AR and ARMA models . . . . .	41
3.4.2	ARX and ARMAX models . . . . .	42
3.4.3	Other empirical models . . . . .	44
3.5	Hybrid models . . . . .	46
3.6	Prediction performance analysis . . . . .	47
<b>4</b>	<b>Stochastic time series modeling</b>	<b>50</b>
4.1	Preliminary concepts . . . . .	50
4.2	Linear time series models . . . . .	57
4.2.1	Autoregressive (AR) process . . . . .	58
4.2.2	Moving average (MA) process . . . . .	60
4.2.3	Mixed autoregressive moving average (ARMA) processes	62
4.2.4	Autoregressive integrated moving average (ARIMA) processes . . . . .	63
4.2.5	Seasonal ARIMA processes . . . . .	63
4.2.6	Processes with exogenous variables . . . . .	64
4.3	The strategy of building models . . . . .	66
4.3.1	Stationarity test . . . . .	67
4.3.2	Model identification . . . . .	69
4.3.3	Model estimation . . . . .	72
4.3.4	Diagnostic checking . . . . .	74
4.3.5	Forecasting . . . . .	76
<b>5</b>	<b>Clustering time series</b>	<b>78</b>
5.1	Clustering time series data . . . . .	78
5.2	Similarity and distance measures . . . . .	81
5.2.1	Similarity in time . . . . .	81
5.2.2	Similarity in shape . . . . .	82
5.2.3	Structural similarity . . . . .	84
5.3	Approaches to time series clustering . . . . .	84
5.3.1	Raw-data-based approaches . . . . .	84
5.3.2	Feature-based approaches . . . . .	86
5.3.3	Model-based approaches . . . . .	86
5.4	Time series clustering algorithms . . . . .	86
5.4.1	Hierarchical clustering . . . . .	88
5.4.2	Partitional clustering . . . . .	90
5.4.3	Cluster validity indices . . . . .	95
5.5	Clustering and local modeling . . . . .	100
5.5.1	Why local modeling? . . . . .	100
5.5.2	Local modeling techniques . . . . .	102
5.6	Clustering applications in type 1 diabetes . . . . .	105



<b>III</b>	<b>Contributions</b>	<b>108</b>
<b>6</b>	<b>Seasonality and glucose prediction</b>	<b>109</b>
6.1	Rationale . . . . .	109
6.2	Research Design and Methods . . . . .	110
6.2.1	Clinical dataset . . . . .	110
6.2.2	Data pre-processing . . . . .	113
6.2.3	SARIMA and SARIMAX models . . . . .	114
6.2.4	Identification procedure . . . . .	116
6.3	Results . . . . .	117
6.3.1	Without exogenous variables . . . . .	117
6.3.2	With exogenous variables . . . . .	120
6.4	Conclusions and margins for further improvement . . . . .	121
<b>7</b>	<b>Seasonal local models</b>	<b>125</b>
7.1	Rationale . . . . .	125
7.2	Research Design and Methods . . . . .	126
7.2.1	Details of procedures . . . . .	126
7.2.2	Meal protocol . . . . .	126
7.2.3	Exercise protocol . . . . .	127
7.2.4	Closed-loop system . . . . .	128
7.2.5	Statistical procedures . . . . .	129
7.2.6	FCM clustering algorithm . . . . .	130
7.2.7	Seasonal models with exogenous signals . . . . .	130
7.2.8	Models integration . . . . .	131
7.2.9	Validation procedure . . . . .	131
7.3	Results . . . . .	132
7.3.1	Data clustering . . . . .	132
7.3.2	Model identification . . . . .	137
7.4	Conclusions and margins for further improvement . . . . .	141
<b>8</b>	<b>Seasonality in free-living conditions</b>	<b>144</b>
8.1	Rationale . . . . .	144
8.2	Research Design and Methods . . . . .	145
8.2.1	Data overview . . . . .	145
8.2.2	Data preprocessing . . . . .	146
8.2.3	Time series clustering algorithm . . . . .	146
8.2.4	SARIMA local models integration for online glucose prediction . . . . .	148
8.2.5	Supervision system and abnormal states detection . . . . .	148
8.2.6	Validation procedure . . . . .	151
8.3	Results . . . . .	151
8.3.1	Time series building . . . . .	153
8.3.2	Data clustering . . . . .	153

8.3.3	Model identification . . . . .	153
8.3.4	Online forecasting validation . . . . .	155
8.3.5	Supervision system validation and abnormal states de- tection . . . . .	157
8.4	Conclusions and margins for further improvement . . . . .	162
<b>9</b>	<b>Conclusions and Future work</b>	<b>166</b>
9.1	Conclusions . . . . .	166
9.2	Future work . . . . .	169
9.3	Publications . . . . .	170
9.3.1	Journal papers . . . . .	170
9.3.2	Conference abstracts and posters . . . . .	171
<b>A</b>	<b>Clarke error grid analysis</b>	<b>173</b>
A.1	Clarke error grid analysis . . . . .	173
A.2	Continuous glucose error grid analysis . . . . .	174



# LIST OF FIGURES

1.1	The glucose-insulin regulatory system. . . . .	3
1.2	IDF regions and global projections for the number of people with diabetes (20-79 years), 2017-2045 [1]. . . . .	5
1.3	The history of insulin therapy [2]. . . . .	8
1.4	Physiological basis of the basal/bolus insulin regimen. Modified from [3]. . . . .	9
1.5	Some external factors that affect the blood glucose level. . . . .	11
1.6	The effectiveness of different exercises in diabetes. Modified from [4]. . . . .	12
1.7	Schematic diagram of the bedside-type artificial endocrine pancreas [5]. . . . .	13
1.8	Timeline of diabetes technology [6]. . . . .	14
1.9	Closed-loop artificial pancreas system. . . . .	15
3.1	Illustration of inter-individual and intra-individual variability to a 60 g mixed meal response in open loop therapy. Data extracted from the study <i>Closedloop4meals</i> (NCT02100488) [7].	27
3.2	Average meal rate of glucose appearance of a low glycemic load (LG) and a high glycemic load (HG) meal matched in carbohydrates. Adapted from [8]. . . . .	29
3.3	Blood glucose and typical hormonal responses to different types and intensities of exercise in individuals with and without T1D. “Cats” stands for “catecholamines”. Modified from [9]. . . . .	30
3.4	Categories of prediction models. . . . .	31
3.5	Block diagram of a glucose-insulin physiological model. . . . .	32
3.6	Block diagram of an artificial neural network as an example of data-driven model. . . . .	33
3.7	Block diagram of a hybrid model. . . . .	34
3.8	BG modeling error (RMSE) versus PH for the different literature methods. . . . .	48

4.1	CGM time series data for specific postprandial periods for a person with T1D. . . . .	51
4.2	Time series for the population of the U.S.A. at ten-year intervals, 1790 – 1980. . . . .	52
4.3	Time series for the Wolfer sunspot numbers, 1770 – 1869. . . . .	52
4.4	Stationary and non-stationary time series. . . . .	53
4.5	Steps for building a good model through Box-Jenkins methodology. . . . .	67
4.6	Leave-one-out cross-validation (LOOCV) strategy. . . . .	70
4.7	The sample ACF and sample PACF of an stationary time series. The blue horizontal lines represent the 5% critical values at $\pm 1.96\sqrt{T}$ , which are used to test whether or not the ACFs and PACFs are significantly different from 0. . . . .	71
4.8	The sample ACF and sample PACF of the residuals. The blue horizontal lines represent the 5% critical values at $\pm 1.96\sqrt{T}$ and are used to test whether or not the ACFs and PACFs are significantly different from 0. . . . .	74
5.1	Illustrative example of the measure of similarity between two time series: (left) Euclidean distance (lock-step) measure; (right) Dynamic Time Warping measure (elastic). . . . .	83
5.2	Time series clustering approaches. . . . .	85
5.3	Time series clustering approaches with similarity measures. . . . .	85
5.4	The major clustering algorithms. . . . .	87
5.5	Agglomerative (AHC) and divisive (DHC) hierarchical clustering. Adapted with modifications from [10]. . . . .	88
5.6	Diagram of global model versus local models with clustering technique. . . . .	104
6.1	Diagram of seasonal CGM time series models. . . . .	110
6.2	CGM time series resulting from the concatenation of twenty 8-h PPs for a same 60g CHO meal. The notation $P_{ij}$ is used to name the different periods, where $i$ is the number of the patient, $i \in \{1, \dots, 10\}$ , and $j$ is the number of the study per patient, $j \in \{1, 2\}$ . Sampling period is 15 minutes, yielding 33 samples per PP. . . . .	114
6.3	Similarity among PPs in the CGM time series as measured by the ED between paired periods. Data is shown according to the color scale in the right. White boxes in the diagonal indicate periods corresponding to a same patient. . . . .	115
6.4	Mean and standard deviation of forecasting measures (MAE (mg/dL), RMSE (md/dL), MAPE (%)) for the 20-fold cross-validation and a 5-h PH. * $P < 0.05$ . . . . .	117

6.5	Mean and standard deviation of MAPE(%) for increasing values of the PH. $*P < 0.05$ . . . . .	118
6.6	Residuals analysis for model SARIMA(4, 0, 4)(1, 0, 1) <sub>33</sub> in best-performing Run 4. . . . .	119
6.7	Forecasting of models ARIMA(4, 0, 4), SARIMA(4, 0, 4)(1, 0, 1) <sub>33</sub> , ARIMAX(4, 0, 4, 2) and SARIMAX(4, 0, 4, 2)(1, 0, 1) <sub>33</sub> for Run 4 considering a 5-h PH. . . . .	120
7.1	The exercises types in the study. Modified from [11]. . . . .	127
7.2	MAAP general structure. . . . .	128
7.3	MAAP control structure for each experiment. Notation: HEA (Hypoglycemia Early Alarm), MD (Meal Detection), H-CR (Hypoglycemia Prediction and Carbohydrate Recommendation), Basal or Bolus (GPC Controlled Variable). . . . .	129
7.4	Diagram summarizing the methodology application: data sets generation (A), local models identification (B), and models integration (C). . . . .	133
7.5	DATA CLUSTERING SUMMARY. Notation: Results presented as mean(SD) for each feature. $*_{i,j}$ indicates SSD ( $p < 0.05$ ) between clusters $i$ and $j$ (if $i, j$ not specified, then differences found between all clusters). $R_1$ is mild hyperglycemia range, $R_2$ is moderate hyperglycemia range, $R_N$ is euglycemia range, $L_1$ is mild hypoglycemia range, and $L_2$ is moderate hypoglycemia range. . . . .	135
7.6	The clustering classifications by FCM algorithm, 4 clusters is the best number of clusters indicated by a CWBI validation index. . . . .	136
7.7	The local glucose predictions $\hat{G}(t)$ for each cluster for a 3-h PH. . . . .	138
7.8	Local predictions at $t = 0$ , memberships at each time instant, and global predictions at $t = 0$ , $t = 60$ , and $t = 90$ by using $\mathbf{M}_{15}$ for a 90-min-ahead prediction. Membership values at the above prediction times are highlighted by the symbol $\circ$ . . . . .	140
7.9	A comparison of the proposed model ( $\mathbf{M}_{15}$ ) with those reported in the literature through the blood glucose modeling error (RMSE) with different PHs, where our prediction errors are marked by the black diamond. . . . .	142
8.1	An illustrative example of an analysis of the fuzzy memberships series for choosing the right local model for use in prediction. . . . .	149

8.2	An illustrative example of using the trust index in order to measure the confidence of using a specific local model for prediction and the normality index in order to determine whether the period trying to estimate is normal or not. . . . .	150
8.3	Diagram summarizing the online methodology application: time series building (A), SARIMA local modeling and identification (B), and online proof of concept validation (C). . .	152
8.4	9 clusters classification performed by PDSFCM algorithm. . .	154
8.5	Cluster prototypes. . . . .	154
8.6	The local glucose predictions $\hat{G}(t)$ for each cluster for a 4-h PH, when using only one event/period as validation example set of the 9 clusters. . . . .	156
8.7	An example of <b>GSM</b> to demonstrate the relation between predictions, memberships, trust, and normality indexes. Local predictions at $t = 0$ , memberships at each time instant, and global predictions at $t = 0$ , $t = 60$ , and $t = 120$ by using $\mathbf{M}_{15}$ for a 120-min-ahead prediction. Membership values at the above prediction times are highlighted by the symbol $\circ$ . . .	158
8.8	An example of <b>GSM</b> with a big prediction error for showing the relationship between predictions, memberships, trust, and normality indexes. Local predictions at $t = 0$ , memberships at each time instant, and global predictions at $t = 0$ , $t = 60$ , and $t = 120$ by using $\mathbf{M}_{15}$ for a 120-min-ahead prediction. Membership values at the above prediction times are highlighted by the symbol $\circ$ . . . . .	159
8.9	An example of <b>GSM</b> with a small prediction error for showing the relationship between predictions, memberships, trust, and normality indexes. Local predictions at $t = 0$ , memberships at each time instant, and global predictions at $t = 0$ , $t = 60$ , and $t = 120$ by using $\mathbf{M}_{15}$ for a 120-min-ahead prediction. Membership values at the above prediction times are highlighted by the symbol $\circ$ . . . . .	160
8.10	An example of how the system behaves in the case of abnormal behaviors. Local predictions at $t = 0$ , memberships at each time instant, and global predictions at $t = 0$ , $t = 60$ , $t = 120$ , $t = 180$ , and $t = 240$ by using $\mathbf{M}_{15}$ for a 60-min-ahead prediction. Membership values at the above prediction times are highlighted by the symbol $\circ$ . . . . .	161
8.11	An example of detecting hypoglycemia through the normality index. . . . .	162
8.12	An example of detecting hypoglycemia and hyperglycemia through the normality index. . . . .	163

8.13 The similarity measure between the normal and abnormal group with the prototypes of 9 clusters by using the Euclidean distance. . . . . 164

8.14 Boxplot of the normal and abnormal group. \* indicates SSD ( $p < 0.05$ ). . . . . 164

A.1 The Clarke Error Grid Analysis for determining BG meter accuracy with five different zones. Modified from [12]. . . . . 174

A.2 The continuous glucose error grid analysis (CG-EGA): (a) rate error grid analysis (R-EGA) and (b) point error grid analysis (P-EGA) to quantify the clinical accuracy of CGM systems with five different zones as the EGA. Adapted from [13]. . . . . 175



# LIST OF TABLES

1.1	Summary of the action of principal hormones that are involved in the glucose regulation system in the body. . . . .	4
1.2	Pharmacokinetics of different insulin types for people with diabetes. . . . .	10
3.1	BG modeling error of different literature methods. * indicates error expressed as MAPE (%). . . . .	49
4.1	Characteristics of theoretical ACF and PACF for specific non-seasonal ARMA models. . . . .	70
4.2	Characteristics of theoretical ACF and PACF for pure SARMA models. *The values at nonseasonal lags are zero. . . . .	71
5.1	Advantages and disadvantages of the hierarchical clustering [14]. . . . .	90
6.1	Demographic information for subjects. . . . .	111
6.2	Performance of SARIMA and ARIMA models by the mean and standard deviation, in parentheses, of different forecasting measures. * $P < 0.05$ in all cases. . . . .	117
6.3	Model parameters for model SARIMA(4, 0, 4)(1, 0, 1) <sub>33</sub> in best-performing Run 4, following notation in (6.1). $\sigma^2$ is the estimate of the error variance from the MLE. . . . .	119
6.4	Ljung-Box Q (LBQ) test for the training residuals of Run 4 Model. . . . .	119
6.5	Model parameters for model SARIMAX(4, 0, 4)(1, 0, 1) <sub>33</sub> in best-performing Run 4, following notation in (6.2). $\sigma^2$ is the estimate of the error variance from the MLE. . . . .	121
6.6	Prediction accuracy measured by MAPE (%) and RMSE (mg/dL), in parentheses, of seasonal versus nonseasonal counterparts for different PHs. * 100 (SARIMA-ARIMA) / ARIMA; ** 100 (SARIMAX-ARIMAX) / ARIMAX. . . . .	122
7.1	Demographic information for subjects. . . . .	127

7.2	Prediction accuracy of seasonal local models. <b>Notation:</b> prediction accuracy results presented as MAPE % (RMSE mg/dL).	139
7.3	Prediction accuracy of <b>GSM</b> . <b>Notation:</b> prediction accuracy results presented as MAPE % (RMSE mg/dL). * indicates SSD ( $p < 0.05$ ) with <b>M</b> and † with <b>ED</b> .	139
7.4	Prediction accuracy of suboptimal <b>GSM</b> . <b>Notation:</b> prediction accuracy results presented as MAPE % (RMSE mg/dL). * indicates SSD ( $p < 0.05$ ) of <b>ED</b> vs <b>ED</b> <sub>15</sub> or <b>M</b> vs <b>M</b> <sub>15</sub> , and * for <b>ED</b> vs <b>M</b> or <b>ED</b> <sub>15</sub> vs <b>M</b> <sub>15</sub> .	141
7.5	Summary of prediction accuracy results (all datasets mean). <b>Notation:</b> prediction accuracy results presented as MAPE % (RMSE mg/dL). * indicates SSD ( $p < 0.05$ ).	141
8.1	Prediction accuracy of seasonal local models, when using only one event/period as validation example set. <b>Notation:</b> prediction accuracy results presented as MAPE % (RMSE mg/dL).	155
8.2	Prediction accuracy of <b>GSM</b> for only one period/event validation example data of the 9 clusters. <b>Notation:</b> prediction accuracy results presented as MAPE % (RMSE mg/dL).	157
8.3	Prediction accuracy of <b>GSM</b> for all periods/events validation data of the 9 clusters. <b>Notation:</b> prediction accuracy results presented as MAPE % (RMSE mg/dL).	157

# LIST OF ACRONYMS

**AADE** American Association of Diabetes Educators

**ACF** Autocorrelation Function

**ACSM** American College of Sports Medicine

**ACVF** Autocovariance Function

**ADA** American Diabetes Association

**ADF** Augmented Dickey-Fuller

**AHC** Agglomerative Hierarchical Clustering

**AIC** Akaike Information Criterion

**ALM** Average-Linkage Metric

**ANN** Artificial Neural Network

**AP** Artificial Pancreas

**AR** Autoregressive

**ARD** Absolute Relative Difference

**ARIMA** Autoregressive Integrated Moving Average

**ARIMAX** Autoregressive Integrated Moving Average with eXogenous In-  
puts

**ARMA** Autoregressive Moving Average

**ARMAX** Auto-Regressive Moving Average with eXogenous Inputs

**ARX** Autoregressive with eXogenous Inputs

**ATP** Adenosine Triphosphate

**BE** Bayesian Estimation

- BG** Blood Glucose
- BIC** Bayesian Information Criterion
- BT** Body Temperature
- CC** Correlation Coefficient
- CCF** Cross-Correlation Function
- CD** Distance Between Two Clusters
- CDSS** Clinical Decision Support System
- CG-EGA** Continuous Glucose Error Grid Analysis
- CGM** Continuous Glucose Monitor
- CHO** Carbohydrate
- CL** Closed-Loop
- CLM** Complete-Linkage Metric
- CoDA** Compositional Data Analysis
- CRC** Clinical Research Center
- CSII** Continuous Subcutaneous Infusion of Insulin
- CVD** Cardiovascular Disease
- CVI** Cluster Validity Index
- CWBI** Compose Within and Between Scattering Index
- DBI** Davies-Bouldin Index
- DCCT** Diabetes Control and Complications Trial
- DFT** Discrete Fourier Transform
- DHC** Divisive Hierarchical Clustering
- DI** Dunn Index
- DKA** Diabetic Ketoacidosis
- DM** Diabetes Mellitus
- DSMP** Diabetes Self-Management Profile
- DTW** Dynamic Time Warping

- DWT** Discrete Wavelet Transform
- EE** Energy Expenditure
- EEG** Electroencephalography
- ED** Euclidean Distance
- EDM** Empirical Dynamic Models
- EGA** Clarke Error Grid Analysis
- ELM** Extreme Learning Machin
- ESOD** Energy of the Second-Order Differences
- FAO** Food and Agriculture Organization
- FCM** Fuzzy C-Means
- FDA** Food and Drug Administration
- FFA** Free Fatty Acids
- FGM** Flash Glucose Monitoring
- FIR** Finite-Impulse Response
- FL** Fuzzy Logic
- fMRI** Functional Magnetic Resonance Imaging
- FPC** Fuzzy Partitional Clustering
- GC** Granger Causality
- GCIIS** Glucose Controlled Insulin Infusion System
- GD** Gestational diabetes
- GI** Glycemic Index
- GL** glyceimic Load
- GMM** Generalized Method of Moments
- GPC** Generalized Predictive Control
- GSM** Global Seasonal Model
- GSR** Galvanic Skin Response
- HbA1c** Glycated Hemoglobin

- HC** Hierarchical Clustering
- H-CR** Hypoglycemia Prediction and Carbohydrate Recommendation
- HEA** Hypoglycemia Early Alarm
- HF** Heat Flux
- HG** High Glycemic Load
- HHS** Hyperosmolar Hyperglycemic Stat
- HIIT** High Intensity Interval Training
- HMM** Hidden Markov Models
- HPC** Hard Partitional Clustering
- IDDM** Insulin Dependent Diabetes Mellitus
- IF** Insulin Feedback
- IID** Independent and Identically Distributed
- IOB** Insulin-On-Board
- ISF** Interstitial Fluid
- IVGTT** Intravenous Glucose Tolerance Test
- IVP** Identifiable Virtual Patient
- JB** Jarque-Bera
- KRLS-ALD** Approximate Linear Dependency Kernel Recursive Least-Squares
- LA** Longitudinal Acceleration
- LBQ** Ljung-Box Q
- LG** Low Glycemic Load
- LOOCV** Leave-One-Out Cross-Validation
- LQ** Linear Quadratic
- LS** Least Squares
- LSD** Fisher's Least Significant Difference
- LSE** Linear Least Squares Estimation
- LSTM** Long-Short-Term Memory

- LV** Latent Variable
- LVX** Latent Variable with eXogenous Inputs
- MA** Moving Average
- MAAP** Multi-Module Multivariable Adaptive Artificial Pancreas
- MAE** Mean Absolute Error
- MARD** Mean Absolute Relative Deviation
- MAPE** Mean Absolute Percentage Error
- MCT** Moderate Continuous Training
- MD** Meal Detection
- MLE** Maximum Likelihood Estimation
- MPC** Model Predictive Control
- MSE** Mean Squared Error
- NCD** Normalized Compression Distance
- NIDDM** Non-Insulin Dependent Diabetes Mellitus
- NLMS** Normalized Least Mean Squares
- NLSE** Non-Linear Least Squares Estimation
- NPH** Neutral Protamine Hagedorn
- NPS** Nearest Prototype Strategy
- ODEs** Ordinary differential Equations
- OCS** Optimal Completion Strategy
- PA** Physical Activity
- PAA** Piecewise Aggregate Approximation
- PACF** Partial Autocorrelation Function
- PC** Partitional Clustering
- PCI** Partition Coefficient Index
- PDS** Partial Distance Strategy
- P-EGA** Point Error Grid Analysis

- PEI** Partition Entropy Index
- PG** Plasma Glucose
- PH** Prediction Horizon
- PID** Proportional Integral Derivative
- PLS** Partial Least-Squares
- PP** Postprandial Period
- RBF** Radial Basis Function
- R-EGA** Rate Error Grid Analysis
- RLS** Recursive Least Squares
- RMSE** Root Mean Square Error
- RMSSTD** Root-Mean-Square Standard Deviation
- RNN** Recurrent Neural Network
- RS** R-Squared
- RT** Resistance Training
- RPE** Rating of Perceived Exertion
- SARIMA** Seasonal Autoregressive Integrated Moving Average
- SARIMAX** Seasonal Autoregressive Integrated Moving Average with exogenous Inputs
- SARMA** Seasonal Autoregressive Moving Average
- SAX** Symbolic Aggregate Approximation
- SBC** Schwarz Bayesian Criterion
- SDE-GB** stochastic differential equation grey-box Model
- SLM** Single-Linkage Metric
- SMBG** Self-Monitoring of Blood Glucose
- SMRC** Sliding Mode Reference Conditioning
- SOFM** Self-Organizing Feature Mapping
- SOM** Self-Organizing Map



- SPR** Semi-Partial R-Squared
- SSD** Statistically Significant Differences
- SSGPE** Sum of Squares of Glucose Prediction Error
- SVD** Singular Value Decomposition
- SVR** Support Vector Regression
- T1D** Type 1 diabetes
- T2D** Type 2 diabetes
- TG** Time Gain
- TS** Takagi-Sugeno Fuzzy Model
- UI** Theil Inequality coefficient
- UKPDS** United Kingdom Prospective Diabetes Study
- WDS** Whole Data Strategy
- WHO** World Health Organization
- XBI** Xie-Beni Index
- XGB** Extreme Gradient-Boosted Tree
- ZOH** Zero-Order Hold



## Part I

# Thesis scope and objectives

# Chapter 1

## Diabetes and the artificial pancreas

The chapter starts in Section 1.1 with an explanation of the glucose-insulin regulatory system in healthy individuals. Then an overview of diabetes mellitus, types, complications, and a description of the main glucose disturbances that affect glucose regulation is presented. Section 1.2 introduces the artificial pancreas system, the latest technology for glucose control in type 1 diabetes consisting in a closed-loop glucose control system, with a first commercial system in 2017 and which is changing the paradigm of glucose management. Finally, the stages of artificial pancreas development to make it more reliable are described.

### 1.1 Glucose regulation

*Glucose* is an energy-rich monosaccharide sugar that is broken down within the body cells to produce Adenosine Triphosphate (ATP), a molecule that stores and transports chemical energy within cells. Glucose is considered to be a primary source of energy for every cell in the body. In spite of that, many organs can use fat and the building blocks of protein (*i.e.*, amino acids) as the source of energy when the blood glucose (BG) level is too low, except the brain and nervous system that consume exclusively glucose to be able to work well [15].

In a healthy person, BG level is typically maintained in a very narrow range (fasting BG of 70-100 mg/dL and post-meal glucose of 120-140 mg/dL, returning to baseline after 2 hours) [15]. Insulin and glucagon are the hormones produced in the pancreas that have opposite actions as shown in Fig. 1.1, composing the so-called glucose-insulin regulatory system (negative feedback loop). These hormones are secreted by the endocrine cells in the islets of Langerhans in the pancreas. At least four cell types have been identified in the islets of Langerhans:  $\alpha$ -cells,  $\beta$ -cells,  $\delta$ -cells, and pancreatic

polypeptide cells (Table 1.1). The majority of the islet cells are  $\beta$ -cells, making up 65-80% of islet mass, and are concentrated in the anterior head, body, and tail of the pancreas [16]. **Glucagon** and **insulin** are secreted by pancreatic  $\alpha$ - and  $\beta$ -cells, respectively.

After a meal intake, BG rises to a high level, since carbohydrates (CHOs) in the food are converted into glucose and absorbed into the bloodstream. This increase in BG induces production of insulin by the pancreas via  $\beta$ -cells. Insulin metabolic effects include [17, 15]: (1) increase of rate of glucose uptake from the bloodstream by the cells, by promotion of glucose transporters into the cells membrane; (2) the promotion of glycolysis (*i.e.*, a process that breaks down glucose in order to produce cellular energy), increasing the cellular utilization of glucose as an energy source; (3) the promotion of glycogenesis by converting glucose into glycogen (a glucose aggregate molecule) for storage in the liver and skeletal muscle cells; (4) the inhibition of lipolysis to release energy. These effects of insulin lead to a BG level in a normal range.

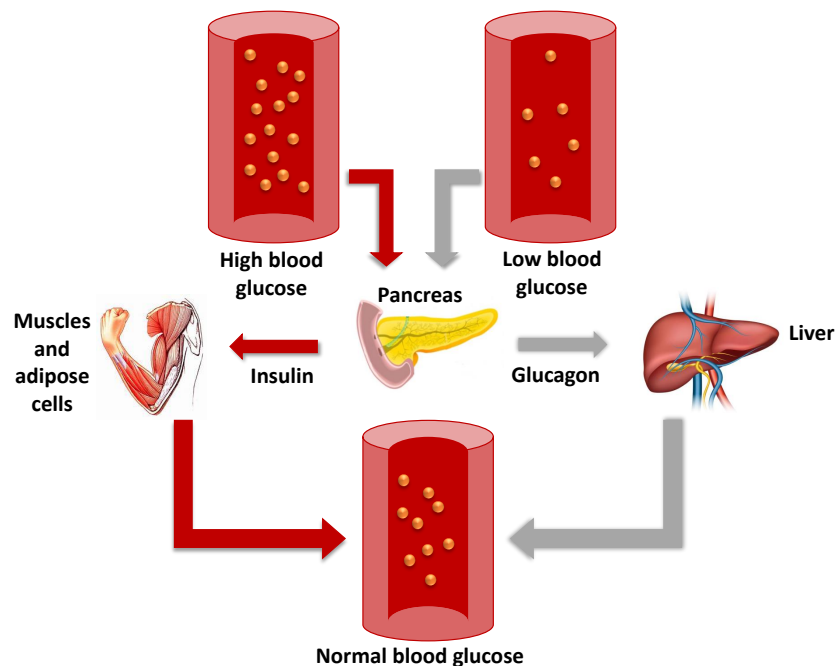


Figure 1.1: The glucose-insulin regulatory system.

On the other side, when the BG level goes low due, for instance, to physical activity or between meals, glucagon is secreted by the  $\alpha$ -cells, with an opposite effect than insulin. Glucagon has an effect on many different cells of the body, such as adipose tissue and kidney, but the main effect is on the *liver cells*. Glucagon effects include [17, 15]: (1) conversion of stored glycogen in the liver cells into glucose and its release into the bloodstream to increase BG and prevent glucose levels from falling too low; (2) stimulation

Hormone	Site of production	Summary of action
Insulin	Pancreatic islets ( $\beta$ -cells)	Stimulates cells to take up glucose from the blood and so lowers BG
Glucagon	Pancreatic islets ( $\alpha$ -cells)	Stimulates hepatocytes to release glucose into the blood and so raises BG
Somatostatin	Pancreatic islets ( $\delta$ -cells)	Reduces gut motility and further absorption of nutrients
Adrenaline (Epinephrine)	Adrenal medulla	Mobilises glycogen and suppresses the release of insulin
Growth hormone	Pituitary	Decreases liver glucose uptake and stimulates gluconeogenesis ( <i>i.e.</i> , new glucose formation)
Cortisol	Adrenal glands (cortex)	Stimulates glycogenolysis ( <i>i.e.</i> , glycogen breakdown) and mobilizes amino acids and ketone bodies

Table 1.1: Summary of the action of principal hormones that are involved in the glucose regulation system in the body.

of gluconeogenesis, *i.e.*, the production of new glucose from lactic acid and other metabolites. In addition to glucagon, there are other hormones that can increase BG levels (the so-called counterregulatory hormones) such as epinephrine (also known as *adrenaline*), cortisol and growth hormone [15, 18, 19]. A summary of actions is presented in Table 1.1.

### 1.1.1 Diabetes overview

*Diabetes mellitus* (DM) is a very common chronic metabolic disease. Described initially in the Egyptian papyrus Ebers in 1500 BC, it results from both genetic predisposition and environmental risk factors and is characterized by disorders of the metabolism of CHO, fat, and protein, which are caused by the loss of endogenous secretion of insulin from the pancreatic  $\beta$ -cells, insulin action, or both [20, 21]. According to the national diabetes statistics report in 2017, diabetes affects 30.3 million people in the United States (9.4% of the US population). More alarming, an estimated 84.1 million more American adults have pre-diabetes. Over half of new diagnosed diabetes cases were in adults 45-64 years old, and nearly 16% of adults diagnosed with diabetes were smokers, nearly 90% were overweight, and more than 40% were physically inactive [22].

Statistically, according to the Diabetes Atlas in 2017, diabetes affects more than 425 million people, of which one-third are people older than 65 years, and it is expected that the number of people with diabetes may rise to 629 million in 2045 as shown in Fig. 1.2. The disease and its complications had a cost of at least USD 727 billion in terms of healthcare expenses in 2017 [1]. Clinically, *glycated hemoglobin* (“HbA1c” or just “A1c”) can be used as a diagnostic test for diabetes, where HbA1c is a surrogate of the average level of BG over the past 2 to 3 months (life period of the red blood cells). The normal range for the HbA1c level for healthy people is between 4% and

5.6%. An HbA1c level between 5.7% and 6.4% indicates pre-diabetes. An HbA1c level of 6.5% means that diabetes is present, so the target HbA1c level for people with diabetes is usually less than 7% as the recommendation of the American Diabetes Association [23, 20].

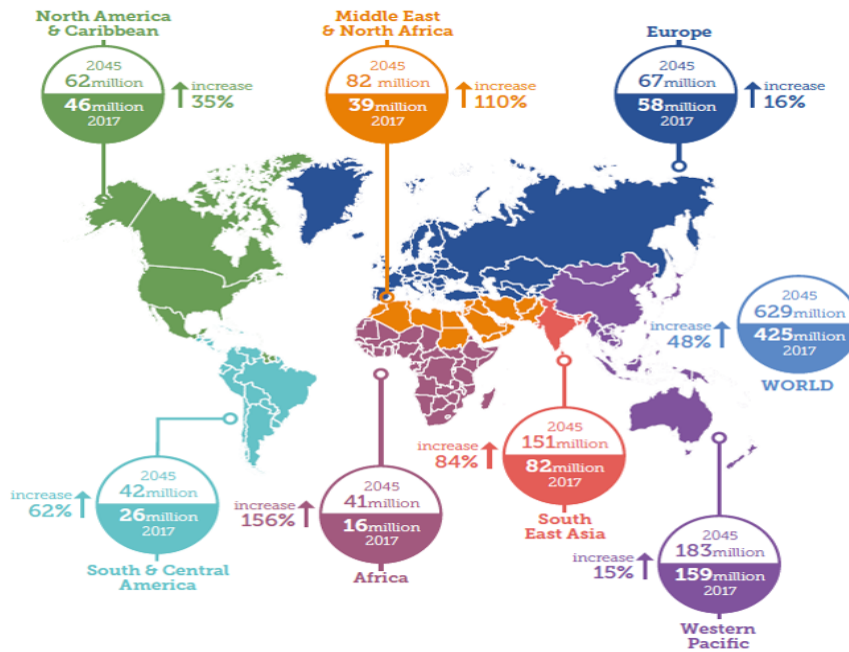


Figure 1.2: IDF regions and global projections for the number of people with diabetes (20-79 years), 2017-2045 [1].

### 1.1.2 Types of diabetes

According to the American Diabetes Association (ADA), the most common forms of DM are **type 1 diabetes**, **type 2 diabetes**, and **gestational diabetes** [20], which are described next.

#### Type 1 diabetes

*Type 1 diabetes mellitus* (T1D) is an immune-mediated chronic disease due to the destruction of  $\beta$ -cells in the pancreatic islets of Langerhans. As a consequence, insulin secretion is reduced to a very low level (down to 10% of normal). About 5-10% of people with diabetes have T1D, which generally develops in childhood and adolescence (thus it is also called “juvenile diabetes”). However, it also can develop in adults. It has a rapid onset and can make the affected person very sick very quickly. Once the BG level goes above the normal range ( $> 180$  mg/dL), the kidneys start to release glucose into the urine. The symptoms of T1D include increased thirst, weight loss,

extreme hunger, and mood changes. Patients with T1D need exogenous insulin administration for survival because nothing is secreted. Therefore, it is also called insulin-dependent DM (IDDM). Patients with T1D are more susceptible to other autoimmune diseases such as Hashimoto's thyroiditis, myasthenia gravis, celiac, and autoimmune hepatitis [24].

### **Type 2 diabetes**

*Type 2 diabetes mellitus* (T2D) is the most common type of diabetes: about 90% to 95% of people with diabetes have T2D. It more often develops in adults, but children can be affected. Once known as adult-onset, it is characterized by insulin resistance and relative lack of insulin. T2D occurs more often in people who have family members with diabetes and have high blood pressure. Obese people (>20% of the ideal body weight) have insulin resistance, so they are most susceptible to T2D [25]. In contrast to T1D, the pancreas is able to produce insulin; however, the cells of the body (*i.e.*, muscles, liver and fat) are inoperative at absorbing the insulin and cannot regulate glucose well. The common symptoms of T2D include increased thirst and frequent urination, and weight loss. Over time, the number of  $\beta$ -cells starts to decrease. This may require more medications in order to increase the insulin sensitivity and eventually, injected insulin could be needed like in patients with T1D. T2D is also called non-insulin dependent DM (NIDDM). Patients with T2D are susceptible to increase the risk of heart disease, macrovascular disease, peripheral vascular disease and stroke [26, 27].

### **Gestational diabetes**

*Gestational diabetes* (GD) is a temporary condition that occurs during pregnancy. GD is caused by not enough insulin to meet the extra needs in the pregnancy period. The regulation of glucose metabolism during pregnancy is complex, where the body produces different types of hormones such as estrogen and progesterone during the pregnancy to allow the extra glucose and nutrients in the body to reach the unborn baby. These hormones make the body insulin-resistant (*i.e.*, decreased insulin sensitivity). It is estimated that between 3-20% of pregnant women develop GD, depending on their risk factors (*i.e.*, overweight, and polycystic ovarian syndrome). GD is typically developing between the 24<sup>th</sup> and 28<sup>th</sup> weeks of pregnancy. GD is divided into two classes: class 1 is controlled through a good diet system, and class 2 is controlled through insulin or oral medications to control their condition. Children of mothers with GD have an increased risk of developing obesity, higher blood pressure, heart disease, and T2D [28, 29]. Therefore, good control of BG is an absolute must to protect the baby [30].



### 1.1.3 Diabetes complications

Diabetes complications are serious health problems that can happen when BG levels are out of normal range a long period of time. Diabetes complications may be disabling or even life-threatening because of late diagnosis and wrong management. Diabetic complications, both *acute* and *chronic*, account for the morbidity and mortality due to this relentless disease [31, 32].

#### Acute complications

*Acute complications* are short-term problems that can happen at any time. The major acute complications are the *hyperglycemic* and *hypoglycemic* emergencies. **Hyperglycemia** is a complex metabolic condition characterized by abnormally high levels of BG ( $> 180$  mg/dL). The symptoms and the first signs of hyperglycemia include frequent urination, increased thirst, and drowsiness and lethargy. Hyperglycemic emergencies include:

- **Diabetic ketoacidosis** (DKA), fundamentally in patients with T1D, but also may occur in patients with T2D. DKA is the most common acute complications of diabetes, and is characterized by elevated hyperglycemia (BG  $> 250$  mg/dL), high levels of ketones in serum or urine and acidosis.
- **Hyperosmolar hyperglycemic state** (HHS), primarily in patients with T2D. HHS is the most serious acute complications of diabetes, and is characterized by severe hyperglycemia (BG  $> 600$  mg/dL), hyperosmolality (osmolality  $> 320$  mOsm/kg), and signs of dehydration in the absence of ketoacidosis. HHS is associated with a significant morbidity and higher mortality than DKA, with a reported mortality for patients with HHS in the range 10% to 20%, which is approximately 10 times higher than the mortality rate for patients with DKA.

**Hypoglycemia** is a condition caused by a very low level of BG, and there are **three levels** of hypoglycemia as follow [33]:

- Level 1 (L1) mild hypoglycemia (54-70 mg/dL).
- Level 2 (L2) moderate hypoglycemia ( $< 54$  mg/dL).
- Level 3 (L3) severe hypoglycemia that characterized by altered mental/physical statuses requiring external assistance for recovery (often below 40 mg/dL).

The symptoms of hypoglycemia include shakiness, sweating, blurred or double vision, seizure, and nervousness. If untreated, hypoglycemia may lead to unconsciousness, coma, and potentially the patient can die [34].

## Chronic complications

*Chronic complications* are long-term problems that can expand gradually with the time, and can lead to significant damage. The most common chronic complications related health problems are [35, 31]:

- **Macrovascular complications:** affect large blood vessels of the heart, brain, and legs.
- **Microvascular complications:** affect small blood vessels, causing problems in the eyes (retinopathy), kidneys (nephropathy), feet, and nerves (neuropathy).

Diabetes can also affect other parts of the body, including the digestive system, the skin, sexual organs, and gum disease and other mouth problems. The Diabetes Control and Complications Trial (DCCT), in 1993 [36], and the UK Prospective Diabetes Study (UKPDS), in 1998 [37], have demonstrated that a good glycemic control decreases diabetic complications, and therefore can reduce morbidity and higher mortality rates due to the disease.

### 1.1.4 Insulin therapy in type 1 diabetes

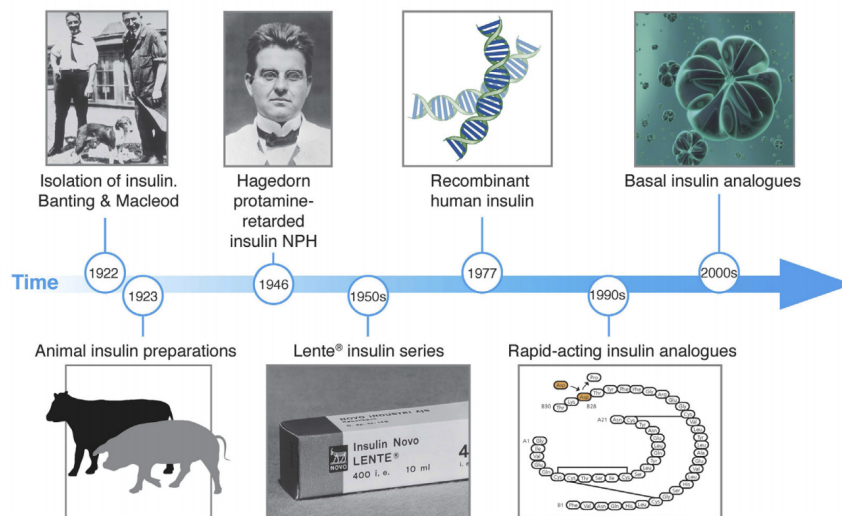


Figure 1.3: The history of insulin therapy [2].

This thesis focuses in T1D, where absence of insulin secretion requires insulin replacement therapies. In 1921-22, Dr. Frederick Banting, Professor John Macleod, and medical student Charles Best carried out many experiments on the pancreas in animals, especially dogs, in Toronto, Canada, in order to isolate the secretions from the islet cells which led to the discovery of *insulin* as a possible treatment for diabetes [2, 3]. It was considered

to be one of the greatest medical breakthroughs in history, with the award of the Nobel Prize to Banting and Macleod in 1923. Then, a slow-acting formulation was developed in the late 1940s, neutral protamine Hagedorn (NPH) insulin. In 1977, recombinant DNA techniques were used to produce synthetic “human” insulin, which distinct from insulin derived from animals, through the biotechnology firm Genentech, to enable production of large amounts of insulin. The first human insulin was marketed both by Eli Lilly and Novo in 1982. Then, rapid-acting insulin analogs were introduced in the 1990s and long-acting basal analogs in the early 2000s as shown in Fig. 1.3.

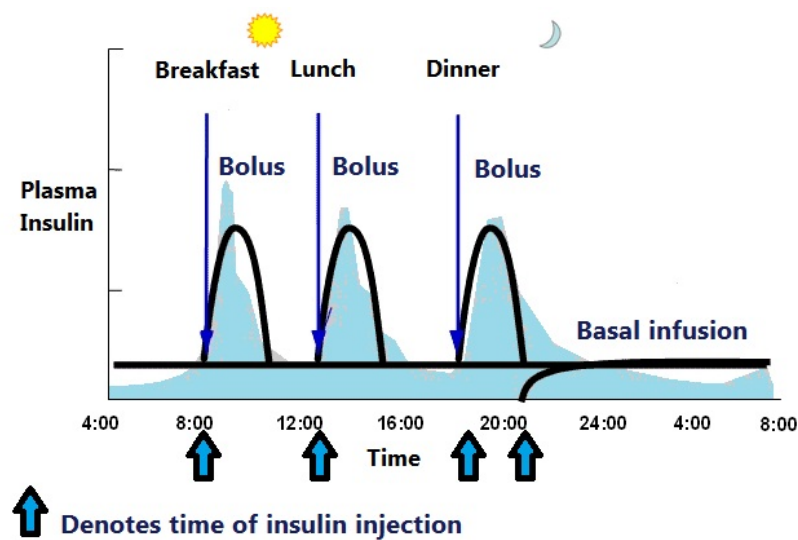


Figure 1.4: Physiological basis of the basal/bolus insulin regimen. Modified from [3].

In the insulin therapy for T1D, patients are treated with daily injections more than once a day depending on the size of a meal, and meal frequency. Patients should be testing their BG levels before each injection for adjusting the insulin dose. The size of the insulin doses must be taken very carefully. Too little insulin leads to rising their BG levels (hyperglycemia), leaving them at risk over time for complications. In contrast, too high doses lead to low BG (hypoglycemia). There are two common types of insulin: basal insulin and bolus insulin, in order to maintain an optimal BG level, “*replacating*” the insulin system in the healthy body as closely as possible. It must be remarked that the insulin system does not exist as an isolated system in the human body, but this engineering abstraction helps us to understand the complexity of the glucose homeostasis. Basal insulin regulates the BG levels in between meals, and bolus insulin is extra units of insulin to manage the BG levels after a meal, as shown in Fig. 1.4. There are many types of

Type	Brand Name	Onset	Peak	Duration
Rapid-acting	Humalog Novolog Apidra	10-30 min	30 min-3 h	3-5 h
Short-acting	Regular (R)	30 min-1 h	2-5 h	Up to 12 h
Intermediate-acting	NPH (N)	1.5-4 h	4-12 h	Up to 24 h
Long-acting	Lantus Levemir	0.8-4 h	Minimal peak	Up to 24 h

Table 1.2: Pharmacokinetics of different insulin types for people with diabetes.

insulin available with different characteristics [38] as displayed in Table 1.2, where insulin analogues are classified according to their duration of action as rapid-acting (*e.g.*, Humalog, Novolog), short-acting (*e.g.*, Regular actrapid), intermediate-acting (*e.g.*, NPH, Lente), and long-acting insulins (*e.g.*, Lantus, Levemir). The onset (onset of action) is the time before insulin reaches the bloodstream, the peak (peak effect) is the time when insulin is most effective, and the duration defines for how long insulin is having an effect.

### 1.1.5 Challenges to glycemic control

In a healthy person, the BG levels stay within a normal range by the glucose homeostasis. However, with diabetes, the BG level can rise higher or go lower than normal. Many factors can affect glycemic control (Fig. 1.5) such as: meal intake with more/fewer CHOs than usual, mismatched insulin doses due to poor estimations of CHOs content in the meal, psychological stress, physical activity, besides the metabolic reaction for each individual giving rise to high intra-patient variability [39].

#### Meals

*Meals* are a major challenge to BG control. The BG (glycemic) response to a meal is the effect that a meal has on BG levels after consumption. Many factors influence glycemic response to food including the nutritional composition, gastric emptying rate, absorption rate, and other physiologic effects [40]. Normally, BG levels increase after a meal, as a result of absorbing digestible CHO from the intestine into the bloodstream. Automatically, the pancreas detects this rise in BG and starts to secrete insulin and then BG level returns gradually to fasting levels within a short period of time. More than that, a cephalic phase exists in healthy individuals, where insulin is secreted in anticipation from the stimulus in the mouth [41].

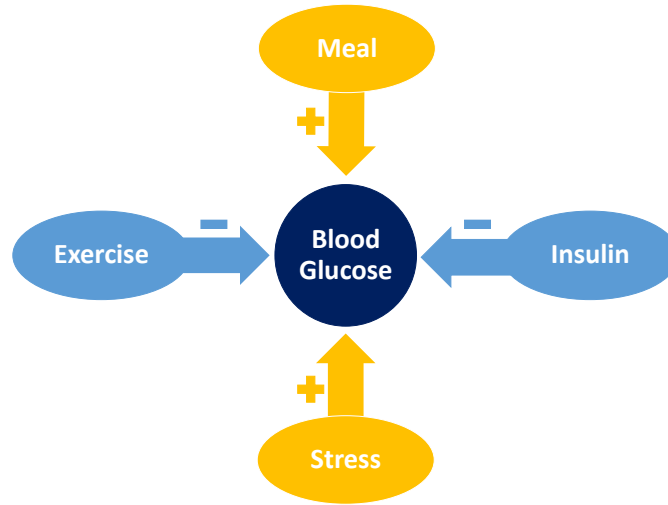


Figure 1.5: Some external factors that affect the blood glucose level.

Particularly, the important problem for people with diabetes is how to reduce the size and duration of rises in BG after meals. Glucose from some foods is absorbed more rapidly than other foods, where dietary fats cause a delay in the absorption process of glucose from CHO, and fiber, especially soluble fiber, cause a slow nutrient absorption and flattens post-meal BG level response [42, 43]. Protein and CHO in a meal cause a reduction of the increase in BG levels and increase insulin secretion [44]. There are different measures that have been developed in order to define the effect of food on the glycemic response such as the glycemic index (GI) and the glycemic load (GL) [45]. Jenkins et al. developed the glycemic index in 1981, in order to measure the effect of CHOs on postprandial glycemic responses [46]. According to the Food and Agriculture Organization (FAO) and the World Health Organisation (WHO), patients with diabetes are recommended to substitute high GI foods with low GI alternatives to help optimize glycemic control [47].

### Physical activity

*Physical activity* (PA) is an important part of any diabetes management plan due to its influence on different BG responses. Physical activity has a significant impact on BG levels in patients with DM, especially T1D, depending on the exercise type (aerobic, anaerobic, mixed), form (continuous or intermittent), intensity (light, moderate, vigorous) and duration (short, moderate, long), besides the physical fitness level and degree of stress of the patient [9, 48]. Regular PA has a significant benefit on insulin sensitivity, improves the blood lipid profile, reduces the risk of heart disease and stroke,

increases muscular strength, and prevents cardiovascular disease (CVD) [49]. The American College of Sports Medicine (ACSM) has published a guideline for PA prescription in T1D [50]. The American Diabetes Association (ADA) and ACSM recommend patients with T1D to keep BG level above 100 mg/dL before, during, after exercise of all types, and below 250-300 mg/dL, and the practice of exercise at an intensity of 40-60% of their maximal oxygen consumption ( $VO_2\text{max}$ ) for 5-7 days/week for a duration of 20-45 minutes [51, 50].

Many experiments demonstrate the effective role of physical activity in body's fuel metabolism, which leads to different glucose responses [52, 48, 53, 54]. Continuous exercise of moderate intensity causes a decline in BG level (Fig. 1.6) due to enhanced insulin sensitivity and accelerated glucose consumption by muscle cells [55, 56], whereas high-intensity exercise ( $\sim 15$  min at 80%  $VO_2\text{max}$ ) can raise the BG level during the exercise and prolonged hyperglycemia during recovery due to the release of stress hormones (*e.g.*, adrenaline) [57, 58].

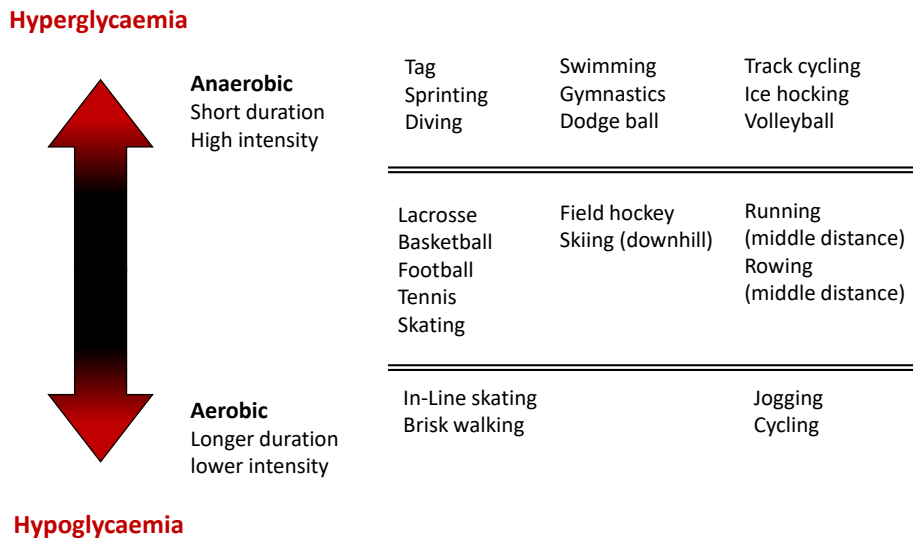


Figure 1.6: The effectiveness of different exercises in diabetes. Modified from [4].

### Psychological stress

*Psychological stress* is typically broken down into two categories: *acute stress* and *chronic stress*. Acute psychological stress is short-term stress and it is a normal part of everyday life. Acute psychological stress is a result of an immediate reaction to sudden unexpected events such as a traffic jam, traffic accident, and loss of job. The stress sources can be physical, like infection or serious illness. Alternatively, they can be mental, like problems in a job, and

health. Chronic psychological stress is long-term stress, as a result of unresolved problems, begins to increase, and lasts for a prolonged period of time such as low social economic status, social isolation, and continual problems in married life. During stressful situations, glucagon and adrenaline levels rise and insulin level goes down, and the liver cells release more glucose in the bloodstream. In the same time, the body tissues like muscle and fat are less sensitive to insulin, as a result of high levels of growth hormone and cortisol. Thus, providing more glucose in the bloodstream. Therefore, psychological stress can cause a significant and prolonged increase in the BG level, and in this case, more insulin is required to lower BG level into the optimal range during stressful situations [59, 60].

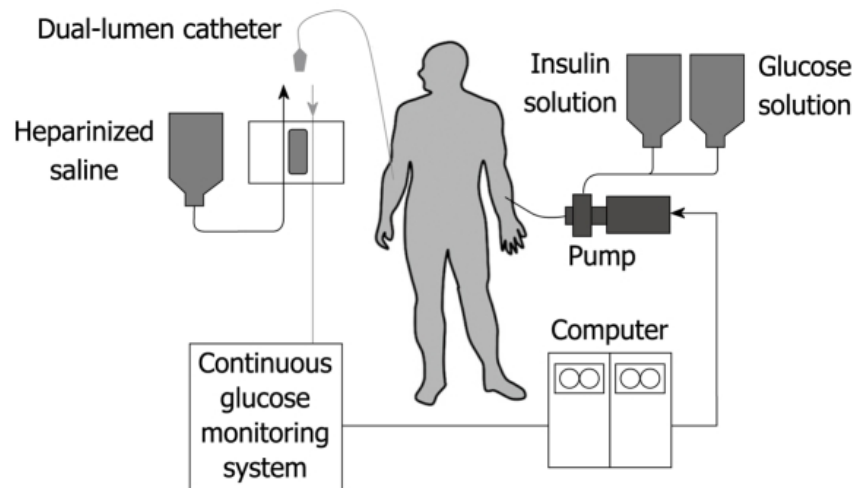


Figure 1.7: Schematic diagram of the bedside-type artificial endocrine pancreas [5].

## 1.2 Closed-loop glucose control in type 1 diabetes

In a healthy individual, glucose is regulated by glucose homeostasis (closed-loop feedback system). In people with T1D, in the absence of insulin, the closed-loop will not be completed. Therefore, one of the technological solutions is a closed-loop control system for glucose regulation to emulate the body's natural ability to infuse insulin. Another way to achieve this would be a clinical solution via the pancreas transplant, but this operation depends on the immune response, that may lead to rejecting the foreign organ. Professor E Perry McCullagh proposed the concept of an implantable artificial endocrine pancreas in 1959, an endocrinologist at The Cleveland Clinic, as a first attempt to create a closed-loop control system for people with T1D [5]. In 1974, Albisser et al. [61, 62], and Pfeiffer et al. [63], were the first to successfully use the artificial endocrine pancreas in a clinical study, which



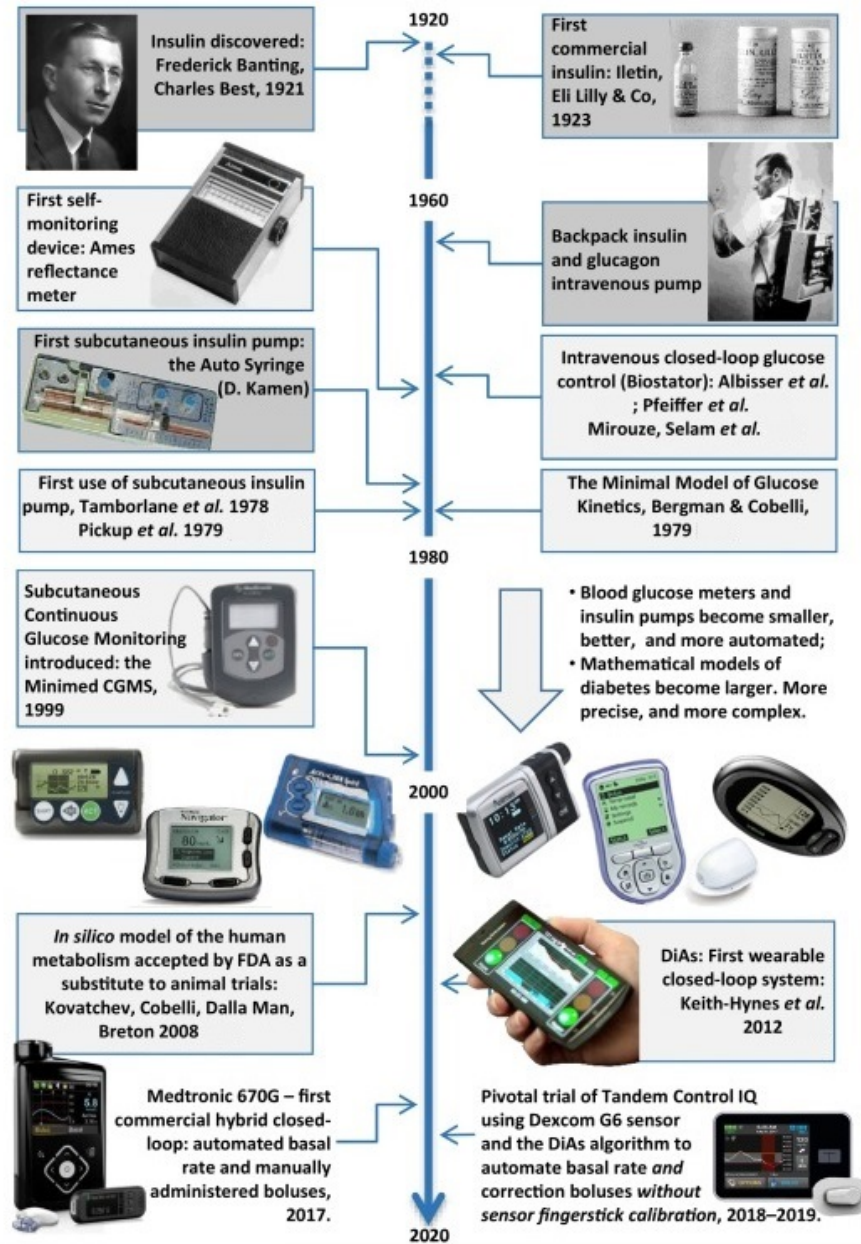


Figure 1.8: Timeline of diabetes technology [6].

led to the emergence of the first commercial device that mimics an artificial pancreas, a bedside-type artificial endocrine pancreas (Fig. 1.7): the Biostator (Miles Laboratory Inc., Elkhart, IN, USA). In 1977, by Clemens *et al.* [64], the Biostator system was developed, a glucose controlled insulin



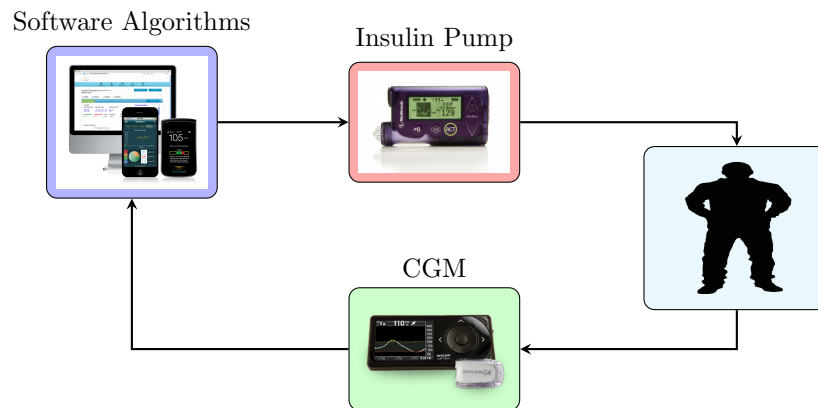


Figure 1.9: Closed-loop artificial pancreas system.

infusion system (GCIIS) for hospital use. After that, advanced algorithms were used to improve the artificial endocrine pancreas through several researchers such as Shichiri et al., 1983 [65], Salzsieder et al., 1985 [66], and Fischer et al., 1990 [67]. After this development, the artificial endocrine pancreas became known as “**artificial pancreas**” (AP) in the clinical research of diabetes. Therefore, the AP is a new ray of hope and a road map for people with diabetes, especially T1D, after a long struggle over the past century (Fig. 1.8).

### 1.2.1 The artificial pancreas

As previously reported, an *artificial pancreas* is a closed-loop control system for glucose regulation to be able to mimic a healthy pancreas to reduce the burden of multiple daily injections of insulin. Therefore, the AP has the potential to transform the lives of people with diabetes, particularly, T1D. The common components of most AP systems are: a continuous glucose monitor (CGM), an insulin pump, and a control algorithm (Fig. 1.9), which determines automatically the required insulin dose based on information received from the CGM for every 5 minutes [68, 69]. The main aim of all AP systems is to compute the “*optimal*” insulin dose, to be applied through the insulin pump, which keeps BG level in normal range (70-180 mg/dL) as much as possible, without causing hypoglycemia.

#### Glucose monitoring

Typically, patients use two main methods to measure BG levels daily to make correct therapeutic decisions: capillary BG measurements and CGM. In the 1970s, capillary BG monitoring was first used as a monitoring tool for diabetic patients, by using glucometers, to give a guide to BG levels

at a specific moment in time. Capillary BG monitoring, also called *self-monitoring of BG* (SMBG), has been used for decades [70]. It has many advantages over venous blood samples: it requires a little amount of blood, it measures whole BG not only BG in plasma, as opposed to the venous test, and it can be tested quickly and in an easy way [71].

A CGM system works through needle sensors that are inserted under the skin, in the subcutaneous adipose tissue, to access the interstitial fluid (ISF). Then, a sensor measures glucose every few minutes and transmits the information wirelessly to a monitor in real time. Monitoring BG levels in real time and continuously can help people with diabetes to make more informed decisions throughout the day about how to balance several factors influencing BG level such as meal intake and physical activity. In the recent years, *flash glucose monitoring* (FGM) has been introduced in the market, where the glucose data curve is sent only when the patient manually “scans” the sensor. However, since FGM is not suitable for an AP systems, it will not be discussed here.

The important advantages of CGM systems over SMBG include:

1. The BG measurements are performed continuously every few minutes while SMBG is only done a few times every day.
2. CGMs can reduce the risks of hypoglycemia and hyperglycemia.
3. CGMs can improve the HbA1c level over time to a larger extent [72].

Whilst, the disadvantages of CGM systems include [73, 74]:

1. CGM systems must be calibrated with SMBG data 2 to 4 times per day as a reference. However, some systems that do not require calibration have recently appeared on the market.
2. The accuracy of the BG measurements is variable and dependent on the calibration significantly.
3. CGM systems monitor glucose changes in ISF and not in blood, but glucose levels in ISF lag temporally BG values in the range of 3-12 min.
4. Glucose sensors have an approximately a 6-month expiration from the date of manufacture, but the sensor inside the patient has to be changed every 3 to 7 days depending on the glucose monitor model, and nowadays CGM sensors are expensive.

CGM accuracy is measured through the *Mean Absolute Relative Deviation* (MARD), which consists in the average relative error between the sensor and a reference glucose value obtained at the same time, and is defined as

$$\mathbf{MARD} := \frac{100\%}{n} \sum_{i=1}^n \left| \frac{CGM_i - YSI_i}{YSI_i} \right|, \quad (1.1)$$

where  $n$  is the number of samples,  $CGM_i$  indicate to the sensor BG reading at time  $i$ , and  $YSI_i$  is the reference BG reading at time  $i$ . Lower MARD values indicate better CGM performance. The main aim is to develop “*smart*” CGM systems able not only to detect hypo- or hyperglycemic episodes, but also to make predictions based on trends that would allow the patient to gain more time to take action to avoid risky events [75].

In 2016, the U.S. Food and Drug Administration (FDA) approved the use of **Dexcom G5 Mobile** (DG5M) CGM with a MARD of 9% for the management of diabetes in individuals age 2 years and older. DG5M is officially the first approved CGM system through FDA, to use without a confirmatory SMBG value to replace traditional BG testing for diabetes treatment decisions (non-adjunctive use) [76]. Other devices have followed, like Abbott Freestyle Libre (a FGM system), Senseonics Eversense (an implantable subcutaneous CGM) and the **Dexcom G6** that received FDA approval in March 2018. Dexcom and Verily are expected to launch the G7 CGM system in 2020 with better specifications (*e.g.*, much thinner, more affordable, and continuously communicating with smartphones) [77].

### Insulin pump

An *insulin pump* is a device that delivers a continuous subcutaneous infusion of insulin 24 hours a day. It consists of a pump reservoir filled with insulin, a computer chip to control the insulin dose, and an infusion set with a small battery, that is designed to administer insulin subcutaneously through a needle or cannula inserted under the skin, usually on the abdomen. The insulin pump, also known as a *continuous subcutaneous infusion of insulin* (CSII), appeared for research purposes at the first time in 1976 by Pickup and Keen [78, 79]. At that time, the size of the CSII was as big as a back bag and over time, technology has developed, nowadays CSII becoming small size and easy to use.

Over time, the number of people with diabetes using CSII is increasing, according to the FDA, who reported that about 375,000 adults with T1D in the US used CSII in 2007, an increase of 245,000 over the number in 2002 [80]. The American Association of Diabetes Educators (AADE) recommends using CSII in the following cases [81]:

- Patients who suffer from wide unpredictable fluctuations in BG levels.
- Patients with severe hypoglycemia and HbA1c level greater than 7%.
- Patients who have hypoglycemic episodes that require assistance or that interfere with work or school.

CSII has several advantages such as: increased flexibility in lifestyle compared to patients on multi-injection therapy, precise insulin delivery and safety options to stop overdosage, suitable for individuals whose meal times are unpredictable, who achieve generally tighter BG control, reducing wide fluctuations in BG, resulting in a better HbA1c. CSII disadvantages include: infusion sets may cause skin infections, DKA can occur rapidly if the pump malfunctions, as short-acting insulin is the only form given or the insulin is inactivated by heat exposure, and they are expensive [82]. Currently, modern pumps feature an insulin-on-board (IOB) calculation, corresponding to the current active insulin or insulin remaining from the previous insulin delivered. IOB can prevent insulin from stacking, therefore can reduce the risk of overdosing, and reduce the risk of hypoglycemia [83].

### Control algorithms

*Control algorithms* are the brain of the AP to keep the BG level in the target range and preventing patients from hypoglycemia. Several approaches have been tested to design such a control algorithm including model predictive control (MPC), proportional-integral-derivative control (PID), fuzzy logic control (FL) and  $H_\infty$  control among others [84, 85, 86, 87, 88]. The main input to all control algorithms is the signal error, which results from the difference between the desired BG level and the CGM readings.

The main difficulty for glucose control algorithms is the delayed action of insulin since both glucose measurement and insulin delivery is done in the subcutaneous tissue. Besides, there is a substantial variability of insulin pharmacokinetics between individuals. Furthermore, several factors lead to varying insulin sensitivity of a patient [89, 90]. The inaccuracy of the CGMs has to be considered when designing control algorithms, especially at lower and rapid fluctuation BG readings.

Glucose control algorithms should contain more options to avoid the risk of hypoglycemia such as: calculating the IOB to take the delayed action of insulin into account, using algorithms to predict hypoglycemia thus sending out alert messages to the patient and stopping insulin infusion [91, 92]. Furthermore, manual announcements such as announcement of meals or exercise, are often parts of control algorithms to be able to minimize post-prandial hyperglycemia and prevent hypoglycemia. Regarding the safety issues, the security alarm is a major challenge in glucose control algorithms; it must be able to detect any critical state to prevent patients from risk.

### 1.2.2 State of development

Rapid progress has been made through last years in the development of AP systems compared with the first closed-loop control system, Kadish 1964 [93, 94]. Currently, several research groups worldwide are developing the

AP systems, either *single hormone* (insulin delivery alone) or *dual hormone* (bihormonal) using insulin and glucagon to mimic the real pancreas and minimize the burden of BG management [95]. The use of insulin-only or insulin and glucagon in AP systems each has its own advantages and disadvantages [96]. In many studies, dual-hormone systems have been proven to be superior to single-hormone systems, especially, in preventing hypoglycemia in response to meals and exercise [97, 98].

Several clinical studies have established the efficacy and safety of both insulin only [99, 100, 101] and dual-hormone APs [97, 102, 103]. These clinical studies have proved that AP systems can improve glycemic control, improve HbA1c level, and therefore prevent long-term complications of diabetes. Weisman et al. [104] demonstrated this improvement through a metaanalysis of clinical outcomes in outpatient studies comparing AP systems with open-loop therapy in adults and children with T1D. The use of AP systems lead to an increase in target BG range by 12.59%, approximately equal to 3-h per day. Besides that, HbA1c could be reduced by a minimum of 0.3% when using AP systems, although HbA1c was assessed in only two studies. In addition, the use of AP systems further reduced hypoglycemia and the burden of disease management [104].

Besides the control algorithms, additional modules have been developed to make the AP system more reliable. Meal detection modules [105, 106, 107] are capable to detect the meal intake automatically without any interaction of the user in order to determine the early insulin bolus to prevent potential hyperglycemia. Exercise modules [4, 108] are applied to AP systems to detect exercise and change therapy to prevent potential hypoglycemia during the exercise and recovery periods. IOB modules [109, 110] are developed to estimate the size of insulin that is still active in the body to avoid insulin stacking that causes hypoglycemia. Hypoglycemia early alarm modules [111, 112, 113] are able to warn from potential hypoglycemia, gaining more time to take action to avoid hypoglycemia.

In 2017, the FDA approved **Medtronic MiniMed 670G** hybrid closed-loop control system as the first commercialized product available in the market today. The approval came after evaluated data for the MiniMed 670G from a clinical trial that involved 124 participants with T1D with mean age 37.8 years, for 3 months after a 2-week run-in period [114]. The MiniMed 670G still needs patient's manual interaction, announcing meals and exercise. This is just the first step, therefore, paving the way for further studies and experiments to reach the best, and to increase the marketing capacity of AP systems.

## Chapter 2

# Thesis objectives

### 2.1 Glucose prediction in the artificial pancreas

The recent commercialization of the first artificial pancreas, the Medtronic MiniMed 670G hybrid (semi-automatic) closed-loop glucose control system, is only a first step towards a fully automatic control system and several challenges still remain open. One of the main challenges, that can leverage better control and supervision systems, is to obtain better models for the prediction of future BG levels as an integral part of artificial pancreas systems. The ability to predict glucose along a given prediction horizon (PH), and estimation of future glucose trends, is the most important feature of any AP, in order to be able to take preventive actions to entirely avoid risk to the patient. Glucose prediction can appear as part of the control algorithm itself, such as in systems based on MPC techniques [115, 116, 117], or as part of a monitoring subsystem to avoid hypoglycemic episodes [118, 119, 120].

However, accurate glucose prediction poses many challenges, as a result of the effect of many factors on glucose dynamics such as meal intake, insulin action, physical activity, stress, and inter- and intra-patient variability. Therefore, ideally, prediction of BG should use many inputs; certainly BG measurements by the CGM sensor, but also estimated ingested CHOs and infused insulin play a major role, besides physical activity as measured by biometric sensors (*e.g.*, galvanic skin response, heart rate variability, skin temperature, energy expenditure).

The generation of risk alerts is the most immediate application of glucose prediction to creating safer control algorithms by allowing the patient to take appropriate therapeutic actions, *e.g.*, snack to counterbalance hypoglycemia, or applying correction insulin boluses to deal with hyperglycemia, based on the predicted BG value [121, 122, 123].

## 2.2 General objective

The modeling error in glucose prediction of available type 1 diabetes models is large, specially for large prediction horizons and under disturbances. This is due to the large intra-patient variability and the unavoidable free-living changing conditions. Therefore, the challenge of improving the accuracy and prediction horizon of the current state of the art models is going to be addressed in this work then being the main objective of the thesis the **“development of new techniques for identifying individualized patient glucose prediction models in type 1 diabetes”**.

## 2.3 Specific objectives

Estimation of future BG concentrations is a critical task for diabetes management, in order to save diabetic people from diabetes complications. An accurate prediction of glycemia can be used for early hypoglycemic and hyperglycemic alarms. In addition, adjustments of insulin amount/rate is a major aim of the AP systems, and is guided by glycemia predictions. Furthermore, predicted glucose values can be used for detecting the abnormal behavior of the patient or the AP system.

The goal of the **first sub-objective** of this thesis is **to improve current models by developing better stochastic models from available historical CGM data of a patient and use the models developed to predict the future BG variations**. This sub-objective involves the analysis of CGM glucose concentration data collected from patients, as well, the inter- and intra-individuals variability.

Glucose-insulin dynamics exhibits great variability from individual to individual. As well, the metabolism of individuals with diabetes differs from the metabolism of healthy individuals. Furthermore, metabolic changes resulting from illness, stress or changes in insulin sensitivity may also cause critical variations in glucose-insulin dynamics within the same individual. Therefore, the models developed should be able to capture the dynamic characteristic of diabetic people behavior, and also inter- and intra-individuals variability in the BG homeostasis.

Several external disturbances such as meal consumption and physical activity also cause large BG fluctuations, especially, in postprandial periods and after exercise. Therefore, the postprandial period is the most challenging period, due to the presence of many external factors that affect the behavior of patients, such as the meal intake and amount of insulin. Changes in exercise habits, and the physiological variability of the patient also lead to huge variations. A reliable model for BG prediction should also perform efficiently under such conditions without demanding any *a priori* disturbance announcement or information.

In order to cope to all these sources of variability, this thesis will explore the CGM data driven development of personalized stochastic models such as *stochastic seasonal time series models*. Therefore, time series analysis techniques with seasonality variation, with success in other areas, are going to be explored. Seasonal phenomena may stem from various factors, such as weather and holidays, which affects many business and economic activities like tourism, and custom events like Christmas. In seasonality, the mean of the observations is not constant but instead develops according to a cyclical pattern. For example, in the monthly temperatures time series in Europe, the mean temperature is not constant since it varies by month, but for the same month in different years, we can expect a constant average value. Hence, seasonality in *postprandial* periods or *nocturnal* periods should be explored.

The proposed stochastic seasonal model can simplify current diabetes management, where this model can be used to provide early hypoglycemia and hyperglycemia alarms and also can be used as a guide for their daily insulin dose and associated adjustments. This model could capture the dynamic characteristic of the patient behavior and it can be able to predict BG concentrations for larger PH. The glucose prediction through the proposed stochastic seasonal model in the sub-objective 1 could be improved by using exogenous inputs such as *insulin infusion* and *physical activity*.

However, the best results of seasonal models appear when similar behaviors are considered. On the other hand, when a nonlinear complex system is very difficult to model, a common approach is to divide the complex modeling problem into simpler modeling subproblems. Therefore, the concept of local modeling comes to solve this problem and the glucose prediction seasonal model can be improved when similar glyceemic behaviors are used (*e.g.*, similar postprandial periods). Then, **Sub-objective 2** copes with the **development of stochastic seasonal local models, and subsequently integrate them to have a better glucose prediction.**

In order to do so, similarity measures techniques will be explored to reach the similarity required in the concept of seasonality. Clustering time series techniques can obtain similar groups (clusters) of glyceemic behaviors as a basis of seasonality. Therefore, clustering techniques can be used to reach similar groups of glyceemic behaviors before applying the seasonal modeling. The system complexity is divided into a set of simpler systems through local seasonal models (*i.e.*, *a family of stochastic seasonal models*), and a global model could be obtained by integrating these local models.

Once a suitable technique for accurate glucose prediction with large prediction horizon is available, it must be tested under free living conditions, and used for control and/or supervision. **Sub-objective 3** will deal with this task, where a **framework for the use of the stochastic seasonal prediction models methodology under free-living online conditions**



must be developed. Seasonality conditions does not exist naturally in real data, and prediction models could only be used in normal life by enforcing the concept of seasonality. In real life, periods have different lengths in the same daily events (*e.g.*, the meal-to-meal time can have a different duration each day) and it must be fixed for using seasonality. Once this problem is solved, the prediction model can be used to design an online *monitoring system* that informs the patient of the BG prediction, how much can he/she trust in the prediction, or whether an abnormal behavior is being detected. That may help subjects with diabetes to mitigate the effects of the glycemic control variability in a safe/robust way in case of presence of meal intake/exercise/unknown situations. The long-term complications of diabetes could then be reduced by the patient taking actions to keep BG concentrations within the narrow range (*i.e.*, 70-180 mg/dL).

Summarizing, three sub-objectives have been devised in order to reach the main objectives of the thesis of developing new techniques for identifying individualized patient glucose prediction models in type 1 diabetes:

1. Development of methods for the accurate and long term prediction of glycemia through personalized stochastic seasonal time series models using real-time CGM data.
2. Improve the methodologies by stochastic seasonal local models, and subsequently integrate them to have a better glucose prediction, in order to cope better with variability and allow seasonality techniques.
3. Design the framework for the use of the previous techniques under free living conditions, including the design of a monitoring system providing in real time to the patient: an accurate glucose prediction with an associated confidence measure, and warning alarms when detecting anomalous states.

## Part II

# State of the art

## Chapter 3

# Glucose prediction in type 1 diabetes

As previously stated, accurate glucose prediction is key for a good diabetes management, either in decision-aid systems for optimizing insulin therapy or automatic systems like the artificial pancreas. However, it remains a big challenge due to the great variability of the glycemic response in T1D individuals. This chapter reviews the different modeling approaches in literature for glucose prediction, with special emphasis in the analysis of the achieved prediction performance. The chapter starts in Section 3.1 with the main challenges of glucose prediction and the factors that affect glycemic response and, therefore, glycemic control. Section 3.2 displays the main categories of glucose prediction models such as physiological models, data-driven models and hybrid models, which will be reviewed in Sections 3.3, 3.4 and 3.5, respectively. This chapter ends with an analysis in Section 3.6 of the current performance limitations in glucose prediction, motivating the approach followed in this thesis.

### 3.1 Challenges of blood glucose prediction

BG prediction in people with T1D is a research problem that has been widely studied since the availability of CGM time series data, especially in combination with insulin pump therapy (see the recent review [124] and references therein). As stated in Section 2.1, prediction of future BG levels is an integral part of AP systems, either as part of the control algorithm itself such as in MPC control and hypoglycemia mitigation modules, or as part of supervision systems. However, glucose prediction has shown to be challenging mainly due to the large *inter- and intra-individual variability* exhibited by T1D patients and the complexity of disturbances like *meals* and *exercise*.

### 3.1.1 Inter- and intra-individual variability

As an illustration, Fig. 3.1(a) shows the response of two different patients to a same 60 g CHO meal in open loop therapy and Fig. 3.1(b) the response of a same patient to the same experiment in two different days. Data are extracted from the study CLOSEDLOOP4MEALS (NCT02100488), performed by the research group at Hospital Clínic de Barcelona and Hospital Clínico Universitario de Valencia, where 20 patients underwent 4 mixed meal studies, two in open loop and two in closed loop [7]. Despite a previous insulin feedback phase to bring the patients about the same glucose value before the meal at  $t = 0$ , variability of responses is significant, both inter- and intra-individual. Remark that in the case of Fig. 3.1(b), the same meal and insulin is supplied, whereas in Fig. 3.1(a), basal and bolus insulin settings are according to each patient’s standard therapy optimized prior to the study. This illustrates the difficulty experienced by patients to achieve a good glucose control.

Clearly, it can be stated from clinical practice that “each patient has its own diabetes”, which is reflected into a wide range of insulin requirements in the T1D population. In [125], a retrospective analysis of age-related variability in insulin needs during 3-week 24-h closed-loop glucose control is performed. A cohort of 20 young children (1-6 years old), 21 children (7-12 years old), 15 adolescents (13-17 years old) and 58 adults (> 18 years old) were included. Total insulin delivery at each age group was  $0.9 \pm 0.21$ ,  $1.01 \pm 0.21$ ,  $0.94 \pm 0.19$  and  $0.62 \pm 0.15$  U/Kg/day leading to percentages of time in range 70-180 mg/dL of  $70.5 \pm 7.4$ ,  $65.2 \pm 6.2$ ,  $69.2 \pm 6.9$  and  $69.0 \pm 9.6$ , respectively. The coefficient of variation of insulin delivery was statistically significantly higher in young children compared with adults, with a mean difference of 10.7 percentage points at night and 6.4 percentage points during the day, as well as compared to adolescents, with a mean difference of 10.2 percentage points at night and 7.0 percentage points during the day. In terms of glucose prediction, this means that model individualization will be needed to cope with large inter-individual variability. Thus, model complexity and its identifiability properties will be significant for a successful model individualization.

Regarding the large intra-individual variability suffered by patients, its sources are not fully understood. This is an important challenge to glucose prediction, since information from all these disturbances won’t be available. Key players in intra-individual variability are the variability of subcutaneous insulin absorption [126, 127], changes in insulin sensitivity due to circadian rhythms and external stimuli [128, 129] and premenstrual periods in women [130], besides the highly variable glycemic effect of meals and exercise.

An intra-individual variability of 27% was reported for time-to-peak plasma insulin concentration in an insulin aspart pharmacokinetic study

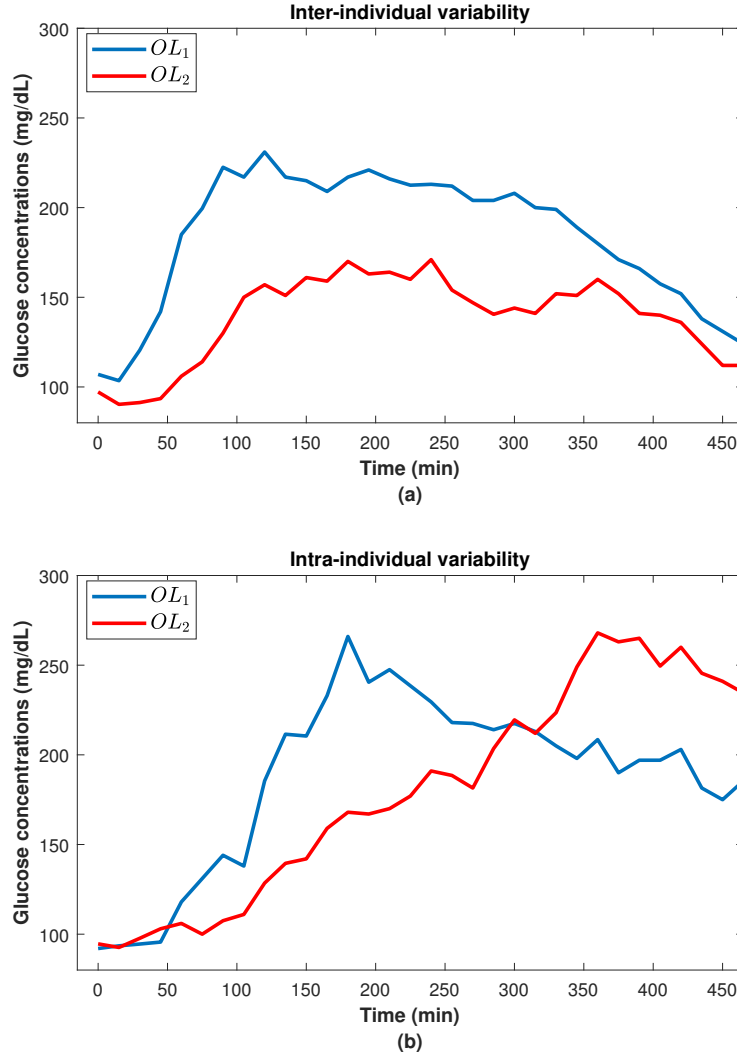


Figure 3.1: Illustration of inter-individual and intra-individual variability to a 60 g mixed meal response in open loop therapy. Data extracted from the study Closed-loop4meals (NCT02100488) [7].

[126], of which, nearly 40% was attributed to variations in depth of cannula insertion, insulin site age and local tissue perfusion. In [131], a significant effect of lypohypertrophy is reported with a 3- to 5-fold increment of the coefficient of variance of the area-under-the-curve of plasma insulin after an insulin injection in lypohypertrophic adipose tissue as compared to normal adipose tissue, reaching up to 65%. This damage of the adipose tissue consists in an abnormal mass of fat under the skin and appears when an infusion

set is located repeatedly at the same place. Its prevalence ranges from 28% to over 64%, depending on the country [131].

Circadian rhythms consist in 24-h cycles in behavioral and physiological processes, anticipating and adapting to environmental changes [129]. The circadian system regulates the daily rhythm of glucose metabolism. A higher insulin sensitivity at breakfast than dinner has been demonstrated in healthy individuals [132]. In [90], the existence of such circadian pattern in T1D was addressed. In contrast to healthy individuals, patterns of insulin sensitivity in T1D showed to be specific of the individual, not being possible an extrapolation to the T1D population. In average, the opposite pattern was found, with lower insulin sensitivity at breakfast than lunch and dinner. Based on these data, in [89, 133], seven possible classes of insulin sensitivity circadian variability were considered for its representation in a T1D simulator.

Many other secondary factors may contribute to variability. Caffeine and fatty food intake, weight gain, and illnesses could lead to an increase in BG levels. In contrast, aging, intense periods of concentration, and some medications are factors that could lower BG levels. Alcohol consumption and impaired digestion could lead to unpredictable glycemic deviation [134].

### 3.1.2 Postprandial response

Glycemic response to a meal is affected by many factors, aside the CHO content which is the only input considered in standard insulin therapy and in hybrid closed-loop systems. In [135], the impact of fat, protein and GI on postprandial glucose response in T1D is analyzed, revealing that all these factors modify the postprandial glucose pattern. It is concluded a late hyperglycemic effect of fat and consistent indication of higher insulin needs for high-fat/protein meals compared to low-fat/protein with identical CHO content. In [8], postprandial glycemic patterns of a low glycemic load (LG) and a high glycemic load (HG) meal, matched in CHOs (121 g), were analyzed using tracer experiments to estimate glucose absorption from the meal. LG resulted in a slower and sustained elevation of plasma glucose with an unpronounced peak within  $153 \pm 104$  min, whereas HG showed a clearer and earlier peak at  $98 \pm 29$  min. Meal absorption of both meals presented different patterns, with absorption of LG continuing beyond the 8-h duration of the study and the one of HG being completed within 6 hours (see Fig. 3.2). The LG meal resulted in at least 15% higher apparent bioavailability of CHOs. In [136] the effect of consuming white rice versus pasta with high or low protein, matched in the CHO contents, was studied, showing higher postprandial peak and area under the curve of rice than pasta. No statistical difference between the two kinds of pasta was found.

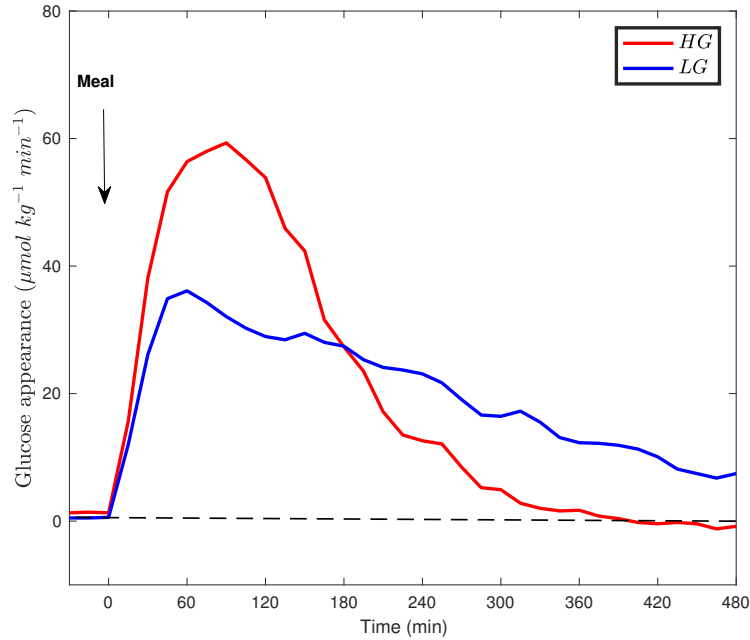


Figure 3.2: Average meal rate of glucose appearance of a low glycemic load (LG) and a high glycemic load (HG) meal matched in carbohydrates. Adapted from [8].

### 3.1.3 Exercise response

Another challenge to diabetes management, including the AP, is exercise [9]. Regular physical activity is beneficial to health. However, it is also a source of glucose fluctuations which will change depending on the type of exercise, intensity and duration. Fig. 3.3 represents different typical exercise modalities and their glycemic impact in individuals with and without T1D. In healthy individuals, a reduction of insulin secretion happens at the onset of exercise, which together with an increase of counterregulatory hormones balance the increased glucose disposal. Only in vigorous intensity aerobic exercise and brief high intensity anaerobic exercise (like a sprint), a rise in glucose will happen which is later corrected by an increase in insulin secretion. However, these mechanisms are not possible in T1D. Moderate intensity aerobic exercise will tend to produce hypoglycemia due to the impossibility of reducing portal insulin concentration at the onset of exercise and the acceleration of insulin absorption provoked by exercise, among other factors. High variability and even unpredictability is observed in the case of T1D.

Of especial relevance is moderate aerobic exercise, due to the increased risk of hypoglycemia. Indeed, in [137], a drop of 68 mg/dL is reported in a 40-min continuous treadmill walking exercise session, despite an insulin

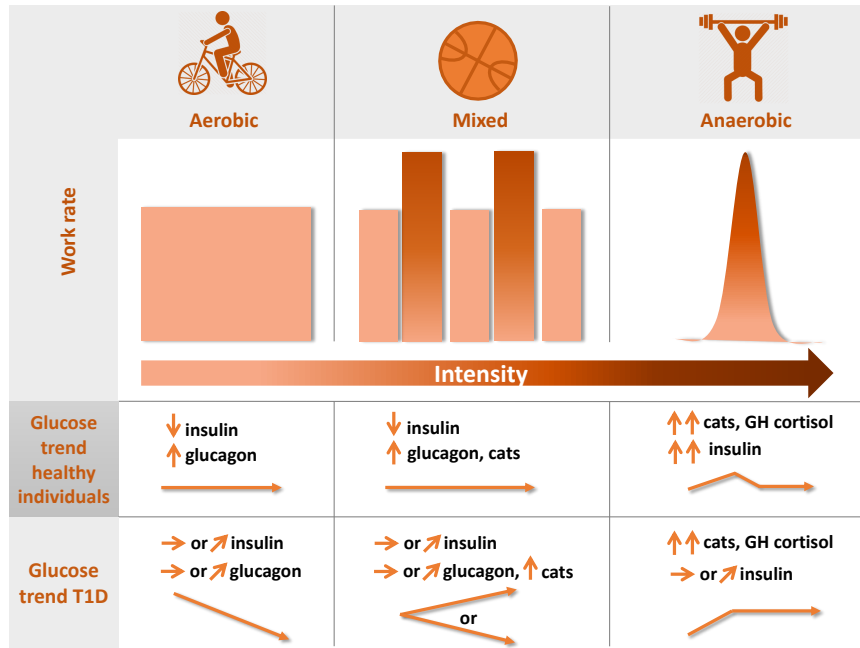


Figure 3.3: Blood glucose and typical hormonal responses to different types and intensities of exercise in individuals with and without T1D. “Cats” stands for “catecholamines”. Modified from [9]

pump shut off at the start of exercise, which is the most aggressive action a single hormone AP can do for an unannounced exercise. Such an impact would result in hypoglycemia unless the patient has a high enough glucose concentration prior to exercise.

In [9], different hurdles are identified related to exercise and the development of the AP, of which the following can be extrapolated to glucose prediction:

- Exercise comes in a variety of forms and intensities, yielding to highly variable (from 2 to 10-fold) increase of BG turnover rates during exercise.
- Exercise effects change from person to person and from day to day.

In order to cope with this high variability, additional signals sensing exercise might be necessary for a more accurate glucose prediction. In [138], an analysis of relevant signals from a wearable device (SenseWear<sup>©</sup> physical activity monitor, Body Media, Pittsburgh, PA, USA) for a variety of exercise types is investigated from 26 clinical experiments. Seven exercise types are tested: exercise stress test, treadmill exercise, treadmill exercise-interval training, bike exercise, weight lifting, exercise with workout video, maximal resistance training and submaximal resistance training. Measurement of



skin temperature is found to be relevant in 6 out of the 7 types of exercises, followed by energy expenditure and heart rate. Galvanic skin response was found to be the most important variable during treadmill exercise-interval training and maximal resistance training sessions.

Despite all these challenges, the most important feature that must be available in an AP system (any diabetes management in general) is the ability to minimize hypoglycemic episodes, accurately determine a correct amount of insulin according to the different meals intake, ability to generate alerts before BG problems, which must be adjustable for different clinical practices, and usability [75, 139]. Thus, the development of models that are effective enough to characterize key physiological aspects in T1D is needed [140]. An effective prediction model should capture the dynamic characteristics of the patient’s behavior in face of insulin (and glucagon when applicable) delivery and external disturbances such as meals and exercise, at least, which determine the inputs that may be available from devices or information provided manually by patients (the so-called “patient’s announcement”).

### 3.2 Categories of prediction models

Prediction models can be classified into three categories (see Fig. 3.4): *physiological models*, *data-driven models*, and *hybrid models* [124, 141], which are described next.

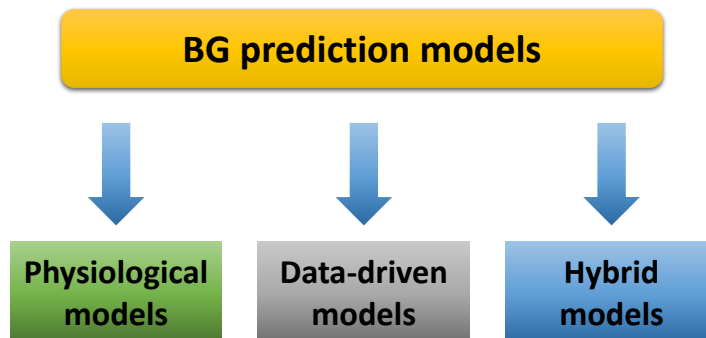


Figure 3.4: Categories of prediction models.

#### Physiological models

Physiological models are mathematical dynamic models that approximate the behavior of an actual physiological system. Therefore, these models require *a priori* knowledge on physiology, in our case the regulatory mechanisms of glucose and other metabolites. These models try to capture the

dynamics of glucose-relevant variables within various systems in the body (*e.g.*, the process of converting CHO intake to glucose in the blood and then its interaction with many organs within the body such as liver and muscle). They are usually compartmental models as shown in Fig. 3.5, since they are generated by dividing the body into homogeneous compartments representing the distribution of glucose and insulin in various organs [142]. A glucose-insulin physiological model able to deal with main events in a patient's daily life necessarily needs sub-models describing *CHO digestion and absorption*, *subcutaneous insulin absorption* and *exercise* [143]. Expected inputs to these models include an estimation of CHO intake, insulin delivered by the insulin pump, and exercise information, such as signals from a wearable if available or manual descriptive information from the patient.

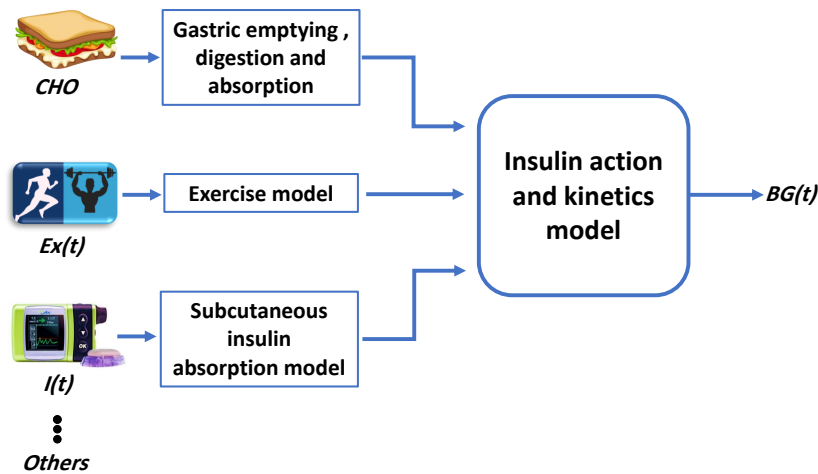


Figure 3.5: Block diagram of a glucose-insulin physiological model.

These models can be classified into two types according to their complexity. The first type is *minimal models* characterized by the few number of parameters and equations that describe the main components of the system's functionality, focused on representing the crucial processes of glucose metabolism and insulin action. The second type is *maximal or comprehensive models*, which include all the available physiological knowledge about the system's functionality. However they suffer from an excess of parameters raising indentifiability issues if model personalization is pursued. These models are used to build simulators for *in silico* experiments to assess controllers and treatment strategies, demonstrating matching between *in silico* and clinical results at a *population level* [69].

### Data-driven models

Data-driven models, also referred to as empirical dynamic models, are black box models developed from input/output data without considering any physiological knowledge (Fig. 3.6). Therefore, there is not the need to know how the physiological mechanisms led to the emergence of input/output behavior. Despite these models cannot be used to explain the mechanisms of the glucose-insulin regulatory system, they can provide good predictions about BG future dynamic behavior [141].

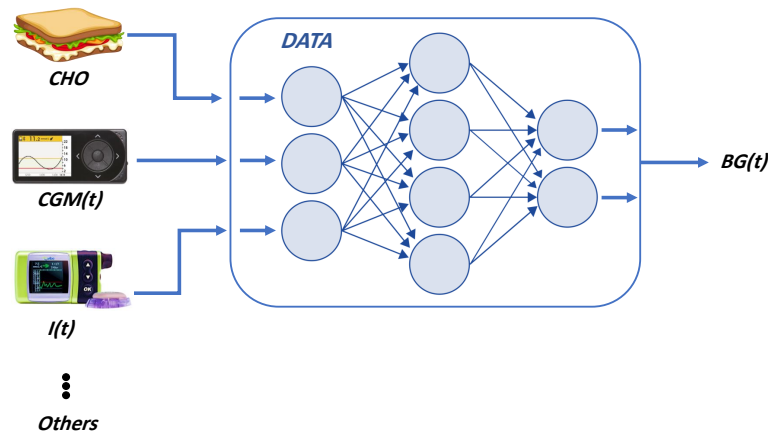


Figure 3.6: Block diagram of an artificial neural network as an example of data-driven model.

Usually, the output of these models is the future BG values, while the input data can be a diversity of information that can influence the BG levels such as previously measured BG values, food composition (*e.g.*, amount of CHOs), exercise, insulin therapy, body temperature, heart rate, energy expenditure, etc. Furthermore, these models are often supported by a big family of techniques such as time series models, fuzzy logic models, support vector models, and artificial neural network (ANN) models, among others [141].

Data-driven models are characterized by many advantages such as [141]:

- they need minimal user input,
- they can be developed in a short time,
- easy implementation of controllers,

which make them more attractive than physiological models for our purpose.

## Hybrid models

Hybrid models typically use physiological models to provide inputs for a data-driven model to predict the BG value or to indicate BG problems like a risk of hypoglycemia as shown in Fig. 3.7. The data-driven model component in these models captures the relationship between the physiological model’s output and future BG outcomes [124]. It is expected that integration of partial physiological knowledge will lead to increased prediction accuracy of the overall model, although this would depend on the quality of the fitting of such model components, whose outputs can be unmeasurable such as glucose absorption rate from a meal model, for instance, unless complex tracer experiments are carried out.

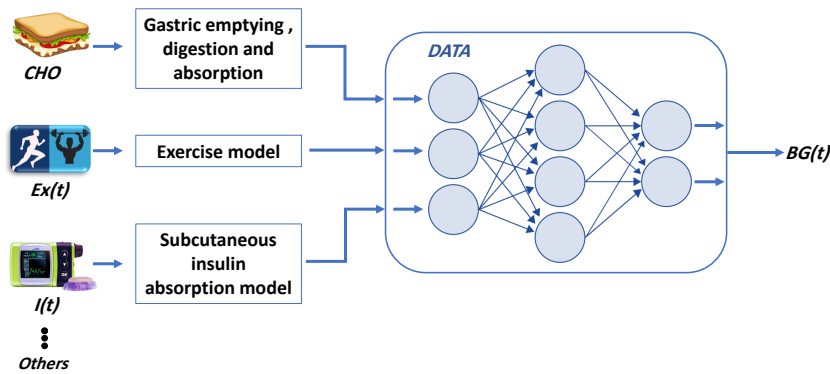


Figure 3.7: Block diagram of a hybrid model.

A review of how the different approaches have been used for prediction and estimation in the context of the AP follows. It is out of the scope of this review to present all models reported in the literature, which is vast. Good reviews in this sense are the ones in [69, 141, 124], and the reader is referred to them and references therein.

## 3.3 Physiological models

### 3.3.1 Minimal model based approaches

Bergman minimal model originated in 1979 [144] to describe the intravenous glucose tolerance test (IVGTT), which aims at analyzing how the body metabolizes glucose. Glucose is infused intravenously and plasma glucose and insulin concentration are measured before and after such infusion. Then, the model is fitted to data to get key parameters such as insulin sensitivity. As a model “to measure”, no subcutaneous insulin absorption or meal ingestion

is considered. Taking this into consideration, the model is described as

$$\dot{X}(t) = -p_2X(t) + p_3(I(t) - I_b) \quad (3.1)$$

$$\dot{G}(t) = -(p_1 + X(t))G(t) + p_1G_b + \frac{U_G(t)}{V_G} \quad (3.2)$$

It consists in a second order non-linear model relating plasma insulin concentration  $I(t)$  (input) and plasma glucose concentration  $G(t)$  (output). Insulin effect,  $X(t)$ , is modeled with a first-order system representing a lagged action of insulin. Plasma glucose concentration,  $G(t)$ , is inhibited by glucose itself and insulin effect. A constant hepatic glucose production  $p_1G_b$  drives the system to equilibrium for a basal insulin concentration  $I_b$ . Finally, the term  $U_G(t)/V_G$  describes the intravenous glucose infusion (or glucose rate of appearance from the meal when a meal model is interconnected). Parameters  $p_1$ ,  $p_2$  and  $p_3$  are kinetic parameters. Parameter  $p_1$  is called “glucose effectiveness” and it represents glucose auto-inhibition. Insulin sensitivity is computed as  $p_3/p_2$ .

Due to its simplicity and identifiability properties, Bergman model (or equivalent model formulations) is widely used outside its original scope, in combination with a subcutaneous insulin pharmacokinetic and a CHO absorption model. An example is the so-called “Identifiable Virtual Patient” (IVP) model [145], with equations

$$R_A(t) = \frac{D_G A_G t e^{-t/t_{maxG}}}{V_G t_{maxG}^2} \quad (3.3)$$

$$\dot{I}_{SC}(t) = -\frac{I_{SC}(t)}{\tau_1} + \frac{u_{SC}(t)}{\tau_1 K_{cl}} \quad (3.4)$$

$$\dot{I}(t) = -\frac{I(t)}{\tau_2} + \frac{I_{SC}(t)}{\tau_2} \quad (3.5)$$

$$\dot{I}_E(t) = -p_2 I_E(t) + p_2 S_I I(t) \quad (3.6)$$

$$\dot{G}(t) = -(GEZI + I_E(t))G(t) + EGP + R_A(t) \quad (3.7)$$

CHO absorption is described as the explicit impulse response of a second order linear system with double pole at  $-1/t_{maxG}$ , to the impulse input  $D_G/V_G\delta(t)$ , where  $D_G$  is the CHO intake and  $V_G$  is the glucose volume of distribution. Thus,  $R_A(t)$  represents a concentration rate, instead of a mass rate. Parameter  $A_G$  is the CHO bioavailability, that is, the fraction of CHO in the meal that will finally reach blood. The mass-rate version of this model originates in the Hovorka model, which is described in the next Section. Insulin action and glucose metabolism are equivalent to the Bergman model, with the difference that plasma insulin  $I(t)$  is considered, instead of its deviation with respect to basal insulin  $I_b$ . State  $I_E(t)$  and parameter  $GEZI$  (Glucose Effectiveness extrapolated to Zero Insulin) are thus

equivalent to  $X(t)$  and  $p_1$ , respectively, with the change in its meaning due to the above difference. Parameter  $EGP$  is equivalent to  $p_1G_b$  and describes the hepatic glucose production. The subcutaneous insulin pharmacokinetic model consists in a second order linear model with time constants  $\tau_1$  and  $\tau_2$ . Parameter  $K_{cl}$  is insulin clearance, that is, how much volume of plasma is cleared of insulin per unit time.

Several extensions of the Bergman model have appeared in literature including the effects of free fatty acids (FFA) [146], exercise [147], counterregulatory response [148] and glucagon PK/PD in the context of dual-hormone AP systems [149].

Regarding its use in the AP domain, *minimal models* have been used in MPC formulations and state and disturbance observers. In [150], an MPC controller is designed based on the Bergman minimal model as part of parametric programming algorithms. The generation of an explicit relationship between the constrained optimal control profile and the current BG concentration is addressed. As well, in [151], an MPC controller is designed based on the Bergman minimal model, which is tested in simulation using a 19-state nonlinear model as virtual patient (Sorensen model [152]). In this study, a meal disturbance was applied in order to test the system's disturbance rejection performance. The authors demonstrate the efficiency of the estimation-based MPC strategy, despite the plant-model mismatch.

In [153], the Bergman minimal model is used in combination with different filters (symmetric Unscented Kalman Filter, Extended Kalman Filter, simplex Unscented Kalman Filter, and the Particle Filter) in order to estimate plasma insulin concentration. Endogenous insulin secretion was also considered since data from non-diabetic individuals was used. Lie derivatives [154] were used for evaluating observability and linear interpolation was applied to get glucose measurements in a time grid of 1 min. The designed observers were validated against IVGTT data in non-diabetic individuals using RMSE as performance metrics. The symmetric Unscented Kalman Filter exhibited better results than the Extended Kalman Filter, with an RMSE value of 10.277 mU/L versus 13.533 mU/L, respectively.

A Luenberger observer [155] is designed in [156] using the Bergman minimal model for the estimation of remote insulin and plasma insulin compartments. Disturbances were represented in the form of an exogenous glucose infusion rate. Results were tested in closed-loop simulations with a linear-quadratic (LQ) controller.

In [157], the IVP model (*i.e.*, extended minimal model with meal absorption and subcutaneous insulin infusion) is used as an initial white-box model for the formulation of stochastic differential equation-based grey-box models (SDE-GBs) with diffusion. Clinical data from 4 T1D patients were used for parameter estimation. In addition, the likelihood-ratio test was used

for determining the optimal SDE-GB. Variation in the “*time to peak meal response*” parameter was tracked. The results obtained demonstrated that the use of SDE-GB instead of ordinary differential equations (ODEs) can provide a meaningful improvement in uncorrelated errors and prediction.

In [158], a modified version of the Bergman minimal model is used in the context of a meal detection algorithm. An Unscented Kalman Filter was developed for state and parameter estimation. The estimated rate of appearance of glucose was used to detect a meal and trigger meal boluses. This algorithm (IMA-APv2) was tested in both simulated and clinical environments, with 14 experiments on 7 T1D patients. The authors successfully demonstrated the ability of this algorithm to decrease the number and intensity of hypoglycemia in individuals with T1D without any intervention or manual input from individuals.

### 3.3.2 Medium complexity models

The most representative medium complexity model in literature is Hovorka model [115, 159], which consists in a 7-*th* order non-linear model given by the equations

$$U_G(t) = \frac{D_G A_G t e^{-t/t_{maxG}}}{t_{maxG}^2} \quad (3.8)$$

$$\dot{S}_1(t) = u_{SC}(t) - \frac{S_1(t)}{t_{maxI}} \quad (3.9)$$

$$\dot{S}_2(t) = \frac{S_1(t)}{t_{maxI}} - \frac{S_2(t)}{t_{maxI}} \quad (3.10)$$

$$\dot{I}(t) = \frac{S_2(t)}{t_{maxI} V_I} - k_e I(t) \quad (3.11)$$

$$\dot{X}_1(t) = -k_{a1} X_1(t) + k_{b1} I(t) \quad (3.12)$$

$$\dot{X}_2(t) = -k_{a2} X_2(t) + k_{b2} I(t) \quad (3.13)$$

$$\dot{X}_3(t) = -k_{a3} X_3(t) + k_{b3} I(t) \quad (3.14)$$

$$\begin{aligned} \dot{Q}_1(t) = & -X_1 Q_1(t) + k_{12} Q_2(t) - F_{01}^c(t) - F_R(t) \\ & + U_G(t) + EGP_0(1 - X_3(t)) \end{aligned} \quad (3.15)$$

$$\dot{Q}_2(t) = X_1 Q_1(t) - k_{12} Q_2(t) - X_2(t) Q_2(t) \quad (3.16)$$

$$G(t) = \frac{Q_1(t)}{V_G} \quad (3.17)$$

CHO digestion and absorption are described by the impulse response of a second order linear model, with input  $D_G$  the amount of CHO in the meal, and output the glucose rate of appearance in blood,  $U_G(t)$ . Subcutaneous insulin pharmacokinetics is a 3-*rd* linear model, with  $u(t)$  the insulin infusion from the pump and states  $S_1(t)$  and  $S_2(t)$  representing mass of insulin along subcutaneous tissue, with transport time constant  $t_{maxI}$ . Insulin rate of

appearance in blood is  $S_2(t)/t_{maxI}$ , that, once converted to concentration rate by dividing by the insulin distribution volume  $V_I$ , enters the first-order kinetic equation defining plasma insulin concentration,  $I(t)$ . Parameter  $k_e$  is the fractional elimination rate. Plasma insulin exerts three actions, following a first-order dynamics (representing a “remote” action):  $X_1(t)$  is the effect of insulin on glucose distribution/transport,  $X_2(t)$  the effect on glucose disposal (it promotes glucose entering muscle and adipose tissue cells) and  $X_3(t)$  the effect endogenous glucose production (it inhibits glucose production by the liver). Constants  $k_{ai}$ ,  $i = 1, 2, 3$ , are kinetic constants and  $k_{bi}/k_{ai}$  are the insulin sensitivities for each effect. Finally, states  $Q_1(t)$  and  $Q_2(t)$  represent the glucose mass in plasma and interstitial tissue, respectively. Transport of glucose between both compartments happens at rates  $k_{12}$  and  $X_1(t)$ , depending on the direction. Besides these transport flows, glucose from the meal,  $U_G(t)$ , and hepatic glucose production  $EGP_0(1 - X_3(t))$ , with  $EGP_0$  the production extrapolated to zero insulin, are input flows to the plasma compartment; insulin-independent consumption (mainly by brain),  $F_{01}^c(t)$ , and renal glucose elimination,  $F_R(t)$  are output flows. For the sake of simplicity, the reader is referred to [115] for the definition of  $F_{01}^c(t)$  and  $F_R(t)$ . Regarding the interstitial compartment, aside the inter-compartment transport, glucose uptake by muscle and adipose tissue,  $X_2(t)Q_2(t)$  is an output flow. Plasma glucose concentration (model output) is finally given by  $Q_1(t)/V_G$ , where  $V_G$  is the glucose distribution volume.

Hovorka model is the basis of the AP system developed by the Hovorka’s team at University of Cambridge (Florence system), based on a non linear MPC controller. This system has been extensively evaluated in clinical trials [160]. The system includes a Kalman filter to update two model parameters: (1) a glucose flux quantifying model misspecification; and (2) CHO bioavailability [161]. Competing models in a stochastic framework to get the best possible explanation of sensor data are used, with models differing in the rate of subcutaneous insulin absorption and action, as well as CHO absorption.

As well, in [162], a Hovorka model based Extended Kalman Filter is built for real-time estimation of plasma insulin concentration from CGM measurements. Estimation of uncertain pharmacokinetic parameters (*i.e.*, the fractional elimination rate from plasma,  $k_e$ , and time-to-maximum insulin absorption,  $t_{maxI}$ ) is also considered, with an state extension. An *in silico* study comprising 100 T1D subjects was carried out for evaluation, as well as clinical data from 12 T1D subjects who underwent four mixed meal studies each. In this case, actual plasma insulin measurements were used to analyze the observer performance. The obtained results demonstrated that this approach may be helpful to increase the efficiency and safety of control algorithms for the AP system, as compared to using a population model to compute, for instance, insulin-on-board.



In line with the previous work, in [163], the Hovorka model is used in combination with different estimation techniques (continuous-discrete Extended Kalman Filter, Unscented Kalman Filter and Moving Horizon Estimator) in order to perform a real-time estimation of plasma insulin concentration. Moreover, the effect of meal intake,  $U_G$ , is considered as a time-varying model parameter and also estimated, together with  $k_e$  and  $t_{maxI}$ , as in [162], to cope with intra-individual variability. These three parameters are included as extended states in the Hovorka model. Partial least-squares (PLS) models are used based on the demographic information (*e.g.*, height, weight, HbA1c, and total daily insulin dose) of each subject to determine suitable initial values for these unknown model parameters.

In [107], the Hovorka model is used as an *in silico* subject in order to develop and test a closed-loop AP system based on an MPC controller. A new procedure for “human-friendly” identification is proposed by using multisine inputs to estimate appropriate model parameters.

In [143], modal interval analysis [164] is applied for computing tight envelopes of glucose concentration for the prediction of postprandial glucose under parametric uncertainty using the Hovorka model. This approach was compared with Monte Carlo simulations, demonstrating a guarantee of the inclusion of the actual response into the computed envelopes, with a much less computational effort. However, only some specific parameters of the Hovorka model were considered uncertain in this study (*i.e.*, insulin sensitivities, glucose absorption, and food intake). In contrast, in [165], monotone systems techniques are used to compute glucose concentration envelopes of the Hovorka model considering uncertainty in all the parameters and initial conditions.

### 3.3.3 Maximal models

The most representative maximal model in AP research is Dalla Man model [166], developed originally for healthy subjects and later adapted to T1D [167]. This model is the core of the UVA-Padova T1D simulator [168, 133] accepted by the FDA in USA as a substitute of animal trials. This simulator is extensively used in the field for *in silico* studies of new controllers. Along time, the model in [166, 167] has been refined incorporating new features like nonlinearities happening during hypoglycemia and glucagon PK/PD [168], and later a better representation of insulin sensitivity intra-patient variability and intradermal and inhaled insulin absorption [89, 133]. Here, for the sake of brevity, only the model in [166, 167] will be described in detail, as well because only in its simplest version this model has been used as part of MPC AP systems. Differences with Hovorka model will be highlighted.

Dalla Man model in [166, 167] is described by the following equations:

$$Q_{sto}(t) = Q_{sto1}(t) + Q_{sto2}(t) \quad (3.18)$$

$$\dot{Q}_{sto1}(t) = -k_{gri}Q_{sto1}(t) + D\delta(t) \quad (3.19)$$

$$\dot{Q}_{sto2}(t) = -k_{empt}(Q_{sto}(t))Q_{sto2}(t) + k_{gri}Q_{sto1}(t) \quad (3.20)$$

$$\dot{Q}_{gut}(t) = -k_{abs}Q_{gut}(t) + k_{empt}(Q_{sto}(t))Q_{sto2}(t) \quad (3.21)$$

$$R_a(t) = \frac{fk_{abs}Q_{gut}(t)}{BW} \quad (3.22)$$

$$\dot{I}_{sc1}(t) = -(k_d + k_{a1})I_{sc1}(t) + IIR(t) \quad (3.23)$$

$$\dot{I}_{sc2}(t) = k_d I_{sc1}(t) - k_{a2}I_{sc2}(t) \quad (3.24)$$

$$R_{ai}(t) = k_{a1}I_{sc1}(t) + k_{a2}I_{sc2}(t) \quad (3.25)$$

$$\dot{I}_l(t) = -(m_1 + m_3)I_l(t) + m_2I_p(t) \quad (3.26)$$

$$\dot{I}_p(t) = -(m_2 + m_4)I_p(t) + m_1I_l(t) + R_{ai}(t) \quad (3.27)$$

$$I(t) = \frac{I_p(t)}{V_I} \quad (3.28)$$

$$\dot{I}_1(t) = -k_i(I_1(t) - I(t)) \quad (3.29)$$

$$\dot{I}_d(t) = -k_i(I_d(t) - I_1(t)) \quad (3.30)$$

$$EGP(t) = k_{p1} - k_{p2}G_p(t) - k_{p3}I_d(t) \quad (3.31)$$

$$\dot{X}(t) = -p_{2U}X(t) + p_{2U}(I(t) - I_b) \quad (3.32)$$

$$U_{id}(t) = \frac{(V_{m0} + V_{mx}X(t))G_t(t)}{(K_{m0} + K_{mx}X(t)) + G_t(t)} \quad (3.33)$$

$$\dot{G}_p(t) = EGP(t) + R_a(t) - U_{ii}(t) - E(t) - k_1G_p(t) + k_2G_t(t)$$

$$\dot{G}_t(t) = -U_{id}(t) + k_1G_p(t) - k_2G_t(t) \quad (3.34)$$

$$G(t) = \frac{G_p(t)}{V_G} \quad (3.35)$$

Meal absorption is represented by state variables  $Q_{sto1}(t)$ ,  $Q_{sto2}(t)$  and  $Q_{gut}(t)$ , denoting liquid and solid phase in the stomach and gut compartment, respectively. The gastric emptying rate is represented by the nonlinear function  $k_{empt}(\cdot)$  (see [167] for its explicit expression), which intends to mimic the impact of fat in gastric emptying. As compared to Hovorka model, equation (3.8), now meal absorption is nonlinear. Subcutaneous insulin absorption and kinetics also differ. In Dalla Man model subcutaneous insulin absorption is represented by state variables  $I_{sc1}(t)$  and  $I_{sc2}(t)$ , with  $IIR(t)$  the exogenous insulin infusion rate. As compared to Hovorka model, equations (3.9)-(3.10), now insulin from both compartments reach the plasma,  $I_p(t)$ , at a rate  $R_{ai}(t)$ . Dalla Man model also represents insulin at the liver,  $I_l(t)$ , where insulin hepatic extraction happens. The rest of equations in Dalla Man model represent glucose kinetics. States  $G_p(t)$  and  $G_t(t)$  are equivalent to  $Q_1(t)$  and  $Q_2(t)$  in Hovorka model. The main difference observed is that, whereas in Hovorka model three effects of insulin

are considered (glucose transport, hepatic production and peripheral glucose uptake), in Dalla Man model no effect of insulin on glucose transport among plasma and peripheral tissues is considered. Effect on the inhibition of hepatic glucose production is represented by  $EGP(t)$ , where insulin acts with a lag represented by the two-compartmental system given by  $I_1(t)$  and  $I_d(t)$ . Effect of insulin on the promotion of peripheral glucose uptake is represented by  $U_{id}(t)$ , which in this case follows a Michaelis-Menten dynamics with a lag represented by  $X(t)$ . Variables  $U_{ii}(t)$  and  $E(t)$  represent insulin-independent glucose uptake (which is constant) and renal glucose excretion, which is modeled similarly to Hovorka model (the reader is referred to [166] for its expression).

Models of such complexity suffer from identifiability issues which makes model personalization a hurdle. Despite this, an MPC AP system based on a linearization of the Dalla Man model is proposed in [116]. A mean linearized model is used, and a single aggressiveness factor in the cost index is used for individualization [169]. A Kalman filter is also used for state estimation using a linear model. In [170], the authors propose a model individualization based on average linearized models for four virtual subpopulations classified according to insulin-to-CHO ratios ranges. None of these publications report information on the quality of the predictions in the MPC controller with clinical data.

### 3.4 Data-driven models

Recent studies have considered an assortment of linear Empirical Dynamic Models (EDM) identified from *in silico* and, more significantly, from clinical data, which is our interest. Linear EDM include AR, ARMA, ARX, and ARMAX models, among others (see Chapter 4, Section 4.2). These models may be more attractive than physiological models for capturing the glucose dynamic behavior due to the advantages stated in Section 3.2.

#### 3.4.1 AR and ARMA models

The AR and ARMA models without exogenous variables assume that future BG concentration can be expressed as a linear function of previous BG measurements without any prior information (*e.g.*, meal or insulin information). Reifman *et al.* [171] proposed a time-invariant (*i.e.*, constant parameter) AR model of order 10 (*i.e.*, AR(10)). The model was identified on CGM data of nine T1D subjects collected over a continuous 5-day period, with a sampling time of 1 min. The model parameters were optimized using regularized least squares (LS) and the quality of predictions was assessed through: (1) the RMSE and (2) the continuous glucose error grid analysis (CG-EGA) (see Appendix A), considering PH until 120 min. Both subject-specific and

cross-subject models were evaluated obtaining comparable results, for example, for a 30-min PH, the CG-EGA for subject-specific models yield 97.6 to 100.0% of data in the clinically acceptable zones A and B, with an average RMSE of 22.3 mg/dL, whereas the CG-EGA for cross-subject models yield 95.8 to 99.7% of data in zones A and B, with an average RMSE of 24.8 mg/dL.

Gani *et al.* [172] identified 30th-order AR models, AR(30), from CGM data with 1-min sampling time from 9 T1D subjects for approximately 5 consecutive days. Three cases were considered: (a) *scenario 1*, in which raw glucose data were used to obtain unregularized AR coefficients; (b) *scenario 2*, in which glucose data were smoothed before computing unregularized AR coefficients; and (c) *scenario 3*, in which smoothing and regularization were used to generate regularized AR coefficients. The model parameters were determined by using least squares and the models were assessed on different PHs in terms of RMSE and time lags. Only scenario 3 guaranteed accurate predictions and a clinically acceptable time lag until 60-min PH, *i.e.*, an average RMSE of 12.6 mg/dL was reported for a 60-min PH and average lag of 12.3 min. The first two scenarios displayed undesirable accuracy and large time lags.

Sparacino and colleagues [173] identified a 1st-order time-varying AR model based on data from 28 T1D subjects wearing a microdialysis system with a 3-min sampling time for 48 consecutive hours, in normal daily life conditions. The model parameters were estimated at each time step using recursive least squares (RLS). They demonstrated the feasibility of predicting hypoglycemic events 20-25 min ahead in time, considering a 30-min PH. A median RMSE ranging from 18.33 to 20.32 mg/dL, depending on the selection of a forgetting factor, was reported for that PH.

Low-order AR and ARMA models were considered by Eren-Oruklu and associates [174] considering PHs up to 30 minutes. Two different databases were used: one formed by 22 healthy hospitalized individuals, and the other one formed by 14 T2D subjects in free daily life conditions. Both databases consist of glucose concentration data were monitored with the Gold<sup>TM</sup> CGMS<sup>®</sup> system, with a 5-min sampling time, for 48 consecutive hours. The model parameters were estimated using weighted RLS. A sum of squares of glucose prediction error (SSGPE) ranging between 10.32 and 12.55 mg/dL was reported, depending on the study, for a 30-min PH.

### 3.4.2 ARX and ARMAX models

Extension of AR and ARMA models is widely used in order to exploit CGM and other available information like insulin infusion and other input signals (exogenous variables), leading to ARX and ARMAX models as introduced in Section 4.2. This allows to better capture the genuine patterns in CGM

behavior and improve the prediction accuracy and PH.

Finan and colleagues [175, 176] evaluated ARX models from simulated and clinical ambulatory data from 9 T1D, with 5-min sampling time. Each dataset spanned 2 to 8 days. Insulin pump information and CHO quantity of meals were used as exogenous variables in the models. The authors concluded that 60 minutes was a maximum achievable PH in terms of model prediction accuracy. An average RMSE of 26, 34, and 40 mg/dL was reported for 30-, 45-, and 60-min PH, respectively [177]. This corresponds to an improvement of 9%, 5%, and 5%, respectively, compared to a zero-order-hold (ZOH) predictor. Castillo *et al.* [178] identified a time-varying ARX model based on data from 15 T1D subjects. A CGM data was monitored using a FreeStyle Navigator CGM device with a 1-min sampling time, for 76 hour. Meal and insulin information were used as exogenous variables in the model. The model parameters were estimated recursively using a normalized least mean squares (NLMS) algorithm and exploiting a physiological gain adaptation rule. The quality of predictions was assessed by the CG-EGA in different glycemic ranges, considering PH up to 50 min.

Eren-Oruklu and associates [179] identified a time-varying multivariate ARMAX model, with exogenous variables including meal intake, lifestyle information, and PA with seven signals to describe the activity of all subjects and emotional conditions: energy expenditure (EE), average longitudinal acceleration (LA), transverse acceleration peaks, transverse acceleration mean of absolute difference, near-body temperature (BT), heat flux (HF) and galvanic skin response (GSR). The model based on data from 5 T2D subjects under free-living conditions. Glucose was monitored with the MMT-7012 Medtronic CGM sensor for about 24 days, with 5-min sampling time. The physiological signals were measured with the body monitoring system SenseWear Pro3 (BodyMedia Inc., Pittsburgh, PA). The model parameters were identified online using the weighted RLS method with a change detection strategy in case of severe glycemic disturbance. A mean absolute percentage error (MAPE) of 4.24 % was reported for 30-min PH. Furthermore, the authors also preliminarily evaluated the ability of the ARMAX model of detecting hypoglycemia.

Turksoy and colleagues [180] proposed an ARMAX model with an alarm algorithm for prediction of hypoglycemia and prevention of its occurrence, with exogenous inputs including IOB, and physical activity information (*i.e.*, EE, GSR). The model and the alarm algorithm were tested on data of 14 T1D subjects. Glucose concentration was monitored with the iPro CGM device, with 5-min sampling time, while the physiological signals were monitored with the SenseWear Pro3 armband system. A Savitzky-Golay filter and a Kalman filter were used to reduce noise in the dataset. An RMSE of 1.86, 7.18, 18.55, and 48.94 mg/dL was reported for 5-, 15-, 30-, and 60-min PH, respectively. The prediction error was decreased significantly,

when physical activity information was added to the model as exogenous inputs. Furthermore, the ARMAX model exhibited the ability to predict the hypoglycemic episodes successfully with satisfactory results.

A variety of linear and nonlinear time-series models were evaluated by Ståhl and Johansson [181] from clinical data from one subject, with non-uniform and sparse sampling (fingerstick measurements) with spline interpolation, in order to produce a short-term BG predictors for up to two-hour-ahead BG prediction. However, many difficulties were met not achieving the required accuracy.

### 3.4.3 Other empirical models

On the other side, another popular approach in the empirical modeling for glucose prediction is the application of artificial intelligence and machine learning techniques [182] such as ANN models and support vector regression (SVR) models, among many others.

In [183], a latent variable (LV)-based statistical analysis was proposed to model glucose dynamics and predict BG concentrations for T1D subjects. The authors identified the proposed model based on simulated data from 10 T1D virtual subjects, in an scenario of 7 days with a sampling time of 5 min, and also based on clinical data from 7 T1D subjects. In this case, CGM data were monitored with the Dexcom SEVEN PLUS device (5-min sampling time). CGM, meal and insulin information were used as inputs to the model. In this study, the proposed model (LVX) was compared with time-invariant AR and ARX models and also with a latent variable based only on CGM data. The RMSE and CG-EGA were used for evaluating the prediction accuracy. On simulated and clinical data, the LVX model proved to be superior to other reference methods. An average RMSE of 8.6, and 14.0 mg/dL was reported for 30-, and 60-min PH, respectively in *silico* subjects. As well, an average RMSE of 11.1, 18.7, 24.4, and 29.2 mg/dL was reported for 15-, 30-, 45-, and 60-min PH, respectively in *clinical* subjects.

In [184], an ANN model was used in parallel with a first-order polynomial extrapolation algorithm to describe the nonlinear and the linear components of glucose dynamics. The proposed model was evaluated on 15 T1D subjects for short-time glucose prediction. CGM data, the glucose rate of appearance after a meal, and the information of CHO intake were used in the proposed model. The obtained results have shown satisfying prediction errors, with an RMSE of 14.0 mg/dL for 30-min PH.

In [185], an ANN model was applied for online predicting future BG levels from CGM data. The evaluation of the proposed prediction algorithm was performed using data from the Guardian Real-Time and the FreeStyle Navigator CGM Systems. The Guardian dataset includes data from nine T1D subjects, while six T1D subjects were monitored using the FreeStyle

Navigator. The results obtained have shown that the prediction accuracy is better with Guardian than the FreeStyle Navigator. An average RMSE of 9.7, 17.5, and 25.1 mg/dL was reported for 15-, 30-, and 45-min PH, respectively for the Guardian system. An average RMSE of 10.4, 19.5, and 29.1 mg/dL was reported for 15-, 30-, and 45-min PH, respectively for the FreeStyle Navigator system. In [186], nonlinear regression was performed in a reproducing kernel Hilbert space through the approximate linear dependency kernel recursive least-squares (KRLS-ALD) algorithm in order to achieve a sparse model structure. Several feature sets were used such as CGM data, CHO intake estimates, insulin regime, and physical activity. Fifteen T1D subjects were used to evaluate the proposed glucose prediction model under free-living conditions. An RMSE of 4.0, 10.5, 18.8, 25.8, and 31.8 mg/dL was reported for 5-, 15-, 30-, 45-, and 60-min PH, respectively.

In [187], several prediction algorithms (AR, extreme learning machine (ELM), and SVR methods) were combined by using adaptive weights inversely proportional to the prediction error of each model used. Ten T1D subjects were used for testing the methodology. Subjects were monitored either with the Dexcom SEVEN PLUS or with a Medtronic CGM device (both with 5-min sampling time). The RMSE, the relative error, the EGA and the J index [188] were used to assess the prediction accuracy. The combined model performed better than the individual algorithms. For instance, the RMSE of the AR, ELM, and SVR methods and the proposed combined model were 26.1, 20.5, 50.9 and 19.0, respectively for 30-min PH.

The deep learning network also has been applied to the glucose prediction problem. In [189], a consecutive model with one long-short-term memory (LSTM) layer, one bidirectional LSTM layer, and many perfectly connected layers was applied for predicting future BG levels. The proposed approach was identified on CGM data of 20 T1D subjects. An RMSE of 11.63, 21.73, 30.22, and 36.92 mg/dL was reported for 15-, 30-, 45-, and 60-min PH, respectively.

In [190], an online adaptive ANN model was proposed for a comparison study with AR, and ARX models. The proposed model was identified/evaluated on 23 T1D subjects under sensor-augmented pump therapy. The obtained results showed that the prediction accuracy of the ANN model were better than the AR and ARX models: RMSE of 11.9, 18.9, and 26.1 mg/dL was reported for 15-, 30-, and 45-min PH, respectively for the ANN model. As well, a feedforward NN model was proposed in [191] for predicting the future BG levels with a PH of 75 min. However, they reported big prediction errors, with an RMSE of 43.09 mg/dL for 75-min PH. In [192], a recurrent neural network (RNN) model was proposed for short-term predictions of glucose levels. Data from four T1D children were used for the development and testing of the proposed model. An RMSE of 13.65 mg/dL was reported for 5-min PH. An Extended Kalman Filter identified state-



space model, as a non-optimal solution to nonlinear problems, is associated with a higher MAPE regarding 30-min PH [193]. Although, RNN models are characterized by the universal approximator of dynamical systems but it is difficult to use in online applications due to their high complexity of the training algorithm and non-convex objective function.

### 3.5 Hybrid models

Hybrid models use a physiological model followed by a data-driven model for glucose prediction or as a signal of BG problems, as stated before. According to the review in [124], Dalla Man meal model is considered to be the most popular model for integrating meal glucose absorption into hybrid models, followed by the Lehmann and Deutsch model [166, 194]. The most popular model for insulin kinetics, when information about insulin therapy is used, is Berger model [195], followed by the Dalla Man. In addition, ANN models are the most common data-driven components in hybrid models for glucose prediction.

In [196], an adaptive hybrid model is proposed for BG prediction by integrating helpful generic physiological features into a machine learning model. The modified Lehmann and Deutsch glucose absorption model was used for this proposal together with a SVR model. Data from 5 T1D subjects with their daily events such as insulin boluses and meals were used in this study. The experimental results in this study demonstrated a significant improvement in RMSE computation when incorporating physiological features into an SVR model. An RMSE of 19.5, and 35.7 mg/dL was reported for 30-, and 60-min PH, respectively.

In [178], a linear model with physiological input signals is used for approximating the glucose-insulin system in order to improve the BG prediction in T1D patients. The Dalla Man meal model, with Verdonk plasma insulin model [197], and an ARX model were used for this purpose. The model was tested on data of 15 T1D subjects. CGM data and exogenous inputs (insulin doses information and the CHO intake) were used in the model. The NLMS algorithm with a variable gain were used for coefficients estimation. The experimental results have shown satisfying performance through the FIT value until 45-min PH, and also through the CG-EGA analysis. The percentage of clinically accurate predictions that lead to benign errors is 97.35% at hypoglycemia, 99.8% at euglycemia and 98.55% at hyperglycemia.

In [198], a personalized hybrid glucose-insulin metabolism model is developed for T1D patients. The personalized model is based on the combined use of the Dalla Man glucose absorption model and Dalla Man insulin absorption model and a Self Organizing Map (SOM). The model was evaluated using data of 12 T1D subjects (CGM and insulin pump information, as well



as the time and amount of CHOs ingested). The predictive performance of the model was evaluated for a 30- and 60-min PH. The RMSE, the correlation coefficient (CC) and the CG-EGA were used to measure the accuracy of glucose prediction. An RMSE (mean  $\pm$  standard deviation (SD)) of  $14.10 \pm 4.57$  and a CC of  $0.94 \pm 0.02$  were reported for 30-min PH. An RMSE of  $23.19 \pm 6.40$  and a CC of  $0.84 \pm 0.05$  were reported for 60-min PH. Regarding the CG-EGA analysis, for 30-min PH, a high percentage of BG predictions were accurate readings. For 60-min PH, most of the BG predictions were accurate readings in the euglycemia and hyperglycemia ranges while most erroneous errors were reported in the hypoglycemia range.

In [199], a short-term prediction model based on time-series modeling is proposed for improving glucose prediction. The Dalla Man glucose absorption model together with a jump NN model is used for this study. CGM data of 10 T1D subjects and ingested CHOs information were used for evaluating this model. RMSE, time gain (TG) [188] and energy of the second-order differences (ESOD) [173] were used for the assessment of prediction accuracy. For a 30-min PH, an average RMSE of 16.6 mg/dL, average time gain of 18.5 min and average ESOD of 9.6 are reported. The use of a hybrid configuration improved the prediction as compared to the previous work by the authors using neural networks [184].

### 3.6 Prediction performance analysis

Despite the widespread use of physiological models in MPC formulations, observers, and other components in the AP, data on performance of the models involved are scarce, if any. Besides, model identifiability is key for a successful model personalization, which is a must due to the large inter-subject variability. Even in simpler models like the IVP (see Section 3.3.1), problems can be found such the need of measurements not available in clinical practice like plasma insulin, aside glucose. This enforces the consideration of population values in a subset of model parameters, which might not represent well inter-subject variability, yielding to a compromise in model accuracy. The above applies as well to hybrid models, where physiological models' output correspond to signals not available, like plasma insulin (unless in a lab setting) or meal glucose absorption rate.

In addition, current physiological models do not represent well important aspects of a patient's daily life, like meals with a variety of nutritional compositions and more especially exercise. In these cases, data-driven models can be more suited to capture the input-output data without the need of physiological knowledge. Indeed, data-driven models have shown satisfying results for predicting BG during and after exercise using physical activity signals as inputs [200, 180].

Thus, in this thesis, our investigation will focus on data-driven models for glucose prediction, especially linear empirical dynamic models, which were reviewed in Section 3.4. Table 3.1 summarizes the performance reported in literature for such models, which will serve to highlight the needed improvements that will be addressed in this thesis. Only results with clinical data are considered.

The most common PHs evaluated are 5, 15, 30, 45 and 60 min, with punctual studies for 75 min and 90 min. Mean(SD) values of RMSE for these common PHs are: 6.5(6.28) mg/dL for 5 min; 9.06(1.74) mg/dL for 15 min; 15.63(5.33) mg/dL for 30 min; 21.21(4.33) mg/dL for 45 min and 30.69(10.31) mg/dL for 60 min (see Fig. 3.8).

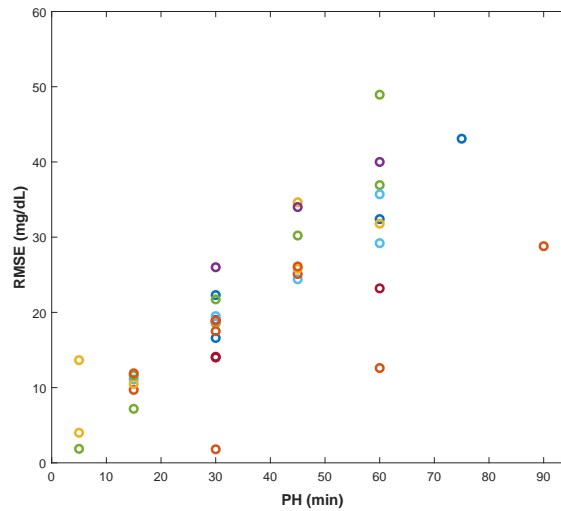


Figure 3.8: BG modeling error (RMSE) versus PH for the different literature methods.

In this thesis, the development of new modeling strategies to increase prediction accuracy and extend the prediction horizon will be addressed. As stated before, the focus will be put on linear empirical dynamical models or stochastic time series. Our hypothesis is that an appropriate management of variability will lead to a clustering problem characterizing local behaviors and the exploitation of historical similar data, which can be formalized by the concept of “seasonal local models”. The next two chapters will be thus devoted to a review of stochastic time series modeling and clustering time series. Then, the main contributions of this thesis will be presented.

Study	Method	Inputs	Database	PH (min)	RMSE (mg/dL)
Reifman <i>et al.</i> [171]	AR	CGM data	9 T1D	30 60	22.3 32.4
Gani <i>et al.</i> [172]	AR	CGM data	9 T1D	30 60 90	1.8 12.6 28.8
Sparacino <i>et al.</i> [173]	AR	CGM data	28 T1D	30 45	18.78 34.64
Finan <i>et al.</i> [177]	ARX	CGM data, insulin, meal	9 T1D	30 45 60	26 34 40
Turksoy <i>et al.</i> [180]	ARMAX	CGM data, insulin on board, energy expenditure, galvanic skin response	14 T1D	5 15 30 60	1.86 7.18 18.55 48.94
Zhao <i>et al.</i> [183]	Latent variable-based model without or with extra inputs (LV, LVX)	CGM data, insulin, meal	7 T1D	15 30 45 60	11.1 18.7 24.4 29.2
Zecchin <i>et al.</i> [184]	Feed-Forward NN and first-order polynomial model	CGM data, glucose rate of appearance after a meal	15 T1D	30	14.0
Zecchin <i>et al.</i> [199]	Jump neural network model	CGM data, glucose rate of appearance after a meal	20 T1D	30	16.6
Pérez-Gandía <i>et al.</i> [185]	Feed-Forward NN	CGM data	9 T1D	15 30 45	9.7 17.5 25.1
Georga <i>et al.</i> [186]	KRLS-ALD	CGM data, CHO, insulin, Physical activity	15 T1D	5 15 30 45 60	4.0 (1.8%* ) 10.5 (5.2%* ) 18.8 (10.0%* ) 25.8 (14.3%* ) 31.8 (18.0%* )
Wang <i>et al.</i> [187]	AR, ELM, and SVR	CGM data	10 T1D	30	19.0
Qingnan <i>et al.</i> [189]	RNN with LSTM and Bi-LSTM	CGM data	20 T1D	15 30 45 60	11.63 21.73 30.22 36.92
Bunescu <i>et al.</i> [196]	Adaptive hybrid	CGM data	5 T1D	30 60	19.5 35.7
Zarkogianni <i>et al.</i> [198]	Hybrid model	CGM data, insulin, CHO	12 T1D	30 60	14.1 23.19
Wang <i>et al.</i> [193]	Time-varying state-space model Extended Kalman Filter	CGM data FIR modeling of subcutaneous insulin absorption and meal absorption	5 T1D	30	20.31%*
Pappada <i>et al.</i> [191]	Feed-Forward NN	CGM data, insulin, nutrition, Lifestyle/Emotional Factors, time	17 T1D	75	43.09
Daskalaki <i>et al.</i> [190]	Online adaptive RNN	CGM data, insulin pump infusion rate	23 T1D	15 30 45	11.9 18.9 26.1
Mougiakakou <i>et al.</i> [192]	RNN	CGM data, insulin, CHO	4 T1D	5	13.65

Table 3.1: BG modeling error of different literature methods. \* indicates error expressed as MAPE (%).

## Chapter 4

# Stochastic time series modeling

Stochastic time series models are an alternative approach for glucose prediction, other than physiological models. Characteristics of stochastic time series models can increase the ability to accurately predict glucose and, therefore, can help in achieving improved glycemic control in people with diabetes when used in a monitoring system or an artificial pancreas framework. This chapter introduces linear empirical dynamic models, or stochastic time series models, which will be relevant for glucose prediction. Section 4.1 explains the basic concepts of time series modeling. Section 4.2 displays different types of linear time series models; then the importance of exogenous variables is discussed, and how to deal with them. Section 4.3 describes the methodology used for building appropriate models.

### 4.1 Preliminary concepts

A *stochastic process* is a statistical phenomenon that develops in time according to some probabilistic laws. It is composed of an indexed collection of a random variables ordered in time  $\{X_t, t \in T\}$ , where  $T$  is a given time span. A *time series* is a realization or an example (sample function) from a certain stochastic process. Thus, time series can be defined as an chronologically-ordered sequence of successive measurements (observations),  $\{X_t, t \in T_0, T_0 \subset T\}$ , done typically over time  $t$  [201, 202].

A time series may be *univariate* or *multivariate*. In a univariate time series there is only one time-dependent variable, such, for example, data collected from a CGM sensor measuring the BG levels every 5 minutes. A multivariate time series has more than one time-dependent variable (*i.e.*, multiple variables vary over time). Therefore, each variable has some dependency on other variables, which are called *predictor* or *explanatory* variables. This dependency is used for predicting future values. An example is data

collected from a CGM sensor, together with physiological data collected from biometric sensors such as heart rate, skin temperature and galvanic skin response.

A time series can be either *discrete* or *continuous*. A discrete time series is obtained when the measurements are typically taken at successive equally spaced points in time, such as hourly, daily, weekly, monthly or yearly time separations. A continuous time series is obtained when the measurements are taken continuously over some time interval. Any time series, in general, is supposed to be affected by some or all of the following components: trend, cyclical fluctuations, seasonal variations, and irregular variations. However, these components can be characterized from the observed data set, when analyzing the time series data. There are many different examples of time series in different fields like engineering, science, sociology, and economics as shown in Fig. 4.1 to Fig. 4.3.

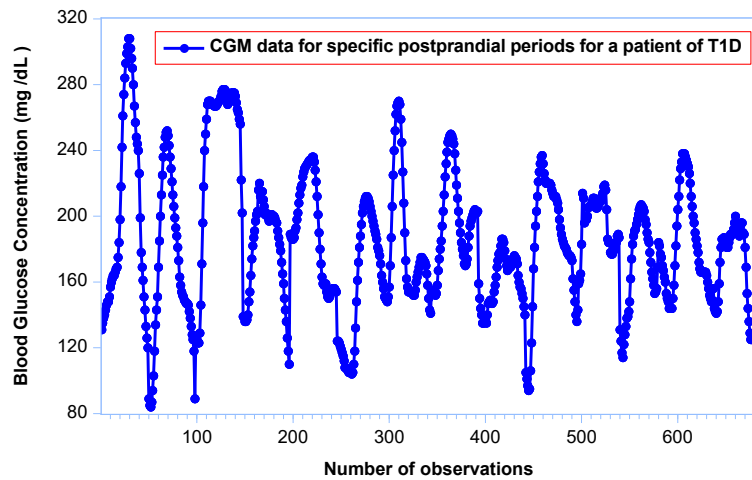


Figure 4.1: CGM time series data for specific postprandial periods for a person with T1D.

*Time series analysis* is the use of statistical methods to analyze time series data and extract meaningful statistics facts and other characteristics of the data. Many objectives can be satisfied by analyzing a time series, which they can all be classified as descriptive, explanatory, predictive, or control [203]. For example, time series analysis can be used in several domains for forecasting such as statistics, finance, geophysics, and meteorology. Moreover, it can be used for signal detection and estimation in signal processing and control engineering field.

A simple way to describe a stochastic process  $\{X_t, t \in T\}$  is through its first- and second-order moments. The first-order moment corresponds to

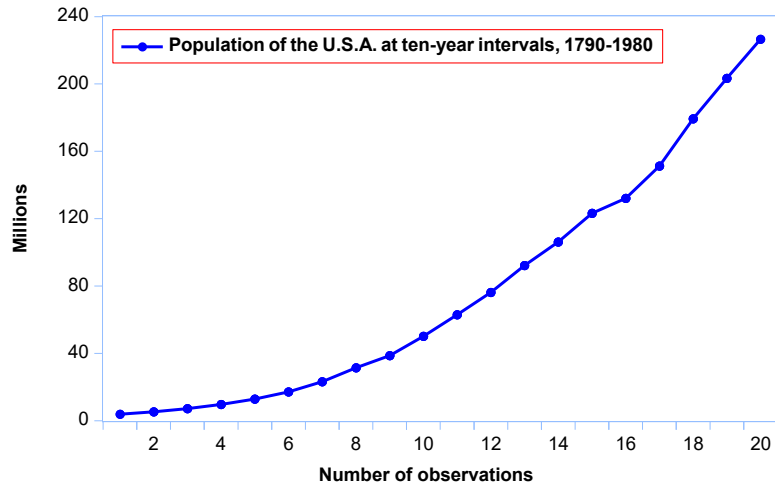


Figure 4.2: Time series for the population of the U.S.A. at ten-year intervals, 1790 – 1980.

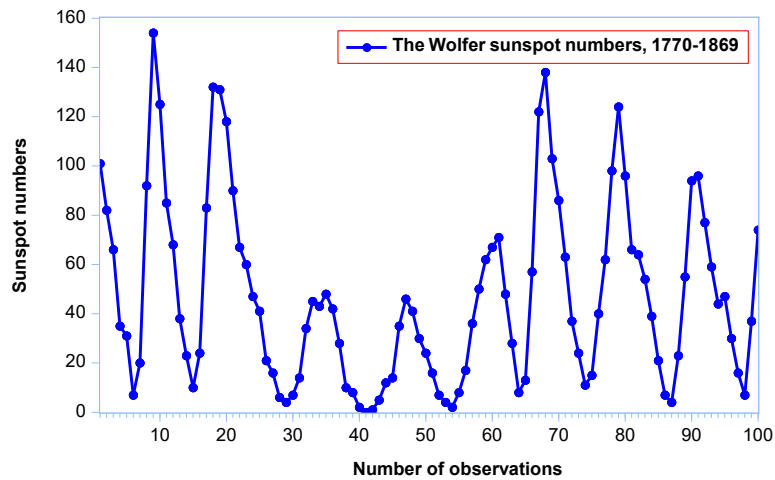


Figure 4.3: Time series for the Wolfer sunspot numbers, 1770 – 1869.

the expectation function, and the second-order moment to the autocovariance function (ACVF), which are defined next:

- The expectation function,  $\mu_X(t)$ , is defined for all  $t$  by

$$\mu_X(t) = \mathbb{E}(X_t). \quad (4.1)$$

where  $\mathbb{E}$  denotes the expected value of the random process  $X_t$ .

- The variance function  $\sigma_X^2(t)$  is defined for all  $t$  by

$$\sigma_X^2(t) = \text{Var}(X_t) = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \quad (4.2)$$

- The autocovariance function  $\gamma_X(t, s)$  is defined as

$$\begin{aligned} \gamma_X(t, s) &= \text{Cov}(X_t, X_s) \\ &= \mathbb{E}\left[(X_t - \mu_X(t))(X_s - \mu_X(s))\right] \\ &= \mathbb{E}(X_t X_s) - \mu_t \mu_s \quad \forall t, s \in T. \end{aligned} \quad (4.3)$$

Obviously, from (4.3), it follows that  $\sigma_X^2$  is a special case of the ACVF when  $t = s$ . Therefore, through the ACVF, which measures the degree of second-order variation between two data points in a random process (*i.e.*, time series) at two different times (*e.g.*,  $t, s$ ), it can be known how observations are related to each other in time, that is, the key aspect in any time series. In contrary, the  $\sigma_X^2$  alone is not enough to assign the second-order moments of a sequence of random variables [204].

*Stationary* processes forms the basis for a massive proportion of time series analysis methods. Thus, stationarity is the most vital and common assumption in time series techniques. The basic idea of stationarity is that the statistical properties controlling the process do not change with time (Fig. 4.4), *i.e.*, the mean and the variance of the implied process are constant and the ACVF depends only on the time difference  $|t - s|$ , which will be referred as *lag*,  $k$ . A stationary stochastic process can be of two types: strictly (strong or complete), and weakly (covariance) stationary process [202, 205].

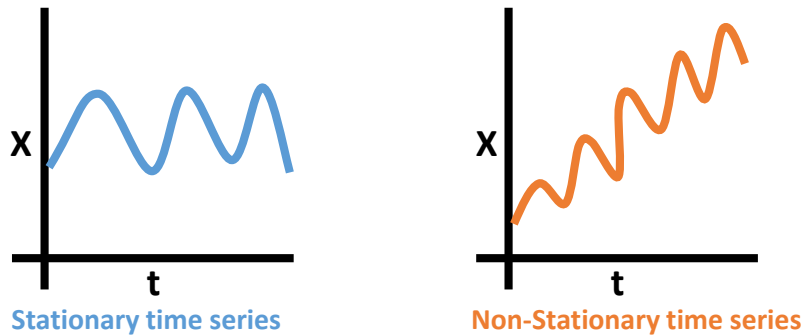


Figure 4.4: Stationary and non-stationary time series.

A time series  $\{X_t\}$  is said to be a *strictly stationary* if it satisfies the following properties:

- The joint probability distribution of  $\{X_{t_1}, \dots, X_{t_n}\}$  for any set of times  $\{t_1, \dots, t_n\}$  and any value of  $n$ , is the same as the joint distribution of

$\{X_{t_1+k}, \dots, X_{t_n+k}\}$ , that is, shifting the time origin by an amount  $k$  has no effect on the joint distributions. That means that the joint distribution only depends on the “difference”  $k$ , not on the time  $(t_1, \dots, t_n)$ .

- (b) The expected value of the time series is constant over time, and not a function of time:

$$\mathbb{E}(X_t) = \mathbb{E}(X_{t+k}) \Rightarrow \mu_t = \mu_{t+k} = \mu \quad \forall t, k.$$

- (c) The variance of the time series is constant over time (homoscedasticity):

$$\text{Var}(X_t) = \text{Var}(X_{t+k}) \Rightarrow \sigma_X^2(t) = \sigma_X^2(t+k) = \sigma^2 \quad \forall t, k.$$

- (d) The ACVF is not constant, does not depend on time, but on the time difference ( $h = |t - s|$ ):

$$\begin{aligned} \text{Cov}(X_t, X_s) &= \text{Cov}(X_{t+k}, X_{s+k}) \Rightarrow \gamma_X(t, s) = \gamma_X(t+k, s+k) \\ &\Rightarrow \gamma_X|t-s| = \gamma_X|t+k-s-k| = \gamma_h \quad \forall s, t, k. \end{aligned}$$

- (e) The autocorrelation function (ACF) is also not constant, does not depend on time, but on the time difference (time interval):

$$\begin{aligned} \text{Corr}(X_t, X_s) &= \text{Corr}(X_{t+k}, X_{s+k}) \Rightarrow \rho_X(t, s) = \rho_X(t+k, s+k) \\ &\Rightarrow \rho_X|t-s| = \rho_X|t+k-s-k| = \rho_h \quad \forall s, t, k. \end{aligned}$$

A time series  $\{X_t\}$  is said to be a *weakly stationary* (i.e., stationary in wide sense) if it satisfies the following properties:

- (a) The expected value is constant over time:

$$\mathbb{E}(X_t) = \mu < \infty \quad \forall t.$$

- (b) The variance is constant over time:

$$\text{Var}(X_t) = \sigma^2 \quad \forall t.$$

- (c) The ACVF is a function of the time difference only, i.e., the ACVF only depends on  $k$ , the distance between observations:

$$\text{Cov}(X_t, X_{t+k}) = \gamma_k \quad \forall t, k.$$

- (d) The ACF is a function of the time difference only.

$$\text{Corr}(X_t, X_{t+k}) = \frac{\text{Cov}(X_t, X_{t+k})}{\sqrt{\text{Var}(X_t)}\sqrt{\text{Var}(X_{t+k})}} = \frac{\gamma_k}{\gamma_0} = \rho_k \quad \forall t, k.$$



Clearly, for a weakly (second-order) stationary process, the autocovariance function  $\gamma_k$ , and the autocorrelation function  $\rho_k$ , have the following properties:

1.  $\gamma_0 = \text{Var}(X_t)$ .
2.  $\rho_0 = 1$ .
3.  $|\gamma_k| \leq \gamma_0$ ;  $|\rho_k| \leq 1$ .
4. They are even functions and symmetric about the lag  $k = 0$ , i.e.,  $\gamma_k = \gamma_{-k}$  and  $\rho_k = \rho_{-k} \quad \forall k$ .
5. They are positive semidefinite in the sense that

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \gamma_{|t_i - t_j|} \geq 0,$$

and

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \rho_{|t_i - t_j|} \geq 0,$$

for any set of times  $\{t_1, \dots, t_n\}$  and any real numbers  $\{\alpha_1, \dots, \alpha_n\}$ .

Strict stationarity is too strict for time series data in normal life applications. Therefore, whenever using the term “stationary” this shall mean weakly (second-order) stationary, unless otherwise indicated, and the following statement from [206] is sometimes quoted:

“Experience with real-world data, however, soon convinces one that both stationarity and Gaussianity are fairy tales invented for the amusement of undergraduates.”

The process  $\{X_t\}$  is said to be a *purely random* process if it consists of a sequence of uncorrelated random variables from a fixed distribution with the following properties:

1.  $\mathbb{E}(X_t) = 0 \quad \forall t$ .
2.  $\text{Var}(X_t) = \sigma_X^2 < \infty \quad \forall t$ .
3.  $\text{Cov}(X_t, X_{t+k}) = \gamma_k = 0 \quad \forall t, k \neq 0$ .

Such a sequence is indicated to as *white noise*, with mean 0 and variance  $\sigma_X^2$ , and can be written by the notation  $\{X_t\} \sim WN(0, \sigma_X^2)$ . Therefore, a white noise process  $\{X_t\}$  is a stationary process with the ACVF

$$\gamma_k = \begin{cases} \sigma_X^2 & k = 0, \\ 0 & k \neq 0, \end{cases}$$

and the ACF

$$\rho_k = \begin{cases} 1 & k = 0, \\ 0 & k \neq 0. \end{cases}$$

If the random variables in the sequence are independent and identically distributed (i.i.d. or iid or IID), with mean 0 and variance  $\sigma_X^2$  (i.e., strictly white noise), then it can be written by the notation  $\{X_t\} \sim IID(0, \sigma_X^2)$ . Moreover, every  $IID(0, \sigma_X^2)$  sequence is  $WN(0, \sigma_X^2)$  but not conversely. If a process  $\{X_t\}$  is Gaussian (i.e., the distribution functions of  $\{X_t\}$  are all multivariate normal), uncorrelated implies independence. Therefore, Gaussian white noise is just i.i.d.  $N(0, \sigma_X^2)$ . White noise processes are beneficial in many situations and play a significant role as basic building blocks for more complicated processes such as moving average processes [202, 205].

For any stationary stochastic process  $\{X_t\}$ , with  $\mathbb{E}(X_t) = \mu$  for all  $t$ , the real numbers  $\{X_1^{(1)}, X_2^{(1)}, \dots, X_T^{(1)}\}$  could be one possible result of a sample of length  $T$  of this process. Therefore,  $\{X_t^{(1)}\}$  would be just one realization of this stochastic process. In practice (i.e., time series applications), only a unique realization of each time series (random variable) is available. This implies that it is not possible to estimate  $\mu$  by computing the sample mean (ensemble average) of any random variable by

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_t^{(i)}. \quad (4.4)$$

where the  $X^{(i)}$ 's are different realizations of the random variable. Instead, the sample mean is computed by a time average as follows

$$\bar{X} = \frac{1}{T} \sum_{t=1}^T X_t. \quad (4.5)$$

The stationary process  $\{X_t\}$  is said to be *ergodic for the mean* if (4.5) converges in probability to  $\mathbb{E}(X_t)$  as  $T$  gets large (i.e.,  $\bar{X} \xrightarrow{P} \mu$ ), and *ergodic for the second moments* if

$$\frac{1}{T-k} \sum_{t=k+1}^T (X_t - \mu)(X_{t-k} - \mu) \xrightarrow{P} \gamma_k, \quad (4.6)$$

Alternatively, this can also be written as  $\hat{\gamma}_k \xrightarrow{P} \gamma_k$ , where  $\hat{\gamma}_k$ , is the lag  $k$  *sample autocovariance*, and is defined as

$$\hat{\gamma}_k = \frac{1}{T} \sum_{t=k+1}^T (X_t - \bar{X})(X_{t-k} - \bar{X}) \quad \text{for } k = 0, 1, \dots, T-1, \quad (4.7)$$

and the same with ACF,  $\hat{\rho}_k \xrightarrow{p} \rho_k$ , where  $\hat{\rho}_k$ , is the lag  $k$  *sample autocorrelation*, and is defined as

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad \forall k. \quad (4.8)$$

The plot of  $\hat{\rho}_k$  versus  $k$  times lags, is called the *correlogram*, and it can be useful as an indicator of nonstationarity. Besides that, if the autocovariance  $\gamma_k$  for a stationary process  $\{X_t\}$ , satisfy

$$\sum_{k=0}^{\infty} |\gamma_k| < \infty, \quad (4.9)$$

then  $\{X_t\}$  is ergodic for the mean [207].

## 4.2 Linear time series models

In order to design time series models, a clarification of the structure of the variation in the time series is needed, arising many questions such as:

- What are the main characteristics (descriptive statistics, persistence, and autocorrelation) of the time series?
- How can the time series be modeled, and how can the dynamic pattern of a real time series be found?
- How to forecast future events based on previous information in the time series, and how to derive hidden insights to make informed decision-making in a real time series?

Wold's decomposition theorem [208] represents the main tenet of linear time series models. Moreover, it has a vital role in time series analysis, by means of the ability to linearly forecast any stationary process, through the weighted average of past forecast errors. According to Wold's decomposition theorem, any zero-mean stationary process  $\{X_t\}$ , can be expressed as the sum of two mutually uncorrelated processes, a *deterministic part* (*i.e.*, the optimal linear combination of past observations of the process  $\{X_t\}$ ), and a *stochastic part*, which can be represented as an infinite moving-average process (*i.e.*, MA( $\infty$ )). Therefore, the stationary process  $\{X_t\}$  can be written as follows

$$X_t = A_t + B_t, \quad (4.10)$$

where  $A_t$  is linearly deterministic, and  $B_t$  is an indeterministic (stochastic) process that can be written as a linear combination of lags of a white noise process  $\varepsilon_t$ :

$$B_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad (4.11)$$

where

1.  $\psi_0 = 1, \sum_{j=0}^{\infty} \psi^2 < \infty,$
2.  $\{\varepsilon_t\} \sim WN(0, \sigma_\varepsilon^2),$
3.  $\{\psi_j\}$  and  $\{\varepsilon_t\}$  are unique,
4.  $\mathbb{E}(A_t \varepsilon_s) = 0 \quad \forall t, s,$
5.  $Cov(\varepsilon_s, A_t) = 0 \quad \forall s, t.$

Furthermore, the stationary process  $\{X_t\}$ , is called purely nondeterministic (stochastic) if  $A_t = 0$ . There are two common sub-groups of linear processes: the autoregressive (AR), and the moving average (MA) processes [207].

### 4.2.1 Autoregressive (AR) process

Autoregressive (AR) processes are one of the most useful processes used in time series modeling. AR process models are created with a property that the observation at time  $t$  is a linear combination of observations at times  $t-1, t-2, \dots, t-p$ . An AR process model that depends on  $p$  past observations is called an AR model of degree  $p$ , denoted by  $AR(p)$ , and it satisfies the following stochastic difference equation

$$X_t = c + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t, \quad (4.12)$$

where  $\phi_i$  are the parameters of the model,  $\phi_p \neq 0$ ,  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ , and

$$c = \left(1 - \sum_{i=1}^p \phi_i\right) \mu, \quad (4.13)$$

where  $\mu$  is the unconditional mean of the process. Moreover, it can be shown that an  $AR(p)$  model is *stationary* and *ergodic* provided that the roots of the characteristic equation

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p = 0, \quad (4.14)$$

lie outside the unit circle. Therefore, a necessary condition for stationarity is that  $|\phi_1 + \phi_2 + \dots + \phi_p| < 1$ , otherwise, the time series  $\{X_t\}$  will exhibit sinusoidal behavior. The presentation of time series models is simplified using lag (back-shift) operator notation. In lag operator polynomial notation,  $z^{-n} X_t = X_{t-n}$ , for any time series  $\{X_t\}$ . Therefore, by using the lag (back-shift) operator ( $z^{-1}$ ), the equation of an  $AR(p)$  can be written as

$$\phi(z^{-1}) X_t = c + \varepsilon_t \Rightarrow X_t = \mu + \psi(z^{-1}) \varepsilon_t, \quad (4.15)$$

where

$$\phi(z^{-1}) = (1 - \phi_1 z^{-1} - \phi_2 z^{-2} - \dots - \phi_p z^{-p}),$$

is the AR operator, and hence,

$$\psi(z^{-1}) = (\phi(z^{-1}))^{-1} = (1 + \psi_1 z^{-1} + \psi_2 z^{-2} + \dots),$$

is an infinite-degree lag operator polynomial, with  $\sum_{j=0}^{\infty} |\psi_j| < \infty$ . From (4.13), the equation of the AR( $p$ ) model (4.12) can be written as

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \varepsilon_t, \quad (4.16)$$

The ACVF of the AR( $p$ ) process is found by multiplying both sides of (4.16) by  $(X_{t-k} - \mu)$ , and taking expectations

$$\gamma_k = \begin{cases} \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p} & \text{for } k > 0, \\ \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma_\varepsilon^2 & \text{for } k = 0. \end{cases} \quad (4.17)$$

The ACF of the AR( $p$ ) process is found by dividing (4.17) by  $\gamma_0$  to obtain

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad \text{for } k > 0. \quad (4.18)$$

Specifying (4.18) for  $k = 1, \dots, p$ , a system of  $p$  equations is obtained that connect the first  $p$  autocorrelations with the parameters of the process. This system of equations is called *Yule-Walker* equations [209, 210, 201].

The partial autocorrelation function (PACF), is an important tool for identifying the order  $p$  of the AR models. The partial autocorrelation at lag  $k$  is the autocorrelation between  $X_t$  and  $X_{t-k}$  after removing any linear dependence on  $X_1, X_2, \dots, X_{t-k+1}$ . The PACF is based on estimating the sequence of AR models in consecutive orders:

$$\begin{aligned} Y_t &= \phi_{11} Y_{t-1} + \varepsilon_{1t} \\ Y_t &= \phi_{21} Y_{t-1} + \phi_{22} Y_{t-2} + \varepsilon_{2t} \\ &\dots \\ Y_t &= \phi_{k1} Y_{t-1} + \phi_{k2} Y_{t-2} + \dots + \phi_{kk} Y_{t-k} + \varepsilon_{kt}, \end{aligned} \quad (4.19)$$

where  $Y_t = X_t - \mu$ , the coefficients  $\phi_{kk}$  for  $k = 1, \dots, p$  are called *partial autocorrelation* coefficients of order  $k$ , and their sequence form the PACF. In practice, they are not derived in this way but, through the ACF as follows: multiply (4.19) by  $Y_{t-k}$ , take expectations and divide by the variance of  $Y_t$ . By repeating the same operation with  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k}$ , successively, the following set of  $k$  equations is obtained (Yule-Walker) [209, 210]:

$$\begin{aligned} \rho(1) &= \phi_{k1} + \phi_{k2} \rho(1) + \dots + \phi_{kk} \rho(k-1) \\ \rho(2) &= \phi_{k1} \rho(1) + \phi_{k2} + \dots + \phi_{kk} \rho(k-2) \\ &\dots \\ \rho(k) &= \phi_{k1} \rho(k-1) + \phi_{k2} \rho(k-2) + \dots + \phi_{kk}. \end{aligned}$$

By using Cramer's rule, the coefficients  $\phi_{kk}$  can be defined as

$$\phi_{kk} = \frac{\det(N)}{\det(M)}, \quad (4.20)$$

where

$$N = \begin{pmatrix} 1 & \rho(1) & \dots & \rho(1) \\ \rho(1) & 1 & \dots & \rho(2) \\ & & \dots & \\ \rho(k-1) & \rho(k-2) & \dots & \rho(k) \end{pmatrix},$$

and

$$M = \begin{pmatrix} 1 & \rho(1) & \dots & \rho(k-1) \\ \rho(1) & 1 & \dots & \rho(k-2) \\ & & \dots & \\ \rho(k-1) & \rho(k-2) & \dots & 1 \end{pmatrix}.$$

Thus, for an AR( $p$ ) model, the first  $p$  partial autocorrelation coefficients are non-zero, and the rest are zeros for lags greater than the order of the AR model [201]. Moreover, Cramer rule also can be used for estimating the coefficients  $\phi_{kk}$  by replacing  $\rho_k$  in the *Yule-Walker* equations by its estimate  $\hat{\rho}_k$  from (4.8). The estimated coefficients,  $\hat{\phi}_{kk}$ , are then called the *sample partial autocorrelation* coefficients of order  $k$ . Besides Cramer's rule, there are many methods for estimating the coefficients  $\hat{\phi}_{kk}$  like Durbin's method [211] and Pagano's technique [212]. The sample PACF of an AR( $p$ ) process has many properties such as [213]:

- $\hat{\phi}_{pp}$  converges to  $\phi_p$  as the sample size  $T$  goes to infinity.
- $\hat{\phi}_{kk}$  converges to zero for all  $k > p$ .
- $\hat{\phi}_{kk}$  from lag  $(p+1)$  onwards are approximately normally independently distributed with variance  $1/T$  [214].

#### 4.2.2 Moving average (MA) process

A Moving Average (MA) model has the property that the observation at time  $t$  is a linear combination of past stochastic errors, also called *shocks* or *innovations*, at times  $t-1, t-2, \dots, t-q$ . A time series  $\{X_t\}$  is said to follow a moving average process of order  $q$  or MA( $q$ ) process if

$$X_t = c + \sum_{i=0}^q \theta_i \varepsilon_{t-i}, \quad (4.21)$$

where  $\theta_i$  are the parameters of the model,  $\theta_0 = 1$ ,  $\theta_q \neq 0$ ,  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ , and  $c$  is the unconditional mean of  $\{X_t\}$ . An MA( $q$ ) model is always *stationary* because it is a finite linear combinations of a white noise sequence.

Furthermore, it is *invertible* if all of the roots of the MA characteristic polynomial

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q = 0, \quad (4.22)$$

lie outside the unit circle. Moreover, the equation of an MA( $q$ ) (4.21) can be written by using the lag notation as

$$X_t = \mu + \theta(z^{-1})\varepsilon_t, \quad (4.23)$$

where  $\theta(z^{-1}) = (1 + \theta_1 z^{-1} + \theta_2 z^{-2} + \dots + \theta_q z^{-q})$  is the MA operator. By using (4.21), the moments of an MA( $q$ ) process are:

- The expectation of an MA( $q$ ) process is

$$\mathbb{E}(X_t) = \mu + \sum_{i=0}^q \theta_i \mathbb{E}(\varepsilon_{t-i}) = \mu. \quad (4.24)$$

- The variance of an MA( $q$ ) process is

$$\gamma_0 = \mathbb{E}(X_t - \mu)^2 = \mathbb{E}\left(\sum_{i=0}^q \theta_i \varepsilon_{t-i}\right)^2. \quad (4.25)$$

Since the  $\varepsilon$ 's are uncorrelated, the variance is

$$\begin{aligned} \gamma_0 &= \sigma_\varepsilon^2 + \sum_{i=1}^q \theta_i^2 \sigma_\varepsilon^2 \\ &= \sigma_\varepsilon^2 \left(1 + \sum_{i=1}^q \theta_i^2\right). \end{aligned} \quad (4.26)$$

- The ACVF of an MA( $q$ ) process is

$$\begin{aligned} \gamma_k &= \mathbb{E}\left[\left(\sum_{i=0}^q \theta_i \varepsilon_{t-i}\right)\left(\sum_{i=0}^q \theta_i \varepsilon_{t-k-i}\right)\right] \\ &= \mathbb{E}\left[\sum_{i=0}^q \theta_{k+i} \theta_i \varepsilon_{t-k-i}^2\right], \end{aligned} \quad (4.27)$$

because all the terms involving the expected value of different  $\varepsilon$ 's are zero because of the WN assumption. Thus,

$$\gamma_k = \begin{cases} \sigma_\varepsilon^2 \sum_{i=0}^{q-k} \theta_i \theta_{k+i} & \text{for } k = 1, 2, \dots, q, \\ 0 & \text{for } k > q. \end{cases} \quad (4.28)$$

Dividing (4.28) by  $\gamma_0$  yields the ACF of the MA process

$$\rho_k = \begin{cases} \frac{\sum_{i=0}^{q-k} \theta_i \theta_{k+i}}{1 + \sum_{i=1}^q \theta_i^2} & \text{for } k = 1, 2, \dots, q, \\ 0 & \text{for } k > q. \end{cases} \quad (4.29)$$

Therefore, an MA( $q$ ) process is stationary for any value of  $\theta_i$ . Moreover, the process is also ergodic for the first moment since  $\sum_{-\infty}^{\infty} |\gamma_k| < \infty$ . If  $\varepsilon_t$  is also *Gaussian* then the process is ergodic for all moments.

### 4.2.3 Mixed autoregressive moving average (ARMA) processes

An ARMA process is a linear combination of two linear processes (*i.e.*, AR, MA), to produce a more sophisticated linear process. A time series,  $\{X_t\}$ , is said to follow an *autoregressive moving average* process of order  $p, q$ , denoted as ARMA( $p, q$ ), if

$$X_t = c + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=0}^q \theta_i \varepsilon_{t-i}, \quad (4.30)$$

or, in lag operator notation,

$$\phi(z^{-1})X_t = c + \theta(z^{-1})\varepsilon_t \Rightarrow X_t = \mu + \psi(z^{-1})\varepsilon_t, \quad (4.31)$$

where  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$  and

$$\psi(z^{-1}) = \frac{\theta(z^{-1})}{\phi(z^{-1})} = (1 + \psi_1 z^{-1} + \psi_2 z^{-2} + \dots),$$

is a rational, infinite-degree lag operator polynomial. The *unconditional mean* of the ARMA( $p, q$ ) process is given by

$$\mu = \frac{c}{\phi(z^{-1})}.$$

An ARMA model can be written in terms of deviations from the mean as follows

$$X_t - \mu = \sum_{i=0}^p \phi_i (X_{t-i} - \mu) + \sum_{i=0}^q \theta_i \varepsilon_{t-i}. \quad (4.32)$$

By multiplying both sides of (4.32) by  $(X_{t-k} - \mu)$  and taking expectations, the ACVF's are

$$\gamma_k = \sum_{i=1}^p \phi_i \gamma_{k-i} \quad \text{for } k \geq (q+1), \quad (4.33)$$

and hence,

$$\rho_k = \sum_{i=1}^p \phi_i \rho_{k-i} \quad \text{for } k \geq (q+1). \quad (4.34)$$



Therefore, after  $q$  lags the ACVF and ACF follow the  $p$ -th order homogeneous difference equation controlled by the AR parameters (*i.e.*, they follow the pattern of the AR component). Moreover, the ACF of an ARMA process *tails off* after lag  $q$ , as in an AR process, which depends only on the AR parameters in the model. Note that (4.33) does not hold for  $k \leq q$ , thus, the ARMA process will have more complicated ACVFs in the range  $1 \leq k \leq q$  compared to the corresponding AR process [207].

#### 4.2.4 Autoregressive integrated moving average (ARIMA) processes

If  $d$  is a nonnegative integer, then a process  $\{X_t\}$  is said to follow an *Integrated ARMA process*, denoted by  $\text{ARIMA}(p, d, q)$ , if

$$\nabla^d X_t = (1 - z^{-1})^d X_t, \quad (4.35)$$

is a *causal*  $\text{ARMA}(p, q)$  process, where

$$\nabla X_t = (1 - z^{-1})X_t = X_t - X_{t-1},$$

is the backward difference operator. “*Differencing*” by  $\nabla$  operator, is an excellent way of handling and transforming a nonstationary series to a stationary one. In general, an ARIMA process can be written as follows

$$\phi(z^{-1})\nabla^d X_t = c + \theta(z^{-1})\varepsilon_t, \quad (4.36)$$

where  $c$  is the constant term defined in the AR process,  $\phi(z)$  and  $\theta(z)$  are polynomials of degrees  $p$  and  $q$ , respectively,  $\phi(z) \neq 0$  for all  $|z| \leq 1$ , and  $\{\varepsilon_t\} \sim WN(0, \sigma_\varepsilon^2)$ . The letter ‘**I**’ in an ARIMA acronym stands for *integrated* and  $d$  denotes the number of differencing operators. In practice,  $d$  is usually a small number ( $d = 0, 1$  or  $2$ ). If  $d = 0$ , then  $\{X_t\}$  is a stationary ARMA process; it is called an integrated ARMA process of order zero and is denoted as  $X_t \sim I(0)$  [202].

#### 4.2.5 Seasonal ARIMA processes

Given  $d$  and  $D$  nonnegative integers, a process  $\{X_t\}$  is said to follow a *seasonal ARIMA process*, denoted by  $\text{SARIMA}(p, d, q)(P, D, Q)_s$ , if the differenced series

$$Y_t = \nabla_s^D \nabla^d X_t, \quad (4.37)$$

is a causal ARMA process defined by

$$\phi_p(z^{-1})\Phi_P(z^{-s})Y_t = c + \theta_q(z^{-1})\Theta_Q(z^{-s})\varepsilon_t, \quad (4.38)$$

where

$$\begin{aligned}\phi_p(z^{-1}) &= 1 - \phi_1 z^{-1} - \phi_2 z^{-2} - \dots - \phi_p z^{-p}, \\ \Phi_P(z^{-s}) &= 1 - \Phi_s z^{-s} - \Phi_{2s} z^{-2s} - \dots - \Phi_{Ps} z^{-Ps}, \\ \theta_q(z^{-1}) &= 1 + \theta_1 z^{-1} + \theta_2 z^{-2} + \dots + \theta_q z^{-q}, \\ \Theta_Q(z^{-s}) &= 1 + \Theta_s z^{-s} + \Theta_{2s} z^{-2s} + \dots + \Theta_{Qs} z^{-Qs},\end{aligned}$$

are polynomials in  $z^{-1}$  of degree  $p$ ,  $q$ ,  $P$ ,  $Q$ . Moreover,  $p$  is the order of the non-seasonal AR component,  $P$  is the order of the seasonal AR (SAR) component,  $q$  is the order of the non-seasonal MA component,  $Q$  is the order of the seasonal MA (SMA) component,  $s$  is the seasonal length (*i.e.*, time span of repeating seasonal pattern),  $d$  is the non-seasonal integration order,  $D$  is the seasonal integration order (*i.e.*, number of seasonal differences, where  $\nabla_s X_t = (1 - z^{-s})X_t = X_t - X_{t-s}$ ), and  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ . Remark that  $\nabla_s \neq \nabla^s$ . Furthermore,  $c$  is the constant term that is defined in the AR process [202].

The process  $\{X_t\}$  is causal if and only if  $\phi_p(z) \neq 0$  and  $\Phi_P(z) \neq 0$  for  $|z| \leq 1$ . In practice, a SARIMA model can also be defined as a *regression with seasonal ARIMA errors*, giving rise to the formulation

$$\begin{aligned}\nabla_s^D \nabla^d X_t &= c + w_t, \\ \phi_p(z^{-1}) \Phi_P(z^{-s}) w_t &= \theta_q(z^{-1}) \Theta_Q(z^{-s}) \varepsilon_t.\end{aligned}\tag{4.39}$$

where  $w_t$  is the disturbance series.

#### 4.2.6 Processes with exogenous variables

The main aim of using exogenous variables in time series models is to improve the fit and prediction performance of the chosen model. However, the qualitative characteristics and method of generation of an exogenous variable are not specified by the model builder. There exist different approaches for integrating exogenous variables into a model. Denoting as  $E_t$  the exogenous variable, a term  $\eta_r(z^{-1})E_t$  is commonly added to the right-hand-side of (4.15), (4.31), (4.36), and (4.38), respectively, yielding the so-called ARX, ARMAX, ARIMAX, or SARIMAX models depending on the considered structure, where

$$\eta_r(z^{-1}) = \eta_0 + \eta_1 z^{-1} + \dots + \eta_r z^{-r}\tag{4.40}$$

is the exogenous polynomial in  $z^{-1}$  of degree  $r$ , yielding to a *finite-impulse-response* (FIR) filter. In practice, as in many programs such as Matlab, R, and Eviews, exogenous variables are considered as *explanatory variables* into a linear regression model with a stochastic error process of certain structure. For instance, a SARIMAX model can be expressed in the form

$$\begin{aligned}\nabla_s^D \nabla^d X_t &= c + \eta_r(z^{-1})E_t + w_t, \\ \phi_p(z^{-1}) \Phi_P(z^{-s}) w_t &= \theta_q(z^{-1}) \Theta_Q(z^{-s}) \varepsilon_t.\end{aligned}\tag{4.41}$$

The model (4.41) can be expressed as  $\text{SARIMAX}(p, d, q, r)(P, D, Q)_s$ .

The main question that frequently appears in time series models with exogenous variables is “*how to determine the exogenous variables that correlate well with the endogenous variable and can be useful in the forecasting given a number of lags?*”. There are many ways to address this question like the *Granger causality* (GC) test. The GC (or “G-causality”), is a statistical concept to examine causality between two time series in terms of predictability [215, 216]. The main idea behind the GC is that time series  $X_t$  “*Granger causes*” time series  $Y_t$  (*i.e.*,  $X_t \rightarrow Y_t$ ) if lagged values of  $X_t$  contains useful knowledge (significantly) that increases the accuracy of the prediction of  $Y_t$  with respect to a forecast considering only lagged values of  $Y_t$ . The mathematical formulation of the GC test is based on linear regression modeling of stochastic processes. Let  $X_t, Y_t$  be two stationary time series. A simple causal model can be written as an AR process. Consider

$$Y_t = c_1 + \sum_{i=1}^L \alpha_i Y_{t-i} + \varepsilon_1 t, \quad (4.42)$$

as a *restricted* regression model, and

$$Y_t = c_2 + \sum_{i=1}^L \alpha_i Y_{t-i} + \sum_{i=1}^L \beta_i X_{t-i} + \varepsilon_2 t, \quad (4.43)$$

as an *unrestricted* regression model, with  $L$  the maximal time lag. The time series  $X_t$  Granger-causes time series  $Y_t$  if (4.43) is statistically significantly better than (4.42) (*i.e.*, if any or all of  $\beta_1, \dots, \beta_L$  are statistically significant). In the GC test, the null hypothesis is  $H_0 : \beta_1 = \beta_2 = \dots = \beta_L = 0$ , *i.e.*,  $X_t$  does not Granger-cause  $Y_t$ . By using the *ordinary least squares regression* (**OLS!**) and conduct an  $F$ -test of the null hypothesis

$$F = \frac{(RSS_R - RSS_{UR})/J}{RSS_{UR}/(T - k)} \sim F_{J, T-k}, \quad (4.44)$$

where  $RSS_R = \sum_{t=1}^T \varepsilon_1 t^2$ ,  $RSS_{UR} = \sum_{t=1}^T \varepsilon_2 t^2$ ,  $J$  is the number of restrictions,  $T$  is the number of observations in the time series, and  $k$  is number of explanatory variables in the unrestricted regression (including the constant). If the test statistic ( $F$ -value) is greater than the critical  $F$ -value then the hypothesis is rejected; otherwise, it is accepted. If the null hypothesis is rejected in both directions (*i.e.*,  $X_t \rightarrow Y_t, Y_t \rightarrow X_t$ ), then there is a *feedback* relation between both time series. In order to choose the lag length in the GC test, Akaike Information Criterion (AIC) or Schwarz Bayesian Criterion (SBC) can be examined to reach the optimal lag [217, 218].

The *cross-correlation function* (CCF) is another way to determine whether one series “*causes*” changes in another [219]. The CCF between two time

series helps to identify the nature of the relationship and how they are correlated in time. The *cross-covariance function* of two jointly stationary processes  $\{X_t\}$  and  $\{Y_t\}$  is

$$\gamma_{XY}(k) = \mathbb{E}[(X_{t+k} - \mu_X)(Y_t - \mu_Y)] \quad \forall k, \quad (4.45)$$

where  $\gamma_{XX}(k) = \gamma_X(k)$ ,  $\gamma_{XY}(k) \neq \gamma_{XY}(-k)$  (i.e.,  $\gamma_{XY}(k)$  is not an even function), but  $\gamma_{XY}(k) = \gamma_{YX}(-k)$ . Therefore, the *cross-correlation function* of jointly stationary processes  $\{X_t\}$  and  $\{Y_t\}$  is defined as follows

$$\rho_{XY}(k) = \frac{\gamma_{XY}(k)}{\sqrt{\gamma_X(0)\gamma_Y(0)}} \quad \forall k. \quad (4.46)$$

This function has many properties such as

- (a)  $\rho_{XY}(k) = \rho_{YX}(-k)$ ,
- (b)  $|\rho_{XY}(k)| \leq 1$ ,
- (c)  $\rho_{XX}(k) = \rho_X(k)$ , i.e.,  $\rho_{XY}(k)$  and  $\gamma_{XY}(k)$  are generalizations of ACVF, and ACF.

The CCF is estimated by the *sample cross-correlation function*

$$\hat{\rho}_{XY}(k) = \frac{\hat{\gamma}_{XY}(k)}{\sqrt{\hat{\gamma}_X(0)\hat{\gamma}_Y(0)}}, \quad (4.47)$$

where

$$\hat{\gamma}_{XY}(k) = \begin{cases} (1/T) \sum_{t=1}^{T-k} (X_t - \mu_X)(Y_{t+k} - \mu_Y) & k \geq 0, \\ (1/T) \sum_{t=1-k}^T (X_t - \mu_X)(Y_{t+k} - \mu_Y) & k < 0. \end{cases}$$

Through a plot of the sample CCF, a *leading* or *lagging* relation between time series can be identified. The time series  $X_t$  *leads* time series  $Y_t$ , if there are one or more  $X_{t+k}$ , with  $k$  *negative*, which are predictors of  $Y_t$ , whilst  $X_t$  *lags*  $Y_t$ , if there are one or more  $X_{t+k}$ , with  $k$  *positive*, which are predictors of  $Y_t$ .

### 4.3 The strategy of building models

The main aim of time series models is to capture reality and high accuracy in future predictions. In time series modeling, the most crucial steps are to identify and construct a model based on the available dataset by a specific methodology. *Box-Jenkins methodology* is the widest method for building time series models, proposed by the statisticians George Box and Gwilym Jenkins [220, 221]. It indicates a set of procedures for *identifying*, *estimating*, and *checking* time series models within the class of ARIMA models and as well seasonal ARIMA models as shown in Fig. 4.5. Therefore, finding suitable models for any time series is a non-trivial assignment. There are five steps to building a good model through the Box-Jenkins methodology:

- (1) Checking for stationarity or non-stationarity and differencing the data, if necessary;
- (2) Identifying and selecting an appropriate model structure;
- (3) Estimating the parameters of the chosen model;
- (4) Diagnostic checking of the chosen model (*Is the model adequate?*); and
- (5) Forecasting, and re-identifying a new model if necessary.

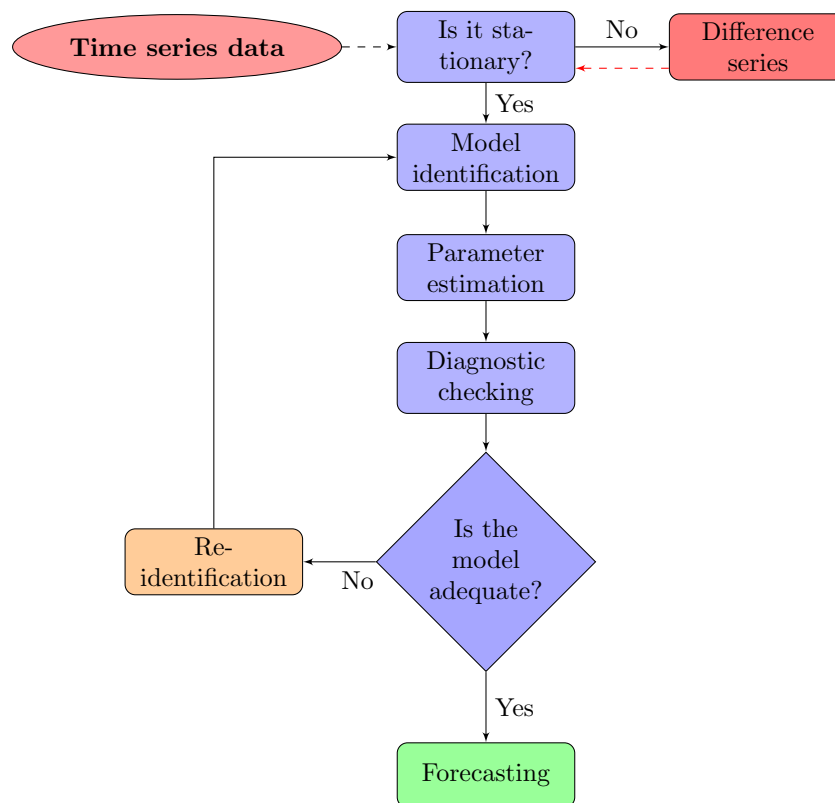


Figure 4.5: Steps for building a good model through Box-Jenkins methodology.

### 4.3.1 Stationarity test

A popular assumption in many time series techniques is that the time series data is stationary. Thus, the main question in this step is “*Is the time series stationary?*”. There are two ways to approach this question. The first one is a graphical procedure, *i.e.*, to examine the time series plot or examine the ACF( $\rho_k$ ) plot. If the ACF decreases to zero linearly and slowly (slow-decaying), then the series is non-stationary. The second way is a test

procedure, *i.e.*, by using a *unit root test*, which is able to determine whether a time series data is stationary or not, as well as whether the series contains a deterministic or stochastic trend (*i.e.*, the shocks have permanent or transitory effects). There are several tests for this purpose such as Phillips-Perron test, KPSS test, Dickey-Fuller test, and an augmented Dickey-Fuller test [222, 221]. The commonly used test for a unit root is the *augmented Dickey-Fuller* (ADF) test, proposed by the statisticians *David Dickey* and *Wayne Fuller* [223]. The ADF test is similar to the original Dickey-Fuller test, but the main difference between the two tests is that the ADF is used for larger and more complex time series models with unknown orders in order to accommodate general ARMA( $p, q$ ) models. Therefore, the original Dickey-Fuller test is a special case of the ADF test [224, 223].

One of the main factors affecting the stationarity of the time series is *seasonal variations* and the *trend* over time. They affect the value of the time series at different times. If the time series  $\{X_t\}$  is not stationary, it can be transformed into a stationary one by means of:

- (1) **Differencing:** Taking differences is one possible way to obtain a stationary time series. Differencing is typically performed to get rid of the non-constant mean of a series and as well for removing the stochastic trend from the series. Mathematically, differencing can be defined as in (4.35). The effect of differencing  $d$  times can be calculated by expanding the corresponding polynomial in  $z^{-1}$  as

$$\nabla^d X_t = (1 - z^{-1})^d X_t = \sum_{i=0}^d \binom{d}{i} (-1)^i X_{t-i}, \quad (4.48)$$

where  $d$  indicates the number of differencing transformations required by the time series to obtain a stationary series. Usually, one difference ( $d = 1$ ) is enough to get a stationary time series.

- (2) **Seasonal differencing:** Taking seasonal difference is a popular way to remove a seasonal effect from the time series. Mathematically, a seasonal difference of order  $D$  is defined as

$$\nabla_s^D X_t = (1 - z^{-s})^D X_t,$$

where  $s$  refers to the seasonal period,  $D$  takes positive integer values. Typically,  $D = 1$  or  $2$  is sufficient to obtain seasonal stationarity.

- (3) **Transformation:** Transformation methods are used to stabilize the non-constant variance of a series (*e.g.*, a series whose variance increases linearly with the mean) and also can suppress large fluctuations in the series. Common transformation methods include *power transform*, *square*

*root, cube root, and logarithm transform.* In general, each of these transformation methods is a member of the family of *Box-Cox transformations* [225]:

$$Y_t = \begin{cases} (X_t^\lambda - 1)/\lambda & \text{if } \lambda \neq 0, \\ \log(X_t) & \text{if } \lambda = 0. \end{cases} \quad (4.49)$$

where the parameter  $\lambda$  is to be set by the analyst. Therefore, different values of  $\lambda$  yield different transformation methods. Popular choices of the parameter  $\lambda$  are:  $\lambda = 0$ , which corresponds to a logarithmic transformation, and  $\lambda = 1/2$ , which corresponds to a square-root transformation. If the variance of the series seems to *increase quadratically* with the mean, the *logarithmic transformation* is a suitable choice. If the variance *increases linearly* with the mean, the *square-root transformation* is a suitable choice.

### 4.3.2 Model identification

Once the stationarity problem has been handled through different techniques in the first step, the next step is to identify and select the adequate model from the linear time series family models (*e.g.*, ARIMA, and SARIMA). Therefore, from the first step, appropriate values for  $d$  or  $D$  in ARIMA( $p, d, q$ ) or SARIMA( $p, d, q$ )( $P, D, Q$ )<sub>s</sub> models can be chosen, based on the results of the stationary test. The main question that arises in this step is “*How to choose suitable values for  $p$ , and  $q$  or  $P$ , and  $Q$  for a given series?*”. Besides that, “*how to select/determine the most satisfactory ARIMA or SARIMA model to represent a time series?*”.

Applying a specific model to some real or simulated time series, it is a difficult task and not trivial. Usually, a raw time series data is split into two sets. The first one represents a *training set*, which is used for identifying and constructing the desired model, and the second one represents the *test set*, which is used to test the adequacy of the model (*e.g.*, train 80% of the data and test the other 20%). In addition, the *leave-one-out cross-validation* (LOOCV) technique can be used as an ideal way for handling short time series, where the problem lies in how to split a dataset (Fig. 4.6). LOOCV is used as well for assessing how the results of the model will generalize to the independent dataset, and to estimate the accuracy of the performance of the predictive model.

The sample ACF and PACF are efficient tools for model identification (Fig. 4.7). An appropriate choice for the orders of  $p$ ,  $q$ ,  $P$  and  $Q$  can be obtained by matching the sample ACF( $\hat{\rho}_k$ ) and sample PACF( $\hat{\phi}_{kk}$ ) of the original time series (or transformed and differenced series when necessary) with the characteristic behavior of the theoretical ACF and PACF for MA process, AR process, mixed processes, and seasonal mixed processes, as shown in Table 4.1 and Table 4.2. In practice, a minimum of  $T = 50$



Figure 4.6: Leave-one-out cross-validation (LOOCV) strategy.

Process	ACF	PACF
AR( $p$ )	Tails off as exponential decay or damped cycle (sine wave)	Cuts off after lag $p$
MA( $q$ )	Cuts off after lag $q$	Tails off as exponential decay or damped cycle (sine wave)
ARMA( $p, q$ )	Tails off after lag $(q - p)$	Tails off after lag $(p - q)$
<b>White noise</b>	No spike	No spike

Table 4.1: Characteristics of theoretical ACF and PACF for specific non-seasonal ARMA models.

observations need to be available in order to determine an appropriate model and  $T/4$  to calculate the number of sample lag- $k$  autocorrelations and partial autocorrelations [205].

In practice as well, identification of the perfect orders  $p$ ,  $q$ ,  $P$  and  $Q$  of the desired model is difficult, especially, in the SARIMA model, because the sample ACF and PACF in real time series data rarely exhibit simple patterns. To overcome this problem, several different models can be identified with different orders, and then distinguish among all models by using an information criterion like *Akaike information criterion* (AIC) or *Bayesian information criterion* (BIC), also known as the *Schwarz criterion* [217, 218]. AIC and BIC can be defined as follows

$$\mathbf{AIC}(\mathbb{M}) = -2 \log L(\mathbb{M}) + 2K(\mathbb{M}), \quad (4.50)$$

$$\mathbf{BIC}(\mathbb{M}) = -2 \log L(\mathbb{M}) + K(\mathbb{M}) \log(T), \quad (4.51)$$



Process	ACF*	PACF*
AR( $P$ )	Tails off at seasonal lags	Cuts off after lag $P$
MA( $Q$ )	Cuts off after lag $Q$	Tails off at seasonal lags
ARMA( $P, Q$ )	Tails off at seasonal lags	Tails off at seasonal lags

Table 4.2: Characteristics of theoretical ACF and PACF for pure SARMA models.  
\*The values at nonseasonal lags are zero.

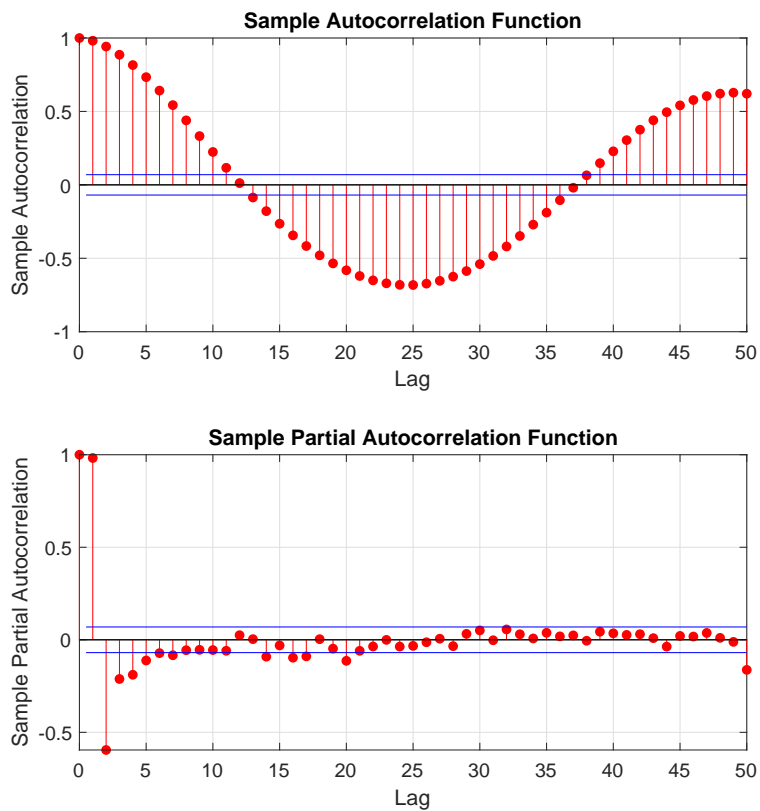


Figure 4.7: The sample ACF and sample PACF of an stationary time series. The blue horizontal lines represent the 5% critical values at  $\pm 1.96\sqrt{T}$ , which are used to test whether or not the ACFs and PACFs are significantly different from 0.

where  $L(\mathbb{M})$  is the likelihood function of the parameters in model  $\mathbb{M}$  evaluated at the MLE (Maximum Likelihood Estimators),  $T$  is the number of observations, and  $K$  is the total number of parameters estimated. A good model is the one that has *minimum* AIC or BIC between all the other models  $\mathbb{M}$  [217, 218].

One of the basic ideas in the model building is to keep the model as simple as possible. Besides that, a *good statistical fit* to the time series data is desired. Therefore, keeping the number of model parameters to a minimum (*i.e.*, *parsimonious model*) is the better choice when building the model depending on the measures of fit, which compares the quality of a set of models (*e.g.*, AIC, BIC), in order to reach a fit model and adequate data description. In practice, one of the key features of the ARMA or SARMA model is that it is able to represent an observed time series with less parameters than AR or MA models; thus, it is characterized by the parsimonious principle in model building [201].

### 4.3.3 Model estimation

After a tentative model has been specified, the next step is to estimate the model parameters from the time series data. There are many approaches to fit ARIMA or SARIMA models such as non-linear/linear least squares estimation (NLSE/LSE), generalized method of moments (GMM), Bayesian estimation (BE), Kalman filtering through state-space models, and maximum likelihood estimation (MLE). Model estimation is a quite complicated non-linear estimation problem, which can be solved by efficient algorithms depending to the statistical program used. However, all these estimation techniques do not perform equally [207, 226].

Because of its many attractive properties (*e.g.*, consistency, asymptotic normality, and asymptotic optimality), the *maximum likelihood estimation* (MLE) approach has been used for estimation unknown parameters in the ARIMA or SARIMA models as the most popular approach to parameter estimation. One of the main advantages of MLE is that it uses all of the information in the data in order to answer the following question:

“Which parameter values generate a greater probability for the observed data?”

The *likelihood function*  $L$  is the joint density function of the data, but treated as a function of the unknown parameters  $\beta$ . Given the observed data  $X_1, X_2, \dots, X_T$ :

$$L(X|\beta) = L(X_T, X_{T-1}, \dots, X_1|\beta). \quad (4.52)$$

Therefore,  $L$  is a function of the unknown parameters  $\beta$  in the model with the observed data held fixed. The MLE is a method of estimating the unknown parameters of a model, given observations, and its estimates are the values of the parameters that maximize this likelihood function:

$$\hat{\beta}_{MLE} = \arg \max_{\beta} L(X|\beta). \quad (4.53)$$

In most cases it is easier to work with the natural logarithm of  $L$ , given by

$$\text{Log}L(X|\beta) = l(X|\beta). \quad (4.54)$$

This function is commonly referred to as the *log-likelihood*. Since the logarithm is a *monotonic transformation* the values that maximize  $L(X|\boldsymbol{\beta})$  are the same as those that maximize  $l(X|\boldsymbol{\beta})$ , that is

$$\hat{\boldsymbol{\beta}}_{MLE} = \arg \max_{\boldsymbol{\beta}} L(X|\boldsymbol{\beta}) = \arg \max_{\boldsymbol{\beta}} l(X|\boldsymbol{\beta}), \quad (4.55)$$

but the *log-likelihood* is computationally more convenient. The  $\hat{\boldsymbol{\beta}}$  is the parameter vector that maximizes the probability for the observed sample  $X_1, X_2, \dots, X_T$ . Thus, the MLE satisfies the so-called *likelihood equations*, which are obtained by differentiating  $l(X|\boldsymbol{\beta})$  with respect to each of the unknown parameters of the vector  $\boldsymbol{\beta}$  and setting the derivatives to zero.

$$\frac{\partial l(X|\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0. \quad (4.56)$$

In many applications of ARMA or SARMA models the *conditional likelihood function* is an alternative to the *exact likelihood function*. In that case, the first  $p$  observations of a Gaussian ARMA or SARMA process are deterministic and are equal to its observed values  $\{X_p, \dots, X_1\}$ . The residual  $\varepsilon_t$  is a Gaussian white noise (*i.e.*,  $\varepsilon_t \sim i.i.d.N(0, \sigma_\varepsilon^2)$ ), computed from the previous initial values of  $\varepsilon_{t-i}$  for  $i = 1, \dots, p$ , with  $\varepsilon_t = 0$ , for  $t \leq p$ . The *log-likelihood function* for this setting has the following form, with  $\boldsymbol{\beta} = (\boldsymbol{\beta}', \sigma_\varepsilon^2)$ , where  $\boldsymbol{\beta}'$  is the transposed vector of the unknown parameters  $\boldsymbol{\beta}$ , and  $\sigma_\varepsilon^2$  is the variance of the innovations.

$$l(\boldsymbol{\beta}) = -\frac{1}{2}(T-p)\log 2\pi - \frac{1}{2}(T-p)\log \sigma_\varepsilon^2 - \frac{S(\boldsymbol{\beta}')}{2\sigma_\varepsilon^2}, \quad (4.57)$$

where

$$S(\boldsymbol{\beta}') = \sum_{t=p+1}^T (\varepsilon_t(\boldsymbol{\beta}'))^2, \quad (4.58)$$

denotes the sum of squares function, and the notation  $\varepsilon_t(\boldsymbol{\beta}')$  confirms that  $\varepsilon_t$  is no longer a disturbance, but a *residual* which depends on the value taken by the variables in  $\boldsymbol{\beta}'$ . Maximizing the log-likelihood function with respect to  $\boldsymbol{\beta}'$  is equivalent to minimizing the conditional sum of squares function  $S(\boldsymbol{\beta}')$ . Using the condition (4.56), this leads to

$$\sum_{t=1}^T \frac{\partial(\varepsilon_t(\boldsymbol{\beta}'))}{\partial \boldsymbol{\beta}'} \varepsilon_t = 0. \quad (4.59)$$

Thus, the MLE of the vector  $\boldsymbol{\beta}'$  can be obtained by minimizing (4.58), which yields an approximate MLE of  $\hat{\boldsymbol{\beta}}'$ . Furthermore, the MLE of the parameter  $\sigma_\varepsilon^2$  can be obtained by

$$\hat{\sigma}_\varepsilon^2 = \frac{S(\hat{\boldsymbol{\beta}}')}{T-p}. \quad (4.60)$$

### 4.3.4 Diagnostic checking

Once a model has been fit, the next step is determining whether the identified and estimated model is adequate. Hence,

“how to check the appropriateness of the fitted model?”

In particular, the *residuals* should be small, without systematic or predictable patterns (*i.e.*, white noise). Moreover, *independence*, *normality*, and *homoscedasticity* (literally, the same variance) should be achieved for the residuals. Therefore the adequacy of the selected model is evaluated by performing assorted statistical tests on the residuals or errors, such as:

- (i) **Graphical procedure:** examining ACF and PACF plots for the residuals to test the presence of any significant spikes in the confidence interval (*i.e.*, the spikes are within the confidence limits) indicating that the residuals seem to be uncorrelated (Fig. 4.8).

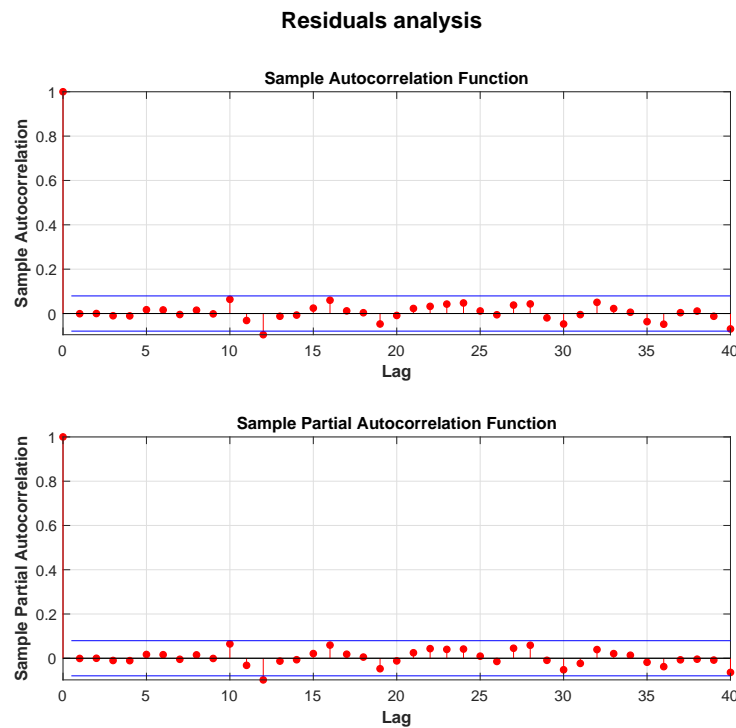


Figure 4.8: The sample ACF and sample PACF of the residuals. The blue horizontal lines represent the 5% critical values at  $\pm 1.96\sqrt{T}$  and are used to test whether or not the ACFs and PACFs are significantly different from 0.

- (ii) **Formal testing procedure:** testing the randomness at each distinct lag by using the portmanteau test of Ljung and Box, which is a statistical test for testing the goodness of fit of the model and is a modified

version of the *Box-Pierce* test [227, 228]. The *Ljung-Box Q* (LBQ) test examines  $m$  autocorrelations of the residuals and is defined as

$$Q(m) = T(T+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2(\hat{\varepsilon})}{T-k}, \quad (4.61)$$

where  $\hat{\rho}_k(\hat{\varepsilon})$  is the sample ACF of the estimated residuals at lag  $k$ ,  $T$  is the sample size, and  $m$  is the number of lags being tested. In the LBQ test, the null hypothesis is rejected (*i.e.*, the model does not exhibit lack of fit) if

$$Q(m) > \chi_{1-\alpha, \nu}^2,$$

where  $\chi_{1-\alpha, \nu}^2$  is the chi-square distribution table value with  $\nu$  degrees of freedom and significance level  $\alpha$ . For instance, when the LBQ test is applied to the residuals of an ARIMA( $p, q$ ) model, the degrees of freedom  $\nu$  must be equal to  $m - p - q$ .

- (iii) **Normality test procedure:** testing the normality of the residuals, by using the *Jarque-Bera* (JB) test, proposed by Carlos Jarque and Anil K. Bera [229]. The JB test is a goodness-of-fit test of whether the residuals follow a normal distribution and is defined as

$$JB = \frac{T}{6} \left( S^2 + \frac{(K-3)^2}{4} \right) \sim \chi_{\alpha, 2}^2, \quad (4.62)$$

where  $S$ ,  $K$ , and  $T$  denote the *sample residual skewness*, the *sample residual kurtosis*, and the *sample size*, respectively.  $S$  and  $K$  are defined as

$$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{T} \sum_{i=1}^T \hat{\varepsilon}_t^3}{\left( \frac{1}{T} \sum_{i=1}^T \hat{\varepsilon}_t^2 \right)^{3/2}}, \quad (4.63)$$

$$K = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\frac{1}{T} \sum_{i=1}^T \hat{\varepsilon}_t^4}{\left( \frac{1}{T} \sum_{i=1}^T \hat{\varepsilon}_t^2 \right)^2}, \quad (4.64)$$

where  $\hat{\mu}_3$  and  $\hat{\mu}_4$  are the estimates of third and fourth central moments, respectively, and  $\hat{\sigma}^2$  is the estimate of the second central moment (*i.e.*, the variance). Under the assumption of normality,  $S$  has an asymptotic normal distribution with mean 0 and variance  $6/T$ . The case of  $K$  is the same, but with mean 0 and variance  $24/T$ . The test statistic (JB) is compared to a chi-squared distribution with 2 degrees of freedom. Normality is rejected if the JB value is greater than the chi-squared value.

If the tentative model is inadequate, a new model must be re-identified (step 2) in order to build a most adequate model. On the contrary, if the tentative model passes the diagnostic tests, then the model is fitted to be used for *forecasting*.

### 4.3.5 Forecasting

Once a model has been fitted to the training data (*in-sample period*), the next step is to use this model for *forecasting*. In this case, validation data (*out-sample period*) are used to evaluate forecasting performance [230]. Although the fitted model is required to generate forecasts, the best-fitted model does not necessarily give us the best forecasting. The good forecasts capture the truthful patterns and dynamic characteristic, which exist in the training set, but do not *replicate* past episodes that will not happen again. The ability to forecast an event (episode) depends on many factors including:

- (i) How much data are available in the training set for the model identification/selection and parameters estimation?
- (ii) How well do we understand the factors that contribute to it?

The forecasts of the time series  $X_t$ , can be defined as

$$\hat{X}_T(h) = \hat{X}_{T+h|T} = \mathbb{E}(X_{T+h}|X_T, X_{T-1}, \dots, X_1), \quad (4.65)$$

where  $(X_T, X_{T-1}, \dots, X_1)$  represent the previous observations in the training period up to time  $T$ .  $\hat{X}_{T+h|T}$  is the conditional expectation of  $X_{T+h}$  given the information available at time  $T$ , and it denotes the *h-step ahead forecast* (*i.e.*, the forecasted value  $X_{T+h}$ ), where  $h$  is a positive integer referred to as the *prediction horizon* (PH). Therefore, there are 3 types of forecasting based on the prediction horizon ( $h$ ): *short-term*, *medium-term*, and *long-term* forecasts.

#### Forecast accuracy measures

Of relevance are the accuracy measures used to assess the forecasting performance of the model, which are addressed here. The *forecast error* is the difference between the actual and forecast value, stated as an absolute value or as a percentage. It can be defined as  $e_i = A_i - F_i$ , where  $A_i$  denote the  $i^{th}$  actual observation and  $F_i$  is a forecast of  $A_i$ . There are several statistical measures that used to measure the accuracy of the forecasting such as:

1. **Mean Squared Error (MSE)**. The *mean squared error* (MSE) or *mean squared deviation* (MSD), is a popular measure for forecasting errors and its variance, and can be computed as

$$MSE := \frac{1}{n} \sum_{i=1}^n e_i^2. \quad (4.66)$$

Where  $n$  is the number of observations for a given time period. MSE penalizes a forecast much more for large errors than it does for small errors because all errors are squared. In addition, it is more sensitive to outliers due to the squared function.

2. **Root Mean Squared Error (RMSE).** The *root mean squared error* (RMSE) or *root mean squared deviation* (RMSD), is the square root of calculated MSE. Therefore, it is computed as follows

$$RMSE := \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} = \sqrt{MSE}, \quad (4.67)$$

3. **Mean Absolute Error (MAE).** The *mean absolute error* (MAE) or *mean absolute deviation* (MAD), is the average of the absolute forecast errors, as given by

$$MAE := \frac{1}{n} \sum_{i=1}^n |e_i|. \quad (4.68)$$

MAE displays the magnitude of the overall error, but is not sensitive to outliers.

4. **Mean Absolute Percentage Error (MAPE).** The *mean absolute percentage error* (MAPE) or *mean absolute percentage deviation* (MAPD), is the average of the absolute percentage errors of forecasts. MAPE is computed as

$$MAPE := \frac{100}{n} \sum_{i=1}^n \left| \frac{e_i}{A_i} \right|. \quad (4.69)$$

MAPE is independent of the scale of measurement and easy to interpret, but does not penalize extreme errors.

5. **Theil inequality coefficient (UI).** The *Theil inequality coefficient* (UI) is a normalized measure of the total forecast error, and can be computed as follows

$$UI = \frac{RMSE}{\sqrt{\frac{1}{n} \sum_{i=1}^n A_i^2 + \frac{1}{n} \sum_{i=1}^n F_i^2}}. \quad (4.70)$$

A value of 0 for UI indicates *perfect forecast*, while a value of 1 corresponds to *maximum inequality* or negative proportionality between the actual and forecast values [231].

## Chapter 5

# Clustering time series

Complexity of factors affecting glucose response in T1D suggests the exploration of clustering techniques and local modeling approaches aiming at better characterizing scenarios with similar responses. This chapter introduces techniques for time series clustering, applications to local modeling and a survey of related work in the field of diabetes. The chapter starts in Section 5.1 with a definition of clustering time series data, then explains the difference between the two categories of clustering algorithms, and the importance of clustering time series techniques in different domains with some examples. Section 5.2 displays different similarity measures to clarify which similarity measure is appropriate to use in the clustering process. Section 5.3 explains different approaches to time series clustering, and how to handle with the dataset. Section 5.4 determines which clustering algorithm should be used for time series clustering, and introduces different cluster validity indices to determine the number of clusters in the dataset. Section 5.5 clarifies the effective role of clustering techniques for local models in order to deal with complex systems. Section 5.6 displays some applications of clustering techniques in T1D.

### 5.1 Clustering time series data

Over the past two decades, the idea of discovering hidden information in datasets has been widely explored, due to the tremendous growth in data size and communication technology. This discovery has mainly been focused on data mining, classification and clustering techniques [232, 233].

As mentioned previously in Section 4.1, a time series is defined as a sequence of measurements indexed in time order. Most often, the measurements are made at regular time intervals. Time series is the most common type of dynamic data, appearing naturally in a variety of different domains, such as statistics, signal processing, pattern recognition, econometrics, weather forecasting, earthquake prediction, control engineering, among



many others. They pose some challenging issues, due to the direct manipulation of continuous, high dimensional datasets in a dynamic way, which is extremely difficult. Besides forecasting, which was addressed in Chapter 4, time series analysis can also be used for clustering and classification applications in data mining and machine learning fields.

Clustering, also called *cluster analysis*, or *unsupervised classification*, is the task of an unsupervised grouping of a dataset (unlabeled patterns, observations, tuples, transactions or feature vectors) into groups (clusters) so that the similar objects will be in the same cluster and objects in different clusters are quite distinct [233]. Application of clustering techniques is important when no labeled data are available regardless of whether the data are binary, categorical, numerical, interval, ordinal, temporal, image, multimedia, or mixtures of the above data types. Clustering plays a key role in discovering and extracting beneficial knowledge from large databases. Time series data clustering has been used in diverse problem domains:

- *Financial markets*: In financial markets, the stocks data represent time series, which continually change with time. Cluster analysis of such time series can provide many insights into the trends in the data like determining structural (dis)similarities in economic processes for economic forecasting.
- *Medical science*: Various types of medical data such as the electroencephalography (EEG) readings and CGM measurements are in the form of time series. Cluster analysis of such time series can provide a comprehension of the common patterns in the data. Moreover, common characteristics from these patterns can be extracted and linked to different types of diseases or warning from long complications.
- *Earth science*: Several applications are found in earth science, such as temperature and pressure studies. Cluster analysis of such time series can provide information about the common climactic trends in the data for climatic forecasting.

In general, clustering algorithms can be classified into two categories: *hard clustering* and *fuzzy clustering*, which are described next.

### Hard clustering

Hard (crisp) clusters mean non-overlapping (un-nested) clusters: the data points (objects, patterns) are partitioned into distinct clusters, where each data point (object, pattern) belongs to one and only one cluster.

Mathematically, in hard clustering, algorithms assign a group label  $L_i \in \{1, 2, \dots, k\}$ , to each data point (object)  $x_i$ , to identify its cluster group, where  $k$  is the number of clusters. Each object is assumed to belong to one

and only one cluster. Therefore, the result of hard clustering algorithms can be represented by a  $k \times n$  matrix

$$U = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ u_{21} & u_{22} & \dots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{k1} & u_{k2} & \dots & u_{kn} \end{pmatrix}, \quad (5.1)$$

where  $n$  represents the number of data points (objects),  $k$  represents the number of clusters, each row describes a cluster  $j$ , each column corresponds to a data point  $i$ ,  $U = [u_{ij}]$  is a partition matrix, and  $u_{ij}$  may only take a value of either 0 or 1. If a data point (object)  $i$  belongs to cluster  $j$ , then  $u_{ij} = 1$ . Otherwise,  $u_{ij} = 0$ . Therefore,  $u_{ij}$  satisfies the following natural conditions:

$$u_{ij} \in \{0, 1\}, \quad \forall i \in \{1, \dots, n\}, \quad \forall j \in \{1, \dots, k\}, \quad (5.2)$$

$$\sum_{j=1}^k u_{ij} = 1, \quad \forall i, \quad (5.3)$$

$$\sum_{i=1}^n u_{ij} > 0, \quad \forall j. \quad (5.4)$$

Condition (5.3) provides that data point  $i$  may only belong to one cluster (*i.e.*, only one entry may take the value of 1 within any certain column). To satisfy condition (5.4), there must be no empty clusters (*i.e.*, each row must contain an entry of value 1).

### Fuzzy clustering

In fuzzy (soft) clustering, the assumption is relaxed so that data points (objects, patterns) belong to more than one cluster, and associated with each data point (object, pattern) is a set of membership levels/degrees.

Mathematically, the result of fuzzy clustering algorithms can also be represented by a  $k \times n$  matrix  $U$ , as defined in the hard clustering, but with the following relaxed restrictions:

$$u_{ij} \in [0, 1], \quad \forall i \in \{1, \dots, n\}, \quad \forall j \in \{1, \dots, k\}, \quad (5.5)$$

$$\sum_{j=1}^k u_{ij} = 1, \quad \forall i, \quad (5.6)$$

$$\sum_{i=1}^n u_{ij} > 0, \quad \forall j. \quad (5.7)$$

Aspects to consider in time series clustering include the selection of a distance measure for the characterization of (dis)similarity, the prototype extraction function, the appropriate choice of the clustering methodology and the clusters evaluation for a right number of clusters. Time series clustering will depend on whether the whole series, a sub-sequence or particular time points are to be clustered. As a result, a time series is sought that effectively epitomize the important characteristics of all series in a given cluster, which must be similar to each other. This series is the so-called *prototype series* (or average series or centroid), which contains useful information for a specific cluster. All these aspects will be reviewed next.

## 5.2 Similarity and distance measures

Mathematically, let be two time series  $X$  and  $Y$  vectors of length  $n$ , and  $x_i$  and  $y_i$  the  $i^{\text{th}}$  values of  $X$  and  $Y$ , respectively. Then,  $X$  and  $Y$  are defined as *similar* if and only if  $D(X, Y) < \epsilon$ , where  $D(X, Y)$  is a distance measure characterizing (dis)similarity between  $X$ ,  $Y$  and  $\epsilon$  is a specified threshold value.

Comparison of time series is widely used in several research domains and many applications. Similarity measures are fundamental to solve many pattern recognition problems like in clustering and classification problems. Therefore, distance measures are one of the cornerstones of any time series clustering algorithm, in order to compare the (dis)similarity between the pairwise time series. However, it is not simple to derive a method that can measure the similarity of time series efficiently due to the characteristics of time series data, which are naturally noisy and include outliers and shifts, besides the high dimensionality. So, a key question is “*which similarity measure is most appropriate to use?*”. The choice of the similarity measure depends on whether similarity in time, similarity in shape or structural similarity is sought [232].

### 5.2.1 Similarity in time

Similarity in time is defined as the similarity between pairwise time series data based on the similarity at each time step.  $L_p$ -norm distance is the most common measure of similarity that reflects similarity in time [233]. As a distance metric,  $L_p$ -norm fulfills the properties of non-negativity, identity, symmetry, and triangle inequality. This distance is a strict metric that can only compare pairwise time series *of the same length*. The  $L_p$ -norm distance (also referred as Minkowski distance) is defined as follows

$$d(X, Y) = \left( \sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}, \quad (5.8)$$

where  $p$  is the order of Minkowski distance. Therefore, there are many different distances derived from the Minkowski distance depending on the order value ( $p$ ):

- **Euclidean Distance:** Euclidean distance (*i.e.*, straight line distance) is the most often used distance/metric in many applications. It is a special case of the Minkowski distance when  $p = 2$  ( $L_2$ -norm) and it is defined as:

$$d(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (5.9)$$

- **Manhattan Distance:** The Manhattan distance/metric was named based on the grid-like layout of the streets of the New York borough of Manhattan (*i.e.*, city block distance). It is a special case of the Minkowski distance when  $p = 1$  ( $L_1$ -norm) and it is defined as:

$$d(X, Y) = \left( \sum_{i=1}^n |x_i - y_i| \right). \quad (5.10)$$

- **Maximum Distance:** Maximum distance (*i.e.*, Chebyshev distance or Tchebychev distance) is defined to be the absolute maximum value of the distances between the coordinates of a pair of time series (objects). It is a special case of the Minkowski distance when  $p = \infty$  ( $L_\infty$ -norm) and it is defined as:

$$d(X, Y) = \lim_{p \rightarrow \infty} \left( \sum_{i=1}^n |x_i - y_i|^p \right)^{1/p} = \max_{0 < i \leq n} |x_i - y_i|. \quad (5.11)$$

### 5.2.2 Similarity in shape

The time of occurrence of the time series is not important when similarity in shape is sought. The concept of similarity in shape is based on the similarities between subsequences of the pairwise time series or their common trends. Therefore, similarity in time can be considered as a special case of similarity in shape. Adequate measures of similarity in shape are elastic measures like the *dynamic time warping* (DTW) measure.

DTW measure stems from the context of sound processing [234], and it is one of the most proper algorithms for measuring the similarity between different time series with irregular lengths. DTW replaces the one-to-one time point comparison that is used in Euclidean distance (ED), with a many-to-one (and vice-versa) comparison (see Fig. 5.1). However, DTW does

not fulfill the triangle inequality condition, and thus, it is not a proper mathematical metric or distance function. Despite this, it has the main feature of allowing the recognition of similar shapes and it is widely used [235].

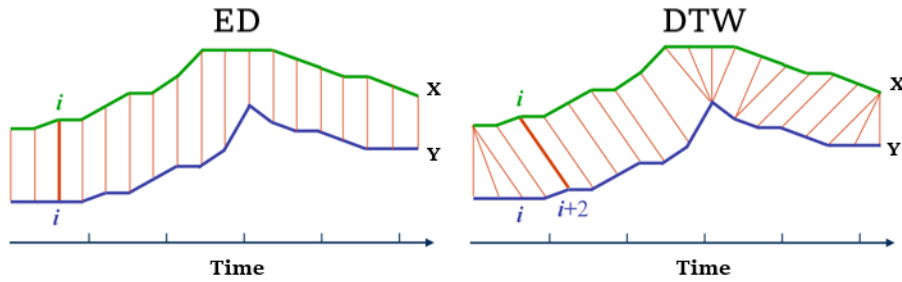


Figure 5.1: Illustrative example of the measure of similarity between two time series: (left) Euclidean distance (lock-step) measure; (right) Dynamic Time Warping measure (elastic).

DTW is computed as follows. Given two time series  $X = \{x_1, x_2, \dots, x_n\}$  and  $Y = \{y_1, y_2, \dots, y_m\}$  of length  $n$  and  $m$ , respectively, the following  $n \times m$  matrix is constructed

$$D = \begin{pmatrix} d(x_1, y_1) & d(x_1, y_2) & \dots & d(x_1, y_m) \\ d(x_2, y_1) & d(x_2, y_2) & & \\ \vdots & & \ddots & \\ d(x_n, y_1) & & & d(x_n, y_m) \end{pmatrix}, \quad (5.12)$$

where  $D_{i,j} = d(x_i, y_j)$ , with  $1 \leq i \leq n$  and  $1 \leq j \leq m$  is typically the ED between the data points of the time series under comparison. The DTW objective is to compute the warping path  $W = \{w_1, w_2, \dots, w_k, \dots, w_K\}$  of neighboring elements in matrix  $D$ , with  $\max(n, m) < K < m + n - 1$ , such that it minimizes the following function:

$$DTW(X, Y) = \min \left( \sqrt{\sum_{k=1}^K w_k} \right). \quad (5.13)$$

The warping path is a contiguous set of matrix elements which defines a mapping between  $X$  and  $Y$  that satisfies the following constraints:

1. **Boundary conditions:**  $w_1 = (1, 1)$  and  $w_K = (m, n)$ , where  $K$  is the length of the warping path, so the first and the last element of the warping path are in diagonally opposite corners of the matrix.
2. **Continuity:** if  $w_i = (a, b)$  then  $w_{i-1} = (a', b')$ , where  $a - a' \leq 1$  and  $b - b' \leq 1$ . Therefore, cells of the matrix denoted by adjacent elements of the warping path must be neighbours.

3. **Monotonicity:** if  $w_i = (a, b)$  then  $w_{i-1} = (a', b')$ , where  $a - a' \geq 0$  and  $b - b' \geq 0$ . Therefore, the warping path must be monotonically, non-decreasing along the time axis.

In practice, the DTW distance between two time series can be calculated through an  $O(n^2)$  dynamic programming [236] to prevent the construction of the whole matrix. Define  $\gamma(i, j)$  to be the cumulative distance. Then

$$\gamma(i, j) = d(x_i, y_j) + \min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\}. \quad (5.14)$$

### 5.2.3 Structural similarity

One of the most important components in any time series is a stochastic persistence component, that is present in several time series domains like economic time series. Therefore, it is important to consider this component when defining the similarity between different time series. This component can be captured by modeling it by stochastic models like ARMA models.

Structural similarity is characterized from measures on the parameters of a fitted model to each time series. If parameters of two models are close, then the corresponding time series have a large probability to behave in a similar way. The most common modeling methods for this objective are Hidden Markov Models (HMM) [237] and ARMA process [238]. Such a similarity measure can be used to extract information about the dynamic behavior of the time series from others belonging to the same cluster [239]. This approach is considered to be appropriate for long time series, not for short time series [240].

## 5.3 Approaches to time series clustering

Time series clustering approaches have been divided into three main categories as shown in Fig. 5.2 depending upon whether they dealing directly with raw time series data without modification, indirectly with features vectors of lower dimension extracted from the raw time series data, or indirectly with a number of models parameters built from the raw time series data [241, 242].

### 5.3.1 Raw-data-based approaches

In raw-data-based approaches, the clustering algorithm is directly applied on raw time series data without any pre-treatment, or changes in the original data. The main modification is replacing the distance or similarity measures for static data with a suitable one for time series, and minor transformations (if necessary) like normalization of the raw data. Raw-data-based approaches are relatively simple and generally lead to high dimensional spaces.

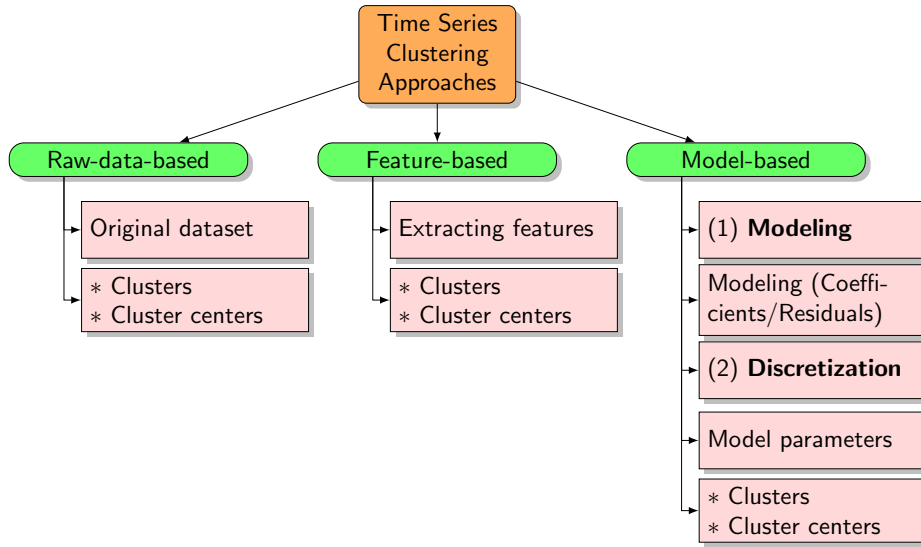


Figure 5.2: Time series clustering approaches.

It can be applied in the time or frequency domain.  $L_p$ -norm distance is the simplest and most widely used method in these approaches. Moreover, the DTW measure is also widely used in the case of varying lengths of the time series (Fig. 5.3). However, this kind of method will not apply when the size of the time series data is too large and data is highly noisy. Clustering the DNA microarray raw data and functional magnetic resonance imaging (fMRI) for brain activity are some known applications in these approaches in order to detect similar regions of activation [243, 244].

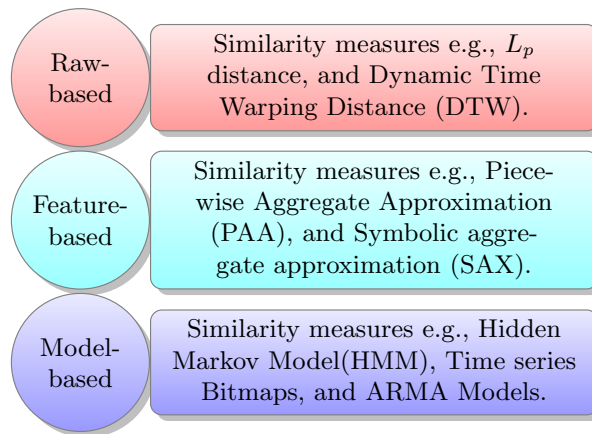


Figure 5.3: Time series clustering approaches with similarity measures.

### 5.3.2 Feature-based approaches

Due to the characteristics of the time series, in the feature-based approaches, the raw time series is converted into a feature vector of a lower dimension. Therefore, it is necessary to find an appropriate data representation to simplify the dimension of the time series data and to reduce the computational and storage cost [245].

Different feature-based representations have been explored depending on statistical measures like the mean, standard deviation, skewness, and kurtosis of the time series [246]. Moreover, there are several common methods for these approaches such as spectral decomposition through Discrete Fourier Transform (DFT) [247], Discrete Wavelet Transform (DWT) [248], Singular Value Decomposition (SVD) [249], Piecewise Aggregate Approximation (PAA) [250], and Symbolic Aggregate Approximation (SAX) [251].

The methods for the extraction of features are generic in nature, however, the extracted features are usually application-dependent. That is, not a single set of features works well for all applications. For example, the extracted features from the fMRI time series include cross-correlation measures in order to match brain regions with similar activation patterns [252, 253]. This feature is not necessarily appropriate for another application. The advantages of feature-based approaches are that they are less sensitive to missing values and they can handle time series of different lengths. However, they have the disadvantage that sometimes it loses significant information of the original data after transformation and reduction.

### 5.3.3 Model-based approaches

Model-based approaches represent a time series in a stochastic manner. The basic idea of this approach is to model the time series first by stochastic models such as Markov models and Hidden Markov model (HMM) [254, 255], statistical models, time series bitmaps [256], and Auto-Regressive Moving Average (ARMA) [238, 257], among others. Then clustering is applied to the model parameters, and different statistical information like fitting residual.

The advantage of model-based approaches is that they can handle time series which are noisy and with high dynamic changes with time. However, they have the disadvantage that the process is too complicated, requiring firstly to reach a good fit model and then extract a useful statistical information or model parameters for time series clustering.

## 5.4 Time series clustering algorithms

As mentioned previously, there are many different ways of measuring the similarity of time series data. Therefore, there are many different ways



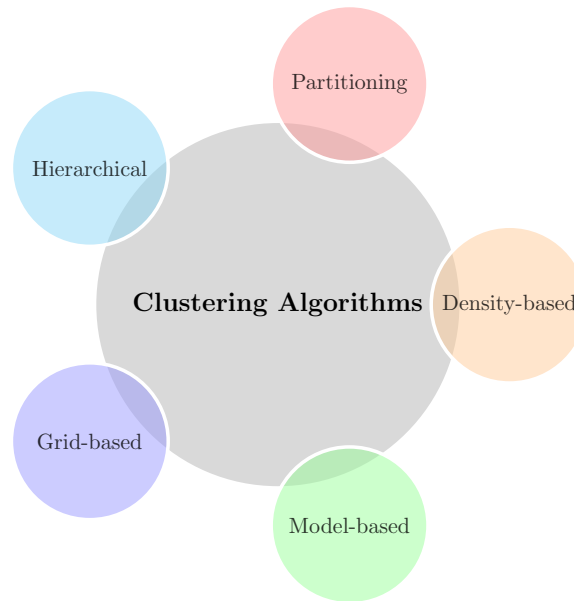


Figure 5.4: The major clustering algorithms.

to cluster the time series data into groups (clusters), each having different application areas. A time series clustering algorithm must answer to the following questions:

- Which data points belong to which cluster?
- How many clusters are there in a dataset?

There is no single answer to these questions, but many depending on the similarity measure, the clustering algorithm and the clustering validity indices used. An understanding of both the clustering problem and the clustering technique is required in order to apply an appropriate algorithm to a given problem.

The main properties of clustering algorithms include the ability to work with a high dimensional dataset, ability to get similar groups (clusters) of mixed shapes, ability to handle with a different types of attributes (features, variables, component, fields, dimensions), and the ability to handle with outliers [14]. The major clustering algorithms can be classified into five categories: partitioning, hierarchical, density-based, grid-based and model-based methods (Fig. 5.4). Traditionally, clustering algorithms are based on two common techniques known as hierarchical (nested) clustering and partitional (un-nested) clustering, based on the characteristics of the clusters

generated (*i.e.*, depending on the type of structure imposed on the data) [258].

### 5.4.1 Hierarchical clustering

Hierarchical clustering (HC) is one of the oldest and common clustering methods. HC is a method for transforming a similarity/proximity matrix for different data points (objects) into a set of nested clusters that are organized as a hierarchy/tree shape and can be visualized by a “*dendrogram*”. HC methods are classified into agglomerative (bottom-up) and divisive (top-down) (see Fig. 5.5):

- **Agglomerative hierarchical clustering (AHC):** It is a bottom-up approach that starts with each data point in its own cluster (singleton) and recursively merges the most similar (closest) pair of clusters until only a single cluster remains.
- **Divisive hierarchical clustering (DHC):** It is a top-down approach that starts with all the data points in one cluster, recursively partitioning the cluster into the most proper smaller clusters, until each cluster consists of one data point.

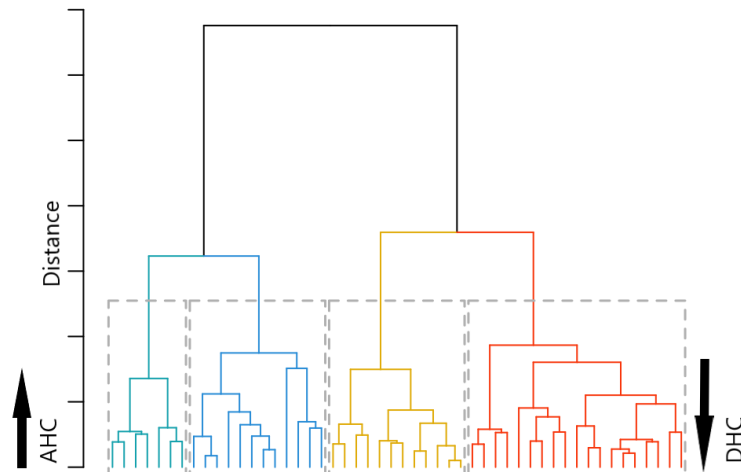


Figure 5.5: Agglomerative (AHC) and divisive (DHC) hierarchical clustering. Adapted with modifications from [10].

In practice, agglomerative clustering is most common than divisive clustering [14, 258]. AHC algorithms include the concept of measuring and comparing distances between clusters for merging the similar or closest clusters up to a stopping criterion and an optimal number of clusters in the hierarchical technique. To do so, there are many important measures to determine

the cluster distance such as single-linkage, average-linkage and complete-linkage (the so-called *linkage metrics* in the AHC literature [14]), which are reviewed next:

1. **Single-linkage metric (SLM)**: It is one of the most known cluster distance for cluster amalgamation, and it is also known as connect- edness, minimum method, or nearest neighbour cluster distance. In SLM, the distance  $d$  between two clusters,  $C_1$  and  $C_2$ , is defined as the minimum distance (*i.e.*, shortest distance) between the clusters objects, *i.e.*:

$$d(C_1, C_2) = d_{min}(C_1, C_2) = \min_{x \in C_1, y \in C_2} d(x, y). \quad (5.15)$$

If the data depend on the similarities, the distance, in this case, is equal to the greatest similarity between the clusters objects.

2. **Average-linkage metric (ALM)**: The distance  $d$  between two clusters,  $C_1$  and  $C_2$  is defined as the average distance between clusters objects, *i.e.*:

$$d(C_1, C_2) = d_{avg}(C_1, C_2) = \frac{1}{|C_1||C_2|} \sum_{x \in C_1} \sum_{y \in C_2} d(x, y), \quad (5.16)$$

where  $|C_i|$  is the size of cluster  $i$ . ALM does not impose a structure on the clustering effect.

3. **Complete-linkage metric (CLM)**: It is one of the most popular measures for cluster distance, and it is also called the diameter, maximum method, or furthest neighbour cluster distance. In CLM, the distance  $d$  between two clusters,  $C_1$  and  $C_2$  is defined as the maximum (greatest) distance between the clusters objects, *i.e.*:

$$d(C_1, C_2) = d_{max}(C_1, C_2) = \max_{x \in C_1, y \in C_2} d(x, y). \quad (5.17)$$

Therefore, CLM produces compact clusters with a small maximum distance between any two objects of the cluster, where objects within a specific cluster are supposed to be close to each other and outlying objects are not integrated.

All of the above linkage metrics can be derived through the *Lance-Williams* dissimilarity update formula [259]. If objects  $i$  and  $j$  are agglomerated into cluster  $C_i \cup C_j$ , then we must simply specify the new (dis)similarity between the cluster and all other clusters  $C_k$ . The formula is

$$d(i \cup j, k) = \alpha_i d(i, k) + \alpha_j d(j, k) + \beta d(i, j) + \gamma |d(i, k) - d(j, k)|, \quad (5.18)$$

Advantages	Disadvantages
Embedded flexibility regarding the extent of granularity	The time complexity of at least $O(n^2 \log n)$ is required, where $n$ represents the number of data points: not suitable for large datasets
Ease of handling any form of similarity or distance	They are static, <i>i.e.</i> , objects assigned to a specific cluster cannot transition to another cluster
Applicability to various attribute types	Very sensitive to noise and outliers
Easy to implement	The vagueness of selecting the right stopping criteria
There is no need to be specified the number of clusters in advance	The undeniable fact that most hierarchical algorithms do not revisit

Table 5.1: Advantages and disadvantages of the hierarchical clustering [14].

where  $\alpha_i$ ,  $\alpha_j$ ,  $\beta$ , and  $\gamma$  define the agglomerative criterion. In the case of the SLM, using  $\alpha_i = \alpha_j = 1/2$ ,  $\beta = 0$ , and  $\gamma = -1/2$  yields to

$$d(i \cup j, k) = \frac{1}{2}d(i, k) + \frac{1}{2}d(j, k) - \frac{1}{2}\gamma|d(i, k) - d(j, k)|,$$

which, can be rewritten as

$$d(i \cup j, k) = \min \left\{ d(i, k), d(j, k) \right\}.$$

Due to the need to structure taxonomies, HC methods are popular in many fields like social and biological sciences. Moreover, HC algorithms have many advantages like implementation in an easy way and ease of handling any form of similarity measures [14]. In contrast, it has disadvantages such as sensitivity to noise and outliers, and not being suitable for large datasets (Table 5.1).

#### 5.4.2 Partitional clustering

Partitional clustering (PC) method, is one of the most popular clustering algorithms. It consists in a partition process of data points (objects) into subsets of disjoint clusters (un-nested clusters) without a hierarchical structure, in an attempt to retrieve natural groups that exist in the data. Furthermore, to increase the valuation value of clustering, each subset (cluster) optimizes a clustering criterion, which can be local or global. A global criterion represents the partition clusters according to the similarity between different prototypes, whilst a local criterion represents the partition clusters via the local structure of the data such as high-density areas [258].

Given a data set of  $N$  points (patterns) in a  $d$ -dimensional metric space, a PC method partitions the data points into  $K$  clusters of data ( $N \geq K$ ). That is, it classifies the data into  $K$  clusters by achieving the following requirements: (i) each cluster contains at least one point, and (ii) each point belongs to exactly one cluster or more rely on the partition types. There are two types of PC methods: hard (crispy) partitional clustering (HPC), and fuzzy (soft) partitional clustering (FPC).

### Hard partitional clustering (HPC)

As mentioned previously, in hard clustering, each data point strictly belongs to one and only one cluster. One of the most widely used algorithms of HPC is the  $K$ -means algorithm that solve the clustering problem in a simple and easy way [260].  $K$ -means make exactly  $K$  different clusters of the greatest possible distinction from  $N$  unlabeled data points (objects) based on features/attributes of the data. The partition by  $K$ -means is carried out by minimizing the sum of squares of the distance (typically the ED) between data points and the corresponding cluster centroid (prototype), where the centroid is (typically) the mean of the data points in the cluster.

Mathematically, the main aim of the  $K$ -means algorithm is to minimize the total intra-cluster variance, *i.e.*, the squared error function:

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2, \quad (5.19)$$

where  $\|x_i^{(j)} - c_j\|^2$  is a distance measure between data points  $x_i^j$  and the cluster centroid  $c_j$ , and  $n$  represents the number of data points in  $j^{\text{th}}$  cluster. The algorithm consist of the following steps:

1. Place  $K$  points into the space represented by the data points that are being clustered. These points represent initial cluster centroids.
2. Assign each data point to the cluster that has the closest centroid.
3. When all data points have been assigned, recompute the positions of the  $K$  centroids.
4. Repeat Steps 2 and 3 until convergence (centroids no longer move and none of the cluster assignments change). This makes a partition of the data points into clusters from which the metric to be minimized can be calculated.

This algorithm needs to specify the number of clusters *a priori*. Therefore, different initial partitions can produce different final clustering. Besides, the algorithm can converge to local minima. Therefore, the result

found by this algorithm is not necessarily the most optimal one. That is one of the drawbacks of this method, besides the fact that it can only handle numerical features and it is highly sensitivity to outliers. Nevertheless, the computational time in this method is faster and can produce tighter clusters as compared to HC methods [261].

### Fuzzy partitional clustering (FPC)

Especially in time series clustering, it is very difficult to determine to which cluster a data point belongs to exactly. Moreover, in practice and real-life applications, clusters may overlap, and data points may belong partially to more than one cluster, which is the main problem in HPC methods. In FPC methods, one data point belongs to more than one cluster based on its membership value/degree, which is defined as the degree of likelihood of one data point belonging to several clusters. The sum of membership values of each data point to all clusters is equal to one. In several applications, the FPC methods have been demonstrated to be better methods than HPC when dealing with discrimination of similar structures [262] and datasets with significant outliers [263].

The most widely used method in the FPC methods is the *Fuzzy C-means algorithm* (FCM). FCM is a fuzzified version of the  $K$ -means algorithm, developed by Dunn in 1973 and improved by Bezdek in 1981 [264, 265]. FCM attempts to find the most characteristic point that characterizes each cluster, which can be considered as the *centroid* of the cluster and, then, it determines the membership degree of each data point to each cluster. To this end, FCM algorithm partitions data points into  $c > 1$  clusters by minimizing the following objective function:

$$J_m(U, V; X) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m d_{ik}^2(x_k, \nu_i), \quad (5.20)$$

where  $X = \{x_1, x_2, \dots, x_n\}$  denotes the unlabeled data points (objects) of size  $n$ ,  $V = (\nu_1, \nu_2, \dots, \nu_c)$  is a vector of unknown cluster centers (prototypes),  $\nu_i \in \mathbb{R}^s$ , and  $U = [u_{ik}]$ ,  $i = \{1, 2, \dots, c\}$ ,  $k = \{1, 2, \dots, n\}$  is the partition matrix, where  $u_{ik} \in [0, 1]$ ,  $\sum_{i=1}^c u_{ik} = 1$ ,  $\forall k$  and  $0 < \sum_{k=1}^n u_{ik} < n \forall i$ ,  $d$  is a distance function (ED function between  $k^{th}$  data points and  $i^{th}$  cluster center) and  $m \in [1, \infty)$  is the fuzziness parameter.

FCM algorithm consists in the following steps:

1. Compute the fuzzy membership,  $u_{ik}$ , using:

$$u_{ik} = \frac{1}{\sum_{l=1}^c \left( \frac{d_{lk}^2}{d_{ik}^2} \right)^{\frac{1}{m-1}}}. \quad (5.21)$$

2. Compute the fuzzy centers,  $c_i$ , using:

$$c_i = \frac{\sum_{k=1}^n u_{ik}^m x_k}{\sum_{k=1}^n u_{ik}^m}. \quad (5.22)$$

3. Repeat the last two steps until the minimum of the objective function  $J_m$  is achieved or  $\|U^{l+1} - U^l\| < \beta$ , where  $\beta$  is the termination criterion between 0 and 1, whereas  $l$  are the iteration steps.

The lower the value of  $\beta$  the better the results, but at the expense of a more number of iterations, which increases the computational time required by this algorithm. This is one of the main drawbacks of this method. However, FCM has the main advantage that it gives the best results for overlapped data points, and it is more robust against outliers. Therefore, it is considered comparatively better than HPC algorithms.

### FCM Clustering of incomplete data

The FCM algorithm is a useful technique for clustering real-dimensional data, but it cannot be directly applied to incomplete data (missing data), since all data is required to compute the cluster prototypes (centroids) and the distance measures. There are different strategies for adapting the original FCM algorithm to missing/incomplete data, which are reviewed next [266]. Some previous definitions follow.

Let  $X = \{x_1, x_2, \dots, x_n\}$  be an incomplete data set in  $s$ -dimensional real space;  $X_W = \{x_k \in X | x_k \text{ is a complete datum}\}$  be the **whole-data** subset of  $X$ ;  $X_P = \{\mathbf{x}_{jk} \text{ for } 1 \leq j \leq s, 1 \leq k \leq n | \text{the value for } \mathbf{x}_{jk} \text{ is present in } x_k \text{ for } 1 \leq k \leq n\}$  be the set of the **available feature** values; and  $X_M = \{\mathbf{x}_{jk} = ? \text{ for } 1 \leq j \leq s, 1 \leq k \leq n | \text{the value for } \mathbf{x}_{jk} \text{ is missing from } x_k \text{ for } 1 \leq k \leq n\}$  be the set of the **missing feature** values.

FCM algorithm can be adapted to missing data as follows:

1. **Whole data strategy FCM (WDSFCM)**. The *whole data strategy* (WDS) is a simple technique for clustering incomplete data, if the proportion of incomplete data is small, *i.e.*, approximately, no less than 75% of complete data items are present in the data set. This algorithm simply deletes all incomplete data and applies the original FCM algorithm on the complete data subset  $X_W$  (complete-case analysis). Then, the prototypes and the memberships of the data vectors in  $X_W$  can be computed via an alternate optimization of (5.21) and (5.22). Afterwards, the cluster membership for each incomplete data item can be estimated by using a nearest-prototype classification scheme based on the partial distances.

2. **Partial distance strategy FCM (PDSFCM)**. The *partial distance strategy* (PDS) is a popular strategy to FCM clustering of incomplete data proposed by Dixon [267], which uses the *partial distance function* for measuring the distances between two data items. The PDS algorithm is applicable to the case where the proportion of missing values is larger. In contrast to the WDS method, the PDS algorithm ignores the missing feature values in  $X_M$  and use all available feature values in  $X_P$  to cluster incomplete data. The partial distance between the incomplete item and cluster prototype is thus defined as

$$D_{ik} = \frac{s}{I_k} \sum_{j=1}^s (\mathbf{x}_{jk} - \nu_{ji})^2 I_{jk}, \quad (5.23)$$

where

$$I_{jk} = \begin{cases} 0 & \text{if } \mathbf{x}_{jk} \in X_M, \\ 1 & \text{if } \mathbf{x}_{jk} \in X_P, \end{cases} \quad (5.24)$$

for  $1 \leq j \leq s$ ,  $1 \leq k \leq n$ , and  $I_k = \sum_{j=1}^s I_{kj}$ . Hence, the PDS version of FCM algorithm (PDSFCM), can be obtained by making two modifications of the original FCM algorithm:

- Calculate  $d_{ik}$  in (5.21) for incomplete data according to (5.23) (*i.e.*, replace  $d_{ik}$  with  $D_{ik}$ ).
- The cluster prototypes are then updated as follows

$$\nu_{ji} = \frac{\sum_{k=1}^n u_{ik}^m I_{jk} \mathbf{x}_{jk}}{\sum_{k=1}^n u_{ik}^m I_{jk}}, \quad \text{for } 1 \leq i \leq c, 1 \leq j \leq s. \quad (5.25)$$

3. **Optimal completion strategy FCM (OCSFCM)**. The *optimal completion strategy* (OCS) is one of the most effective techniques for clustering incomplete data. The main idea of this technique is to compute the missing values in  $X_M$  as additional variables. Therefore, the clustering objective function (5.20) of the original FCM is rewritten as follows

$$J_m(U, V; X_M) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m d_{ik}^2(x_k, \nu_i). \quad (5.26)$$

Zeroing the gradient of  $J_m$  with respect to the missing features values of  $x_k$  leads to

$$\mathbf{x}_{jk} = \frac{\sum_{i=1}^c u_{ik}^m \nu_{ji}}{\sum_{i=1}^c u_{ik}^m} \quad \text{for } 1 \leq j \leq s, 1 \leq k \leq n. \quad (5.27)$$



By this manner, the missing values are imputed by the *weighted means* of all cluster centers during each iteration.

4. **Nearest prototype strategy FCM (NPSFCM).** The nearest prototype strategy (NPS) can be considered as a simple modification of the OCS algorithm. The idea is to replace the missing value by the corresponding feature value of the nearest cluster prototype during each iteration, yielding to the following expression

$$\mathbf{x}_{jk}^{(l)} = \nu_{ji}^{(l)}, \quad D_{ik} = \min\{D_{1k}, D_{2k}, \dots, D_{ck}\}, \quad (5.28)$$

to update the missing values based on the  $D_{ik}$  between the incomplete datum and the prototype at each iteration  $l = 1, 2, \dots$ .

### 5.4.3 Cluster validity indices

Important features for a good clustering method/algorithm are the ability to generate clusters which maximize inter-cluster distances (good separation) while minimizing intra-cluster distances (high homogeneity) and achieving high connectedness between neighboring data points (patterns). This subsection addresses questions such as “*how many clusters are there in the data?*” (patterns existence) and “*how good are the partitions?*” (quality of clustering). The number of clusters is usually determined by *cluster validity indices* (CVIs) [233]. In general, a CVI is an important matter in clustering analysis, in order to determine the quality of clustering [268]. Therefore, an optimal number of clusters can be decided based on the largest or smallest CVI value, depending on the validity indices.

A CVI represents two indicators, namely *compactness* and *separation*. Compactness indicates the concentration of data points (pattern) in a specific cluster, and it is measured by the distance between each data point (pattern) and its centroid (prototype) in the same cluster. The smaller the distance, the better the compactness of the cluster. Separation indicates the degree of separation between clusters, and it is measured by the distance between cluster centroids (prototypes). In this case, the largest the distance, the stronger the separation of clusters. Therefore, the balance between the two indicators is important for designing the CVIs [269].

A review of CVIs used in different clustering contexts is presented next.

#### Cluster validity indices for HC

To determine the optimal number of clusters in the data points by the HC algorithms, there are four validity indices which have to be used simultaneously at each step [270, 271]:

- Root-mean-square standard deviation (RMSSTD) of the new cluster.
- Semi-partial R-squared (SPR).
- R-squared (RS).
- Distance between two clusters (CD).

The root mean square standard deviation (**RMSSTD**) index is a statistical evaluation method used to measure the homogeneity within clusters at each step of a HC algorithm. RMSSTD is the pooled standard deviation of all the attributes/features within each cluster. At any hierarchical step, larger values of RMSSTD indicate that the clusters are not well separated (less homogeneity). When a new clustering scheme is formed at a given step in the HC algorithm, this is an indication of worsening. Whilst, the lower the value of RMSSTD, the better the separation of clusters (high homogeneity).

The sum of squares of the data points  $\mathcal{D}$  in a  $d$ -dimensional space is defined as

$$SS = \sum_{i=1}^n (x_i - \bar{y})^2 = \sum_{i=1}^n \sum_{j=1}^d (x_{ij} - \bar{y}_j)^2, \quad (5.29)$$

where  $\bar{y}$  is the mean of the data points. Along with this, some additional symbolism used is:

- $SS_w$ : The sum of squares within a cluster.
- $SS_b$ : The sum of squares between clusters.
- $SS_t$ : The total sum of squares of the whole data set (*i.e.*,  $SS_t = SS_b + SS_w$ )

Then the RMSSTD is given by

$$\begin{aligned} RMSSTD &= \left( \frac{SS_w}{d(n-k)} \right)^{1/2} \\ &= \left[ \frac{\sum_{i=1}^k \sum_{x \in C_i} \sum_{j=1}^d (x_j - \mu_{ij})^2}{d(n-k)} \right]^{1/2}, \end{aligned} \quad (5.30)$$

where  $\mu_{ij}$  is the  $j^{th}$  component of  $\mu_i$  which is the mean of points in  $C_i$ , *i.e.*,

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x.$$

At any hierarchical step, in order to form a new cluster through the merging process, the **SPR** index is defined as the difference between the pooled  $SS_w$  of the new cluster and the sum of pooled  $SS_w$ 's of clusters linked for obtaining a new cluster (loss of homogeneity), divided by the  $SS_t$  for the whole data set. Therefore, SPR measures the loss of homogeneity after merging the two clusters at a given step in the HC algorithm. If the SPR value is zero (*i.e.*, the loss of homogeneity equal zero) then the new cluster obtained through the merging process is from two perfectly homogeneous clusters. While, if the SPR value rises then the new cluster is obtained through merging two heterogeneous clusters.

The **RS** index of the new cluster obtained through the merging process is the ratio of  $SS_b$  to  $SS_t$ . Therefore, this index is used to measure the degree of difference between clusters (*i.e.*, (dis)similarity measure). If the value of RS is close to zero, that indicates no difference between clusters or a non-significant difference. On the other hand, when RS value is close to 1, there is an indication of a significant difference between clusters. The RS index is given by

$$\begin{aligned} RS &= \frac{SS_b}{SS_t} \\ &= \frac{SS_t - SS_w}{SS_t} \\ &= \frac{\sum_{x \in \mathcal{D}} \sum_{j=1}^d (x_j - \bar{y}_j)^2 - \sum_{i=1}^k \sum_{x \in C_i} \sum_{j=1}^d (x_j - \mu_{ij})^2}{\sum_{i=1}^n \sum_{j=1}^d (x_{ij} - \bar{y}_j)^2}. \end{aligned} \quad (5.31)$$

Finally, the distance between two clusters **CD**, is an index to measure the (dis)similarity between two clusters that are merged at a given step in the HC algorithm. Therefore, this index depends on the linkage metric that is used to measure the distance (*i.e.*, SLM, ALM, CLM). Using the above four indices, the number of clusters in HC algorithms can be determined.

### Cluster validity indices for HPC

The most commonly used validity indices in the HPC algorithms (*i.e.*,  $K$ -means algorithm) are:

- **Dunn's Index (DI)**: The main aim of this index is to reach the optimal number of clusters, which is characterized by the most compact and well-separated clusters (*i.e.*, maximizing inter-cluster distances and minimizing intra-cluster distances) [272]. The Dunn's index is defined as

$$DI = \min_{1 \leq i \leq c} \left\{ \min_{1 \leq j \leq c, j \neq i} \left( \frac{d(C_i, C_j)}{\max_{1 \leq k \leq c} \Delta(C_k)} \right) \right\}, \quad (5.32)$$

where  $d(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y)$  defines the distance between clusters  $C_i$  and  $C_j$  (inter-cluster distance);  $\Delta(C_k)$  represents the diameter (diam) of cluster  $C_k$  (intra-cluster distance), where  $\text{diam}(C_k) = \max_{x, y \in C_k} d(x, y)$ , and  $c$  is the number of clusters. By this index, the number of clusters that maximizes DI is taken as the optimal number of clusters.

- **Davies-Bouldin Index (DBI)**: It is the average similarity between each cluster and its most similar one [273]. The DBI is defined as

$$DBI = \frac{1}{c} \sum_{i=1}^c \max_{1 \leq j \leq c, j \neq i} \left\{ \frac{\Delta(C_i) + \Delta(C_j)}{d(C_i, C_j)} \right\}, \quad (5.33)$$

where  $i, j$  are indices of clusters which come from the same partitioning,  $d(C_i, C_j)$ ,  $\Delta(C_i)$ ,  $\Delta(C_j)$  and  $c$  are defined as the DI index. Contrary to the DI index, DBI must be minimized to derive the optimal number of clusters.

### Cluster validity indices for FPC

There are two categories of CVIs for the FPC algorithms. The *first category* only uses the membership values,  $u_{ik}$ , from the membership matrix ( $U$ ). On the other hand, the *second category* uses both the original dataset and the membership values, therefore, geometrical properties of the original data are taken into account besides the fuzzy degrees of membership.

#### *Validity indices involving only the membership values*

Commonly used CVIs in this category are the partition coefficient index (PCI) and partition entropy index (PEI) proposed by Bezdek [274]:

- **The partition coefficient index (PCI)**: It evaluates the compactness by measuring the amount of overlap between clusters. The PCI is defined as

$$PCI = PCI|_{[u],c} = \frac{\sum_{i=1}^c \sum_{k=1}^n u_{i,k}^2}{n}, \quad (5.34)$$

where  $c$  and  $n$  are the number of clusters and data set respectively, and  $u_{ik}$  is the membership degree of data point  $x_k$  in the fuzzy cluster  $c_i$ . The PCI takes values in the range  $1/c \leq PCI \leq 1$ . Therefore, the optimal number of clusters ( $c$ ) is achieved when the PCI value is maximized. Moreover, a high PCI value indicates a good compactness of a cluster.

- **The partition entropy index (PEI):** Bezdek proposed the PEI based upon the logarithmic form of the PCI; therefore, the PEI is defined as

$$PEI = PEI|_{[u],c} = -\frac{1}{n} \sum_{i=1}^c \sum_{k=1}^n u_{i,k} \log_a(u_{i,k}), \quad (5.35)$$

where  $a \in (1, \infty)$  is the base of the logarithm,  $u_{i,k} \log_a u_{i,k} = 0$  whenever  $u_{i,k} = 0$ . The PEI takes values in the range  $0 \leq PE \leq \log_a(c)$ , where  $c$  is the number of clusters. Contrary to the PCI, the optimal number of clusters  $c$  is achieved when the PEI value is minimized.

There are many drawbacks of this category of CVIs such as deviating of connection to the geometrical properties of the data set and the monotonic increasing tendency when increasing the number of clusters. Besides, the data set is difficult to handle when there is overlapping between clusters.

***Validity indices involving the membership values and the original data set***

In this category, the CVIs take the geometrical properties (*i.e.*, the structure) of the original data into account besides the membership values in order to avoid the disadvantages of the first category. Common CVIs in this category are:

- **The Xie-Beni index (XBI):** Xie and Beni proposed XBI as an index of fuzzy partitions. It consists in the ratio of total compactness (C) to the separation (S) of the fuzzy partition (*i.e.*, ratio of the total variation of the partition and the centroids ( $U, V$ ) and the separation of the centroids) and it is defined as

$$XBI = \frac{C}{S} = \frac{\sum_{i=1}^c \sum_{k=1}^n u_{i,k}^2 \|x_k - \nu_i\|^2}{n (\min_{i \neq j} \|\nu_i - \nu_j\|^2)}. \quad (5.36)$$

Therefore, the optimal number of clusters  $c$  is obtained by finding the fuzzy partition with the lowest XBI [275].

- **Compose Within and Between Scattering index (CWBI):** It was proposed by Rezaee et al. [276], as an index of fuzzy  $c$ -partitions in order to assess the average compactness and separation of fuzzy partitions and reach the optimal number of clusters ( $c$ ). The CWBI is defined as

$$CWBI = \alpha Scat(c) + Dis(c), \quad (5.37)$$

where  $\alpha$  is a weighting factor to counterbalance both terms in an appropriate manner which equals a distance functional  $Dis(c_{\max})$ , where  $c_{\max}$  represents the maximum number of input clusters; and  $Scat(c)$  indicates the average scatter (variation) for  $c$  clusters (*i.e.*, measures the compactness of the clustering), which is defined as

$$Scat(c) = \frac{\frac{1}{c} \sum_{i=1}^c \|\sigma(\nu_i)\|}{\|\sigma(X)\|}, \quad (5.38)$$

where  $\|x\| = (x^T \cdot x)^{1/2}$ ;  $\sigma(X)$  denotes the variance of data (pattern) set, which is defined as

$$\sigma(X) = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2, \quad \bar{x} = \sum_{k=1}^n \frac{x_k}{n} \quad \forall x_k \in X,$$

while,  $\sigma(\nu_i)$  denotes the fuzzy variation of cluster  $i$ , which is defined as

$$\sigma(\nu_i) = \frac{1}{n} \sum_{k=1}^n u_{ik} (x_k - \nu_i)^2.$$

Therefore, the smaller the value of  $Scat(c)$ , the better the compactness of the clustering. On the other hand, the distance function  $Dis(c)$  measures the total scattering separation between clusters, which is defined as

$$Dis(c) = \frac{D_{\max}}{D_{\min}} \sum_{i=1}^c \left( \sum_{j=1}^c \|\nu_i - \nu_j\| \right)^{-1}, \quad (5.39)$$

where  $D_{\max} = \max\{\|\nu_i - \nu_j\|\}$ ,  $D_{\min} = \min\{\|\nu_i - \nu_j\|\}$ ,  $i, j \in \{2, 3, \dots, c\}$ . The smaller the value of  $Dis(c)$ , the better the separation of clusters. Therefore, the optimal number of clusters  $c$  is obtained by finding the fuzzy partition with the lowest value of the CWBI.

## 5.5 Clustering and local modeling

### 5.5.1 Why local modeling?

Systems in biochemical and chemical engineering are characterized by noisy data and significant complexity, uncertainty and nonlinearity [277], which makes modeling a hard task. One solution to deal with this problem is to divide a complex system into simpler subsystems that can be analyzed

individually. Then a specific model for each subsystem can be identified to approximate the complex system locally: the so-called *local models*. Their combination by smooth interpolation yields to the complete *global model*.

There are several advantages of the *local modeling* approach, such as:

1. The possibility of getting several local models with a simpler structure instead of a single complex model of the system, which may be beneficial for enhancing the interpretability of the model [278] and dynamic changes, and also the model accuracy.
2. The possibility of getting similar regions/patterns if there are several local models and characterize them in search of similar dynamics and similar characteristics. Therefore, the system response could be better understood and explained with this approach.

There are many studies in which the local modeling has been successfully applied for the modeling of complex systems. In [279], the local approximation approach was used to predict the future behavior of a chaotic time series throughout state space. Despite the chaotic system (*e.g.*, chaotic time series) can show complex external behavior, local models have demonstrated an accurate approximation of chaotic dynamical systems. As well, local linear models have been used in many studies on time series prediction problems and often exhibited more accurate predictions than global models [280, 281]. In [282], local linear models based on an unsupervised neural network algorithm, Kohonen self-organizing feature mapping (SOFM) have been successfully used for modeling a non-linear time series (chaotic system).

In [283], local models have been employed with a nonlinear MPC controller for a simulation study of a batch fermentation process. The local models have been constructed to be valid during different parts of a batch cycle. These models have proven that they can represent the whole batch cycle in a simpler way than the global model. In [284], local models have been used as local learning algorithms instead of training a complex system with all data samples. These models have demonstrated their ability to improve the performance of an optical character recognition problem in a simpler way than the complex system. In [285], extended radial basis function (RBF) networks have been used where each local model is represented as a linear function of the input. This approach has been shown high success in control problems. In addition, local models based on FPC algorithms have been successfully applied in many studies in control problems [286, 287].

The multiple local dynamic models have been applied in calibration algorithms for CGM systems in T1D for improving the description of the complexity of the plasma glucose to interstitial glucose dynamics. This approach has demonstrated its efficiency in accuracy improvement of plasma glucose estimations [288].

### 5.5.2 Local modeling techniques

The main goal of local modeling is, firstly, to find a set of  $c$  **local models** in order to reduce the complexity of the system, and secondly, to combine these models by smooth interpolation into a complete **global model** that represents the output of the system [289]. As a result, each local model is simpler than the global model and can be analyzed and interpreted individually. Each local model will be representative of some regions of the input space with similar characteristics. The validity region of each local model can be defined by a weighting factor in the range  $[0, 1]$  that indicates the degree of the validity of each local model [290].

Thus, key concepts in local modeling are **partition** and **interpolation** [291]. The greater the number of local models chosen, the easier it will be for each of them. However, this property faces the so-called “*curse of dimensionality*”. Regarding the interpolation, a smooth transition from one local model to another may be needed as the system moves in its operating space. Unexpected transitions between local models may give some hints on the unknown underlying dynamics of a system. It could be used, for example, for interpreting specific phenomena in some applications.

#### Partitioning of the modeling problem

There are several techniques for the identification of the local models of the system such as:

1. The **gain scheduling control** or **local model network** (LMN) technique. The idea for using this approach is to find multiple local models for the whole operating range of the controlled process. Then, these models are smoothly combined by using the validity function to provide a global model of the system [292].
2. **Fuzzy systems** techniques that have proven to be a significant tool for modeling complex systems. There are many examples of fuzzy systems like rule-based fuzzy systems [293] and fuzzy linear regression models [294]. For example, in rule-based fuzzy systems, a system can be defined as a combination of “**If-Then**” rules with fuzzy relation or fuzzy predicates. These fuzzy rules can be written in the form:

**If** *antecedent proposition* **then** *consequent proposition*,

where the antecedent is always a fuzzy proposition and composed of linguistic variables. Depending on the structure of the consequent, there are three main fuzzy models [295]:



- *Linguistic fuzzy model* [293, 296]: both the antecedent and the consequent are fuzzy propositions.
  - *Takagi-Sugeno fuzzy model (TS)* [297, 298]: the antecedent is a fuzzy proposition; the consequent is a crisp function (local model).
  - *Fuzzy relational model* [299]: one specific antecedent proposition to be correlated with many different consequent propositions through a fuzzy relation.
3. **Clustering techniques**, as introduced earlier in this chapter, are also a common technique for extracting local models from a complex dataset as shown in Fig. 5.6. The output of the clustering techniques to the systems modeling are distinct clusters, each cluster represents a local model [300]. Therefore, similar data points/patterns are represented by the same local model. Thus, the final global model of the whole dataset could be modeled by integrated all the local models. The use of clustering has many advantages versus the other possible techniques for local modeling such as:
- Clustering techniques find automatically clusters of similar behavior; therefore, it can be easier for interpreting the resulting models.
  - The automatic detection of similar clusters/groups by different clustering techniques makes the final global model consists of the best local model that adjusts the proposed structure.
  - Clustering techniques are more flexible offering many possibilities to define similarity measures according to the specific application.

### Local models interpolation

Likewise, there are several techniques for local models interpolation. Besides clustering, that is a technique devoted to partition, most of the local models partitioning techniques have their own tools to ease the process from the partitioning step to models integration. Of course, a different technique can be used for partitioning and for integration, but most authors use the same technique in both steps. Therefore, the two mainly used in literature are:

1. Once a set of models is defined by gain scheduling or local model network technique (one for each predefined operating point), these models can be smoothly combined by using a **validity function** whose values are guided through a **scheduling variable** to provide a global model of the system [291].

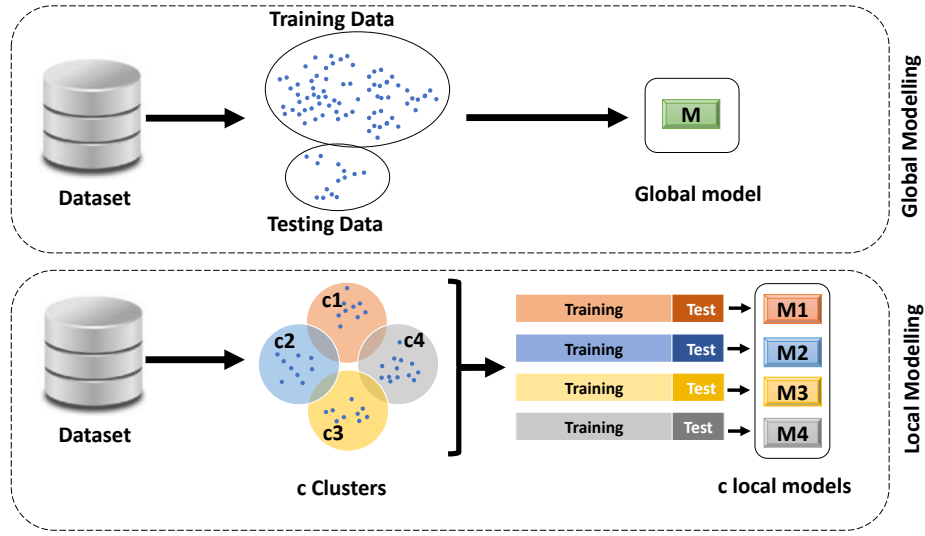


Figure 5.6: Diagram of global model versus local models with clustering technique.

2. A **fuzzy inference** algorithm is needed to compute an output value given an input value when fuzzy models are used. In the case of fuzzy relational models, fuzzy relational calculus (relational composition) can be used [301]. Fuzzy linguistic models can be rewritten as fuzzy relational models or use the Max-Min Mamdani inference [296]. The output of a fuzzy model is a fuzzy value, but in many applications a crisp output is desired and a **defuzzification** method must be used. For example, Mamdani inference scheme often use **center of gravity** defuzzification, whose extension when singletons are used in the consequents can be applied as inference mechanism to Takagi-Sugeno fuzzy models [302].

Both approaches are very similar if the  $c$  local models available in LMN and the consequents of the  $c$  rules in TS have the same form  $LM_i$  (for example, any regression model), resulting in an integrated global model  $GM$ :

$$GM = \sum_{i=1}^c \mu_i \cdot LM_i, \quad (5.40)$$

where the combination of the sum and  $\mu$  can represent the models combination and the validity function in gain scheduling, or the inference process in fuzzy models, respectively. The use of clustering in the partition step gives some advantages, because the  $\mu$  needed for integration in (5.40) can be directly related to the fuzzy membership functions obtained for each cluster in the clustering process. Therefore, the models can smoothly combined

by using the fuzzy memberships as validity or weighting functions to provide a global model [303].

## 5.6 Clustering applications in type 1 diabetes

Machine learning is representing an important branch of artificial intelligence and has been widely applied in T1D in order to identify glyceic patterns and understand the input data so that it can make decisions and predictions based on it. Nowadays, both types of machine learning algorithms (*i.e.*, supervised and unsupervised learning) are used more than ever before, for automatic analysis of high dimensional of the diabetic dataset.

Clustering is representing a common form of *unsupervised learning* as stated before to discover the hidden structure of data and has been applied in many studies in T1D. Clustering techniques were applied in [304] to identify different BG behavior profiles of 10 T1D patients with different scenarios: days without exercise, days with exercise without corrective actions, and days with exercise with corrective actions (snacks before and after exercise and basal insulin reduction). The normalized compression distance (NCD) [305] was used as a metric of the similarity measure. The authors were able to identify different BG behaviors on different days for a same patient, with potential applications to better management of intra-patient variability for an adequate BG control.

Clustering technique was also applied in [306] as a part of a clinical decision support system (CDSS), which combines a classifier of different glyceic profiles with a predictor of BG levels. The clustering technique was used for analyzing the data and building a classifier to identify different glyceic profiles automatically. Therefore, the hybrid method in this study was based on the use of a classifier together with search-based-method (*i.e.*, grammatical evolution method), to identify glyceic profiles in the first step and then predicting glucose levels for each profile (*i.e.*, local prediction). The profiles were simulated through the UVA/Padova T1D simulator [168] yielding to time series from historical data (CGM measurements, CSII, and CHO intake) of 20 T1D virtual patients. Also, NCD was used as a metric of the similarity measure in the clustering algorithm. Although this approach was not working in real-time (*online*) in this study, however, the feasibility of identifying different BG profiles and generating personalize BG prediction models through CDSS was demonstrated through simulated data. The obtained results were reported as the mean of the performance metric (*e.g.*, MSE) of the 20 profiled prediction models, without having to identify a global model for the multiple prediction models due to good classification results (99.7% classification success with the studied dataset). An average MSE of 18.21, 22.20, 22.82 and 23.24 mg/dL was reported for 60-, 120-, 150- and 180-min PH, respectively.

In [307], cluster analysis was applied to identify different profiles of diabetes self-management. A sample of 156 youth (10-16 years) with T1D and their maternal and paternal caregivers was used in this study. They were administered a modified version of the diabetes self-management profile (DSMP), which is a widely used, structured interview in order to assess diabetes self-management behaviors. Cluster analysis was performed independently by SPSS two-step cluster analysis [308] based on youth, maternal, and paternal reports in order to classify subjects based on their profiles of scores in five diabetes self-management areas (*i.e.*, insulin administration, meal planning, exercise, BG testing, and self-care adjustments). This study demonstrated the feasibility of using a multi-component assessment for the development of an experimentally derived classification of self-management patterns. As well, an extended study on the identification of patterns of diabetes self-management in T1D was conducted in [309], where a sample of 239 youth (9-11 years) with T1D and their parent's caregivers was used. HbA1c was used as the indicator of glycemic control in this study. Two-step cluster analysis was performed to classify different patterns of self-management. This study successfully demonstrated that the different patterns are related to glycemic control (HbA1c) in youth with T1D.

Clustering was also applied in Compositional Data Analysis (CoDA) in order to identify groups of similar compositions, which were formed from the discretization of daily glucose profiles into time spent in different glucose ranges based on the different BG levels [310]. Data for 6 T1D patients for eight weeks were expressed in coordinates, where CGM and CSII data were used. Then the k-means algorithm was applied to the coordinates in order to classify different daily profiles. This approach successfully demonstrated the ability to classify the daily profiles/patterns into different groups that reflect the behavior of the patient.

Classification technique is representing a common form of *unsupervised learning* technique besides the regression and also has used recently in T1D [182]. Classification technique was applied in [311] to classify and predict the future postprandial glycemic status to adjust the meal insulin bolus dose. In this study, the extreme gradient-boosted tree (XGB) algorithm [312] was used in order to classify the expected postprandial glycemic status (*i.e.*, hyperglycemia, euglycemia, or hypoglycemia). The algorithm was performed based on the information of CSII, pre-prandial CGM measurements, and CHO intake estimates for 100 virtual patients of T1D. The proposed algorithm has shown good accuracy at classifying the postprandial glycemic status. As well, the proposed algorithm has demonstrated its ability to improve glycemic control and can be used for different purposes (*e.g.*, glucose alarms, decision support on insulin therapy).

Through previous studies, it is clear that the use of machine learning in the diabetes field, especially in T1D, continues to increase with time due to

the increasing data size and its complexity. The use of clustering technique in T1D has been demonstrated its ability to analyze and classify different BG profiles/patterns, especially when using the simulation data. Furthermore, clustering algorithms play an effective and vital role in simplifying complex systems. However, clustering a real/clinical diabetic dataset is still a major challenge. Where this kind of data is characterized by high variability and different glycemic responses, especially during exercises and stress. Moreover, diabetic dataset under free-living conditions may have missing data due to sensor or technical errors. In addition, choosing an appropriate distance metric for measuring the similarity between different data points is also a main challenge in the clustering algorithms.

In this thesis, clustering techniques will be used to analyze, classify, and cluster different BG patterns of specific periods such as postprandial and nocturnal periods of T1D individuals in the clinical and simulated dataset. Clustering methods will be used to identify and cluster different glycemic patterns as local behaviors for reducing the complexity of diabetes time-series data. Local behaviors will then be used in modeling through *seasonal local models* corresponding to each cluster. Where the similarity behaviors represent the cornerstone of using a concept of *seasonality* in time series dataset. Finally, by integrating the local models through an appropriate method, the *global model* will be obtained to represent the whole data.

**Part III**  
**Contributions**

## Chapter 6

# Seasonality and glucose prediction

This chapter introduces the study of seasonal phenomena in CGM time series to build seasonal models for glucose prediction. The starting hypothesis is that, under some similarity assumptions, seasonality properties can stem from preprocessed CGM time series, that can be exploited to improve prediction of key responses like postprandial response. The chapter starts in Section 6.1 with the main rationale of this study. Section 6.2 describes the experimental data and methodology used. Section 6.3 displays the results of our study comparing seasonal and nonseasonal models in order to analyze the effect of seasonality in the model prediction accuracy. Then the effect of using exogenous inputs in both models is addressed. Section 6.4 displays the conclusions and margins for further improvement.

### 6.1 Rationale

A CGM time series includes many dynamic phenomena, being one of the challenges in their analysis to detect patterns that dominate the series behavior. These patterns may vary over time, because of the glycemic variability, which is one of the most challenging issues in the BG control of patients with T1D, especially in the postprandial period and during exercise. Several factors can affect BG concentrations, such as the meal composition, the type, duration and intensity of exercise, the amount of insulin infusion, changes in insulin sensitivity (*e.g.*, circadian hormone secretions, illness) and physiological stress, among others. The glucose response is the result of the complex interaction between all these variables. Therefore, it is rational to explicitly take into account possible changes that could affect the CGM behavior, and therefore, the glucose prediction.

As we have mentioned in Chapter 3, Section 3.3, linear time series models or linear EDMs, also known as data-driven non-parametric models, have

been widely used in the context of glucose prediction, especially in T1D. Despite these models cannot explain the underlying physiology (*i.e.*, they cannot be used to demonstrate the mechanism of glucose-insulin metabolism in detail), they are, however, more attractive than physiological models in the sense that they are simple to implement, computationally cheap, and can provide good predictions about future glucose behavior. EDMs are also widely used in other domains such as business and economic time series. A particular characteristic in these domains is **seasonality**, that is, the existence of regular patterns of changes and fluctuations that repeat periodically [313].

This chapter explores the extension of the concept of *seasonality* for glucose prediction with a proof-of-concept study. The main rationale is that preprocessing of CGM time series (and available additional information) may translate daily events into seasonal phenomena. For instance, glucose concentration tends to peak and then decline in a characteristic way after a meal intake in a particular scenario. In this case, a new preprocessed family of time series can be built from the original CGM data by concatenating postprandial periods (PPs) of fixed length where similarity of behaviors is expected, according to some metrics, which would theoretically produce seasonal time series as shown in Fig. 6.1. This allows for the application of seasonal models that exploit this similarity for more accurate predictions and longer prediction horizons.

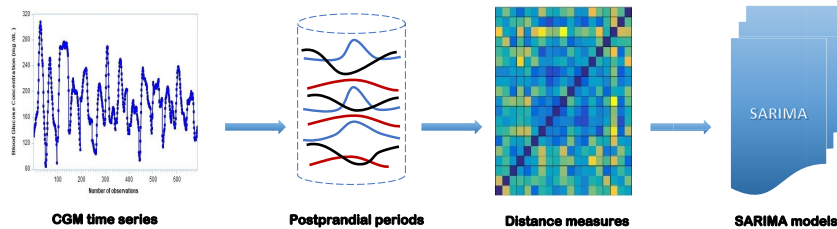


Figure 6.1: Diagram of seasonal CGM time series models.

## 6.2 Research Design and Methods

### 6.2.1 Clinical dataset

#### Study design and subjects

A subset of the data from the CLOSEDLOOP4MEAL study evaluating a new closed-loop system was used for this study. Two closed-loop studies were considered. The protocol was approved by the Ethics Committees of the *Clinic University Hospital of Valencia* [7]. The study was designed as an in-hospital approach fulfilling the regulatory conditions applied in Spain



<b>ID</b>	<b>Duration of Diabetes</b> (year)	<b>BMI</b> ( $kg/m^2$ )	<b>Gender</b>	<b>HbA1c</b> (%)
RAH	38.3	25.8	Female	8.2
MEBR	17.8	25.7	Female	6.9
LFB	13.5	27.7	Female	8.0
NAC	24	30.2	Female	8.1
CGG	17	26.0	Female	8.0
LBG	43	28.7	Female	8.4
MLL	20	31.8	Male	8.3
ACR	21	23.6	Female	7.1
JLS	21	29.6	Female	8.2
JMC	16	26.9	Male	8.2
<b>Average</b>	<b>23.16</b>	<b>27.6</b>	<b>8 Female</b> <b>2 Male</b>	<b>7.94</b>

Table 6.1: Demographic information for subjects.

to this sort of projects before moving to at-home settings. Subjects were eligible to participate if they were between 18 and 60 years of age, had a diagnosis of T1D for at least 1 year, HbA1c between 6.0% and 8.5%, and were on CSII for at least 6 months. Hypoglycemia unawareness was ruled out using a validated questionnaire.

Overall, 10 subjects with T1D completed the clinical closed-loop (CL) study. Table 6.1 shows the characteristics of the participants of the CL studies. Each subject underwent an in-hospital 8-h standardized mixed meal test (60 g CHO) on two occasions of CL (CL1 and CL2). After a meal announcement, an augmented bolus was given, followed by manual adjustments of the basal rate every 15 min according to a CL controller. Patients were instructed to wear a CGM device and follow a structured SMBG protocol during a 6-day period before the first meal test. Data from CGM and SMBG were used to obtain individual estimated insulin sensitivity and a pharmacokinetic/pharmacodynamic model resulting in the calculation of the following parameters: insulin-to-CHO ratio, sensitivity factor, basal insulin needs, and insulin-on-board (IOB).

### Study devices

Insulin infusion was carried out with the Paradigm Veo<sup>®</sup> insulin pump (Medtronic MiniMed, Northridge, CA) and CGM using Enlite-2 sensors<sup>®</sup> (Medtronic MiniMed, Northridge, CA). Two CGM devices were inserted at least 24 h before the meal tests, to improve performance and avoid missing data and problems related to sensor drift. For safety and regulatory reasons, two sensors were used in this phase of development to ensure the algorithm to be fed with the secondary CGM in case of sensor failure. In all

subjects, calibration of CGM was performed using the Contour<sup>®</sup> Next Link (Ascensia Diabetes Care Holdings AG, Basel, Switzerland, formerly Bayer). Glucose concentrations were also measured every 15 min with a reference method YSI 2300 Stat Plus Glucose Analyzer (YSI 2300; YSI Incorporated Life Sciences, Yellow Springs, OH).

The CL system was based on a novel *sliding mode reference conditioning* (SMRC) glucose controller built in a PC. Glucose values from the two CGM devices were introduced manually every 15 min into the controller interface. Manual operation greatly simplified regulatory approval of the system in the first submission of this type of studies in Spain. The system defined a primary and secondary CGM device automatically, based on an accuracy analysis (*absolute relative difference*, ARD, from reference) before the start of the CL controller. Only data from the primary CGM were used, except in case of malfunction, resulting in an automatic switch to the secondary CGM. Malfunction was defined as an ARD between the CGM reading and the plasma glucose (PG) reference greater than 40% at one time point or ARD greater than 30% in two consecutive periods. The insulin infusion rate for the next 15-min time interval was calculated by the controller and manually set by the attending physician/nurse.

The glucose controller used consists of a feed-forward action plus two control loops [314, 7]:

- (a) The feed-forward action is an augmented bolus calculated based on meal announcement.
- (b) The inner control loop is a PID-type controller designed to drive the measured glucose to a target value. It is tuned from the insulin pump settings.
- (c) The outer control loop is based on SMRC and modulates the glucose target value on the estimated IOB, minimizing the impact of controller overcorrection resulting in late hypoglycemia. When the estimated IOB is beyond prespecified limits, a high-frequency discontinuous signal is generated and filtered inducing smooth changes in the target glucose value so that IOB constraints are not violated. Thus, this outer loop acts as a safety supervisory loop. The IOB estimation is calculated using a previous population pharmacokinetic model [314]. Finally, the IOB limit is estimated individually based on 1-week CGM data and previous insulin pump settings. Compared to insulin feedback (IF), also used in combination to PID controllers, SMRC is expected to induce an early pump shut-off due to the augmented bolus administration with a potential benefit in PP control, compared to a later effect by IF driven by the estimated plasma insulin concentration.

### Mixed meal tests

Before the meal test, fasting subjects were admitted to the clinical research unit at 08:00 AM. In a sitting position, two venous lines were prepared, one for arterialized venous blood sampling and the other for insulin/glucose infusion, if required. To ensure comparable metabolic conditions between studies, where appropriate, subjects received an intravenous infusion of regular human insulin in a feedback manner, or glucose, to maintain PG at 90-100 mg/dL until the beginning of the studies. At 12:00 h ( $t=0$ ), a standard mixed meal (530 Kcal, 60 g CHO, 45.3% CHO, 24.2% protein, 30.5% fat) was consumed in 15-20 min. At the same time, insulin was administered following the randomization protocol (CL), and PG was monitored for the ensuing 8 h until the end of study at 20:00 h (time 480 min). If PG fell below 70 mg/dL during two consecutive 15-min periods, oral glucose was administered in fixed amounts of 15 g until recovery from hypoglycemia. Despite meal size was controlled in this in-patient study, this didn't prevent the presence of high intra- and inter-individual variability. These were measured by the coefficient of variance of the area under the curve for the 8-hour duration of the study ( $CV - AUC_{8h}$ ), which was computed with the *trapezoidal rule*.

### 6.2.2 Data pre-processing

The above data set contained 20 8-h postprandial responses which were concatenated to compose the CGM time series, as illustrated in Fig. 6.2. This organization of the data can be carried out as long as the meal event is known, which was assumed here. This is the case in current hybrid artificial pancreas systems. As well, a fixed length of postprandial responses was considered, which can be considered an ideal case. This limitation will be addressed in Chapter 8.

The resulting CGM time series had a mean of 136.1 mg/dL, with a standard deviation of 48.48 mg/dL. Despite the same meal was provided, data exhibited high variability with postprandial peaks ranging from 304 mg/dL (P91) to 125 mg/dL (P42) and the incidence of hypoglycemia in some patients (P11, P51, P52, P71, P101), two of them moderate hypoglycemia (P11, P101), according to CGM values. They were non-normally distributed. Inter-individual variability measured by  $CV - AUC_{8h}$  was 21.52%, whereas intra-individual variability was 9.17%. However, the latter spanned from 3.22% (patient 6) to 18.67% (patient 9), i.e., the worst-case intra-patient variability is close to the inter-patient variability. Since only two studies per patient were available, intra-patient variability might be underestimated.

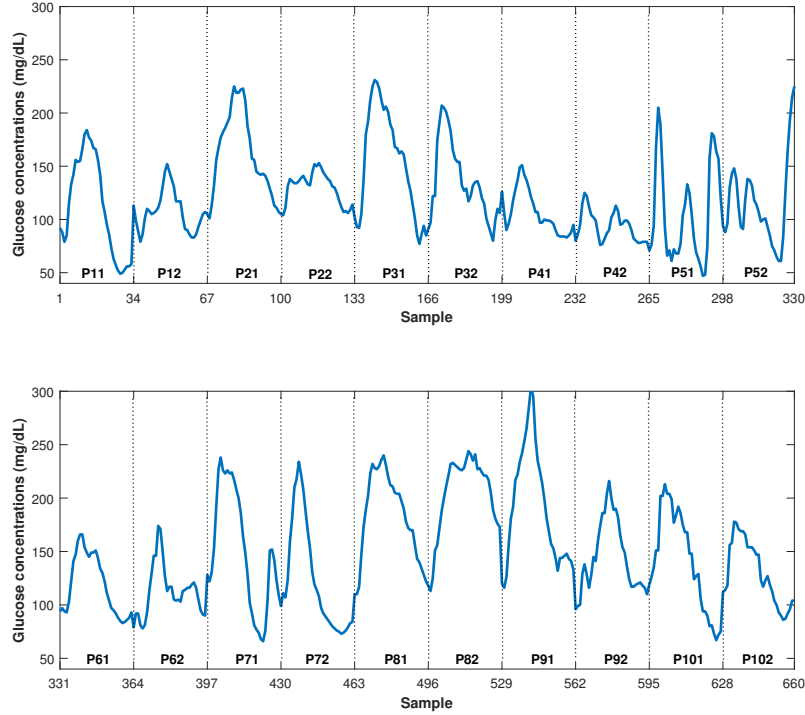


Figure 6.2: CGM time series resulting from the concatenation of twenty 8-h PPs for a same 60g CHO meal. The notation  $P_{ij}$  is used to name the different periods, where  $i$  is the number of the patient,  $i \in \{1, \dots, 10\}$ , and  $j$  is the number of the study per patient,  $j \in \{1, 2\}$ . Sampling period is 15 minutes, yielding 33 samples per PP.

Euclidean distance between each pair of PPs was also computed to analyze the similarity of time series (see Fig. 6.3), providing similar conclusions. Patient 9 is the most dissimilar among studies (green box in P91-P92), only exceeded by comparatively few yellow-red boxes outside the diagonal (between-patient comparisons). P81, P82, and P91 were the most dissimilar with the rest of periods (higher incidence of yellow-red boxes). Total basal insulin infusion in the 8-h period ranged from 5.21U (P31) to 16.40U (P71). An extended bolus computed from the patient's insulin-to-CHO ratio and open-loop basal infusion rate was also administered at meal time.

### 6.2.3 SARIMA and SARIMAX models

As mentioned previously, in an empirical dynamic model, an observation at time  $t$  is expressed as a linear combination of observations at times  $t - 1, t - 2, \dots, t - p$  (previous  $p$  measurements) by the AR component, and

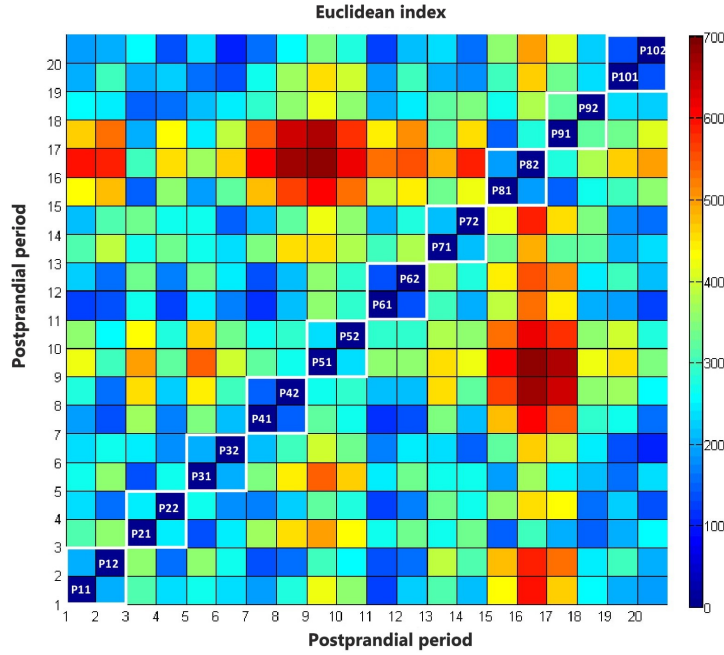


Figure 6.3: Similarity among PPs in the CGM time series as measured by the ED between paired periods. Data is shown according to the color scale in the right. White boxes in the diagonal indicate periods corresponding to a same patient.

as a linear combination of stochastic errors, also called shocks, at times  $t, t-1, t-2, \dots, t-q$  by the MA component. In a SARIMA model, SAR and SMA terms are added so that an observation at time  $t$  depends on previous observations and stochastic errors at times with lags that are multiples of the seasonality period  $s$ . In the context of postprandial glucose prediction, this means that the glucose prediction will depend not only on previous measurements for that PP, but also on previous PPs in the time series. Therefore, given a CGM time series  $\{G_t | t = 1, 2, \dots, k\}$ , a SARIMA model (4.39) can be expressed as

$$\begin{aligned} \nabla_s^D \nabla^d G_t &= c + w_t, \\ \phi_p(z^{-1}) \Phi_P(z^{-s}) w_t &= \theta_q(z^{-1}) \Theta_Q(z^{-s}) \varepsilon_t, \end{aligned} \quad (6.1)$$

where  $G_t$  is the glucose concentration at time  $t$ ,  $w_t$  is the disturbance series.

The exogenous signals that related well with a CGM time series data were used in a SARIMA model to improve the fit and prediction performance of the chosen model. Moreover, for capturing the dynamic characteristics of CGM behavior. In many statistical packages such as R and Eviews, exogenous variables are considered as *explanatory variables* into a linear regression model with a stochastic error process of certain structure. In this case, the SARIMAX model (4.41) can be expressed as

$$\begin{aligned} \nabla_s^D \nabla^d G_t &= c + \eta_r(z^{-1})E_t + w_t, \\ \phi_p(z^{-1})\Phi_P(z^{-s})w_t &= \theta_q(z^{-1})\Theta_Q(z^{-s})\varepsilon_t. \end{aligned} \tag{6.2}$$

The *Granger causality* test was used to test the null hypothesis that the endogenous variable (CGM) does not Granger-cause the dependent variables (exogenous signals) and vice-versa (*e.g.*, the lagged CSII values do not explain the variation in CGM pattern and vice-versa, according to the Granger causality test).

### 6.2.4 Identification procedure

Box-Jenkins methodology was used for model building and evaluation (Chapter (4), Section (4.3)). A LOOCV procedure was considered dividing data into training and test sets (Fig. 4.6). In order to avoid data from a same patient to appear both in training and test, data from the test patient was excluded from the training set. This resulted in 18 PPs for training and 1 for test/validation, since two CL studies per patient were available. PPs in the training set were randomly ordered at each run according to a random sequence generator (<http://www.random.org>). A stationarity analysis was first carried out with the unit-root test (ADF test). The backward-difference operator  $\nabla$  was applied to the time series as many times as necessary (integration order  $d$ ) to remove nonstationarity, if present. The ACF and PACF were used to identify the orders of the AR and MA terms ( $p$  and  $q$ , respectively), as well as identifying seasonality (seasonally differencing the time series with the operator  $\nabla_s$  if necessary). The MLE was used for parameter estimation. As well, the AIC was used for model selection. In this study, Eviews software, version 9.5, was used; in this case, the AIC is defined as

$$\mathbf{AIC}(\mathbb{M}) = \frac{1}{n}(-2L(\mathbb{M}) + 2K(\mathbb{M})), \tag{6.3}$$

where  $L$  is the value of the log-likelihood,  $K$  is the number of free parameters in the model  $\mathbb{M}$ , and  $n$  is the number of observations. Remark the scaling by  $1/n$ .

For *diagnostic checking*, ACF and PACF plots for the residuals were analyzed to test the existence of any significant spikes in the confidence interval, the LBQ test was used for testing randomness at each distinct lag and the JB test was used to test the normality of the residuals. Finally, accuracy of the model forecasting was measured with the following metrics: **MAE**, **RMSE**, and **MAPE**.

Forecasting measures	SARIMA	ARIMA
MAE (mg/dL)	34.56 (19.35)*	47.72 (24.43)
RMSE (mg/dL)	40.02 (21.62)*	55.02 (26.93)
MAPE (%)	22.02 (9.41)*	30.01 (13.05)

Table 6.2: Performance of SARIMA and ARIMA models by the mean and standard deviation, in parentheses, of different forecasting measures. \* $P < 0.05$  in all cases.

## 6.3 Results

### 6.3.1 Without exogenous variables

Both SARIMA and ARIMA models were identified for each run in the cross-validation. Fig. 6.4 shows the forecasting accuracy metrics for a 5-h PH for both cases. A high PH was initially chosen to challenge the model. SARIMA outperformed ARIMA in all metrics as shown in Table 6.2. In the following, the analysis will be restricted to MAPE since the three measures provided the same information.

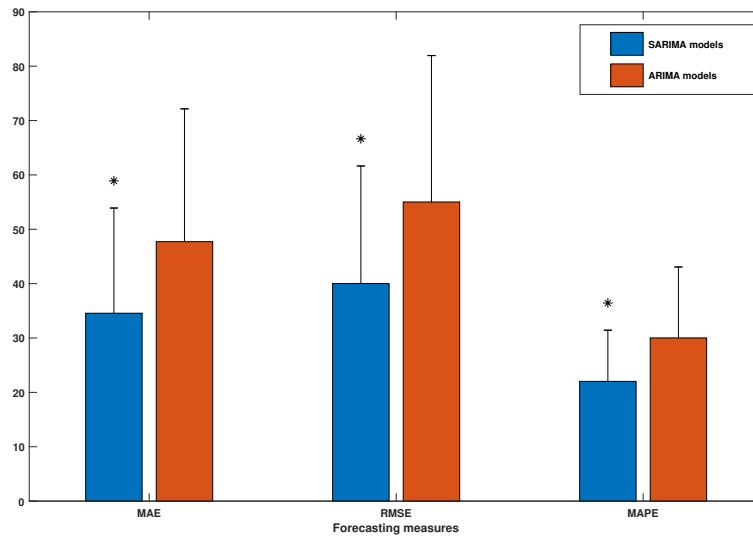


Figure 6.4: Mean and standard deviation of forecasting measures (MAE (mg/dL), RMSE (mg/dL), MAPE (%)) for the 20-fold cross-validation and a 5-h PH. \* $P < 0.05$ .

Fig. 6.5 shows the obtained MAPE as the PH increases from 30 min to 5 hours, consistently outperforming SARIMA. The identified model structure differed slightly between runs, with AR and MA orders up to 4. No time series differentiation was needed for both ARIMA and SARIMA models.

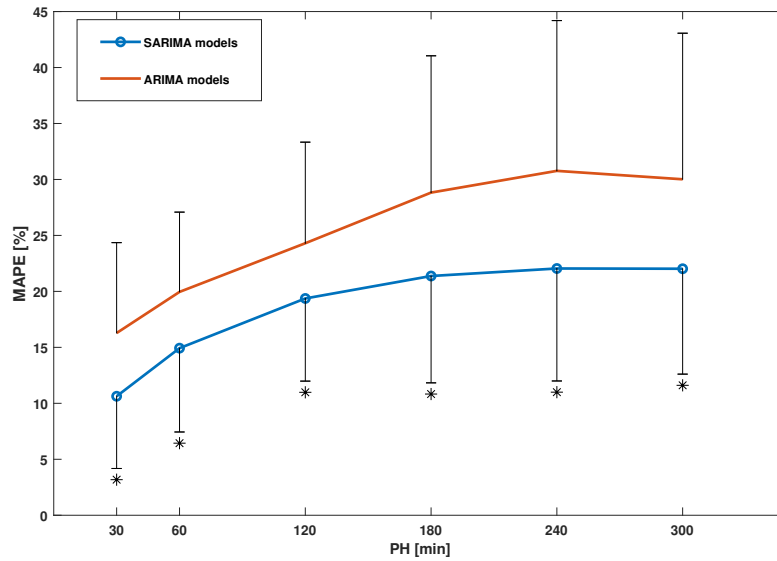


Figure 6.5: Mean and standard deviation of MAPE(%) for increasing values of the PH. \* $P < 0.05$ .

Seasonality with lag 33 (the size of the PP) was obtained in all cases, as expected. SAR and SMA orders were up to 2.

The best performing run was **Run 4**, with validation data P22. In this case, inspection of the ACF revealed data were stationary (the trend had a non-significant  $P$  value of 0.0877) and seasonal at lag 33 with a significant  $P$  value of 0.0000. Seasonally differenced data were stationary with significant  $P$  value of 0.0001, so it was not necessary to take any difference. Model SARIMA(4, 0, 4)(1, 0, 1)<sub>33</sub> was the most appropriate model, with AIC 7.9566. Table 6.3 shows the estimated model parameters using maximum likelihood estimation. All spikes in the residuals ACF were within the significance limits (*white noise*) as shown in Fig. 6.6. Table 6.4 shows the LBQ test for testing randomness at each distinct lag, also demonstrating that the residuals have no remaining autocorrelations. The tests for residual normality showed that the residuals were approximately normal (see Fig. 6.6). A MAPE of 6.73% was obtained for training data. A similar fitting was obtained for ARIMA, with MAPE 7.05%. Fig. 6.7 shows the prediction performance using validation data for a 5-h PH. A MAPE of 6.62% and a RMSE of 10.28 mg/dL were obtained for SARIMA. For ARIMA, prediction metrics were worse with MAPE 9.39% and RMSE 14.39 mg/dL, as it becomes apparent in Fig. 6.7.



Parameter	Value	SE	t-Statistic	P value
$c$	134.1109	14.70842	9.117968	0.0000
$\phi_1$	3.124690	0.008934	349.7433	0.0000
$\phi_2$	-3.783549	0.015602	-242.5068	0.0000
$\phi_3$	2.044158	0.015166	134.7894	0.0000
$\phi_4$	-0.399508	0.008156	-48.98071	0.0000
$\Phi_{33}$	0.912586	0.056899	16.03866	0.0000
$\theta_1$	-1.827147	0.307422	-5.943459	0.0000
$\theta_2$	1.124422	0.347364	3.237018	0.0013
$\theta_3$	-0.177478	0.087565	-2.026818	0.0431
$\theta_4$	0.086584	0.056154	1.541905	0.1236
$\Theta_{33}$	-0.826937	0.078243	-10.56878	0.0000
$\sigma^2$	157.4450	53.98418	2.916503	0.0037

Table 6.3: Model parameters for model  $SARIMA(4, 0, 4)(1, 0, 1)_{33}$  in best-performing Run 4, following notation in (6.1).  $\sigma^2$  is the estimate of the error variance from the MLE.

Lag	12	24	36	48
<b>Q-Stat</b>	2.8661	10.575	21.358	30.513
<b>P-Value</b>	0.239	0.719	0.723	0.801

Table 6.4: Ljung-Box Q (LBQ) test for the training residuals of Run 4 Model.

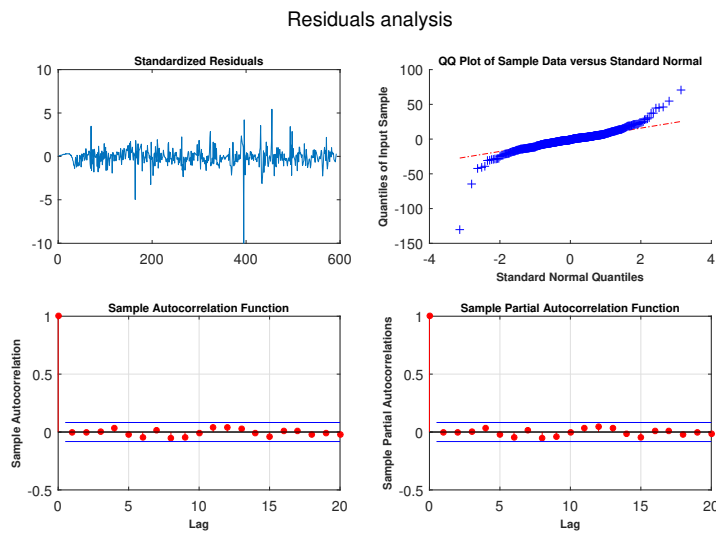


Figure 6.6: Residuals analysis for model  $SARIMA(4, 0, 4)(1, 0, 1)_{33}$  in best-performing Run 4.

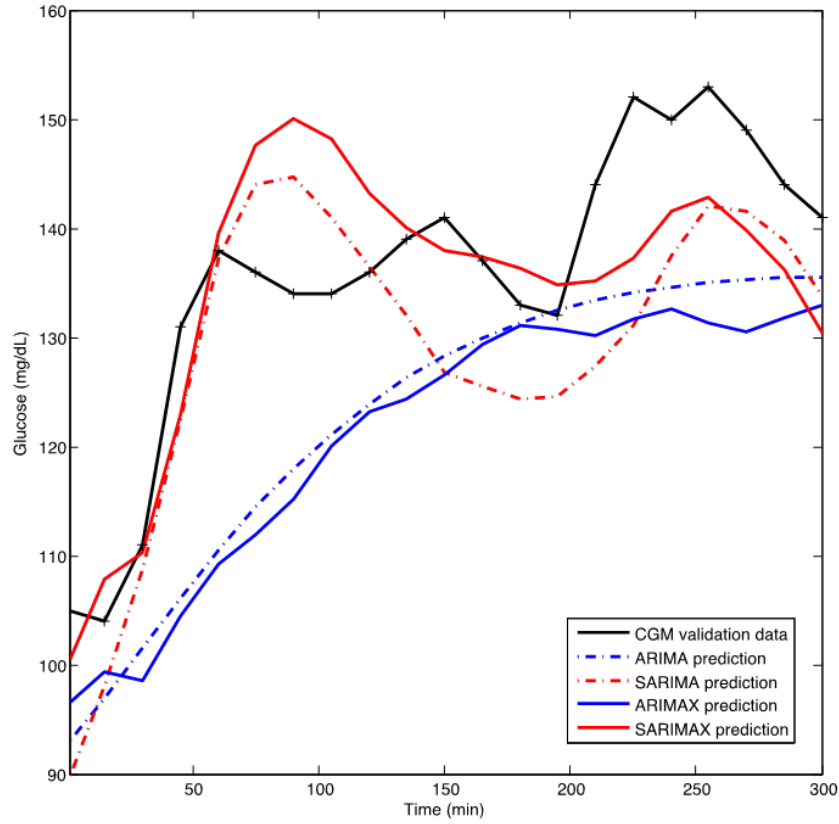


Figure 6.7: Forecasting of models  $ARIMA(4, 0, 4)$ ,  $SARIMA(4, 0, 4)(1, 0, 1)_{33}$ ,  $ARIMAX(4, 0, 4, 2)$  and  $SARIMAX(4, 0, 4, 2)(1, 0, 1)_{33}$  for Run 4 considering a 5-h PH.

### 6.3.2 With exogenous variables

The effect of considering CSII as exogenous variable for performance improvement was investigated. Future insulin infusion information can be available in applications such as MPC, during the optimization process. This analysis was carried out only for Run 4 as the best performing case, challenging further improvement. Insulin infusion signal contained bolus and basal infusion and was expressed in  $U$  per sampling period. Granger causality test was applied to test the null hypothesis that CGM does not “Granger cause” insulin infusion and vice-versa. The null hypothesis was rejected with a significant  $P$  value of 0.0146. Therefore, inclusion of insulin infusion into the model might improve performance. The order of the exogenous polynomial was computed from the cross-correlation plot and AIC, resulting in the model  $SARIMAX(4, 0, 4, 2)(1, 0, 1)_{33}$  with AIC 7.9544. Table 6.5 shows the estimated parameters for this model. The same procedure was used to derive its non-seasonal counterpart resulting in the model

Parameter	Value	SE	t-Statistic	P value
$c$	131.3957	17.53808	7.492023	0.0000
$\eta_0$	1.059158	0.307175	3.448056	0.0006
$\eta_1$	0.933659	0.376779	2.478004	0.0135
$\eta_2$	0.223623	0.323739	0.678182	0.4979
$\phi_1$	3.245163	0.006145	528.1297	0.0000
$\phi_2$	-4.107178	0.007966	-515.5566	0.0000
$\phi_3$	2.356905	0.011389	206.9404	0.0000
$\phi_4$	-0.50644	0.008237	-61.48038	0.0000
$\Phi_{33}$	0.938280	0.042357	22.15150	0.0000
$\theta_1$	-2.004745	0.288557	-6.947493	0.0000
$\theta_2$	1.400848	0.376337	3.722324	0.0002
$\theta_3$	-0.275892	0.112957	-2.442438	0.0149
$\theta_4$	0.043012	0.049410	0.870522	0.3844
$\Theta_{33}$	-0.838930	0.066691	-12.57943	0.0000
$\sigma^2$	154.4792	56.49393	2.734440	0.0064

Table 6.5: Model parameters for model  $SARIMAX(4,0,4)(1,0,1)_{33}$  in best-performing Run 4, following notation in (6.2).  $\sigma^2$  is the estimate of the error variance from the MLE.

ARIMAX(4,0,4,2) with AIC 7.9952. In the forecasting period, a MAPE of 5.12% and a RMSE of 8.47 mg/dL were obtained for the SARIMAX model for a 5-h PH, compared to 6.62% and 10.28 mg/dL for SARIMA, and 10.51% and 16.17 mg/dL for ARIMAX. Differences among the behavior of the different models can be observed in Fig. 6.7. Finally, forecasting performance as measured by MAPE and RMSE at different PHs is presented in Table 6.6. PHs of 30, 60, 120, 180, 240, and 300 minutes were considered.

## 6.4 Conclusions and margins for further improvement

In this chapter, we investigated the concept of **seasonality** in CGM time series for glucose prediction. Training data consisted in a collection of PPs from different patients covering both early and late postprandial phases (8 hours). Time between meals during the day is generally shorter. *Nocturnal period* was not represented by our data. However, PP has shown to be much more challenging than nocturnal period for an AP. Both  $CV - AUC_{8h}$  and ED (Fig. 6.3) showed large inter-individual variability and a large range in intra-individual variability, with its worst-case represented by inter-individual variability. Thus, the concatenated time series defines a challenging scenario with a worst-case highly variable patient. Data variability might be attenuated with the use of classification techniques, collecting

The model	PH					
	30 min	60 min	120 min	180 min	240 min	300 min
<b>No exogenous inputs</b>						
SARIMA(4, 0, 4)(1, 0, 1) <sub>33</sub>	7.61 (9.8955)	5.97 (8.5567)	5.46 (8.1011)	6.07 (8.9592)	6.95 (10.7894)	6.62 (10.2870)
ARIMA(4, 0, 4)	9.00 (9.8259)	13.19 (18.2491)	12.47 (17.3091)	10.51 (15.3702)	9.79 (14.8312)	9.39 (14.3960)
Difference*(%)	15.44 (0.7083)	54.74 (53.1117)	56.21 (53.1975)	42.25 (41.7106)	29.01 (27.2520)	29.50 (28.5427)
<b>CSII as exogenous input</b>						
SARIMAX(4, 0, 4, 2)(1, 0, 1) <sub>33</sub>	2.90 (3.6264)	3.20 (4.6225)	5.86 (9.1595)	4.49 (7.7283)	4.82 (8.2186)	5.12 (8.4743)
ARIMAX(4, 0, 4, 2)	7.97 (9.1372)	12.99 (18.8784)	12.95 (18.4822)	11.10 (16.5420)	10.58 (16.2076)	10.51 (16.1783)
Difference**(%)	63.61 (60.3117)	75.37 (75.5143)	54.75 (50.4415)	59.55 (53.2807)	54.44 (49.2917)	51.28 (47.6193)

Table 6.6: Prediction accuracy measured by MAPE (%) and RMSE (mg/dL), in parentheses, of seasonal versus nonseasonal counterparts for different PHs. \* 100 (SARIMA-ARIMA) / ARIMA; \*\* 100 (SARIMAX-ARIMAX) / ARIMAX.

similar enough postprandial responses into different datasets, with their corresponding prediction model.

A first-order seasonal AR and MA component was identified with seasonality lag 33 in all SARIMA runs due to the concatenated nature of the time series. In all runs, SARIMA outperformed ARIMA revealing a significant role of seasonality. 5-h PH average MAPE was reduced in 26.62%. Considering individual runs, the improvement ranged from 6.3% (Run 7; validation data P41) to 54.52% (Run 3; validation data P21). In the best performing case, according to MAPE (Run 4), this reduction amounted to 29.45%. Prediction improvement by introducing seasonality also becomes apparent from Fig. 6.7 The benefit of seasonality was consistent among different PHs, as illustrated in Fig. 6.5 and Table 6.6 for Run 4. Lower PHs benefited more, with a MAPE reduction over 50% for PHs of 60 and 120 minutes, and over 40% for 180 min. In this case, MAPE was close to 6% and RMSE below 10 mg/dL, doubling these values when seasonality was not considered. In greater PHs benefit of seasonality is still observed, although decreasing due to variability in the time series.

Consideration of CSII rate into the seasonal model further improved performance for Run 4. Although analysis was limited to this case to reduce computational burden, remark it corresponds to the most challenging situation for model improvement since SARIMA model for Run 4 has the best prediction accuracy in the cross-validation study. SARIMAX improved performance as compared to SARIMA with a 61.89% reduction in MAPE (2.90% vs 7.61%) for 30-min PH to a 7.33% reduction at 2-h PH (5.86% vs 5.46%) and reductions over 20% for PHs over 180 min, as shown in Table 6.6. A RMSE below 10 mg/dL was obtained for all PHs. This means that SARIMAX models might allow the increment of PHs in MPC-based AP systems. Table 6.6 also shows that SARIMAX outperformed in all cases its nonseasonal counterpart ARIMAX.

This is a proof-of-concept study and as such it has limitations. It is assumed that mealtime is known, allowing for the construction of concatenated time series with fixed-length PPs. However, to date, meal announcement is a common component of AP systems and, otherwise, meal detection algorithms are incorporated. Remark that although focus was put on PPs, this approach can be applied to other fixed-length time series data subsets representing characteristic scenarios where similarity is expected or learned from classification techniques. Another limitation is the data used, which did not correspond to a single patient, although inter-patient variability in the data was representative of worst-case intra-patient variability defining a challenging scenario. Collections of 18 PPs were used for model training at each cross-validation run. Seasonal components of the identified models were first or second order, which means that current meal depends, at most, on the two previous similar meals, as well the glucose depends not only on

current and previous measurements, but also on previous similar events, depend on the SARIMA model structure. Thus, the length of the data used is considered appropriate for this proof-of-concept study.

Despite the limitations of this study, seasonality has shown to be an important factor to improve model predictive power allowing for the significant extension of PHs. Results reported in Section 6.3 allow us to conclude that the SARIMA and SARIMAX models exhibit relatively higher forecasting accuracy for larger PHs, giving results statistically comparable. The seasonal models with exogenous inputs that relate well with the CGM measurements, could help to improve the prediction accuracy and PH. Improvements of the developed seasonal models would be achieved with longer experimental data for model development, since the efficiency of the linear empirical dynamic models is highly dependent on data used for the training set.

Positive results in this proof-of-concept encourage the research conducted in the next two chapters, where the following open issues are addressed:

1. The classification of periods under scenarios yielding “*similar*” “*enough*” glycemic responses to fully exploit the expected benefit of seasonal models.
2. Adding more exogenous variables into the SARIMAX model allowing to address prediction including exercise.
3. Getting rid of the limitation of fixed length time subseries to address a real life scenario with variable length event-to-event data.

## Chapter 7

# Seasonal local models

This chapter introduces the study of classification of PPs under different scenarios to yield similar enough glycemic responses to fully exploit the expected benefit of seasonal models. Different seasonal local models are then identified for each group, whose predictions are integrated to yield a final global BG prediction. The chapter starts in Section 7.1 with the main rationale of this study. Section 7.2 describes experimental data, materials, and methodology used. Section 7.3 displays the results of our study regarding clustering of different PPs, seasonal local modeling for each cluster, and local models integration. Section 7.4 displays the conclusions and margins for further improvement.

### 7.1 Rationale

An important task in T1D management is the prevention of hypo- and hyperglycemia throughout the day and night, which requires high accuracy BG predictions.

In Chapter 6, it was demonstrated that the use of seasonal time series models can lead to significant improvement in prediction accuracy and PH. As well, we also demonstrated that CSII information might effectively improve prediction performance. However, seasonal models like SARIMA and SARIMAX require similar enough glycemic patterns in CGM historical data to obtain accurate models, which cannot be assumed in free-living CGM data.

This chapter explores the possibility of preprocessing original CGM historical data to obtain sets of similar glycemic profiles (clusters) useful for SARIMA/SARIMAX model identification. As in Chapter 6, focus will be put on PPs due to its impact in glycemic control. First of all, a time series is built from experimental CGM data comprising meal and exercise events by concatenating different PPs of fixed length. Similarity of PPs is not expected due to differences in meals and exercise sessions on different days

and subjects. Therefore, a number of methods for clustering PPs are tested to reach the needed PPs similarity for seasonal modeling. After clusters gathering similar glucose responses are available, a SARIMAX local model is built for each of them (considering insulin infusion rate and energy expenditure as exogenous variables). Finally, a global model for BG predictions is obtained by integrating the local models.

## 7.2 Research Design and Methods

### 7.2.1 Details of procedures

Data from a clinical study conducted at University of Chicago was used. Overall, 10 subjects with T1D completed 18 clinical CL experiments [11]. Table 7.1 shows the characteristics of the participants of the CL studies through three different protocols. The study was explained well for each subject, and informed consent was obtained accordingly the instructions in the protocols. The subjects arrived at the Clinical Research Center (CRC) of the University of Chicago at 07:00 AM on the first day of the experiment. They wore their personal insulin pump, a CGM (Medtronic Guardian-Real Time Continuous Glucose Monitor [Northridge, CA], or Dexcom G4 Platinum [San Diego, CA]) and a BodyMedia SenseWear Pro3 [Pittsburgh, PA] armband reporting physical activity via energy expenditure (EE) and galvanic skin response (GSR). CGM interstitial glucose values and physical activity signals were sent to the AP every 5 minutes. All experiments were performed over 60 hours. Each individual used their personal glucometer for calibration of the CGM and for BG testing during physical activity. Subjects calibrated the CGM at least twice each day according to manufacturer instructions. The BG levels were obtained every 30 minutes for 60 hours, using a YSI (Yellow Spring Instruments) 2300 STAT PLUS glucose analyzer. The YSI or SMBG measurements were collected only for safety purposes and were not used in any parts of the algorithm for insulin dosing.

### 7.2.2 Meal protocol

Meals were provided to the subjects from the CRC Metabolic Kitchen over the 60-hour period. The study diet consisted of a one-day menu of three meals and one evening snack. Identical meals were provided each day of the study. Breakfast provided 25% of calories, lunch 30%, dinner 35% and evening snack 10%. The macronutrient composition of the diet was 40-50% of calories from CHO, 20-30% from protein, and 30-35% from fat. Food planning and analysis were completed using Food Processor SQL from ESHA Research (Elizabeth Stewart Hands and Associates, Salem, OR). All meals were prepared under supervision by a registered dietitian nutritionist in the University of Chicago CRC Metabolic Kitchen. Breakfast, lunch,



ID	Age (year)	Duration of Diabetes (months)	BMI ( $kg/m^2$ )	Gender	Race/Ethnicity
S1	21	136	23.68	Male	White/non-Hispanic
S2	23	70	18.89	Male	White/non-Hispanic
S3	26	194	26.28	Male	White/non-Hispanic
S4	19	195	23.55	Female	White/non-Hispanic
S5	24	206	23.39	Male	African American
S6	24	124	25.5	Male	White/non-Hispanic
S7	33	358	38.62	Female	White/non-Hispanic
S8	22	129	23.75	Male	White/non-Hispanic
S9	34	381	24.78	Male	White/non-Hispanic
S10	25	209	25.48	Female	White/non-Hispanic
<b>Average</b>	<b>25.1</b>	<b>200.2</b>	<b>25.39</b>	<b>7 Male 3 Female</b>	<b>9 White/non-Hispanic 1 African American</b>

Table 7.1: Demographic information for subjects.

dinner meals and evening snacks occurred at the same time each day, with a variety of snacks based on individual food preferences.

### 7.2.3 Exercise protocol

Individuals completed two bouts of exercise per day, for *three* consecutive days as shown in Fig. 7.1. Each bout of exercise was scheduled six hours apart. Days 1 and 3 consisted of *aerobic, moderate continuous training (MCT)*, and *resistance training (RT)*. MCT was specified with the heart rate reserve method ( $HHR = [HR_{max} - HR_{rest}] \times \% \text{ intensity desired} + HR_{rest}$ ) and consisted of a warm up period of 5 minutes, then 20 minutes of 60-80% HRR and finally a 5 minute cool down. RT was leaded by rating of perceived exertion (RPE) of 11-15 (6-20 point Borg Scale), with subjects performing 3 sets of 10 repetitions on 4 exercises (leg extension, seated chest press, seated row, and lat pulldown). Day 2 subjects first completed RT, and then high intensity interval training (**HIIT**) (6 hours post RT). HIIT training included a 5 minute warm up, followed by 3 minutes of treadmill exercise at 60-70% HRR followed by 4 minutes at 80-90% HRR, this was then repeated 3 times; such that 4 bouts of high intensity and 3 bouts of lower intensity were completed for a total of 36 minutes.

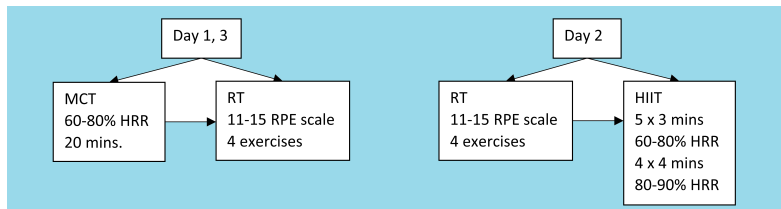


Figure 7.1: The exercises types in the study. Modified from [11].

### 7.2.4 Closed-loop system

The *multi-module multivariable adaptive artificial pancreas* (MAAP) system developed at Illinois Institute of Technology was used in this study [11]. The system operates autonomously, without any manual inputs to be entered by individuals. Insulin infusion is computed based upon CGM values and physical activity signals as shown in Fig. 7.2. The MAAP system is based upon a generalized predictive control (GPC) algorithm. The BG concentration is modeled by an ARMAX model with exogenous inputs GSR, EE, and BG concentration. The appropriate insulin was given as a basal rate (U/h) in the early experiments (S1-E1; S2-E1; S4-E1, E2; S5; S6-E1, E2; S7; S9; S10-E1; see Fig. 7.3), and in the later experiments the GPC calculated insulin boluses and the subject's predefined standard basal rate were used. In experiments where the basal rate of the subject was used as a baseline, the controller relayed periods of pump suspension in the case that the rate was too high, while it added small boluses in the case that the rate was too low. Therefore, the basal rate was still adjusted and organized by the controller based on the calculated needs of each subject in the experiment. Specialized modules were designed to improve the MAAP performance such as a hypoglycemia early alarm (HEA) system, a meal detection (MD) algorithm, and a hypoglycemia prediction and carbohydrate recommendation (H-CR) algorithm. All presented clinical data utilized the HEA algorithm in which CGM, IOB, and physical activity signals are used to predict hypoglycemia 25 minutes in advance. H-CR algorithm suggested an appropriate amount of CHOs based on the current BG concentration and its rate of change to prevent the potential hypoglycemia episode. MD algorithm administered meal boluses based upon CGM measurements, a modified Bergman minimal model and an unscented Kalman filter for state estimates. The modules used in each experiment are shown in Fig. 7.3.

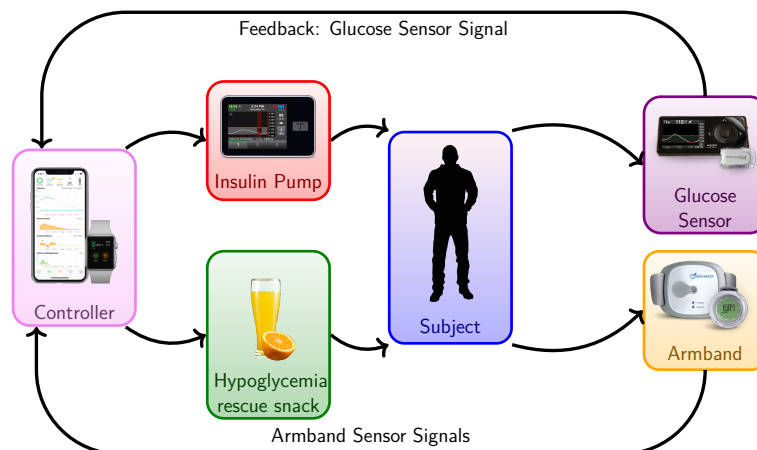


Figure 7.2: MAAP general structure.

		Basal +HEA	Basal +HEA+MD	Basal +H-CR+MD	Bolus+ H-CR+MD
S1	E1				
	E2				
	E3				
	E4				
S2	E1				
	E2				
S3	E1				
	E2				
S4	E1				
	E2				
S5	E1				
S6	E1				
	E2				
S7	E1				
S8	E1				
S9	E1				
S10	E1				
	E2				

Figure 7.3: MAAP control structure for each experiment. Notation: HEA (Hypoglycemia Early Alarm), MD (Meal Detection), H-CR (Hypoglycemia Prediction and Carbohydrate Recommendation), Basal or Bolus (GPC Controlled Variable).

### 7.2.5 Statistical procedures

In order to analyze if different “clusters of PP glucose profiles” exist within the data gathered in the experiments described above, three main hypothesis were tested:

- clusters exist depending on the **exercise type performed**,
- clusters exist depending on the **MAAP control structure**,
- clusters exist depending on the **PP glucose profiles** (shapes).

PPs with a fixed length of 4h starting at mealtime were considered, independently of the meal type. This ensured no interference with the next meal, according to closed-loop controllers postprandial performance. These PPs can be easily reorganized for the first two, but an automatic clustering algorithm is required for detecting similar PPs time profiles. The FCM clustering algorithm presented in Section 5.4 could perform well this task in a simple way.

Variables (features) used to compare hypotheses and prove that clusters really exist were: exercise type, exercise intensity, controller type, meal composition (*i.e.*, the percentage of calories from CHO, protein, and fat), total insulin infused, average glucose concentrations, and the time spent in different glycemic ranges (moderate hypoglycemia ( $< 54$  mg/dL), mild hypoglycemia ( $54 - 70$  mg/dL), euglycemia ( $70 - 180$  mg/dL), mild hyperglycemia ( $180 - 250$  mg/dL), and moderate hyperglycemia ( $> 250$  mg/dL)).

For the purpose of comparing the dissimilarity of the clusters, one-way analysis of variance (ANOVA) was used to determine whether there is statistically significant differences (SSD) between mean values of the variables characterizing clusters, and the Fisher's Least Significant Difference (LSD) post-hoc test was used to confirm SSD. A significance level of 5% (P-value  $< 0.05$ ) was considered in all cases. The same analysis was performed, also in our results, to determine SSD between the accuracy of the different modeling approaches.

### 7.2.6 FCM clustering algorithm

FCM clustering attempts to find the most characteristic point representing each cluster, which can be considered as the prototype or "centroid" of the cluster and determines the membership grade of each data point (PP) to each cluster. FCM algorithm was applied to our CGM data (*i.e.*, 4-h PPs extracted from CGM time series in this case) considering the Euclidean distance (ED) function between  $j^{th}$  CGM time series and  $i^{th}$  cluster prototype. FCM partitions CGM data into  $c > 1$  clusters by minimizing the objective function  $J_m$  (5.20). Memberships are stored in the partition matrix  $U = [u_{ik}]$ ,  $i = 1, 2, \dots, c$ ,  $k = 1, 2, \dots, n$ , where  $u_{ik} \in [0, 1]$ ,  $\sum_{i=1}^c u_{ik} = 1$ ,  $\forall k$  and  $0 < \sum_{k=1}^n u_{ik} < n \quad \forall i$ .

At each iteration, the fuzzy membership values  $u_{ik}$  were computed using (5.21), and the fuzzy centers  $c_i$  using (5.22). These steps were repeated until the minimum objective function  $J_m$  value was achieved. In order to assess the average compactness and separation of fuzzy partitions and reach the optimal number of clusters ( $c$ ), the **CWBI** validity index was used.

### 7.2.7 Seasonal models with exogenous signals

As mentioned previously, a SARIMA model with exogenous inputs is an expanded form of its nonseasonal counterpart ARIMAX model that includes as new model components SAR and SMA terms. Given a CGM time series  $G(t)$  resulting from the concatenation of fixed-length PPs, a SARIMAX model is expressed as (6.2). In practice, the AR, and MA coefficients can be estimated by fitting an AR and MA model to the residual errors  $\varepsilon(t)$  themselves. Thus, increasing the similarity between different glucose responses leads to a better fitted model. The concatenation of PPs in the CGM time

series will enforce seasonality with period  $s$  equal to the PP length. This means that seasonal components of the model (SAR, SMA) will relate to measured behavior in previous PPs, potentially leading to better glucose predictions. If similar enough PPs are available as training set (*i.e.*, PPs that can be grouped into the same cluster), then a SARIMAX model can be identified for each cluster  $i$ , leading to a set of local glucose estimations (prediction)  $\hat{G}_i(t)$  for  $i = \{1, 2, \dots, c\}$ .

### 7.2.8 Models integration

Real time BG prediction  $G(t_p)$  of a PP at time instant  $t_p$  can be solved as a crisp (hard) approach by selecting the  $\hat{G}_i(t_p)$  prediction corresponding to the cluster  $i$  whose prototype  $\nu_i$  is the closest (minimum distance) to the glucose validation data available at time  $t_p$ . If the data to prototype distance (typically Euclidean) is  $D_i(t_p)$ , then:

$$\hat{G}(t_p) = \hat{G}_m(t_p) \mid m = \arg \min_{1 \leq i \leq c} D_i(t_p). \quad (7.1)$$

A fuzzy approach can also be used, where the PP belongs partially (from 0 to 1) to all clusters and the BG prediction is computed by weighting the  $c$  estimations by the fuzzy membership  $u_i(t_p)$  at time  $t_p$  to each cluster  $i$ :

$$\hat{G}(t_p) = \sum_{i=1}^c u_i(t_p) \hat{G}_i(t_p), \quad (7.2)$$

$$u_i(t_p) = \frac{1}{\sum_{l=1}^c \left( \frac{D_l^2(t_p)}{D_i^2(t_p)} \right)^{\frac{1}{m-1}}}. \quad (7.3)$$

Both approaches were tested in this work. Furthermore, the PP cluster membership can be calculated using the PP data available at time  $t_p$  (all available samples from meal intake start), or using a subset of samples (the most recent) in a given time window. The predictions in this study were calculated from the moment the patient takes his/her meal, although there is no real data before the meal (the concatenated time series does not have this information, but if they were available the results presented in this work could be further improved).

### 7.2.9 Validation procedure

The LOOCV technique was used as a reliable approach for handling short time series, as well for assessing whether results could be generalized to other datasets. Clustering was performed over the whole set of PPs, resulting in  $c$  clusters. Then,  $c$  seasonal time series  $P_i(t)$  were created by concatenating those PPs assigned to each cluster (*i.e.*, with similar profiles). For each

$P_i(t)$  a PP was randomly taken (using the random sequence generator in <http://www.random.org>) as validation set  $P_i^V(t)$ , and the remaining data was used as training set  $P_i^T(t)$ .

The process was run four times resulting in four training ( $P_i^{T1}(t), P_i^{T2}(t), P_i^{T3}(t), P_i^{T4}(t)$ ) and four validation ( $P_i^{V1}(t), P_i^{V2}(t), P_i^{V3}(t), P_i^{V4}(t)$ ) sets for each cluster. Regarding the exogenous variables, real values were used in the training set while the mean of training values were used in the validation set (future exogenous variables values are not available in a real-time case, unless specific applications such as the optimization process in MPC control). Furthermore, the metrics used for model forecasting accuracy were **RMSE** (mg/dL) and **MAPE** (%).

## 7.3 Results

The methods presented in the previous section were applied to the available experimental data by following the diagram presented in Fig. 7.4. Firstly, data was clustered and (four) training and validation sets are defined (Part A); secondly, SARIMAX models were identified for each cluster (Part B); and, finally, models were (crisp or fuzzy) integrated and validated (Part C) using RMSE and MAPE.

### 7.3.1 Data clustering

Due to the nature of the experimental data, an analysis of the influence of the exercise type and controller used on CGM profiles (consisting in 85 PPs) was performed, as well as automatic clustering, based in this case on 79 complete PPs, since 6 PPs with missing samples were excluded because FCM is not directly applicable to incomplete data.

#### Exercise type clusters

First of all, the different PPs were manually organized in clusters based on the different exercise type performed after that meal (including one without exercise, after dinner: “No.EX”) with the aim of analyzing its effect on CGM profiles. Results are presented in green rows in Fig. 7.5. It can be observed that the average CGM value of different PPs is slightly higher in RT and HIIT than in the MCT, and slightly higher also in No.EX with values of 158.22, 143.84, 141.83 and 156.40 mg/dL, respectively (no SSD found,  $p = 0.177$ ). SSD is revealed between No.EX and all exercises sessions (*i.e.*, MCT, RT, and HIIT) ( $p = 0.000$ ) in most elements of meal composition (*i.e.*, carbs, calories, protein, and fat). BG values are within the euglycemic range for 77.55%, 68.00%, 73.47% and 67.42% of the time, respectively (no SSD found,  $p = 0.410$ ). In addition, BG values are in moderate hypoglycemia for 1.17%, 1.09%, 2.04% and 1.89% of the time, respectively (no SSD found,

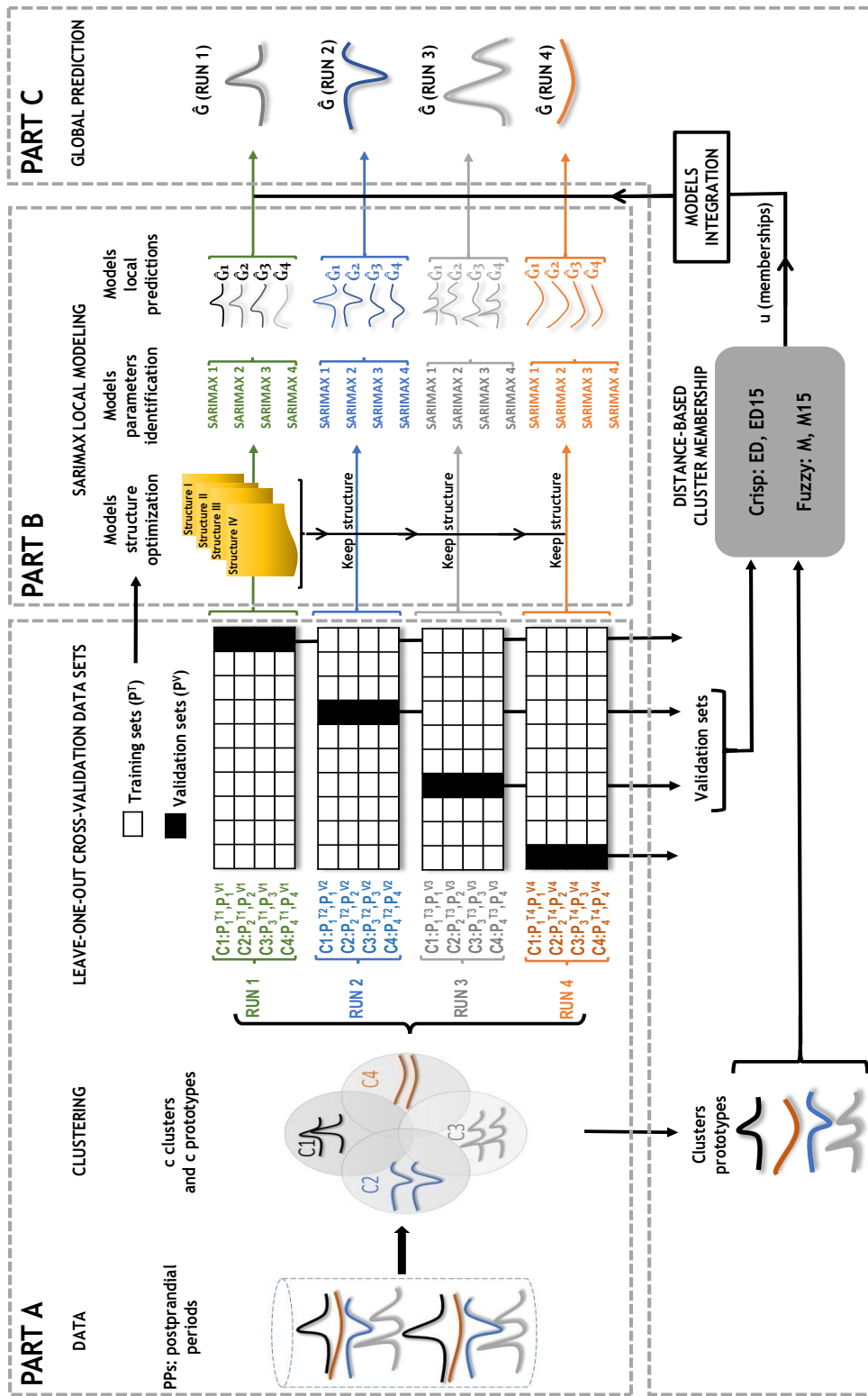


Figure 7.4: Diagram summarizing the methodology application: data sets generation (A), local models identification (B), and models integration (C).

$p = 0.668$ ), and in moderate hyperglycemia for 20.26%, 23.25%, 23.85% and 25.25% of the time, respectively (no SSD found,  $p = 0.876$ ). However, SSD is revealed between MCT, RT, and HIIT exercise sessions ( $p = 0.000$ ) in the area under the curve of the energy expenditure ( $AUC_{EE}$ ) variable. This difference is expected, as the MCT is an entirely aerobic exercise versus the HIIT having heart rate values in the anaerobic range for 39% of the time, and RT is almost entirely anaerobic exercise.

Besides that, BG area under the curve ( $AUC_{BG}$ ) has a negative correlation with  $AUC_{EE}$  in the MCT exercise sessions ( $r = -0.50$ ) because this exercise promotes BG reductions whereas RT and HIIT may promote decrease, increase or lead to stable BG depending on the duration and intensity [315].

### Control structure clusters

Secondly, PPs were reorganized based on the used MAAP control structure (Fig. 7.3). Red rows in Fig. 7.5 show that average CGM value is higher in (Basal+HEA) and (Basal+H-CR+MD) than in the (Bolus+H-CR+MD) and (Basal+HEA+MD) control type, with values of 170.16, 167.77, 139.57 and 145.23 mg/dL, respectively, and there was SSD between them ( $p = 0.000$ ). The same SSD pattern is revealed in BG values within the euglycemic range for 58.16%, 80.20%, 56.99% and 75.71% of the time, respectively ( $p = 0.000$ ), in mild hyperglycemia for 29.87%, 16.63%, 35.16% and 21.84% of the time, respectively ( $p = 0.002$ ), and in moderate hyperglycemia for 11.04%, 1.43%, 8.01% and 1.63% of the time, respectively ( $p = 0.031$ ). However, BG values are in mild hypoglycemia for 0.83%, 2.40%, 0.47% and 1.02% of the time, respectively (no SSD found in this case,  $p = 0.108$ ), and in moderate hypoglycemia for 0.65%, 0.20%, 0.00% and 0.00% of the time, respectively, (no SSD found,  $p = 0.527$ ). As a secondary conclusion, Bolus+H-CR+MD is revealed as the best MAAP module in all experiments. The MD in the MAAP control structure improves the CGM measurements and can keep CGM measurements in the euglycemic range (70-180 mg/dL).

### Automatic clustering

Finally, FCM clustering algorithm was used for automatic classification of PPs. As shown in Fig. 7.6 the best number of clusters to classify each PP into a specific group is 4 clusters via FCM with the CWBI cluster validity index. Blue rows in Fig. 7.5 show that average CGM values are higher in cluster 1, cluster 2, and PPs without cluster (PPs with membership below 0.35) than in cluster 3 and 4 with values of 159.23, 184.67, 172.52, 136.68 and 131.30 mg/dL, respectively (SSD found,  $p = 0.000$ ). The effect of different exercise type on the glycemic behavior via  $AUC_{EE}$  for each cluster is clear, especially in cluster 2 and cluster 4 with SSD ( $p = 0.023$ ). Furthermore,



Data clustering (# PPs)	CGM (mg/dL)	AUC <sub>bg</sub> (mg/dL)·min	Meal composition					AUC <sub>Cef</sub> (Kcal/min)·min	Total insulin (U)	Time spent in different glycemic ranges				
			Carbs (g)	Cals (Kcal)	Prot (g)	Fat (g)	Fiber (g)			R <sub>1</sub> (%)	R <sub>2</sub> (%)	R <sub>N</sub> (%)	L <sub>1</sub> (%)	L <sub>2</sub> (%)
EXERCISE TYPE (85 PPs)	MCT (14 PPs)	34439.7 (5948.6)	63.03 <sup>a,4</sup> (17.87)	528.35 <sup>a,3,4</sup> (141.82)	26.06 <sup>a,3,4</sup> (8.65)	19.71 <sup>a,3,4</sup> (5.27)	6.94 (5.98)	18.96 <sup>*</sup> (7.09)	6.49 (2.82)	20.26 (17.17)	2.04 (4.56)	77.55 (17.87)	1.17 (2.64)	0.0 (0.0)
	RT (28 PPs)	38467.9 (8361.56)	70.50 <sup>a,4</sup> (11.93)	586.83 <sup>a,4</sup> (102.99)	29.54 <sup>a,4</sup> (6.12)	21.52 <sup>a,4</sup> (3.95)	7.12 (3.53)	10.81 <sup>*</sup> (5.73)	6.02 (3.14)	23.25 (17.08)	7.73 (23.13)	68.00 (23.13)	1.09 (2.59)	0.36 (1.89)
	HIIT (16 PPs)	35050.5 (4857.46)	78.63 <sup>a,4</sup> (15.31)	632.29 <sup>1,4</sup> (96.63)	31.93 <sup>a,4</sup> (6.17)	23.78 <sup>a,4</sup> (2.59)	8.83 (2.89)	26.64 <sup>*</sup> (7.50)	5.76 (3.00)	23.85 (15.57)	1.02 (3.95)	73.47 (15.51)	2.04 (3.61)	0.26 (0.99)
	No.EX (27 PPs)	156.40 (27.83)	87.97 <sup>*</sup> (18.29)	736.60 <sup>*</sup> (135.47)	37.97 <sup>*</sup> (9.37)	27.55 <sup>*</sup> (4.99)	9.64 (6.32)	3.87 <sup>*</sup> (3.15)	6.74 (5.43)	25.25 (20.43)	6.12 (15.32)	67.42 (20.96)	1.89 (3.24)	0.30 (1.54)
MAP CONTROL STRUCTURE (85 PPs)	Basal+H-CR+MD (22 PPs)	167.77 <sup>a,4</sup> (37.86)	75.89 <sup>a,4</sup> (4.97)	616.26 (136.92)	30.09 (9.83)	22.61 (5.50)	7.52 <sup>a,3,4</sup> (5.37)	14.25 (10.23)	7.29 <sup>*</sup> (5.53)	29.87 <sup>a,2</sup> (22.42)	11.04 <sup>a,2</sup> (22.21)	58.16 <sup>a,3,4</sup> (27.22)	0.83 (1.95)	0.65 (2.30)
	Bolus+H-CR+MD (40 PPs)	139.57 <sup>a,1,3</sup> (18.86)	70.81 <sup>a,4</sup> (15.88)	627.12 (141.77)	33.68 (9.48)	23.33 (5.12)	6.21 <sup>a,3,4</sup> (2.81)	13.74 (10.50)	4.99 <sup>a,1,3</sup> (1.38)	16.63 <sup>a,1,3</sup> (11.81)	1.43 <sup>a,1</sup> (4.60)	80.20 <sup>a,1,3</sup> (11.75)	2.40 (3.64)	0.20 (1.29)
	Basal+HEA (13 PPs)	170.16 <sup>a,2,4</sup> (22.84)	77.10 (25.59)	612.26 (154.63)	30.91 (6.63)	22.76 (4.94)	12.33 <sup>a,1,2</sup> (5.24)	10.29 (9.13)	10.22 <sup>*</sup> (4.05)	35.16 <sup>a,2</sup> (19.22)	8.01 (11.53)	56.99 <sup>a,3,4</sup> (15.40)	0.47 (1.70)	0.0 (0.0)
	Basal+HEA+MD (10 PPs)	146.23 <sup>a,1,3</sup> (19.32)	90.60 <sup>a,1,2</sup> (13.92)	730.23 (100.36)	31.68 (4.57)	27.61 (3.96)	12.36 <sup>a,1,2</sup> (5.32)	10.19 (9.26)	4.02 <sup>a,1,3</sup> (1.85)	21.84 (17.34)	1.63 (5.16)	75.71 <sup>a,1,3</sup> (17.88)	1.02 (3.23)	0.0 (0.0)
AUTOMATIC CLUSTERS (FCM) (79 complete PPs)	Cluster 1 (38 PPs)	159.23 <sup>a,2,3,4</sup> (13.09)	80.74 <sup>a,4</sup> (21.74)	660.66 (160.27)	33.85 (10.32)	24.19 (5.96)	8.32 (5.21)	10.57 <sup>a,4</sup> (8.88)	7.40 <sup>a,3</sup> (4.89)	26.98 <sup>a,2,3,4</sup> (10.17)	9.07 (10.99)	62.70 <sup>a,2,3,4</sup> (9.99)	1.25 (2.37)	0.6 (2.34)
	Cluster 2 (14 PPs)	184.67 <sup>a,1,2,4</sup> (22.33)	77.57 (18.28)	641.85 (140.95)	32.85 (8.13)	23.96 (5.48)	9.67 (6.61)	9.21 <sup>a,4,5</sup> (8.21)	8.66 <sup>a,3,4</sup> (5.63)	43.73 <sup>*</sup> (22.08)	8.89 (20.06)	48.54 <sup>a,1,3,4</sup> (18.69)	0.0 (0.0)	0.0 (0.0)
	Cluster 3 (15 PPs)	136.68 <sup>a,1,2,5</sup> (13.15)	82.46 <sup>a,4</sup> (13.01)	708.54 (96.50)	35.60 (5.82)	26.29 (3.31)	8.90 (4.86)	10.40 <sup>a,4</sup> (8.33)	4.40 <sup>a,1,2</sup> (1.90)	12.38 <sup>a,1,2,5</sup> (9.81)	0.0 (0.0)	87.48 <sup>a,1,2,5</sup> (8.72)	1.22 (3.14)	0.0 (0.0)
	Cluster 4 (23 PPs)	131.30 <sup>a,1,2,5</sup> (17.26)	66.61 <sup>a,1,3</sup> (16.10)	583.10 (143.23)	30.77 (10.22)	21.93 (5.01)	7.23 (2.57)	16.82 <sup>a,1,2,3</sup> (10.18)	5.17 <sup>a,2</sup> (1.77)	16.24 <sup>a,1,2,5</sup> (6.92)	0.71 (3.33)	80.83 <sup>a,1,2,5</sup> (9.00)	2.75 (3.27)	0.35 (1.66)
	Without Cluster (cluster_5_9 PPs)	172.52 <sup>a,3,4</sup> (33.55)	73.37 (15.65)	590.55 (121.56)	28.01 (6.20)	21.08 (4.70)	6.57 (6.02)	17.95 <sup>a,2</sup> (13.63)	7.13 (2.64)	29.48 <sup>a,2,3,4</sup> (15.81)	9.75 (20.98)	58.96 <sup>a,3,4</sup> (17.85)	1.59 (3.16)	0.45 (1.28)

Figure 7.5: DATA CLUSTERING SUMMARY. Notation: Results presented as mean(SD) for each feature. <sup>a,i,j</sup> indicates SSD ( $p < 0.05$ ) between clusters  $i$  and  $j$  (if  $i, j$  not specified, then differences found between all clusters).  $R_1$  is mild hyperglycemia range,  $R_2$  is moderate hyperglycemia range,  $R_N$  is euglycemia range,  $L_1$  is mild hypoglycemia range, and  $L_2$  is moderate hypoglycemia range.

SSD is revealed between cluster 2 and cluster 4 ( $p = 0.006$ ) in the total amount of insulin. The meal intake and the exercise type are almost equal between cluster 1 and cluster 3, but in this case the insulin sensitivity is an important factor detected in the total amount of insulin, and there is SSD between them ( $p = 0.022$ ). Moreover, SSD is revealed between cluster 4 and cluster 1, and between cluster 4 and cluster 3 ( $p = 0.012$ ,  $p = 0.007$ , respectively) in the total amount of CHOs.

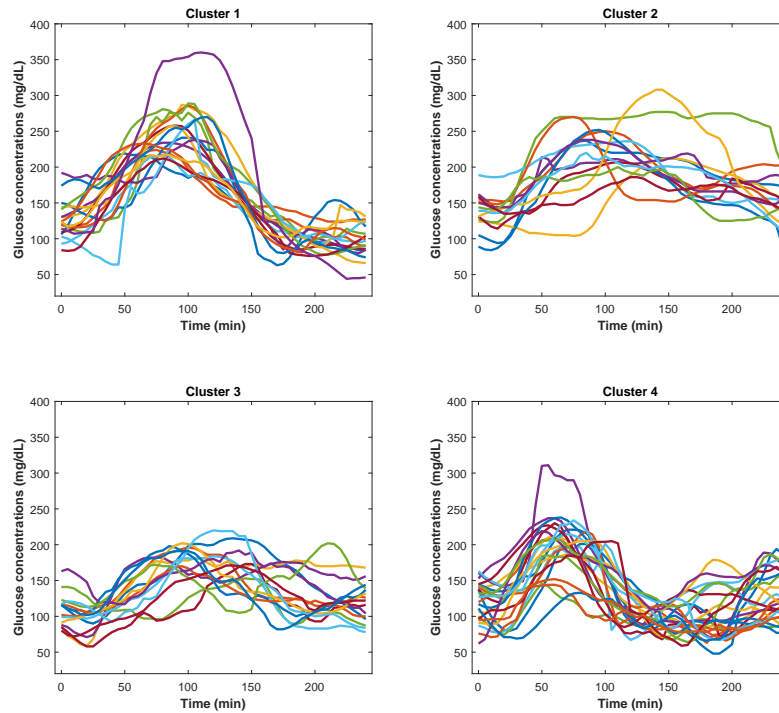


Figure 7.6: The clustering classifications by FCM algorithm, 4 clusters is the best number of clusters indicated by a CWBI validation index.

Besides that, a negative correlation between the total amount of CHOs and the total amount of insulin appears in cluster 3 ( $r = -0.42$ ), as a result of the slow absorption of the meal intake that would lead to a slow BG decrease. In addition, different exercise type effectiveness is detected by the mean difference time between the meal intake and the start time of the exercise in cluster 3 is 2.38 hours, and in cluster 4 is 1.49 hours, with SSD between them ( $p = 0.013$ ). This is not the case for cluster 1 and cluster 2 (2.21 hours and 2.04 hours, respectively) where SSD between them was not found ( $p = 0.840$ ). As well, there is no SSD between cluster 3 and both clusters 1 and 2 with  $p = 0.100$  and  $p = 0.151$ , respectively.

### Clustering analysis

Although differences are detected when clustering PPs using exercise type or control structure, the SSD between clusters results of the automatic clustering overcome the other approaches. Provided that the aim of the clustering in this work is that PPs classified in the same cluster should have similar behaviors or features, FCM objective function (with similarity defined as the Euclidean distance of the PPs profiles) finds an optimal partition of the dataset: SSD is found for all features as opposed to manual partition by exercise and controller types. Therefore, this approach is used in the following as the previous step for seasonal model identification, provided that clusters detected by FCM lead to clusters of PPs with similar behaviors helping in the development of a family of seasonal local models for improved accuracy and prediction horizon.

### 7.3.2 Model identification

#### Seasonal local models

Before model identification for each cluster, some previous work must be done. First of all, the effect of considering exogenous variables for performance improvement was studied, showing up that CSII, defined as the addition of bolus and basal infusion expressed in U per sampling period, and EE are relevant signals, as revealed by the Granger causality test, which is applied to test the null hypothesis that CGM does not Granger cause CSII, EE and vice-versa. Then, seasonality was studied and lag 49 (PPs size) was obtained for all clusters, and SAR and SMA orders were up to 2. With all this information available, a model for each cluster was identified through the *Box-Jenkins methodology* (Fig. 7.4, part B).

In the case of data sets  $P_i^{T1}(t)$  (LOOCV RUN 1), the following model structures were obtained:

- Cluster 1: SARIMAX (2, 0, 1, 2)(2, 1, 1)<sub>49</sub>, with AIC 6.6063.
- Cluster 2: SARIMAX (4, 0, 2, 2)(2, 1, 2)<sub>49</sub>, with AIC 6.7320.
- Cluster 3: SARIMAX (2, 0, 2, 2)(2, 1, 2)<sub>49</sub>, with AIC 6.3191.
- Cluster 4: SARIMAX (4, 0, 2, 2)(2, 1, 1)<sub>49</sub>, with AIC 7.378199.

All the above SARIMAX models incorporated EE and CSII as exogenous variables of order 2. Fig. 7.7 shows the local glucose predictions  $\hat{G}_i(t)$  for  $i = \{1, 2, 3, 4\}$ . In the most challenging scenario of a 3-h PH forecasting period, a MAPE of 3.10% (RMSE of 6.96 mg/dL) is obtained for cluster 1, 4.78% (10.77 mg/dL) for cluster 2, 9.25% (15.76 mg/dL) for cluster 3, and 8.07% (16.31 mg/dL) for cluster 4. Results for different PHs (15, 30,

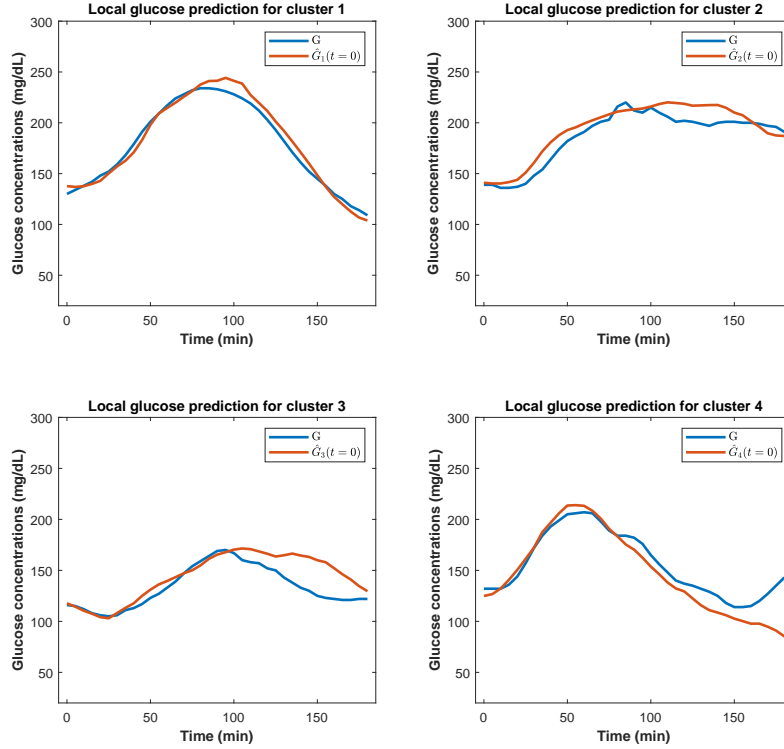


Figure 7.7: The local glucose predictions  $\hat{G}_i(t)$  for each cluster for a 3-h PH.

60, 120, and 180 minutes) using the validation set  $P_i^{V1}(t)$  are presented in Table 7.2.

### Global seasonal model

For building a global seasonal model **GSM**, different strategies for integration of local models were tested here: the crisp approach in (7.1) with the **ED** and the fuzzy approach in (7.2) using Memberships (**M**). The use of a time window for the computation of the PP cluster membership was also tested. A value of 15 minutes was found to be the most appropriate (labeled as **ED**<sub>15</sub> and **M**<sub>15</sub> for the crisp and fuzzy strategy, respectively) and the results in this case are also presented.

Table 7.3 shows the average MAPE (%) and RMSE (mg/dL) values of **GSM** predictions through the different PPs validation data of the 4 clusters found for RUN 1, as outlined in Fig. 7.4. Results for different PHs using all the above explained online cluster membership calculation options are presented. SSD is found between **M** and **M**<sub>15</sub> at 45-min PH ( $p = 0.024$ ),

Cluster	SARIMAX Model	PH				
		15 min	30 min	60 min	120 min	180 min
1	(2, 0, 1, 2)(2, 1, 1) <sub>49</sub>	2.18 (4.21)	1.87 (3.70)	1.73 (4.60)	2.74 (6.79)	3.10 (6.96)
2	(4, 0, 2, 2)(2, 1, 2) <sub>49</sub>	2.43 (3.69)	4.22 (7.40)	4.81 (10.51)	4.19 (10.00)	4.78 (10.77)
3	(2, 0, 2, 2)(2, 1, 2) <sub>49</sub>	0.98 (1.19)	1.38 (1.59)	2.70 (4.81)	3.22 (6.40)	9.25 (15.76)
4	(4, 0, 2, 2)(2, 1, 1) <sub>49</sub>	1.74 (2.85)	2.36 (4.81)	2.43 (5.64)	4.47 (7.11)	8.07 (16.31)

Table 7.2: Prediction accuracy of seasonal local models. **Notation:** prediction accuracy results presented as MAPE % (RMSE mg/dL).

Membership calculation	PH				
	15 min	30 min	45 min	60 min	90 min
<b>ED</b>	3.57 (6.62)	5.07 (9.72)	6.54 (12.63)	7.85 (15.16)	10.51 (19.75)
<b>ED</b> <sub>15</sub>	3.62 (6.65)	4.87 (9.29)	6.06 (11.74)	7.17 (14.04)	9.44 (18.34)
<b>M</b>	3.34 (5.98)	4.73 (8.81)	6.05 (11.30)	7.21 (13.46)	9.07 (16.75)
<b>M</b> <sub>15</sub>	2.94 (5.14)	3.89 (7.75)	5.41 <sup>*,†</sup> (10.00)	6.29 <sup>*,†</sup> (12.17)	8.66 <sup>†</sup> (16.49)

Table 7.3: Prediction accuracy of **GSM**. **Notation:** prediction accuracy results presented as MAPE % (RMSE mg/dL). \* indicates SSD ( $p < 0.05$ ) with **M** and <sup>†</sup> with **ED**.

and 60-min PH ( $p = 0.011$ ). The use of the 15-min window through **M**<sub>15</sub> is SSD when comparing with **ED** at PH bigger than 30-min PH ( $p = 0.012$  for 45-min,  $p = 0.008$  for 60-min, and  $p = 0.022$  for 90-min), therefore, **M**<sub>15</sub> could improve the worst case errors in the integration process.

Fig. 7.8 is included to illustrate the **M**<sub>15</sub> integration process: four  $\hat{G}_i(t)$  are available at each instant and four memberships  $u_i$  are also calculated. It can be seen how memberships change with time and, therefore, weights in (7.2) for  $\hat{G}(t)$  90-min-ahead prediction.

The previous results show the high accuracy for long PH, but extraction of the optimal SARIMAX model structure from the identification data set for each cluster is a time consuming task. Therefore, a more realistic approach was tested for an online environment where human and/or computational resources could be limited. In this new case, the model structure (model orders, exogenous variables, . . .) identified for  $P_i^{T1}(t)$  was maintained (as it was done in a pre-processing off-line step) and only model parameters were

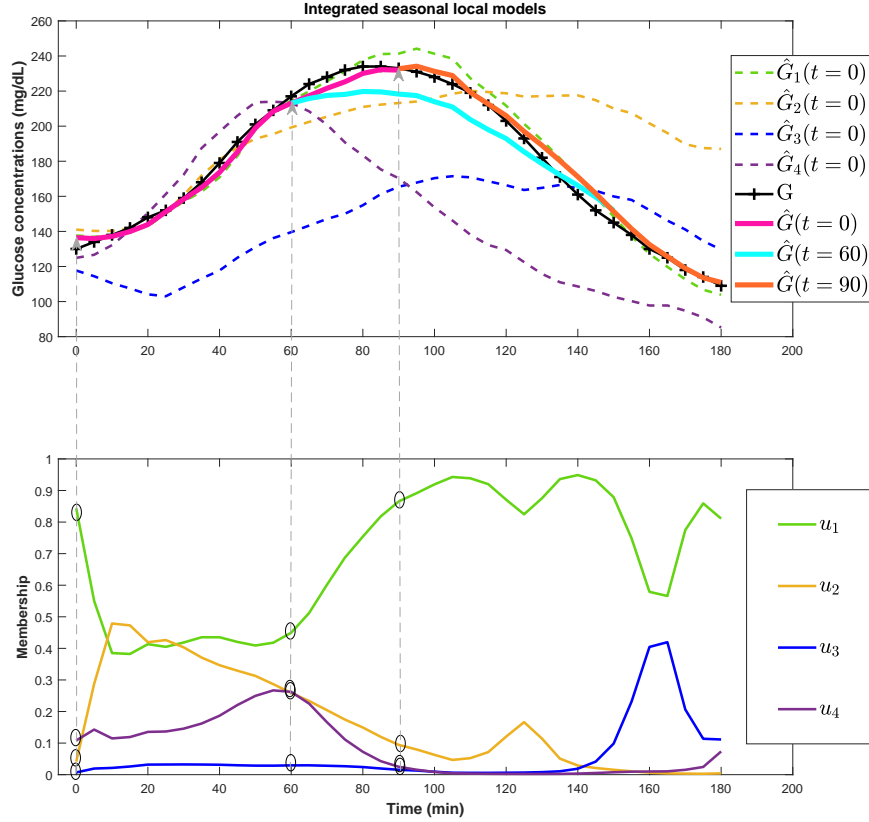


Figure 7.8: Local predictions at  $t = 0$ , memberships at each time instant, and global predictions at  $t = 0$ ,  $t = 60$ , and  $t = 90$  by using  $\mathbf{M}_{15}$  for a 90-min-ahead prediction. Membership values at the above prediction times are highlighted by the symbol  $\circ$ .

re-identified using  $P_i^{T2}(t)$ ,  $P_i^{T3}(t)$ , and  $P_i^{T4}(t)$  (see **part B** of Fig. 7.4).

Mean results of this three suboptimal **GSM** predictions are presented in Table 7.4 where, of course, higher prediction errors appear but results are still good for all PHs. SSD is found between **ED** and **ED<sub>15</sub>** at 15-min PH ( $p = 0.025$ ), and 30-min PH ( $p = 0.040$ ); and between **M** and **M<sub>15</sub>** at 15-min PH ( $p = 0.000$ ), 30-min PH ( $p = 0.000$ ), and 45-min PH ( $p = 0.010$ ). Moreover, SSD is found between **ED** (**ED<sub>15</sub>**) and **M** (**M<sub>15</sub>**) at all PHs.

Table 7.5 summarizes the best (**M<sub>15</sub>**) and worst (**ED**) mean results of the 4 datasets (1 for optimal model structure, and 3 suboptimals) used. In this way, the positive effect of the 15-min window appears by always improving the worst case errors in the integration process (SSD always found). Prediction accuracy of **GSM** overcomes all the literature approaches presented in

Membership calculation	PH				
	15 min	30 min	45 min	60 min	90 min
<b>ED</b>	5.82 (9.24)	8.23 (12.79)	10.24 (19.64)	12.19 (22.17)	13.72 (25.75)
<b>ED<sub>15</sub></b>	5.77* (8.71)	8.93* (11.99)	11.69 (21.80)	14.43 (24.79)	18.32 (29.83)
<b>M</b>	5.38* (7.68)	7.78* (10.28)	9.61* (16.29)	11.20* (18.61)	13.35* (22.36)
<b>M<sub>15</sub></b>	4.64*,* (6.98)	7.09*,* (9.99)	8.79*,* (15.70)	10.87* (19.29)	12.74* (21.82)

Table 7.4: Prediction accuracy of suboptimal **GSM**. **Notation:** prediction accuracy results presented as MAPE % (RMSE mg/dL). \* indicates SSD ( $p < 0.05$ ) of **ED** vs **ED<sub>15</sub>** or **M** vs **M<sub>15</sub>**, and \* for **ED** vs **M** or **ED<sub>15</sub>** vs **M<sub>15</sub>**.

Membership calculation	PH				
	15 min	30 min	45 min	60 min	90 min
<b>ED</b>	5.26 (8.58)	7.44 (12.03)	9.31 (17.89)	11.10 (20.42)	12.92 (24.25)
<b>M<sub>15</sub></b>	4.21* (6.52)	6.29* (9.43)	7.95* (14.27)	9.73* (17.51)	11.72* (20.49)

Table 7.5: Summary of prediction accuracy results (all datasets mean). **Notation:** prediction accuracy results presented as MAPE % (RMSE mg/dL). \* indicates SSD ( $p < 0.05$ ).

Table 3.1. Results obtained for the **M<sub>15</sub>** membership calculation in Table 7.3 are specially remarkable. As well, Fig. 7.9 shows the difference between **GSM** (*i.e.*, **M<sub>15</sub>**) prediction errors and the BG modeling error (RMSE) with a different PH of different literature methods.

## 7.4 Conclusions and margins for further improvement

Physical activity has a significant impact on BG concentrations in patients with T1D. However, in our data each exercise group had a mixture of glucose responses and the total number of MCT, RT, and HIIT exercise sessions was not balanced. As well, a mixture of glucose responses was found for the Bolus+H-CR+MD control structure, despite being the best MAAP module in all experiments. However, clusters gathering similar glucose profiles, as required to gain the potential of seasonal models, were successfully found by automatic clustering, using challenging CGM time series obtained from

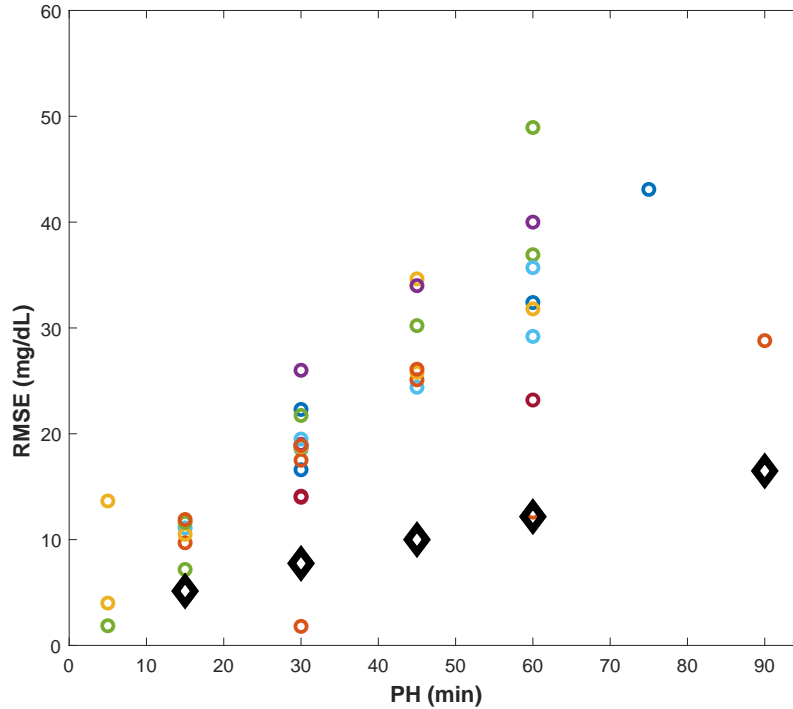


Figure 7.9: A comparison of the proposed model ( $M_{15}$ ) with those reported in the literature through the blood glucose modeling error (RMSE) with different PHs, where our prediction errors are marked by the black diamond.

AP experiments in T1D patients, including a variety of meals and exercise.

Increasing the similarity between different glucose responses in a cluster lead to accurate local models, which is the main idea behind the use of clustering techniques. SARIMAX seasonal local models were identified for each cluster and, subsequently, used online by fuzzy integrating the different available local predictions by means of an Euclidean-based distance to cluster strategy. Results reported in Table 7.3 allow us to conclude that the SARIMAX local models with exogenous inputs (CSII, EE) exhibit relatively higher forecasting accuracy for larger PHs (3-h PH), giving results statistically comparable. As well, the results reported in Table 7.4 allow us to conclude that the **GSM** exhibits high prediction accuracy for larger PHs (a 90-min-ahead prediction). Hence, the **GSM** overcomes BG forecasting of the approaches found in literature, which were summarized in Table 3.1.

Once demonstrated the feasibility of seasonal local modeling through clustering, challenges imposed by free-living data can be addressed, which will be covered in the next Chapter. There, an extension of the work will be carried out to deal with:



1. Handling with longer experimental data for model development.
2. Dealing with missing data as a result of sensor errors or human errors, especially when applying the FCM clustering technique or identifying/building seasonal models.
3. Applying the **GSM** in daily normal life where PPs will have different lengths (varying meal-to-meal time).

## Chapter 8

# Seasonality prediction under free-living conditions

This chapter introduces the use of stochastic seasonal local models in normal daily life for people with T1D. In order to do so, firstly, the data sets generation methodology for seasonal local modeling proposed in the previous chapter is extended to be able to deal with data including variable length PPs, nocturnal periods, and external user actions such as hypoglycemia rescues. Secondly, the new framework provides online glucose predictions for different PHs. The user can employ these glucose predictions taking into account a *confidence measure* also provided at each time instant. Additionally, and based on the available glucose prediction for PHs long enough to take effective therapeutic actions, the framework provides a future patient state and is able to detect abnormal behaviors.

The chapter starts in Section 8.1 with the main rationale of this proof of concept. Section 8.2 describes experimental data and the methodology used in the study. Section 8.3 shows the results of a global seasonal model integrating different seasonal local models for online glucose prediction, patient state forecasting and abnormal behavior detection. As well, displays a supervision system to help the user for detecting the abnormal states. Section 8.4 summarizes the conclusions and highlights the future work for further improvement of the proposed framework.

### 8.1 Rationale

Glucose prediction models can play an important role in future diabetes management systems. As demonstrated previously in Chapter 7, they can provide accurate estimates of future glucose values for large PHs. This can be very helpful for people with diabetes if the PH is long enough to let the subject with diabetes anticipate therapeutic decisions. Therefore, if an accurate online prediction of BG concentration is possible then a complete

*supervision system* based on these values can be devised. This supervision system can provide the users not only the glucose predictions, but this real time information can be processed to provide more specific information such are: future patient glycemic status (*e.g.*, hypo- and hyperglycemia), and abnormal behavior detection (*e.g.*, pump malfunctioning, catheter obstruction, etc.).

Despite **SARIMA** and **SARIMAX** models have demonstrated previously their effectiveness for glucose prediction in long PHs with high accuracy, they are data-hungry and require the existence of seasonality periods with fixed lengths to be applied. Therefore, *can these methods be applied in normal life (free-living conditions with the support of CGM data) including PPs or nocturnal periods with different and variable lengths?*

The aim of this chapter is to explore the possibility of extension of the seasonal local models based methodology presented in Chapter 7 to free-living real-time online use, and the utilization of accurate glucose predictions in a supervision system useful for the patients.

As a proof of concept, the methodology will be tested over 6 months of *in silico* simulated data. First of all, original CGM historical data of different full days, including PPs and nocturnal periods, will be preprocessed. The long-term simulated data is partitioned into a set of “**event-to-event**” time subseries, driven by meals and night periods, to enforce seasonality where original PPs and nocturnal periods are in different lengths. Preprocessed data is then clustered and, once sets of similar glycemic profiles (clusters) useful for building seasonal models are available, a SARIMA local model is built for each of them. Finally, the online BG prediction “**global model**” is obtained by integrating “**local models**”.

A proposal of supervision system is also presented, where the available glucose predictions and glycemic status estimations are, at each time instant, stringed along with a newly defined “**trust index**” (based on the real-time convexity of the fuzzy membership functions used for model integration) giving information about how much the estimation can be trusted. Additionally, a “**normality index**” (based on real-time values of possibility membership functions) giving information about how normal is the actual behavior or if it is an abnormal situation.

## 8.2 Research Design and Methods

### 8.2.1 Data overview

A long-term period of six months data from a *virtual* patient is available. The simulated data were generated using the average adult patient of an extended version of the UVA/Padova simulator [168] with variability sources, and exercise was not considered. Three daily random meals of 40, 90 and

60 g at 7:00, 14:00 and 21:00 hours were considered, and variability sources include: meal-size variability ( $\pm 10\%$ ), meal-time variability ( $\pm 20$  min), uncertainty in CHO estimation (uniform distribution between  $-30\%$  and  $+40\%$ ), meal absorption rate ( $k_{abs} \pm 30\%$ ), CHO bioavailability ( $f \pm 10\%$ ), insulin absorption model parameters ( $k_d, ka_1, ka_2 \pm 30\%$ ), and insulin sensitivity parameters ( $V_{mx}, Kp3 \pm 30\%$ ).

### 8.2.2 Data preprocessing

Once long-term (simulated) data is available from a patient, seasonality must be enforced to apply the seasonal stochastic modeling techniques to the identification data set (20% of available data is saved for validation). Provided that the initial instant, duration, and final instant of the different events considered in the available data (PPs and nocturnal periods) are variable, the original time series data is partitioned into a set of “**event-to-event**” time subseries, driven by event meals. Therefore, each time subseries initial point is the time instant when the event (meal) starts and its final point is the initial point of the next event. A special case is the (after) dinner PPs where 6 hours fixed-length were considered. As well, the night periods start at the last point of dinner PPs and finalizes at breakfast time.

Hence, each day provides four subseries with different lengths, and these lengths change every day. In order to force seasonality, the period with the maximum duration must be detected and its length stored in  $s$ . All the other subseries duration must be fictitiously expanded to  $s$ , by adding “not a number -NaN-” values after the last available value. In this way, all the periods will have the same length ( $s$ ) but most of them will have some *NaNs* in the final positions of the time subseries.

The final preprocessing step is removing all those PPs that can be considered as abnormal. The proposed methodology is focused on finding models representing the normal behavior of the patient. The periods considered as abnormal in the available data are those affected by patients rescue actions (CHOs ingestion) in the case of hypoglycemia. This rescue action greatly modifies the expected postprandial profile and, therefore, the values affected by this action will be deleted from the data set. In this case, the values affected (and deleted) are all those in a PP after a rescue is performed, and the whole PP after those with a rescue.

### 8.2.3 Time series clustering algorithm

The subseries generated in the preprocessing step come from PPs and nocturnal periods with original different lengths but filled with *NaNs* in order to obtain subseries with the same length and, therefore, seasonality equal to that length. Following the procedure explained in Chapter 7, the next step is to cluster the subseries to find similar glycemic profiles useful for building

seasonal models. However, standard clustering algorithms (such as FCM) cannot be directly applied to time series with “*missing values*” (incomplete data) because they are based on distance measures requiring all feature values of each data item. When distances are available, then the prototypes of each cluster and the memberships can be calculated.

In this case, it is impossible to delete incomplete CGM time series as an option to deal with missing values because it would result in deleting most data. However, instead of using the original FCM technique, the partial distance strategy can be applied to FCM through a *partial distance function* resulting in the so called PDSFCM algorithm. This algorithm is able to deal with CGM incomplete data, and PDSFCM can be obtained by making two modifications to the original FCM algorithm:

1. The partial distance between the incomplete CGM data point and cluster prototype can be defined as (5.23).
2. The cluster prototypes are updated to each iteration by (5.25).

Therefore, PDSFCM partitions CGM data into  $c > 1$  clusters by minimizing the objective function  $J_m$  (5.20) as FCM does but with the previous two modifications.

The FCM-based clustering algorithms need the optimal number of clusters ( $c$ ) as an input. The adjusted XBI validity index is used in this study instead of the CWBI used in Chapter 7. This is due to the easier adjustment to handle missing data. The XBI showed good performance in the number of clusters estimation for FCM, but (5.36) can not be directly used and XBI must be also adjusted to be able to use incomplete data. The original Euclidean distance has to be replaced again by the partial distances between the incomplete data points and the cluster prototypes. Including this modification, XBI can be computed for CGM incomplete series as follows:

$$XBI = \frac{C}{S} = \frac{\sum_{i=1}^c \sum_{k=1}^n u_{i,k}^2 \frac{s \sum_{l=1}^s (x_{kl} - \nu_{il})^2 i_{kl}}{\sum_{l=1}^s i_{kl}}}{n (\min_{i \neq j} \|\nu_i - \nu_j\|^2)}, \quad (8.1)$$

with

$$i_{kl} = \begin{cases} 1 & \text{if } x_{kl} \text{ is available,} \\ 0 & \text{otherwise} \end{cases}$$

where all terms have defined previously in Sections 5.4.2 and 5.4.3. The optimal number of clusters  $c$  corresponds to the lowest XBI value.

### 8.2.4 SARIMA local models integration for online glucose prediction

The approach for local modeling and local models integration will be the same tested in Chapter 7, where the seasonal local modeling and the subsequent fuzzy approach for model integration has been proven its good performance above the crisp approach, especially with a window of size 15 minutes ( $\mathbf{M}_{15}$ ).

Therefore, once the  $c$  (number of clusters) sets of similar enough CGM time series (including PPs and nocturnal periods) with the same (enforced) length are available as training sets thanks to the steps described in the previous subsections, a SARIMA model with seasonality period  $s$  (equal to the enforced time series period length) expressed as (6.1) can be identified for each cluster  $i$ , leading to a set of local models for glucose prediction  $\hat{G}_i(t)$  for  $i = \{1, 2, \dots, c\}$ .

Local models integration for real-time BG prediction  $G(t_p)$  for a desired PH at time instant  $t_p$  will be defined by the fuzzy approach, where the available (CGM previous to  $t_p$ ) data belongs partially (from 0 to 1) to all clusters and the BG prediction is computed through (7.2), by using the fuzzy membership  $u_i(t_p)$  at time  $t_p$  to each cluster  $i$  (7.3).

It must be remarked that after a period is finalized (*i.e.*, the next period/event is started), the whole period data is then appended as new data in the cluster with the highest membership. It is not expected that the profile of the modified data set changes too much provided that the aspect of the new time series will be similar to the others in the cluster, but it is fundamental to have this new series in the cluster as SARIMA models use pre-sampling data (previous data of the same subset/cluster) for their predictions. Additionally, it would be interesting to have all these new series stored in the cluster for an eventual online updating of the SARIMA models.

### 8.2.5 Supervision system and abnormal states detection

Based on the results of Chapter 7, the real-time BG prediction  $G(t_p)$  is expected to have low MAPE and RMSE errors for long PHs. Additionally, more real-time information can be given to the user.

On the one hand, a prediction of the glycemic status of the patient can be estimated at each time instant for a predefined PH. The glycemic status will be based on the BG prediction for this PH combined with the standard glycemia ranges defined, for example, in Fig. 7.5.

On the other hand, information can be provided about how similar the present period trying to forecast is to the data base from which models being used where identified. In order to do so, a trust index and a normality index are defined.

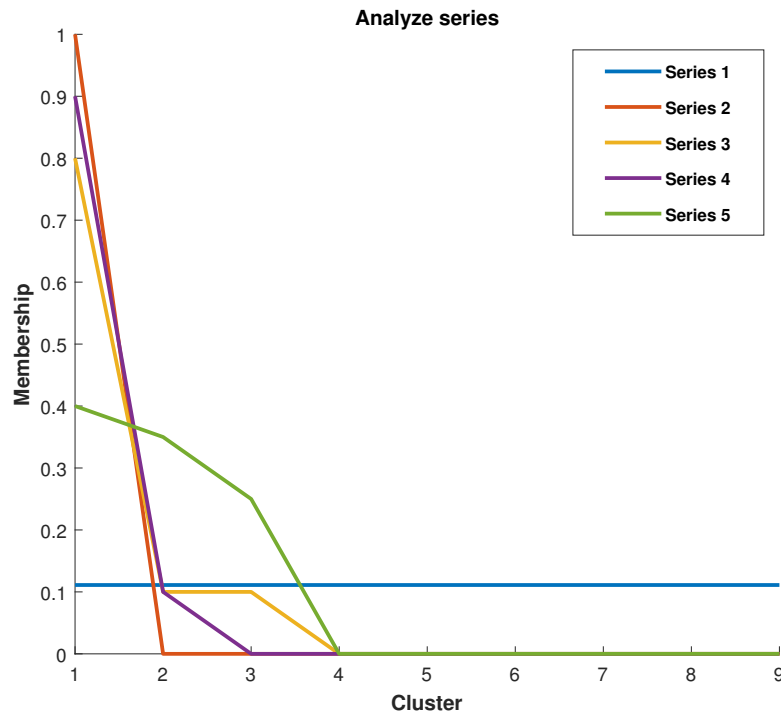


Figure 8.1: An illustrative example of an analysis of the fuzzy memberships series for choosing the right local model for use in prediction.

The “**trust index**” is designed to give information about how much the glucose estimation can be trusted. Provided that the glucose model is a multimodel system, the idea behind the index is that we can rely more on the model if one of the available models match perfectly the system. Fuzziness is included to give robustness by the integration of the local models when one of them is not enough to provide the best estimation, but the ideal case is that the membership is one for one of the local models and zero for the others. In order to check this mathematically in real-time, all online estimated memberships for the clusters are ordered in descending membership value and a new discrete-time series is created. Therefore, the best case will be when the first value is one and zero the others (see series 2 in orange color in Fig. 8.1). However, the worst case will be when all values are equal because it indicates that all the local models participate equally to the glucose estimation (see series 1 in blue color in Fig. 8.1). Mathematically, the property that can help in quantifying the differences is discrete time series convexity: in the best case convexity is high, and very low in the worst case. Hence, a discrete estimation of the series convexity can be easily computed by taking the second differences of the discrete series values. Finally, all terms in the second differences time series are summed

up then giving one as a result in the ideal case, and zero in the worst case (see trust index in the first subfigure of Fig. 8.2).

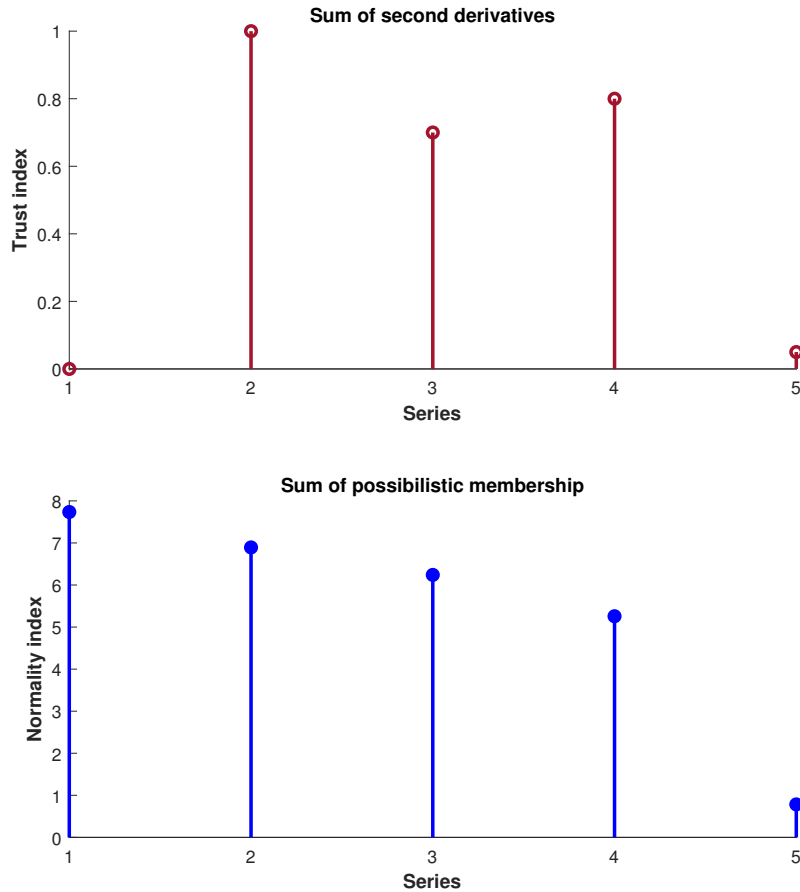


Figure 8.2: An illustrative example of using the trust index in order to measure the confidence of using a specific local model for prediction and the normality index in order to determine whether the period trying to estimate is normal or not.

The “**normality index**” is designed to determine whether the period trying to estimate is normal or not, in the sense that the behavior trying to forecast is among (somewhere in the middle of) the past behaviors (and then estimated by a single local model or a combination of them) or not (*i.e.*, beyond the past behaviors and then extrapolated by some or all the models). The use of possibility memberships instead of fuzzy memberships given by (5.21) can give some hints in this case. The possibility membership  $u_{ik}^p$  can be computed using:



$$u_{ik}^p = \frac{1}{1 + (d_{ik}^2)^{\frac{1}{m-1}}}, \quad (8.2)$$

where  $u_{ik}^p \in [0, 1]$ ,  $\sum_{i=1}^c u_{ik} \leq 1$ ,  $\forall k$  and  $0 < \sum_{k=1}^n u_{ik} < n \quad \forall i$ ,  $d$  is a distance function (the Euclidean distance function between  $k^{th}$  data points and  $i^{th}$  cluster center) and  $m \in [1, \infty)$  is the fuzziness parameter.

It can be observed that the difference between (5.21) and (8.2) is that the values in the first are normalized with the sum of all the memberships. It is very useful in the case of fuzzy membership use for model integration by weighing different glucose estimations. However, if the period trying to estimate is abnormal and far away from all the available local models, the distance will be similar to all prototypes then resulting in all the memberships equal when (5.21) is applied. The same result is obtained when a period trying to estimate is normal and just in the middle of the local models available. Nevertheless, if (8.2) is used the abnormal period will result in very low membership for all the clusters. The sum of all those possibility memberships computed online provides a measure of the abnormality of the period: the sum of the possibility memberships closer to zero the more abnormal period (see normality index in the second subfigure of Fig. 8.2).

When abnormal behaviors appear the user must be aware of the extrapolation being performed, and it will alert about an abnormal situation: hardware problems, extreme hypoglycemia or hyperglycemia, or any behavior beyond the past time series available for local models identification. The abnormal behavior, detected by low values (the threshold is equal to two after extensive simulations) like series 5 in the second subfigure of Fig. 8.2, can provide alerts to the user and highlight the necessity of new models to be learned due to new user behaviors.

### 8.2.6 Validation procedure

The six months available simulated data for the average adult patient of the UVA/Padova simulator [168] with three meals per day and multiple variability sources will be divided into two sets. The first 80% of the data is used as a *training* set for building suitable models, and 20% of the data is used for *testing* the prediction accuracy for these models with **RMSE** (mg/dL) and **MAPE** (%) metrics in different PHs.

## 8.3 Results

The methods presented in the previous section can be applied online by following the diagram presented in Fig. 8.3. The following subsections will apply this methodology to the available 6 months data.

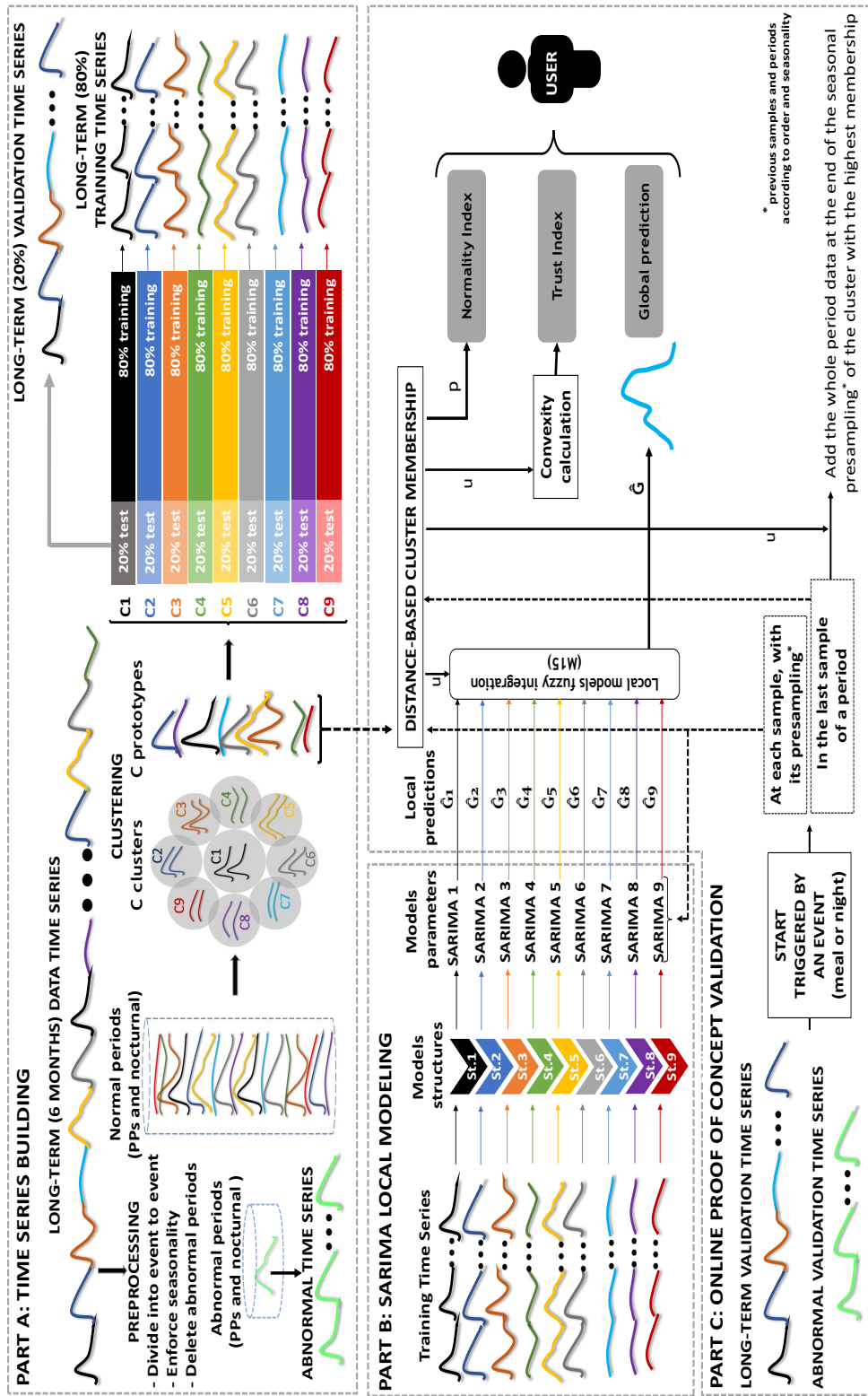


Figure 8.3: Diagram summarizing the online methodology application: time series building (A), SARIMA local modeling and identification (B), and online proof of concept validation (C).

### 8.3.1 Time series building

The long-term time series is preprocessed as explained in 8.2.2 resulting in 648 event-to-event periods (including PPs, and nocturnal periods) with a fixed length (enforced seasonality) of 100 observations, which is the length of the original longest period. Those periods in which hypoglycemia appeared and patients applied rescue actions have been excluded (from the rescue time till the end) from the modeling process, as well as the next period, since it will be affected by the rescue taken. These kind of periods can be used subsequently as an example of abnormal behaviors, because they are not included in the identification process and they should be different to any available cluster/model.

The purpose of the exclusion process of these periods is to measure later the ability of the algorithm to detect any abnormal status and raise a warning alarm for the patient to take the necessary corrective actions. Remark that normality/abnormality is related here to a common/uncommon behavior in the historical data. As far as only normal periods have been used in the modeling process, those periods leading to hypoglycemia will be considered abnormal time series.

### 8.3.2 Data clustering

PDSFCM clustering, in combination with the XBI cluster validity index for number of clusters determination, is performed over the whole set of normal periods, resulting in 9 clusters. Available periods are classified as shown in Fig. 8.4. The profile of each cluster can be observed in the cluster *prototypes* (see Fig. 8.5)

Once all periods are classified, 9 seasonal time series are created by concatenating those periods assigned to each cluster (*i.e.*, with similar profiles). Additionally, pre-sampling data (data before the period starts from the CGM time series) is inserted for each concatenated period. This historical data is needed by the models in a length depending on the models orders. A length of five pre-sampling data is considered enough, provided that models orders obtained in previous chapters were lower in the AR and MA processes. The 9 long-term training series were obtained using the first 80% of the available data, and the remaining 20% is included in a long-term validation time series. This concludes the PART A in Fig. 8.3.

### 8.3.3 Model identification

The PART B in Fig. 8.3 starts with the 9 time series available (including 80% of training data) with 100 samples plus 5 pre-samples per period, then having a 105 enforced seasonality (periods size). The model for each cluster is identified through a modified version of the *Box-Jenkins methodology*

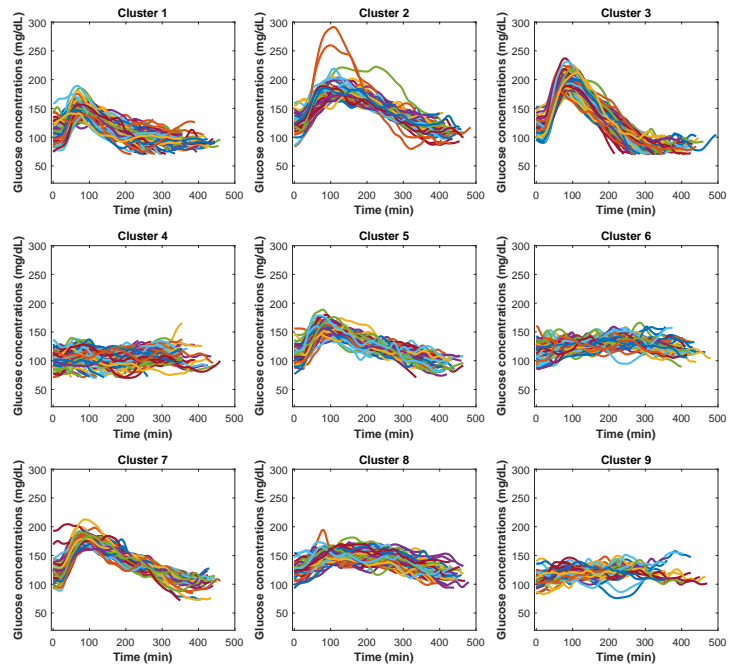


Figure 8.4: 9 clusters classification performed by PDSFCM algorithm.

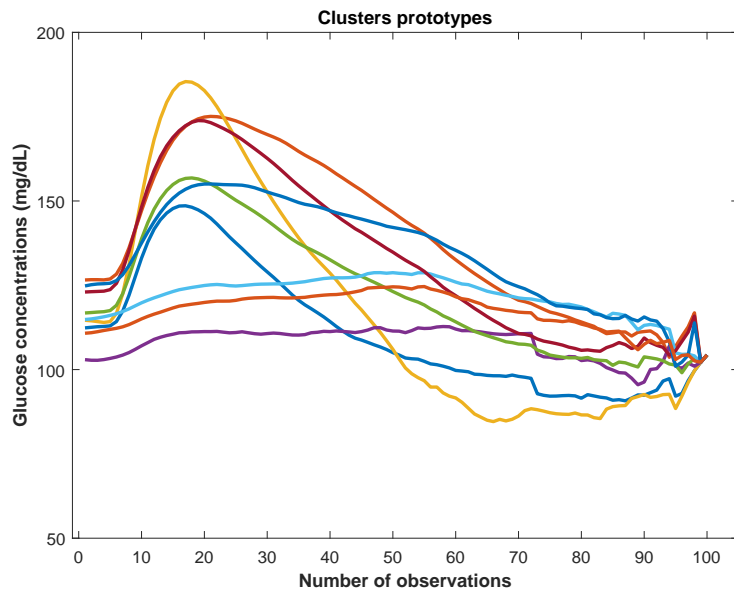


Figure 8.5: Cluster prototypes.

Cluster	SARIMA Model	PH					
		15 min	30 min	60 min	120 min	180 min	240 min
1	$(4, 0, 0)(1, 0, 1)_{105}$	8.15 (13.25)	7.50 (12.26)	6.92 (11.10)	4.19 (8.21)	3.40 (6.90)	3.07 (6.15)
2	$(0, 0, 4)(0, 1, 2)_{105}$	3.30 (4.56)	3.66 (5.21)	2.32 (3.90)	2.64 (5.25)	2.43 (4.91)	2.71 (5.30)
3	$(4, 0, 4)(2, 0, 1)_{105}$	4.26 (4.63)	4.06 (4.62)	4.16 (6.25)	4.22 (7.35)	3.76 (6.60)	4.54 (7.09)
4	$(1, 1, 1)(0, 1, 0)_{105}$	3.22 (3.21)	2.70 (3.01)	3.00 (3.10)	3.32 (4.37)	4.10 (5.23)	3.50 (4.81)
5	$(0, 1, 1)(1, 0, 1)_{105}$	0.62 (0.90)	0.91 (1.10)	1.16 (1.81)	1.29 (2.22)	1.51 (2.52)	1.37 (2.33)
6	$(0, 0, 4)(2, 0, 1)_{105}$	8.00 (8.25)	6.20 (7.13)	4.11 (5.26)	2.77 (4.35)	3.20 (4.70)	2.50 (4.10)
7	$(0, 0, 2)(1, 1, 0)_{105}$	3.11 (4.06)	2.71 (3.60)	2.80 (4.23)	2.61 (4.63)	2.56 (4.62)	3.70 (7.00)
8	$(4, 0, 1)(0, 0, 1)_{105}$	2.07 (2.20)	1.72 (2.00)	3.23 (4.01)	3.10 (4.22)	3.31 (5.40)	4.00 (7.30)
9	$(0, 0, 4)(1, 0, 0)_{105}$	0.60 (0.92)	1.64 (2.54)	2.07 (2.80)	1.60 (2.37)	1.68 (2.40)	1.56 (2.20)

Table 8.1: Prediction accuracy of seasonal local models, when using only one event/period as validation example set. **Notation:** prediction accuracy results presented as MAPE % (RMSE mg/dL).

(Section 3.3) able to deal with incomplete series (including NaNs). In this case, the residuals of missing values are not taken into account in the identification process. The appropriate SARIMA model structure and parameters are identified for each cluster.

Table 8.1 shows the appropriate SARIMA model for each cluster with the prediction accuracy results of the 4-hr PH. In the most challenging scenario of a 4-h PH forecasting period, a MAPE of 3.07% (RMSE of 6.15 mg/dL) is obtained for cluster 1, 2.71% (5.30 mg/dL) for cluster 2, 4.54% (7.09 mg/dL) for cluster 3, 3.50% (4.81 mg/dL) for cluster 4, 1.37% (2.33 mg/dL) for cluster 5, 2.50% (4.10 mg/dL) for cluster 6, 3.70% (7.00 mg/dL) for cluster 7, 4.00% (7.30 mg/dL) for cluster 8, and 1.56% (2.20 mg/dL) for cluster 9. Results for different PHs (15, 30, 60, 120, 180, and 240 minutes) using *only* one event/period as validation set example are presented in Table 8.1. Fig. 8.6 shows the local glucose predictions  $\hat{G}_i(t)$  for  $i = \{1, 2, \dots, 9\}$ , when using *only* one event/period as validation set of the 9 clusters.

### 8.3.4 Online forecasting validation

PART C in Fig. 8.3 is devoted to describe the online procedure proof of concept and its validation. Table 8.2 shows the average MAPE (%) and RMSE (mg/dL) values of **GSM** predictions through *only one* period/event validation data of the 9 clusters. Table 8.3 shows the average MAPE (%) and RMSE (mg/dL) values of **GSM** predictions through **all** the periods/events validation data of the 9 clusters.

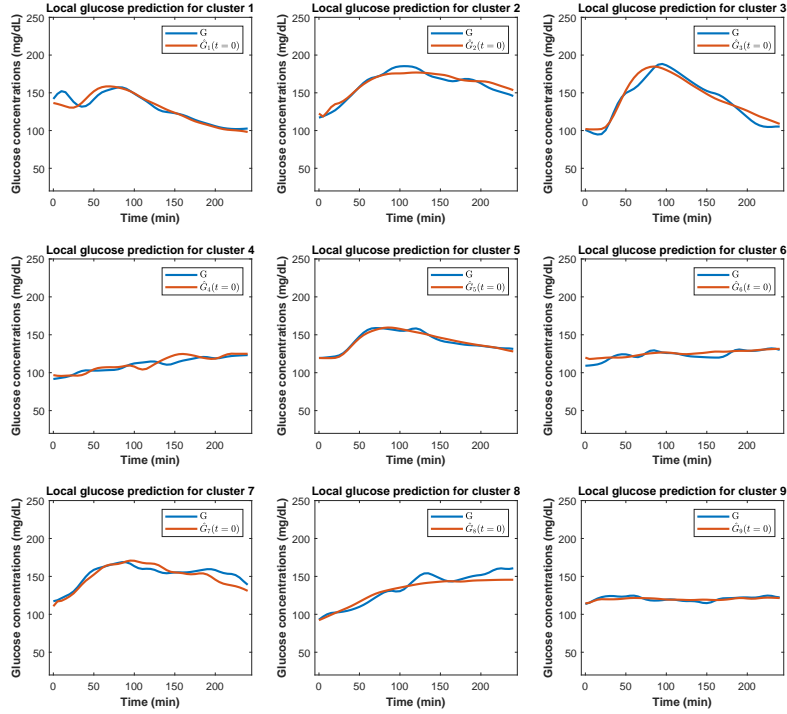


Figure 8.6: The local glucose predictions  $\hat{G}(t)$  for each cluster for a 4-h PH, when using only one event/period as validation example set of the 9 clusters.

The integration process is updated *online* for each validation period/event in order to get the online BG prediction as follows:

1. Calculate the membership for the new period/event available data to the nine clusters.
2. Combine the predictions (4-h PH) of the nine local models using the memberships calculated in the previous step.
3. Repeat the process until the period is finished.
4. Add the finished period to the corresponding cluster (the highest membership) as seasonal pre-sampling.
5. Repeat the previous steps with the new period/event and calculate the mean prediction error for the whole validation data set (129 periods).

Membership calculation	PH					
	15 min	30 min	60 min	120 min	180 min	240 min
$\mathbf{M}_{15}$	3.27 (5.11)	5.07 (7.84)	7.02 (11.72)	9.08 (15.00)	10.43 (17.45)	11.48 (18.50)

Table 8.2: Prediction accuracy of **GSM** for only one period/event validation example data of the 9 clusters. **Notation:** prediction accuracy results presented as MAPE % (RMSE mg/dL).

Membership calculation	PH					
	15 min	30 min	60 min	120 min	180 min	240 min
$\mathbf{M}_{15}$	4.10 (6.46)	5.95 (8.95)	8.43 (13.39)	11.32 (18.61)	13.65 (22.91)	13.97 (23.81)

Table 8.3: Prediction accuracy of **GSM** for all periods/events validation data of the 9 clusters. **Notation:** prediction accuracy results presented as MAPE % (RMSE mg/dL).

### 8.3.5 Supervision system validation and abnormal states detection

In addition to the glucose predictions value, a normality index and a trust index are provided to the user at each time instant as defined in 8.2.5. The whole information available at each instant is: nine  $\hat{G}_i(t)$  local estimations, a global estimation through the  $\mathbf{M}_{15}$  integration process, nine fuzzy memberships  $u_i$ , a trust index, nine possibilistic memberships  $P_i$ , and a normality index. The fuzzy memberships change with time and, therefore, weights in (7.2) for  $\hat{G}(t)$  120-min-ahead prediction in the following examples. As well, the trust index is high (close to one) when a single model represents the behavior for some time then giving confidence in the estimation. The normality index, using for its calculation the changing with time possibilistic memberships, is high (close to the number of clusters, nine in this case) when a lot of the available models can represent the behavior and very low (two or lower in this case) when none of them represent the behavior and, therefore, an abnormal situation is presented.

Fig. 8.7 is included to illustrate graphically the whole information available at each instant for a case with medium MAPE. Four cases can be devised in the plots with respect to trust index:

1. Changes in the local models used for global prediction make decrease trust index (see the models change between 160 and 180 min., and the effect on the trust index).
2. If the same local model is the mainly used for the global prediction, then the trust index increases (see, for example, since 50 min. until

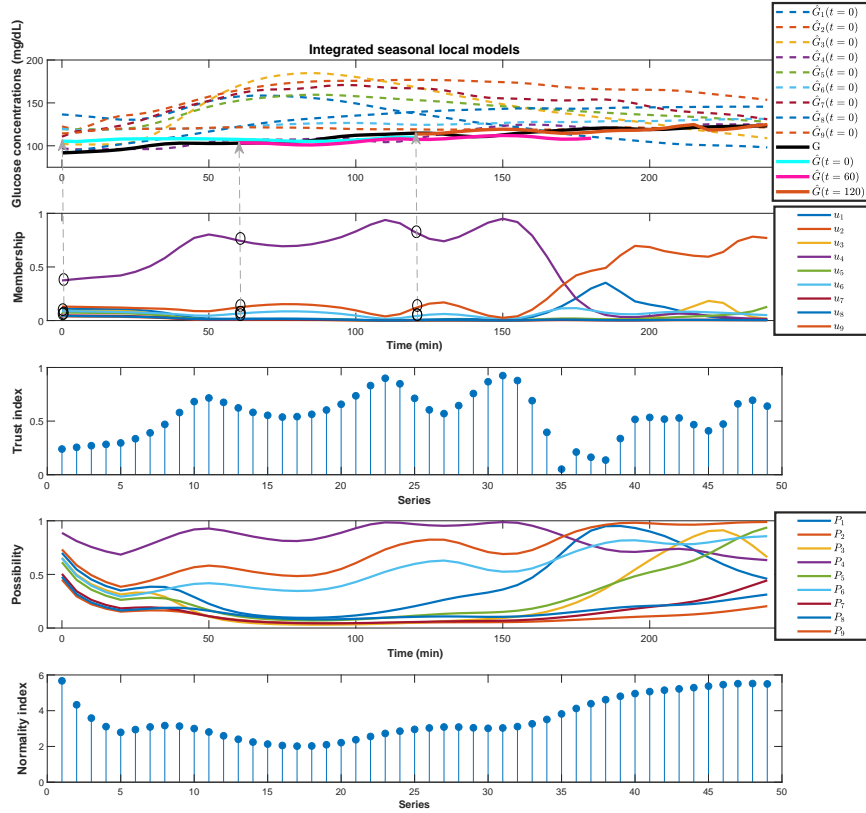


Figure 8.7: An example of **GSM** to demonstrate the relation between predictions, memberships, trust, and normality indexes. Local predictions at  $t = 0$ , memberships at each time instant, and global predictions at  $t = 0$ ,  $t = 60$ , and  $t = 120$  by using  $\mathbf{M}_{15}$  for a 120-min-ahead prediction. Membership values at the above prediction times are highlighted by the symbol  $\circ$ .

150 min. where model  $\hat{G}_4(t)$  can represent well the behavior).

3. When not only one but several models are used for the global prediction, then the trust index decreases (see the beginning till 50 and from the model change in 180 min. until the end).
4. In the case that only one model is used for the global prediction, then the trust index increases (compare the period from 100 to 150 min. where almost only one model is used to the period from 200 min. till the end, where some models are used for the prediction).

In this example, it can be seen that in the possibilistic memberships at least one model is close to one, then leading the normality index to values always above two. Therefore, no abnormal behavior is devised. At the



beginning and at the end (after 3 hours) a lot of the available models can represent the behavior, so normality index value is high.

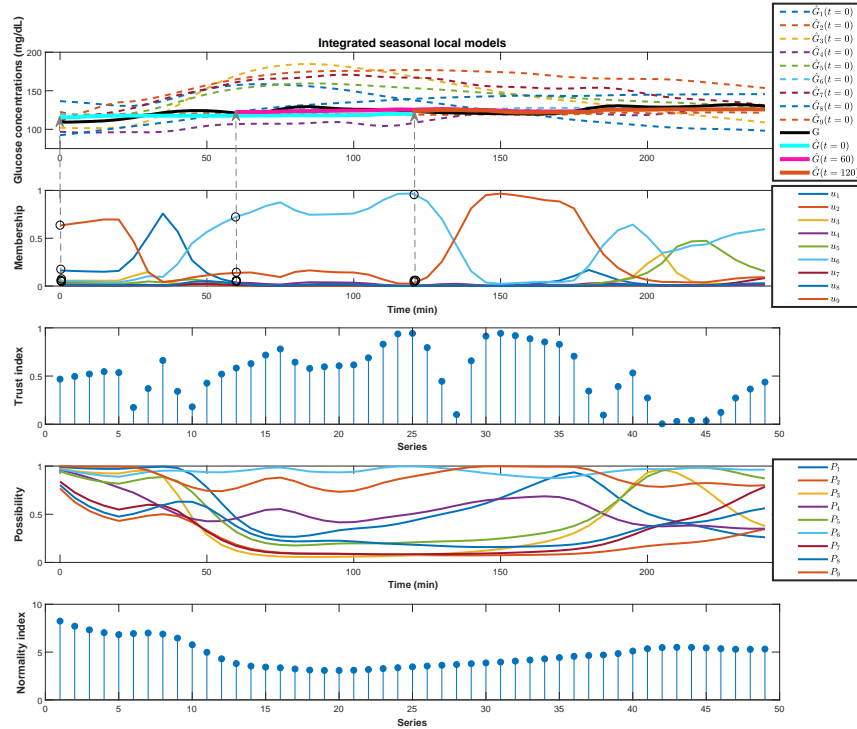


Figure 8.8: An example of **GSM** with a big prediction error for showing the relationship between predictions, memberships, trust, and normality indexes. Local predictions at  $t = 0$ , memberships at each time instant, and global predictions at  $t = 0$ ,  $t = 60$ , and  $t = 120$  by using  $M_{15}$  for a 120-min-ahead prediction. Membership values at the above prediction times are highlighted by the symbol  $\circ$ .

Fig. 8.8 (high MAPE) and Fig. 8.9 (low MAPE) show two more examples to illustrate the relationship between prediction errors, fuzzy memberships and trust index. High values of the trust index must be related to low MAPEs, and vice-versa. A simple comparative analysis of both plots confirms the approach correctness: the number of instant values of trust index that are relatively high (above 0.7, for example) is bigger in the case of low MAPE (more than 30 out of 50 points) than in the case of high MAPE (about 10 out of 50), and trust index also take negative values in this case.

The four behavior cases identified in Fig. 8.7 can also be detected in these two new figures, and the normality index values are also always high although in one of the cases MAPE is high: the error can be big, but the behavior can be considered normal.

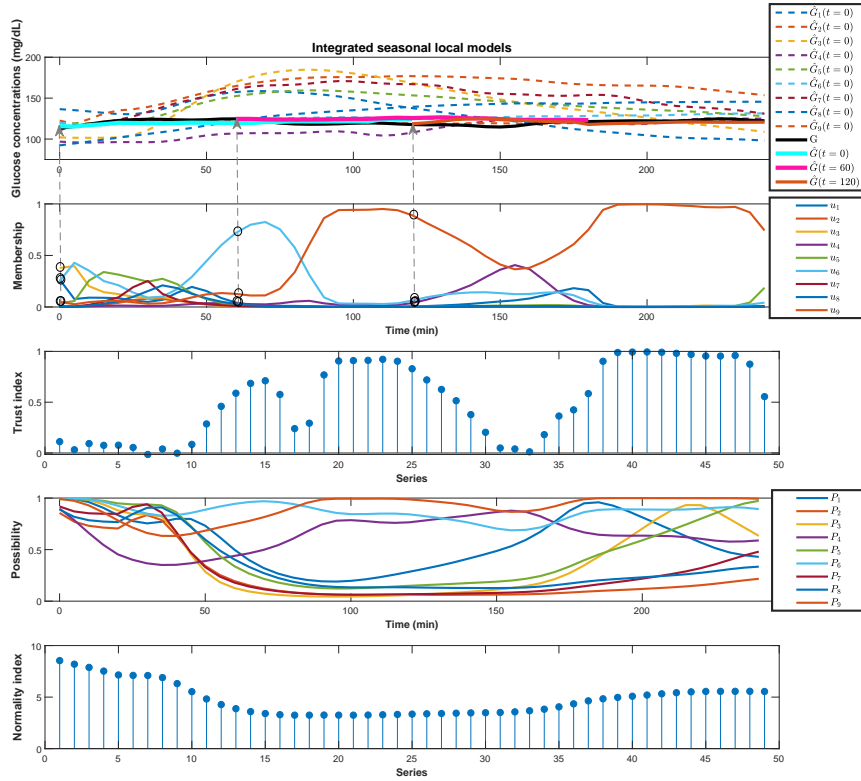


Figure 8.9: An example of **GSM** with a small prediction error for showing the relationship between predictions, memberships, trust, and normality indexes. Local predictions at  $t = 0$ , memberships at each time instant, and global predictions at  $t = 0$ ,  $t = 60$ , and  $t = 120$  by using  $M_{15}$  for a 120-min-ahead prediction. Membership values at the above prediction times are highlighted by the symbol  $\circ$ .

Although events/periods of hypoglycemia have been excluded in this study in the identification process, these abnormal states could be detected through the normality index. In the case of abnormal behavior time series (PPs leading to *hypoglycemia*), the MAPE will be high as expected, since this kind of abnormal behavior was not included for model identification (in the training data). In order to study this case, 13 PPs with abnormal behavior of 5 hours length have been used. A mean MAPE of 34.10% (RMSE of 27.73 mg/dL) is obtained for 60-min PH from the time instant  $t = 240$  until the end. This is a very high value when compared to mean MAPE of 8.43% (RMSE of 13.39 mg/dL) for 60-min PH shown in Table 8.3.

A final Fig. 8.10 is included to illustrate how the system behaves in the case of abnormal behaviors (*i.e.*, the time series leading to hypoglycemia). All the behaviors related to the trust index studied in the three previous

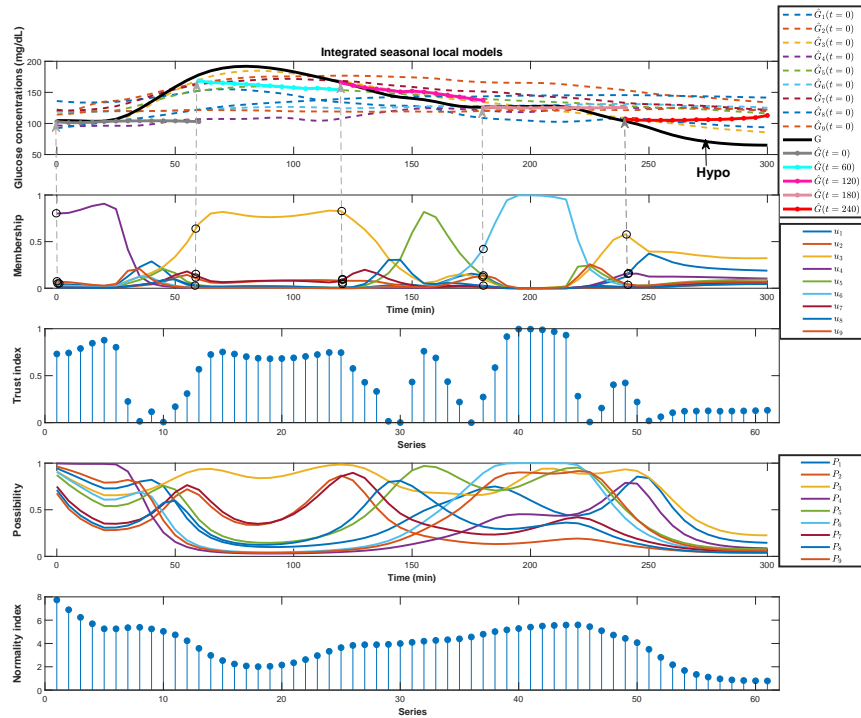


Figure 8.10: An example of how the system behaves in the case of abnormal behaviors. Local predictions at  $t = 0$ , memberships at each time instant, and global predictions at  $t = 0$ ,  $t = 60$ ,  $t = 120$ ,  $t = 180$ , and  $t = 240$  by using  $M_{15}$  for a 60-min-ahead prediction. Membership values at the above prediction times are highlighted by the symbol  $\circ$ .

figures are also present here, and the trust index is very low in different moments: when model changes are detected and also after time instant  $t = 240$ . However, the abnormal state can be detected through the normality index, where the lower values below two (threshold where none of the previously analyzed cases arrived) indicate an abnormal state as shown in the last subfigure 5 Fig. 8.10. This means that no local model can represent the behavior and therefore, an alarm for the user can be launched to take actions before upcoming hypoglycemia (about 30 min. before).

Fig. 8.11 and Fig. 8.12 show two additional examples of different PPs to illustrate the vital role of the normality index for detecting the abnormal states not only hypoglycemia but it can detect both hypoglycemia and hyperglycemia. In order to confirm the good results of the normality index, 118 periods have been tested including normal and abnormal behaviors (not used in the identification process but stored for this validation purpose). The normality index has been calculated for all the PPs and they have been

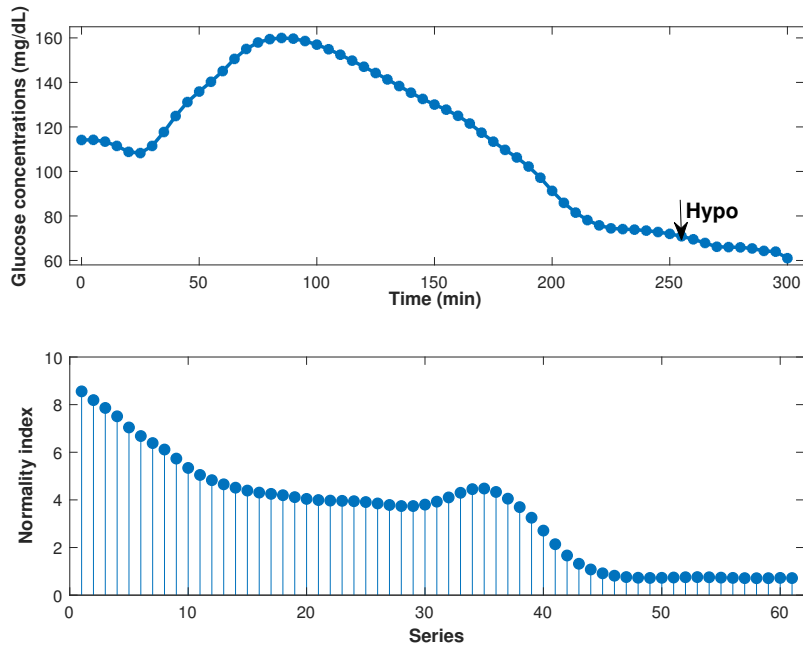


Figure 8.11: An example of detecting hypoglycemia through the normality index.

classified into two groups: those whose normality index is always above 2 (normal) those that go below 2 (abnormal). These two groups of classification lead to 66 normal periods and 52 abnormal periods. The Euclidean distance has been used to measure the real similarity between each PP and the prototypes of 9 clusters characterizing the identified models. The results of the similarity are shown in Fig. 8.13 where it can be seen that the PPs classified online (and in advance) by the normality index as abnormal, are actually farther to the prototypes (*i.e.*, learned behaviors and models) than those classified as normal. Box plot in Fig. 8.14 highlights this difference and both groups are SSD with  $p = 0.000$ .

### 8.4 Conclusions and margins for further improvement

The concept of seasonality has demonstrated successfully its accuracy in the BG prediction models for long PHs, but it is not directly applicable for the use in the normal life. However, the concept of seasonality can be enforced in normal life data then being very useful for developing BG prediction models or for detecting abnormal behaviors of people with diabetes.

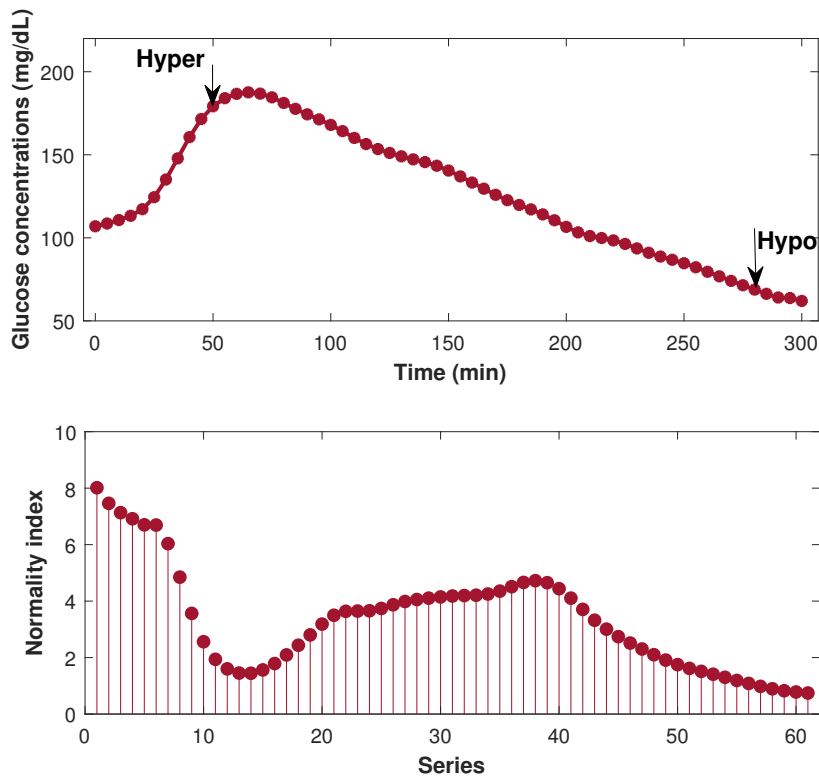


Figure 8.12: An example of detecting hypoglycemia and hyperglycemia through the normality index.

In this chapter, a proof of concept of how to apply the concept of seasonality in normal life has been presented and validated by using a realistic virtual patient under free-living conditions. PPs and nocturnal periods with different lengths were extracted from CGM time series data and pre-processed to fit the seasonality modeling approach. PDSFCM clustering, in combination with the XBI cluster validity index, has been used to detect the similarity between different periods, reach the best number of clusters, and classify the periods in 9 clusters. As well, SARIMA local prediction models have been used to identify a model for each cluster and a global model resulted from the online fuzzy integration of the local models. The model provided accurate predictions and has been augmented at each time instant with a complete supervision system giving a value of the confidence in the prediction (trust index) and a detection of future abnormal behaviors through the normality index (for example, hypoglycemia or hyperglycemia states).

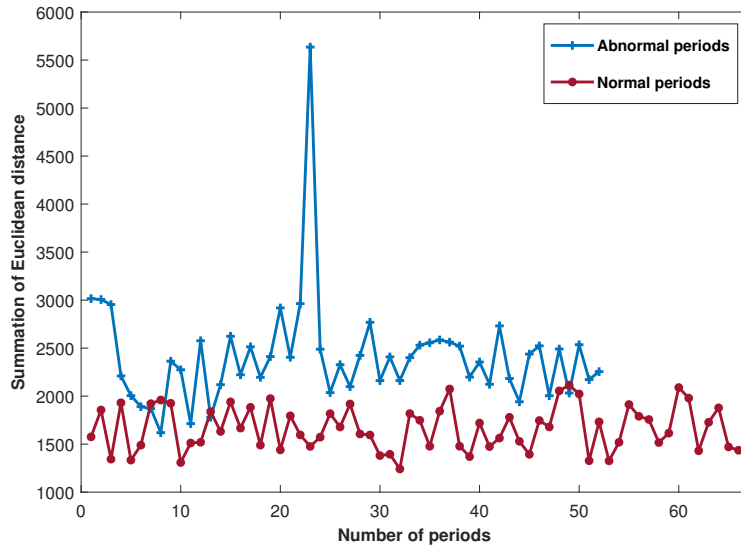


Figure 8.13: The similarity measure between the normal and abnormal group with the prototypes of 9 clusters by using the Euclidean distance.

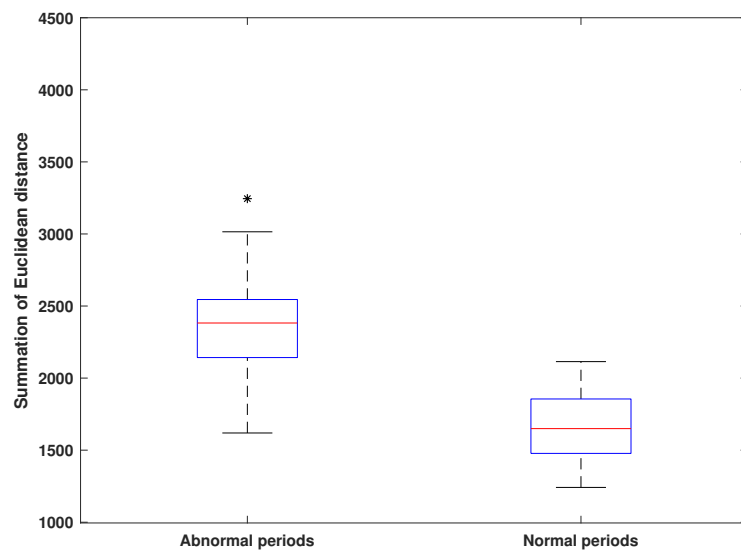


Figure 8.14: Boxplot of the normal and abnormal group. \* indicates SSD ( $p < 0.05$ ).

Results reported in Table 8.1 conclude that the SARIMA local models exhibit relatively higher forecasting accuracy for larger PHs (4-h PH) with lower prediction errors. As well, the results reported in Table 8.3 allow us to conclude that the **GSM** exhibits high prediction accuracy for larger PHs

(a 240-min-ahead prediction), therefore, **GSM** could be helpful to allow the diabetic patients anticipate therapeutic decisions. As well, this study has demonstrated the feasibility of using the *trust index* for measuring the confidence in the estimation and for choosing the appropriate local prediction model and the *normality index* for detecting the abnormal behavior/states before the action occurs without using the abnormal behavior time series in the model identification.

## Chapter 9

# Conclusions and Future work

This chapter introduces the conclusion of the research with some directions for possible future work. Section 9.1 summarizes the main conclusions of the thesis through the contributions presented in the previous chapters. Section 9.2 provides some ideas that could be developed to go beyond the objectives of the thesis but serve the main objective. Section 9.3 displays the results of the research through scientific publications in journals and conferences.

### 9.1 Conclusions

The tight monitoring of BG concentration is a very important and essential task in diabetes management for avoiding diabetes complications due to hypo- and hyperglycemic episodes. The predictor's models of BG concentrations for a long PH (*e.g.*, 60-120 min ahead in time) might improve T1D therapy by allowing the patient to tune the therapy based on BG future values. As well, the long PH allows the patient to avoid or mitigate critical events. The most important issue in the BG prediction is the “**accuracy**” of the prediction of BG concentrations in every glycemic range that significant in closed-loop applications based on MPC, and also in *open-loop* therapy to allow plenty of time for diabetics to take correction action and anticipate therapeutic decisions. This need links directly with the main objective of the thesis: developing new techniques for identifying individualized patient glucose prediction models in T1D.

In the *last decade*, there were several prediction methods proposed in the literature based on glucose prediction models that use only CGM history data as inputs, and have been reviewed in **Chapter 3**. *Recently*, with the technological progress and through many studies on the factors affecting the BG behavior, assorted attempts of using also insulin, CHO and PA information have been proposed as exogenous inputs in these models. In order to capture the dynamic characteristics of BG behaviors and improve prediction accuracy away from the physiological aspects. The main problem of



using exogenous variables in these models is the effects of these variables are influenced by physiological delays and inter- and intra-individuals variability. One of the most important BG predictive models is the linear empirical models that have proven their effectiveness in BG prediction such as AR, ARMA, ARX, and ARMAX models.

Linear time series models described in the literature perform better, in terms of the average RMSE, for 5- and 15-min PH, as compared to the RNN and nonlinear regression methods. However, the average RMSE increases with large PHs as described in **Chapter 4**. The concept of **seasonality** with a linear time series models has been proposed, so that the dynamic characteristics of CGM behavior could be better captured. In this thesis, the stochastic seasonal model (SARIMA) for predicting a patient's future BG concentrations has been developed by concatenating time series with fixed-length PPs. In **Chapter 6**, it has been shown that BG concentrations can be predicted using the recent history of CGM data only by using SARIMA model where the inter-patient variability in the data was representative of worst-case intra-patient variability defining a challenging scenario. As well, BG concentrations can be predicted by using exogenous input (*e.g.*, CSII) by using SARIMAX model.

In **Chapter 6**, despite the limitations of the study deals with the thesis **sub-objective 1**, and seasonality has shown to be an important factor to improve model predictive power allowing for the significant extension of PHs. The seasonal models with exogenous inputs that relate well with the CGM measurements could help to improve the prediction accuracy and PH. Furthermore, SARIMA and SARIMAX models exhibit relatively higher forecasting accuracy for larger PHs, giving results statistically comparable, where a MAPE of 6.62% (RMSE 10.29 mg/dL), and a MAPE of 5.12% (RMSE 8.47 mg/dL) are obtained for 5-hr PH forecasting period respectively. In addition, SARIMA and SARIMAX models are proved highly efficient in BG predicting superior to ARIMA and ARIMAX models.

SARIMA and SARIMAX models require similar enough glycemic patterns in CGM historical data to obtain accurate models, which cannot be assumed in free-living CGM data. Therefore, the classification of periods under scenarios yielding similar enough glycemic responses is the best way to fully exploit the benefit of seasonal models. The clustering techniques revision performed in **Chapter 5** has been used in **Chapter 7** to fulfil **Sub-objective 2**. It has been shown that automatic clustering of CGM time series obtained from AP experiments in T1D patients, including meals and exercise, finds different PPs time profiles and clusters are detected. Increasing the similarity between different glucose responses leads to a better fitted model, that is the main idea of the use of clustering techniques in this work (different PPs time profiles and clusters are detected).

The local modeling approach reveals as a good approach (different models with a simpler structure), instead of a single complex model of the system. Furthermore, the automatic detection of similar clusters/groups leads to a good global model for the whole dataset through the best local model. A SARIMAX seasonal model for each cluster (provided that PPs in the same cluster are similar) can be determined and, subsequently, used online by fuzzy integrating different available predictions by the Euclidean-based distance to cluster. SARIMAX local models with several physiological measurements from a multi-sensor body monitor worn on the arm and that are known to have a significant effect of the whole-body fuel metabolism could improve BG prediction accuracy. In addition, it can be concluded that SARIMAX local models with exogenous inputs (CSII, EE) exhibit relatively higher forecasting accuracy for larger PHs (3-h PH), giving results statistically comparable.

In **Chapter 7**, it has also been shown that **GSM** overcomes BG forecasting of the approaches found in the literature and exhibits high prediction accuracy for larger PHs (a 90-min-ahead prediction) with a MAPE of 8.66%. Therefore, the long-term complications of diabetes can be reduced by using online BG predictions through **GSM**, because the larger PHs through the **GSM** could allow patients to take corrective actions. Hence, **GSM** can be used to provide early hypoglycemia or hyperglycemia alarms. In addition, **PA** has a significant impact on BG concentrations in patients with T1D and, therefore, **PA** needs more attention through diabetes management.

The final challenge in this research is how to apply the concept of seasonality in the daily life, as stated in **Sub-objective 3**. A diabetic patient has PPs or night (nocturnal) periods with different lengths. In **Chapter 8**, it has been shown that the concept of seasonality can be performed in the normal daily life of diabetic people. Therefore, enforcing the concept of seasonality in the normal life of diabetics is very useful for predictor models or detecting an abnormal behavior and, therefore, it is an important advanced towards building a complete supervision system for the diabetes management.

In **Chapter 8**, it has been shown through long-term (six-months) simulated data from a virtual patient (without exercise) that the SARIMA local models exhibit higher forecasting online accuracy under free living conditions than other approaches in the literature, and for larger PHs (up to 240-min-ahead prediction). As well, the feasibility of using a *trust index* for measuring the confidence in the estimation and a *normality index* for detecting abnormal behavior/states has been tested. Therefore, this framework can be used to provide, for example, early hypoglycemia or hyperglycemia alarms in our case of study. Predicted BG values through the **GSM** can also guide patients with their daily insulin dose decisions and adjustments. Additionally, the **GSM** developed can be integrated with a model-based

control algorithm to build fully automated AP systems.

## 9.2 Future work

The main objective of this work has been addressed, but some ideas have raised during the research that are worth to mention. If more time would have been available, experiments with real data could reinforce the preliminary results presented here (especially in those cases where extensive simulations have been used), but experimental studies are time consuming and require lots of time and money, especially in the diabetes field.

Additionally, several open improvements on the proposed techniques could be pursued in the future to serve the main objective of the thesis. These improvements are not limited to data but rather to a deeper analysis of particular mechanisms and new proposals to try different procedures. The possible future work to be done is detailed as follows:

1. **A larger number of diabetic patients are required for building a good stochastic seasonal model.** The main problem found throughout the work has been the few data available for building good prediction models. Data from more patients, and more data for each patient would lead to capture more dynamic characteristics of the BG behavior.
2. **Addition of exogenous inputs to SARIMA local modeling by using clinical data.** In order to improve the prediction accuracy and increase the ability to capture the BG fluctuations more signals are needed. The case of physical activity has been addressed, and could be translated into real life using wearables, but it has not been fully tested with the final proposal. Other exogenous inputs, such as stress or other variables could also be considered.
3. **Alternative fuzzy clustering techniques able to deal with missing CGM data.** A fuzzy clustering approach, with good results, has been used but other techniques, such as those based on cluster dispersion, should be tested to work with missing data and clusters/groups with different dispersion in diabetic real datasets.
4. **Extend the enforcing seasonality in the normal life concept to all kind of periods/events.** In order to measure the ability of the techniques to detect hypoglycemic episodes, more experiments including this kind of behavior must be done. In this work, hypoglycemia has been considered as an abnormal behavior as a proof of concept in the abnormal states detection through the online BG prediction.

5. **Include online improvement of the model.** A dynamic fuzzy clustering technique could be used to, after a big enough number of new (*i.e.*, abnormal up to this moment) periods are detected, to develop a new cluster and a new local model for them.
6. **Integrate GSM into control algorithms in AP systems.** The proposed **GSM** can be integrated with, for example, a MPC algorithm for building a fully automated AP system.
7. **Study the effect of using the concept of seasonality in non-linear stochastic models.** The concept of seasonality could be used with non-linear stochastic models, such as seasonality with ANN models.

### 9.3 Publications

The following *journal papers* and *conference abstracts* have been published or submitted during the scientific research work leading to this thesis:

#### 9.3.1 Journal papers

1. Rossetti, P., Quiros, C., Moscardo, V., Comas, A., Giménez, M., Ampudia-Blasco, F.J., León, F., **Montaser, E.**, Conget, I., Bondia, J., and Vehí, J. (2016). Closed-loop control of postprandial glycemia using an insulin-on-board limitation through continuous action on glucose target. *Diabetes technology & therapeutics*, 19(6), pp.355-362, doi:10.1089/dia.2016.0443.
2. **Montaser, E.**, Díez, J.L., and Bondia, J. (2017). Stochastic Seasonal Models for Glucose Prediction in the Artificial Pancreas. *Journal of diabetes science and technology*, 11(6), pp.1124-1131, doi.org/10.1177/1932296817736074.
3. **Montaser, E.**, Díez, J.L, Rossetti, P., Rashid, M., Cinar, A., and Bondia, J. (2019). Seasonal Local Models for Glucose Prediction in Type 1 Diabetes. *Journal of Biomedical and Health Informatics*, doi:10.1109/JBHI.2019.2956704.
4. **Montaser, E.**, Díez, J.L., and Bondia, J. (2019). Stochastic Seasonal Models for Glucose Prediction in Type 1 Diabetes under Free-Living Conditions. *IEEE Transactions on Control Systems Technology*, Submitted.

### 9.3.2 Conference abstracts and posters

1. Rossetti, P., Quirós, C., Moscardo, V., Comas, A., Giménez, M., Ampudia-Blasco, F.J., León, F., **Montaser, E.**, Conget, I., Bondia, J., and Vehí, J. (2016a). Better postprandial glucose control with a new developed closed-loop control system as compared with open-loop treatment in patients with type 1 diabetes, In *76th American Diabetes Association Scientific Sessions*, New Orleans, Louisiana, USA.
2. Quirós, C., Giménez, M., Rossetti, P., Moscardo, V., Comas, A., Ampudia-Blasco, F.J., León, F., **Montaser, E.**, Conget, I., Bondia, J., and Vehí, J. (2016b). Better postprandial glucose control with a new closed-loop system as compared with open-loop treatment in patients with type 1 diabetes, In *European Association for the Study of Diabetes Annual Meeting*, Munich, Germany.
3. **Montaser, E.**, Díez, J.L., and Bondia, J. (2017). Stochastic seasonal models for glucose prediction in type 1 diabetes, In *10th International Conference on Advanced Technologies & Treatments for Diabetes*, Paris, France.
4. Díez, J.L, **Montaser, E.**, Rashid, M., Cinar, A., and Bondia, J. (2019). Fuzzy clustering based seasonal stochastic local modeling framework for glucose prediction in type 1 diabetes, In *12th International Conference on Advanced Technologies & Treatments for Diabetes*, Berlin, Germany.

# Appendices

# Appendix A

## Clarke error grid analysis

### A.1 Clarke error grid analysis

Two types of accuracy are defined to measure the performance of a BG meter: *analytical accuracy* and *clinical accuracy*. Analytical accuracy is a quantitative method for determining BG meter accuracy through the difference between the BG meter value and the BG reference value. Clinical accuracy is a *qualitative* method for determining BG meter accuracy in terms of impact of wrong measurements in clinical decisions. *Error grid analysis* is the accepted tool for clinical accuracy studies and is presented in this Appendix.

**Clarke Error Grid Analysis** (EGA) [316, 317] was developed in 1987 by Clarke and colleagues at the University of Virginia. Then EGA became one of the acceptable gold standards in order to quantify BG meter accuracy. An error grid with five different zones (Fig. A.1) were defined as follows

1. **Zone A** contains points that are within 20% discrepancy between BG meter values and reference BG values, “clinically accurate”;
2. **Zone B** contains pairs with greater than 20% discrepancy but would not lead to inappropriate treatment, “benign errors”;
3. **Zone C** contains pairs with a discrepancy that may lead to errors in treatment with low risk, therefore this values would command to unnecessary treatment, “overcorrection errors”;
4. **Zone D** contains pairs with a discrepancy that may lead to erroneous to detect hypo- or hyperglycemia, “failure to treat errors”; and
5. **Zone E** contains pairs with a discrepancy that may lead to confusing the treatment of hypoglycemia for hyperglycemia and vice-versa (*i.e.*, an inappropriate treatment), “erroneous treatment errors”.

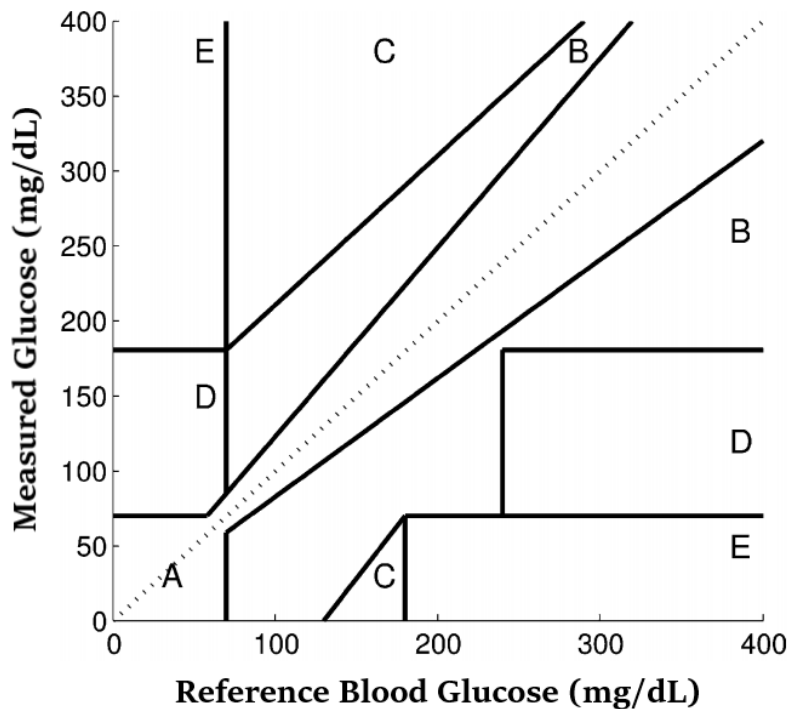


Figure A.1: The Clarke Error Grid Analysis for determining BG meter accuracy with five different zones. Modified from [12].

EGA's output is the percentage of data points that take place into each zone describing the risk of wrong clinical decisions. While this was developed to deal with glucometer data, the advent of CGM has raised criticisms of not taking into account the rate/direction of change, leading to new error grids by the same team such as the Continuous Glucose Error Grid Analysis, which is described next.

## A.2 Continuous glucose error grid analysis

The **continuous glucose error grid analysis** (CG-EGA) was developed to overcome the disadvantage of EGA to determine the clinical accuracy of CGM systems through point and rate/direction accuracy. The CG-EGA is composed of two components: (a) *rate error grid analysis* (R-EGA) and (b) *point error grid analysis* (P-EGA). Both components have a similar clinical meaning to the original EGA as shown in Fig. A.2. Unlike the original EGA, the CG-EGA deals with time-based characteristics of the CGM data. The CG-EGA has the ability to evaluate pairs of reference and sensor readings as a process in time through a bidimensional time series. It also can take



into account the physiological time lag between glucose levels in blood and ISF [317].

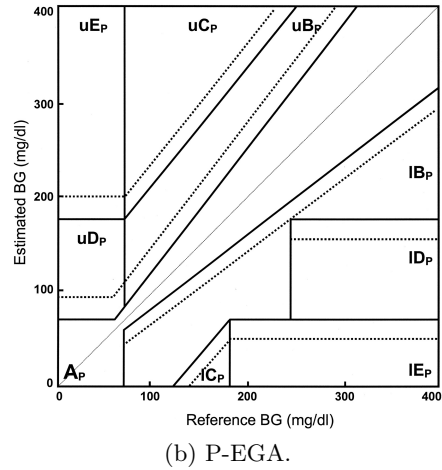
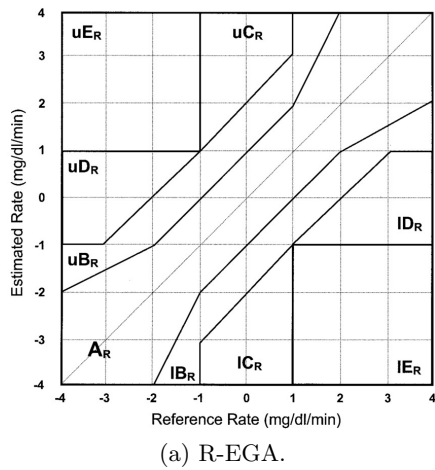


Figure A.2: The continuous glucose error grid analysis (CG-EGA): (a) rate error grid analysis (R-EGA) and (b) point error grid analysis (P-EGA) to quantify the clinical accuracy of CGM systems with five different zones as the EGA. Adapted from [13].

# Bibliography

- [1] Diabetes Atlas. International diabetes federation. idf diabetes atlas, 8th edn. brussels, belgium: International diabetes federation, 2017.
- [2] Joseph M Tibaldi. Evolution of insulin: from human to analog. *The American journal of medicine*, 127(10):S25–S38, 2014.
- [3] Savitha Subramanian, David Baidal, Jay S Skyler, and Irl B Hirsch. The management of type 1 diabetes. In *Endotext [Internet]*. MDText.com, Inc., 2016.
- [4] Kamuran Turksoy, Thiago Marques Luz Paulino, Dessi P Zaharieva, Loren Yavelberg, Veronica Jamnik, Michael C Riddell, and Ali Cinar. Classification of physical activity: information to artificial pancreas control systems in real time. *Journal of diabetes science and technology*, 9(6):1200–1207, 2015.
- [5] Kenro Nishida, Seiya Shimoda, Kenshi Ichinose, Eiichi Araki, and Motoaki Shichiri. What is artificial endocrine pancreas? mechanism and history. *World Journal of Gastroenterology: WJG*, 15(33):4105, 2009.
- [6] Boris Kovatchev. A century of diabetes technology: Signals, models, and artificial pancreas control. *Trends in Endocrinology & Metabolism*, 2019.
- [7] Paolo Rossetti, Carmen Quiros, Vanessa Moscardo, Anna Comas, Marga Giménez, F Javier Ampudia-Blasco, Fabián León, Eslam Montaser, Ignacio Conget, Jorge Bondia, et al. Closed-loop control of postprandial glycemia using an insulin-on-board limitation through continuous action on glucose target. *Diabetes technology & therapeutics*, 19(6):355–362, 2017.
- [8] D Elleri, JM Allen, J Harris, K Kumareswaran, M Nodale, L Leelarathna, CL Acerini, A Haidar, ME Wilinska, N Jackson, et al. Absorption patterns of meals containing complex carbohydrates in type 1 diabetes. *Diabetologia*, 56(5):1108–1117, 2013.

- [9] Michael C Riddell, Dessi P Zaharieva, Loren Yavelberg, Ali Cinar, and Veronica K Jamnik. Exercise and the development of the artificial pancreas: one of the more difficult series of hurdles. *Journal of diabetes science and technology*, 9(6):1217–1226, 2015.
- [10] Alboukadel Kassambara. *Practical guide to cluster analysis in R: Un-supervised machine learning*, volume 1. STHDA, 2017.
- [11] Kamuran Turksoy, Iman Hajizadeh, Nicole Hobbs, Jennifer Kilkus, Elizabeth Littlejohn, Sediqeh Samadi, Jianyuan Feng, Mert Sevil, Caterina Lazaro, Julia Ritthaler, et al. Multivariable artificial pancreas for various exercise types and intensities. *Diabetes technology & therapeutics*, 20(10):662–671, 2018.
- [12] Simone Del Favero, Andrea Facchinetti, and Claudio Cobelli. A glucose-specific metric to assess predictors and identify models. *IEEE Transactions on Biomedical Engineering*, 59(5):1281–1290, 2012.
- [13] Boris P Kovatchev, Linda A Gonder-Frederick, Daniel J Cox, and William L Clarke. Evaluating the accuracy of continuous glucose-monitoring sensors: continuous glucose-error grid analysis illustrated by theasense freestyle navigator data. *Diabetes Care*, 27(8):1922–1928, 2004.
- [14] Pavel Berkhin. A survey of clustering data mining techniques. In *Grouping multidimensional data*, pages 25–71. Springer, 2006.
- [15] John E Hall. *Guyton and Hall textbook of medical physiology e-Book*. Elsevier Health Sciences, 2015.
- [16] Leonid Poretsky et al. *Principles of diabetes mellitus*. Springer, 2010.
- [17] P James and R McFadden. Understanding the processes behind the regulation of blood glucose. *Nursing times*, 100(16):56–58, 2004.
- [18] Derek LeRoith, Simeon I Taylor, and Jerrold M Olefsky. *Diabetes mellitus: a fundamental and clinical text*. Lippincott Williams & Wilkins, 2004.
- [19] Diana W Guthrie, CDE BC-ADM, and Richard A Guthrie. *Management of diabetes mellitus: A guide to the pattern approach*. Springer Publishing Company, 2008.
- [20] American Diabetes Association et al. Diagnosis and classification of diabetes mellitus. *Diabetes care*, 37(Supplement 1):S81–S90, 2014.
- [21] Francesco Belfiore and Carl Erik Mogensen. *New concepts in diabetes and its treatment*. Karger Medical and Scientific Publishers, 2000.

- [22] Centers for Disease Control, Prevention, et al. National diabetes statistics report, 2017. *Atlanta, GA: Centers for Disease Control and Prevention*, 2017.
- [23] David E Goldstein, Randie R Little, Rodney A Lorenz, John I Malone, David Nathan, Charles M Peterson, and David B Sacks. Tests of glycemia in diabetes. *Diabetes care*, 27(7):1761–1773, 2004.
- [24] Aleksandra Krzewska and Iwona Ben-Skowronek. Effect of associated autoimmune diseases on type 1 diabetes mellitus incidence and metabolic control in children and adolescents. *BioMed research international*, 2016, 2016.
- [25] Kelly M West. *Epidemiology of diabetes and its vascular lesions*. Elsevier, 1978.
- [26] M Laakso. Cardiovascular disease in type 2 diabetes: challenge for treatment and prevention. *Journal of internal medicine*, 249(3):225–235, 2001.
- [27] Markku Laakso. Cardiovascular disease in type 2 diabetes from population to man to mechanisms: the kelly west award lecture 2008. *Diabetes care*, 33(2):442–449, 2010.
- [28] Sharon J Herring and Emily Oken. Obesity and diabetes in mothers and their children: can we stop the intergenerational cycle? *Current diabetes reports*, 11(1):20–27, 2011.
- [29] PM Catalano. The impact of gestational diabetes and maternal obesity on the mother and her offspring. *Journal of Developmental Origins of Health and Disease*, 1(4):208–215, 2010.
- [30] American Diabetes Association et al. Gestational diabetes mellitus. *Diabetes care*, 27:S88, 2004.
- [31] Gareth Williams and John C Pickup. *Handbook of diabetes*. Wiley-Blackwell, 2004.
- [32] Kenneth L Becker. *Principles and practice of endocrinology and metabolism*. Lippincott Williams & Wilkins, 2001.
- [33] International Hypoglycaemia Study Group et al. Glucose concentrations of less than 3.0 mmol/l (54 mg/dl) should be reported in clinical trials: a joint position statement of the american diabetes association and the european association for the study of diabetes. *Diabetes Care*, 40(1):155–157, 2017.

- [34] Thomas Danne, Revital Nimri, Tadej Battelino, Richard M Bergental, Kelly L Close, J Hans DeVries, Satish Garg, Lutz Heinemann, Irl Hirsch, Stephanie A Amiel, et al. International consensus on use of continuous glucose monitoring. *Diabetes care*, 40(12):1631–1640, 2017.
- [35] MH Beers et al. *The merck manual of health & aging. new jersey.. merck & co*, 2004.
- [36] Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *New England journal of medicine*, 329(14):977–986, 1993.
- [37] UK Prospective Diabetes Study (UKPDS) Group et al. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (ukpds 33). *The lancet*, 352(9131):837–853, 1998.
- [38] William T Cefalu and Jack L Leahy. *Insulin Therapy*. CRC Press, 2002.
- [39] Ali Cinar and Kamuran Turksoy. *Advances in Artificial Pancreas Systems: Adaptive and Multivariable Predictive Control*. Springer, 2018.
- [40] Julia MW Wong and David JA Jenkins. Carbohydrate digestibility and metabolic effects. *The Journal of nutrition*, 137(11):2539S–2546S, 2007.
- [41] Jaapna Dhillon, Janice Y Lee, and Richard D Mattes. The cephalic phase insulin response to nutritive and low-calorie sweeteners in solid and beverage form. *Physiology & behavior*, 181:100–109, 2017.
- [42] Gérard Slama, Jean-Claude Klein, Anne Delage, Enrique Ardila, Hervé Lemaignan, Laure Papoz, and Georges Tchobroutsky. Correlation between the nature and amount of carbohydrate in meal intake and insulin delivery by the artificial pancreas in 24 insulin-dependent diabetics. *Diabetes*, 30(2):101–105, 1981.
- [43] DJ Jenkins, TM Wolever, Anthony R Leeds, Miguel A Gassull, Peter Haisman, Jang Dilawari, David V Goff, Geoffrey L Metz, and KG Alberti. Dietary fibres, fibre analogues, and glucose tolerance: importance of viscosity. *Br Med J*, 1(6124):1392–1394, 1978.
- [44] Frank Q Nuttall, Arshag D Mooradian, Mary C Gannon, Charles Billington, and Phillip Krezowski. Effect of protein ingestion on the glucose and insulin response to a standardized oral glucose load. *Diabetes care*, 7(5):465–470, 1984.

- [45] Livia SA Augustin, Cyril WC Kendall, David JA Jenkins, Walter C Willett, Arne Astrup, Alan W Barclay, Inger Björck, Jennie C Brand-Miller, Furio Brighenti, Anette E Buyken, et al. Glycemic index, glycemic load and glycemic response: an international scientific consensus summit from the international carbohydrate quality consortium (icqc). *Nutrition, Metabolism and Cardiovascular Diseases*, 25(9):795–815, 2015.
- [46] DJ Jenkins, TM Wolever, Rodney H Taylor, Helen Barker, Hashmeim Fielden, Janet M Baldwin, Allen C Bowling, Hillary C Newman, Alexandra L Jenkins, and David V Goff. Glycemic index of foods: a physiological basis for carbohydrate exchange. *The American journal of clinical nutrition*, 34(3):362–366, 1981.
- [47] Fao. *Carbohydrates in Human Nutrition: Report of a Joint FAO/WHO Expert Consultation*. 1998.
- [48] Pietro Galassetti and Michael C Riddell. Exercise and type 1 diabetes (t1dm). *Compr Physiol*, 3(3):1309–1336, 2013.
- [49] Jacklyn A Pivovarov, Craig E Taplin, and Michael C Riddell. Current perspectives on physical activity and exercise for youth with diabetes. *Pediatric diabetes*, 16(4):242–255, 2015.
- [50] American College of Sports Medicine et al. *ACSM’s health-related physical fitness assessment manual*. Lippincott Williams & Wilkins, 2013.
- [51] BA Franklin, MH Whaley, ET Howley, and GJ Balady. American college of sports medicine: Acsm’s guidelines for exercise testing and prescription, 2000.
- [52] Ludovic Jean Chassin, Malgorzata E Wilinska, and Roman Hovorka. Intense exercise in type 1 diabetes: exploring the role of continuous glucose monitoring. *Journal of diabetes science and technology*, 1(4):570–573, 2007.
- [53] KJ Guelfi, N Ratnam, GA Smythe, TW Jones, and PA Fournier. Effect of intermittent high-intensity compared with continuous moderate exercise on glucose production and utilization in individuals with type 1 diabetes. *American Journal of Physiology-Endocrinology And Metabolism*, 292(3):E865–E870, 2007.
- [54] Cajsa Tonoli, Elsa Heyman, Bart Roelands, Luk Buyse, Stephen S Cheung, Serge Berthoin, and Romain Meeusen. Effects of different types of acute and chronic (training) exercise on glycaemic control in type 1 diabetes mellitus. *Sports medicine*, 42(12):1059–1080, 2012.

- [55] Rémi Rabasa-Lhoret, Josée Bourque, Francine Ducros, and Jean-Louis Chiasson. Guidelines for premeal insulin dose reduction for postprandial exercise of different intensities and durations in type 1 diabetic subjects treated intensively with a basal-bolus insulin regimen (ultralente-lispro). *Diabetes Care*, 24(4):625–630, 2001.
- [56] Juha A Tuominen, S-L Karonen, L Melamies, G Bolli, and VA Koivisto. Exercise-induced hypoglycaemia in iddm patients treated with a short-acting insulin analogue. *Diabetologia*, 38(1):106–111, 1995.
- [57] C Purdon, M Brousson, SL Nyveen, PD Miles, JB Halter, M Vranic, and EB Marliss. The roles of insulin and catecholamines in the glucoregulatory response during intense exercise and early recovery in insulin-dependent diabetic and control subjects. *The Journal of Clinical Endocrinology & Metabolism*, 76(3):566–573, 1993.
- [58] Ronald J Sigal, Cameron Purdon, Simon J Fisher, Jeffrey B Halter, Mladen Vranic, and Errol B Marliss. Hyperinsulinemia prevents prolonged hyperglycemia after intense exercise in insulin-dependent diabetic subjects. *The Journal of Clinical Endocrinology & Metabolism*, 79(4):1049–1057, 1994.
- [59] Sriparna Ghosal, Amanda Nunley, Parinaz Mahbod, Alfor G Lewis, Eric P Smith, Jenny Tong, David A D’Alessio, and James P Herman. Mouse handling limits the impact of stress on metabolic endpoints. *Physiology & behavior*, 150:31–37, 2015.
- [60] W Kenneth Ward, Jessica R Castle, and Joseph El Youssef. Safe glycemic management during closed-loop treatment of type 1 diabetes: the role of glucagon, use of multiple sensors, and compensation for stress hyperglycemia, 2011.
- [61] AM Albisser, BS Leibel, TG Ewart, Z Davidovac, CK Botz, and W Zingg. An artificial endocrine pancreas. *Diabetes*, 23(5):389–396, 1974.
- [62] A Ml Albisser, BS Leibel, TG Ewart, Z Davidovac, CK Botz, W Zingg, H Schipper, and R Gander. Clinical control of diabetes by the artificial pancreas. *Diabetes*, 23(5):397–404, 1974.
- [63] Ernst-Friedrich Pfeiffer, Ch Thum, and AH Clemens. The artificial beta cell—a continuous control of blood sugar by external regulation of insulin infusion (glucose controlled insulin infusion system). *Hormone and Metabolic Research*, 6(05):339–342, 1974.

- [64] AH Clemens, PH Chang, and RW Myers. The development of biostator, a glucose controlled insulin infusion system (gciis). *Hormone and metabolic research= Hormon-und Stoffwechselforschung= Hormones et metabolisme*, pages 23–33, 1977.
- [65] Motoaki Shichiri, R Kawamori, Y Goriya, Y Yamasaki, M Nomura, N Hakui, and H Abe. Glycaemic control in pancreatectomized dogs with a wearable artificial endocrine pancreas. *Diabetologia*, 24(3):179–184, 1983.
- [66] E. Salzsieder, G. Albrecht, U. Fischer, and E. . Freyse. Kinetic modeling of the glucoregulatory system to improve insulin therapy. *IEEE Transactions on Biomedical Engineering*, BME-32(10):846–855, Oct 1985.
- [67] Uwe Fischer, Eckhard Salzsieder, Ernst-Joachim Freyse, and Günther Albrecht. Experimental validation of a glucose-insulin control model to simulate patterns in glucose turnover. *Computer methods and programs in biomedicine*, 32(3-4):249–258, 1990.
- [68] B Wayne Bequette. Challenges and recent progress in the development of a closed-loop artificial pancreas. *Annual reviews in control*, 36(2):255–266, 2012.
- [69] Claudio Cobelli, Chiara Dalla Man, Giovanni Sparacino, Lalo Magni, Giuseppe De Nicolao, and Boris P Kovatchev. Diabetes: models, signals, and control. *IEEE reviews in biomedical engineering*, 2:54, 2009.
- [70] SK Vashist. Continuous glucose monitoring systems: A review, diagnostics 3 (4), 385-412, 2013.
- [71] R Boyd, B Leigh, and P Stuart. Capillary versus venous bedside blood glucose estimations. *Emergency medicine journal*, 22(3):177–179, 2005.
- [72] Miranda Langendam, Yoeri M Luijf, Lotty Hooft, J Hans DeVries, Aart H Mudde, and Rob JPM Scholten. Continuous glucose monitoring systems for type 1 diabetes mellitus. *The Cochrane Library*, 2012.
- [73] B Wayne Bequette. Continuous glucose monitoring: real-time algorithms for calibration, filtering, and alarms. *Journal of diabetes science and technology*, 4(2):404–418, 2010.
- [74] Céline M Girardin, Céline Huot, Monique Gonthier, and Edgard Delvin. Continuous glucose monitoring: A review of biochemical per-



- spectives and clinical use in type 1 diabetes. *Clinical biochemistry*, 42(3):136–142, 2009.
- [75] Giovanni Sparacino, Andrea Facchinetti, and Claudio Cobelli. “smart” continuous glucose monitoring sensors: on-line signal processing issues. *Sensors*, 10(7):6751–6772, 2010.
- [76] FDA Advisory Panel. Fda advisory panel votes to recommend non-adjunctive use of dexcom g5 mobile cgm. *Diabetes Technol. Ther.*, 18:512–516, 2016.
- [77] Chelsea Zimmerman, Anastasia Albanese-O’Neill, and Michael J Haller. Advances in type 1 diabetes technology over the last decade. *European Endocrinology*, 15(2):70, 2019.
- [78] JC Pickup, H Keen, JA Parsons, and KG Alberti. Continuous subcutaneous insulin infusion: an approach to achieving normoglycaemia. *Br Med J*, 1(6107):204–207, 1978.
- [79] John Pickup and Harry Keen. Continuous subcutaneous insulin infusion at 25 years: evidence base for the expanding use of insulin pump therapy in type 1 diabetes. *Diabetes care*, 25(3):593–598, 2002.
- [80] US Food, Drug Administration, et al. General hospital and personal use medical devices panel. *Insulin Infusion Pumps Panel Information*, 2010.
- [81] American Association of Diabetes Educators et al. Insulin pump therapy: guidelines for successful outcomes. *AADE Consensus Summit*, 2008.
- [82] M James Lenhard and Grafton D Reeves. Continuous subcutaneous insulin infusion: a comprehensive review of insulin pump therapy. *Archives of Internal Medicine*, 161(19):2293–2300, 2001.
- [83] Howard Zisser, Lauren Robinson, Wendy Bevier, Eyal Dassau, Christian Ellingsen, Francis J Doyle III, and Lois Jovanovic. Bolus calculator: a review of four “smart” insulin pumps. *Diabetes technology & therapeutics*, 10(6):441–444, 2008.
- [84] Jiming Chen, Kejie Cao, Youxian Sun, Yang Xiao, and Xu Kevin Su. Continuous drug infusion for diabetes therapy: a closed-loop control system design. *EURASIP Journal on Wireless Communications and Networking*, 2008(1):495185, 2007.
- [85] Robert S Parker, Francis J Doyle III, Jennifer H Ward, and Nicholas A Peppas. Robust hinf glucose control in diabetes using a physiological model. *AIChE Journal*, 46(12):2537–2549, 2000.

- [86] Sandra M Lynch and B Wayne Bequette. Model predictive control of blood glucose in type 1 diabetics using subcutaneous glucose measurements. In *American Control Conference, 2002. Proceedings of the 2002*, volume 5, pages 4039–4043. IEEE, 2002.
- [87] Francis J Doyle, Lauren M Huyett, Joon Bok Lee, Howard C Zisser, and Eyal Dassau. Closed-loop artificial pancreas systems: engineering the algorithms. *Diabetes care*, 37(5):1191–1197, 2014.
- [88] Kamuran Turksoy, Nicole Frantz, Laurie Quinn, Magdalena Dumin, Jennifer Kilkus, Brooks Hibner, Ali Cinar, and Elizabeth Littlejohn. Automated insulin delivery: the light at the end of the tunnel. *The Journal of pediatrics*, 186:17–28, 2017.
- [89] Roberto Visentin, Chiara Dalla Man, Yogish C Kudva, Ananda Basu, and Claudio Cobelli. Circadian variability of insulin sensitivity: physiological input for in silico artificial pancreas. *Diabetes technology & therapeutics*, 17(1):1–7, 2015.
- [90] Ling Hinshaw, Chiara Dalla Man, Debashis K Nandy, Ahmed Saad, Adil E Bharucha, James A Levine, Robert A Rizza, Rita Basu, Rickey E Carter, Claudio Cobelli, et al. Diurnal pattern of insulin action in type 1 diabetes: implications for a closed loop system. *Diabetes*, page DB\_121759, 2013.
- [91] Fraser Cameron, B Wayne Bequette, Darrell M Wilson, Bruce A Buckingham, Hyunjin Lee, and Günter Niemyer. A closed-loop artificial pancreas based on risk management. *Journal of diabetes science and technology*, 5(2):368–379, 2011.
- [92] Daniela Elleri, Janet M Allen, Kavita Kumareswaran, Lalantha Leelarathna, Marianna Nodale, Karen Caldwell, Peiyao Cheng, Craig Kollman, Ahmad Haidar, Helen R Murphy, et al. Closed-loop basal insulin delivery over 36 hours in adolescents with type 1 diabetes: randomized clinical trial. *Diabetes care*, 36(4):838–844, 2013.
- [93] AH Kadish. Automation control of blood sugar. i. a servomechanism for glucose monitoring and control. *The American journal of medical electronics*, 3:82–86, 1964.
- [94] Claudio Cobelli, Eric Renard, and Boris Kovatchev. Artificial pancreas: past, present, future. *Diabetes*, 60(11):2672–2682, 2011.
- [95] Sara Trevitt, Sue Simpson, and Annette Wood. Artificial pancreas device systems for the closed-loop control of type 1 diabetes: what systems are in development? *Journal of diabetes science and technology*, 10(3):714–723, 2016.

- [96] Parkash A Bakhtiani, Lauren M Zhao, Joseph El Youssef, Jessica R Castle, and W Kenneth Ward. A review of artificial pancreas technologies with an emphasis on bi-hormonal therapy. *Diabetes, Obesity and Metabolism*, 15(12):1065–1070, 2013.
- [97] Ahmad Haidar, Laurent Legault, Maryse Dallaire, Ammar Alkhatieb, Adèle Coriati, Virginie Messier, Peiyao Cheng, Maude Millette, Benoit Boulet, and Rémi Rabasa-Lhoret. Glucose-responsive insulin and glucagon delivery (dual-hormone artificial pancreas) in adults with type 1 diabetes: a randomized crossover controlled trial. *Canadian Medical Association Journal*, 185(4):297–305, 2013.
- [98] Ahmad Haidar, Laurent Legault, Virginie Messier, Tina Maria Mitre, Catherine Leroux, and Rémi Rabasa-Lhoret. Comparison of dual-hormone artificial pancreas, single-hormone artificial pancreas, and conventional insulin pump therapy for glycaemic control in patients with type 1 diabetes: an open-label randomised controlled crossover trial. *The lancet Diabetes & endocrinology*, 3(1):17–26, 2015.
- [99] Jort Kropff, Simone Del Favero, Jerome Place, Chiara Toffanin, Roberto Visentin, Marco Monaro, Mirko Messori, Federico Di Palma, Giordano Lanzola, Anne Farret, et al. 2 month evening and night closed-loop glucose control in patients with type 1 diabetes under free-living conditions: a randomised crossover trial. *The lancet Diabetes & endocrinology*, 3(12):939–947, 2015.
- [100] Lalantha Leelarathna, Sibylle Dellweg, Julia K Mader, Janet M Allen, Carsten Benesch, Werner Doll, Martin Ellmerer, Sara Hartnell, Lutz Heinemann, Harald Kojzar, et al. Day and night home closed-loop insulin delivery in adults with type 1 diabetes: three-center randomized crossover study. *Diabetes Care*, 37(7):1931–1937, 2014.
- [101] Stacey M Anderson, Dan Raghinaru, Jordan E Pinsker, Federico Boscari, Eric Renard, Bruce A Buckingham, Revital Nimri, Francis J Doyle, Sue A Brown, Patrick Keith-Hynes, et al. Multinational home use of closed-loop control is safe and effective. *Diabetes Care*, 39(7):1143–1150, 2016.
- [102] Firas H El-Khatib, Steven J Russell, David M Nathan, Robert G Sutherland, and Edward R Damiano. A bihormonal closed-loop artificial pancreas for type 1 diabetes. *Science translational medicine*, 2(27):27ra27–27ra27, 2010.
- [103] Firas H El-Khatib, Steven J Russell, Kendra L Magyar, Manasi Sinha, Katherine McKeon, David M Nathan, and Edward R Damiano. Autonomous and continuous adaptation of a bihormonal bionic pancreas

- in adults and adolescents with type 1 diabetes. *The Journal of Clinical Endocrinology & Metabolism*, 99(5):1701–1711, 2014.
- [104] Alanna Weisman, Johnny-Wei Bai, Marina Cardinez, Caroline K Kramer, and Bruce A Perkins. Effect of artificial pancreas systems on glycaemic control in patients with type 1 diabetes: a systematic review and meta-analysis of outpatient randomised controlled trials. *The lancet Diabetes & endocrinology*, 5(7):501–512, 2017.
- [105] Kamuran Turksoy, Sediqeh Samadi, Jianyuan Feng, Elizabeth Littlejohn, Laurie Quinn, and Ali Cinar. Meal detection in patients with type 1 diabetes: a new module for the multivariable adaptive artificial pancreas control system. *IEEE journal of biomedical and health informatics*, 20(1):47–54, 2016.
- [106] Eyal Dassau, B Wayne Bequette, Bruce A Buckingham, and Francis J Doyle. Detection of a meal using continuous glucose monitoring (CGM): implications for an artificial  $\beta$ -cell. *Diabetes care*, 31(2):295–300, 2008.
- [107] Hyunjin Lee, Bruce A Buckingham, Darrell M Wilson, and B Wayne Bequette. A closed-loop artificial pancreas using model predictive control and a sliding meal size estimator. *Journal of Diabetes Science and Technology*, 3(5):1082–90, 2009.
- [108] Eyal Dassau, Fraser Cameron, Hyunjin Lee, B Wayne Bequette, Howard Zisser, Lois Jovanovič, H Peter Chase, Darrell M Wilson, Bruce A Buckingham, and Francis J Doyle. Real-time hypoglycemia prediction suite using continuous glucose monitoring: a safety net for the artificial pancreas. *Diabetes care*, 33(6):1249–1254, 2010.
- [109] Chiara Toffanin, Howard Zisser, Francis J Doyle III, and Eyal Dassau. Dynamic insulin on board: incorporation of circadian insulin sensitivity variation. *Journal of diabetes science and technology*, 7(4):928–940, 2013.
- [110] Desmond Barry Keenan, John J Mastrototaro, Benyamin Grosman, Neha J Parikh, and Anirban Roy. Insulin on board compensation for a closed-loop insulin infusion system, June 14 2016. US Patent 9,364,609.
- [111] Elif Seyma Bayrak, Kamuran Turksoy, Ali Cinar, Laretta Quinn, Elizabeth Littlejohn, and Derrick Rollins. Hypoglycemia early alarm systems based on recursive autoregressive partial least squares models. *Journal of diabetes science and technology*, 7(1):206–214, 2013.

- [112] B Wayne Bequette. Hypoglycemia prevention using low glucose suspend systems. In *Data-driven Modeling for Diabetes*, pages 73–89. Springer, 2014.
- [113] Yan Nian Wang, Yan Rui Shen, Yong Qiang Yong, Quan Zhong Li, and Chang Qing Sun. Glucose prediction and hypoglycemia alarms based on adaptive model. In *Advanced Materials Research*, volume 971, pages 275–279. Trans Tech Publ, 2014.
- [114] Richard M Bergenstal, Satish Garg, Stuart A Weinzimer, Bruce A Buckingham, Bruce W Bode, William V Tamborlane, and Francine R Kaufman. Safety of a hybrid closed-loop insulin delivery system in patients with type 1 diabetes. *Jama*, 316(13):1407–1408, 2016.
- [115] Roman Hovorka, Valentina Canonico, Ludovic J Chassin, Ulrich Haueter, Massimo Massi-Benedetti, Marco Orsini Federici, Thomas R Pieber, Helga C Schaller, Lukas Schaupp, Thomas Vering, et al. Non-linear model predictive control of glucose concentration in subjects with type 1 diabetes. *Physiological measurement*, 25(4):905, 2004.
- [116] Lalo Magni, Davide M Raimondo, Luca Bossi, Chiara Dalla Man, Giuseppe De Nicolao, Boris Kovatchev, and Claudio Cobelli. Model predictive control of type 1 diabetes: an in silico trial, 2007.
- [117] Benyamin Grosman, Eyal Dassau, Howard C Zisser, Lois Jovanovič, and Francis J Doyle III. Zone model predictive control: a strategy to minimize hyper-and hypoglycemic events. *Journal of diabetes science and technology*, 4(4):961–975, 2010.
- [118] Cesar C Palerm, John P Willis, James Desemone, and B Wayne Bequette. Hypoglycemia prediction and detection using optimal estimation. *Diabetes technology & therapeutics*, 7(1):3–14, 2005.
- [119] Fraser Cameron, Günter Niemeyer, Karen Gundy-Burlet, and Bruce Buckingham. Statistical hypoglycemia prediction. *Journal of diabetes science and technology*, 2(4):612–621, 2008.
- [120] Rebecca A Harvey, Eyal Dassau, Howard Zisser, Dale E Seborg, Lois Jovanovič, and Francis J Doyle III. Design of the health monitoring system for the artificial pancreas: low glucose prediction module. *Journal of diabetes science and technology*, 6(6):1345–1354, 2012.
- [121] Bruce Buckingham, Erin Cobry, Paula Clinton, Victoria Gage, Kimberly Caswell, Elizabeth Kunselman, Fraser Cameron, and H Peter Chase. Preventing hypoglycemia using predictive alarm algorithms and insulin pump suspension. *Diabetes technology & therapeutics*, 11(2):93–97, 2009.

- [122] Bruce Buckingham, H Peter Chase, Eyal Dassau, Erin Cobry, Paula Clinton, Victoria Gage, Kimberly Caswell, John Wilkinson, Fraser Cameron, Hyunjin Lee, et al. Prevention of nocturnal hypoglycemia using predictive alarm algorithms and insulin pump suspension. *Diabetes care*, 33(5):1013–1017, 2010.
- [123] Colleen S Hughes, Stephen D Patek, Marc D Breton, and Boris P Kovatchev. Hypoglycemia prevention via pump attenuation and red-yellow-green traffic lights using continuous glucose monitoring and insulin pump data. *Journal of diabetes science and technology*, 4(5):1146–1155, 2010.
- [124] Silvia Oviedo, Josep Vehí, Remei Calm, and Joaquim Armengol. A review of personalized blood glucose prediction strategies for t1dm patients. *International journal for numerical methods in biomedical engineering*, 33(6):e2833, 2017.
- [125] Klemen Dovc, Charlotte Boughton, Martin Tauschmann, Hood Thabit, Lia Bally, Janet M Allen, Carlo L Acerini, Sabine Arnolds, Carine de Beaufort, Richard M Bergenstal, et al. Young children have higher variability of insulin requirements: Observations during hybrid closed-loop insulin delivery. *Diabetes Care*, 42(7):1344–1347, 2019.
- [126] Ahmad Haidar, Daniela Elleri, Kavita Kumareswaran, Lalantha Leerathna, Janet M Allen, Karen Caldwell, Helen R Murphy, Malgorzata E Wilinska, Carlo L Acerini, Mark L Evans, et al. Pharmacokinetics of insulin aspart in pump-treated subjects with type 1 diabetes: reproducibility and effect of age, weight, and duration of diabetes. *Diabetes care*, 36(10):e173–e174, 2013.
- [127] Lutz Heinemann. Variability of insulin absorption and insulin action. *Diabetes technology & therapeutics*, 4(5):673–682, 2002.
- [128] Gary Scheiner and Bret A Boyer. Characteristics of basal insulin requirements by age and gender in type-1 diabetes patients using insulin pump therapy. *Diabetes research and clinical practice*, 69(1):14–21, 2005.
- [129] Jingyi Qian and Frank AJL Scheer. Circadian system and glucose metabolism: implications for physiology and disease. *Trends in Endocrinology & Metabolism*, 27(5):282–293, 2016.
- [130] MC Jolly, R Hovorka, I Godsland, R Amin, N Lawrence, V Anyaoku, D Johnston, and S Robinson. Relation between insulin kinetics and insulin sensitivity in pregnancy. *European journal of clinical investigation*, 33(8):698–703, 2003.

- [131] Susanne Famulla, Ulrike Hövelmann, Annelie Fischer, Hans-Veit Coester, Lidia Hermanski, Matthias Kaltheuner, Lars Kaltheuner, Lutz Heinemann, Tim Heise, and Laurence Hirsch. Insulin injection into lipohypertrophic tissue: blunted and more variable insulin absorption and action and impaired postprandial glucose control. *Diabetes Care*, 39(9):1486–1492, 2016.
- [132] Ahmed Saad, Chiara Dalla Man, Debashis K Nandy, James A Levine, Adil E Bharucha, Robert A Rizza, Rita Basu, Rickey E Carter, Claudio Cobelli, Yogish C Kudva, et al. Diurnal pattern to insulin secretion and insulin action in healthy individuals. *Diabetes*, 61(11):2691–2700, 2012.
- [133] Roberto Visentin, Enrique Campos-Náñez, Michele Schiavon, Dayu Ly, Martina Vettoretti, Marc Breton, Boris P Kovatchev, Chiara Dalla Man, and Claudio Cobelli. The uva/padova type 1 diabetes simulator goes from single meal to single day. *Journal of diabetes science and technology*, 12(2):273–281, 2018.
- [134] JM Schneider. Think like a pancreas: A practical guide to managing diabetes with insulin (book). *Library Journal*, 129(13):108–108, 2004.
- [135] Kirstine J Bell, Carmel E Smart, Garry M Steil, Jennie C Brand-Miller, Bruce King, and Howard A Wolpert. Impact of fat, protein, and glycemic index on postprandial glucose control in type 1 diabetes: implications for intensive diabetes management in the continuous glucose monitoring era. *Diabetes care*, 38(6):1008–1015, 2015.
- [136] Stamatina Zavitsanou, Jennifer Massa, Sunil Deshpande, Jordan E Pinsker, Mei Mei Church, Camille Andre, Francis J Doyle III, Alicia Michelson, Jamie Creason, Eyal Dassau, et al. The effect of two types of pasta versus white rice on postprandial blood glucose levels in adults with type 1 diabetes: A randomized crossover trial. *Diabetes Technology & Therapeutics*, 2019.
- [137] Dessi Zaharieva, Loren Yavelberg, Veronica Jamnik, Ali Cinar, Kamuran Turksoy, and Michael C Riddell. The effects of basal insulin suspension at the start of exercise on blood glucose levels during continuous versus circuit-based exercise in individuals with type 1 diabetes on continuous subcutaneous insulin infusion. *Diabetes technology & therapeutics*, 19(6):370–378, 2017.
- [138] Kamuran Turksoy, Colleen Monforti, Minsun Park, Garrett Griffith, Laurie Quinn, and Ali Cinar. Use of wearable sensors and biometric variables in an artificial pancreas system. *Sensors*, 17(3):532, 2017.

- [139] Jorge Bondia, Josep Vehí, César C. Palerm, and Pau Herrero. Artificial pancreas: automatic control of insulin infusion in type 1 diabetes mellitus (in spanish). *Revista Iberoamericana de Automática e Informática Industrial RIAI*, 7(2):5–20, 2010.
- [140] Garry M Steil, Brian Hipszer, and Jaques Reifman. Update on mathematical modeling research to support the development of automated insulin delivery systems, 2010.
- [141] Naviyn Prabhu Balakrishnan, Gade Pandu Rangaiah, and Lakshminarayanan Samavedham. Review and analysis of blood glucose (bg) models for type 1 diabetic patients. *Industrial & Engineering Chemistry Research*, 50(21):12041–12066, 2011.
- [142] Ewart Carson and Claudio Cobelli. *Modelling methodology for physiology and medicine*. Newnes, 2013.
- [143] Remei Calm, M García-Jaramillo, Jorge Bondia, MA Sainz, and Josep Vehí. Comparison of interval and monte carlo simulation for the prediction of postprandial glucose under uncertainty in type 1 diabetes mellitus. *Computer methods and programs in biomedicine*, 104(3):325–332, 2011.
- [144] Richard N Bergman, Y Ziya Ider, Charles R Bowden, and Claudio Cobelli. Quantitative estimation of insulin sensitivity. *American Journal of Physiology-Endocrinology And Metabolism*, 236(6):E667, 1979.
- [145] Sami S Kanderian, Stuart A Weinzimer, and Garry M Steil. The identifiable virtual patient model: comparison of simulation and clinical closed-loop study results. *Journal of diabetes science and technology*, 6(2):371–379, 2012.
- [146] Anirban Roy and Robert S Parker. Dynamic modeling of free fatty acid, glucose, and insulin: An extended” minimal model”. *Diabetes technology & therapeutics*, 8(6):617–626, 2006.
- [147] Anirban Roy and Robert S Parker. Dynamic modeling of exercise effects on plasma glucose and insulin levels, 2007.
- [148] Vanessa Moscardó, Paolo Rossetti, F Javier Ampudia-Blasco, and Jorge Bondia. Modelling of adrenaline counterregulatory action during hypoglycaemia in type 1 diabetes. In *2016 IEEE Conference on Control Applications (CCA)*, pages 430–435. IEEE, 2016.
- [149] Pau Herrero, Pantelis Georgiou, Nick Oliver, Monika Reddy, Desmond Johnston, and Christofer Toumazou. A composite model of glucagon-glucose dynamics for in silico testing of bihormonal glucose controllers. *Journal of diabetes science and technology*, 7(4):941–951, 2013.



- [150] Pinky Dua, Francis J Doyle, and Efstratios N Pistikopoulos. Model-based blood glucose control for type 1 diabetes via parametric programming. *IEEE Transactions on Biomedical Engineering*, 53(8):1478–1491, 2006.
- [151] Sandra M Lynch and B Wayne Bequette. Estimation-based model predictive control of blood glucose in type i diabetics: a simulation study. In *Proceedings of the IEEE 27th Annual Northeast Bioengineering Conference (Cat. No. 01CH37201)*, pages 79–80. IEEE, 2001.
- [152] John Thomas Sorensen. *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, 1985.
- [153] Claudia Eberle and Christoph Ament. The unscented kalman filter estimates the plasma insulin from glucose measurement. *Biosystems*, 103(1):67–72, 2011.
- [154] Robert Hermann and Arthur Krener. Nonlinear controllability and observability. *IEEE Transactions on automatic control*, 22(5):728–740, 1977.
- [155] David Luenberger. Observers for multivariable systems. *IEEE Transactions on Automatic Control*, 11(2):190–197, 1966.
- [156] Levente Kovács, Béla Paláncz, and Zoltán Benyó. Design of luenberger observer for glucose-insulin control via mathematica. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 624–627. IEEE, 2007.
- [157] Anne Katrine Duun-Henriksen, Signe Schmidt, Rikke Meldgaard Røge, Jonas Bech Møller, Kirsten Nørgaard, John Bagterp Jørgensen, and Henrik Madsen. Model identification using stochastic differential equation grey-box models in diabetes. *Journal of diabetes science and technology*, 7(2):431–440, 2013.
- [158] Kamuran Turksoy, Iman Hajizadeh, Sediqeh Samadi, Jianyuan Feng, Mert Sevil, Minsun Park, Laurie Quinn, Elizabeth Littlejohn, and Ali Cinar. Real-time insulin bolusing for unannounced meals with artificial pancreas. *Control Engineering Practice*, 59:159–164, 2017.
- [159] Malgorzata E Wilinska, Ludovic J Chassin, Carlo L Acerini, Janet M Allen, David B Dunger, and Roman Hovorka. Simulation environment to evaluate closed-loop insulin delivery systems in type 1 diabetes. *Journal of diabetes science and technology*, 4(1):132–144, 2010.

- [160] Eleni Bekiari, Konstantinos Kitsios, Hood Thabit, Martin Tauschmann, Eleni Athanasiadou, Thomas Karagiannis, Anna-Bettina Haidich, Roman Hovorka, and Apostolos Tsapas. Artificial pancreas treatment for outpatients with type 1 diabetes: systematic review and meta-analysis. *BMJ*, 361:k1310, April 2018.
- [161] Malgorzata E Wilinska, Erwin S Budiman, Gary A Hayter, Marc B Taub, and Roman Hovorka. Method of overnight closed-loop insulin delivery with model predictive control and glucose measurement error model, November 19 2013. US Patent 8,585,637.
- [162] Diego de Pereda, Sergio Romero-Vivo, Beatriz Ricarte, Paolo Rossetti, Francisco Javier Ampudia-Blasco, and Jorge Bondia. Real-time estimation of plasma insulin concentration from continuous glucose monitor measurements. *Computer methods in biomechanics and biomedical engineering*, 19(9):934–942, 2016.
- [163] Iman Hajizadeh, Mudassir Rashid, Kamuran Turksoy, Sediqeh Samadi, Jianyuan Feng, Nicole Frantz, Mert Sevil, Eda Cengiz, and Ali Cinar. Plasma insulin estimation in people with type 1 diabetes mellitus. *Industrial & Engineering Chemistry Research*, 56(35):9846–9857, 2017.
- [164] Ernest Gardenes, Miguel A Sainz, Lambert Jorba, Remei Calm, Rosa Estela, Honorino Mielgo, and Albert Trepap. Model intervals. *Reliable Computing*, 7(2):77–111, 2001.
- [165] Diego De Pereda, Sergio Romero-Vivo, Beatriz Ricarte, and Jorge Bondia. On the prediction of glucose concentration under intra-patient variability in type 1 diabetes: A monotone systems approach. *Computer methods and programs in biomedicine*, 108(3):993–1001, 2012.
- [166] Chiara Dalla Man, Robert A Rizza, and Claudio Cobelli. Meal simulation model of the glucose-insulin system. *IEEE Transactions on biomedical engineering*, 54(10):1740–1749, 2007.
- [167] Chiara Dalla Man, Davide M Raimondo, Robert A Rizza, and Claudio Cobelli. Gim, simulation software of meal glucose-insulin model, 2007.
- [168] Chiara Dalla Man, Francesco Micheletto, Dayu Lv, Marc Breton, Boris Kovatchev, and Claudio Cobelli. The uva/padova type 1 diabetes simulator: new features. *Journal of diabetes science and technology*, 8(1):26–34, 2014.
- [169] Paola Soru, Giuseppe De Nicolao, Chiara Toffanin, Chiara Dalla Man, Claudio Cobelli, Lalo Magni, AP@ home consortium, et al. Mpc based artificial pancreas: strategies for individualization and meal compensation. *Annual Reviews in Control*, 36(1):118–128, 2012.

- [170] Mirko Messori, Matthew Ellis, Claudio Cobelli, Panagiotis D Christofides, and Lalo Magni. Improved postprandial glucose control with a customized model predictive controller. In *2015 American Control Conference (ACC)*, pages 5108–5115. IEEE, 2015.
- [171] Jaques Reifman, Srinivasan Rajaraman, Andrei Gribok, and W Kenneth Ward. Predictive monitoring for improved management of glucose levels. *Journal of diabetes science and technology*, 1(4):478–486, 2007.
- [172] Adiwinata Gani, Andrei V Gribok, Srinivasan Rajaraman, W Kenneth Ward, and Jaques Reifman. Predicting subcutaneous glucose concentration in humans: data-driven glucose modeling. *IEEE Transactions on Biomedical Engineering*, 56(2):246–254, 2009.
- [173] Giovanni Sparacino, Francesca Zanderigo, Stefano Corazza, Alberto Maran, Andrea Facchinetti, and Claudio Cobelli. Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series. *IEEE Transactions on biomedical engineering*, 54(5):931–937, 2007.
- [174] Meriyan Eren-Oruklu, Ali Cinar, Laretta Quinn, and Donald Smith. Estimation of future glucose concentrations with subject-specific recursive linear models. *Diabetes technology & therapeutics*, 11(4):243–253, 2009.
- [175] Daniel A Finan, Howard Zisser, Lois Jovanovic, Wendy C Bevier, and Dale E Seborg. Practical issues in the identification of empirical models from simulated type 1 diabetes data. *Diabetes technology & therapeutics*, 9(5):438–450, 2007.
- [176] Daniel A Finan, Cesar C Palerm, Francis J Doyle, Howard Zisser, Lois Jovanovic, Wendy C Bevier, and Dale E Seborg. Identification of empirical dynamic models from type 1 diabetes subject data. In *2008 American Control Conference*, pages 2099–2104. IEEE, 2008.
- [177] Daniel A Finan, Francis J Doyle III, Cesar C Palerm, Wendy C Bevier, Howard C Zisser, Lois Jovanovič, and Dale E Seborg. Experimental evaluation of a recursive model identification technique for type 1 diabetes. *Journal of diabetes science and technology*, 3(5):1192–1202, 2009.
- [178] Giovanna Castillo Estrada, Luigi del Re, and Eric Renard. Nonlinear gain in online prediction of blood glucose profile in type 1 diabetic patients. In *49th IEEE Conference on Decision and Control (CDC)*, pages 1668–1673. IEEE, 2010.

- [179] Meriyan Eren-Oruklu, Ali Cinar, Derrick K Rollins, and Laretta Quinn. Adaptive system identification for estimating future glucose concentrations and hypoglycemia alarms. *Automatica*, 48(8):1892–1897, 2012.
- [180] Kamuran Turksoy, Elif S Bayrak, Laretta Quinn, Elizabeth Littlejohn, Derrick Rollins, and Ali Cinar. Hypoglycemia early alarm systems based on multivariable models. *Industrial & engineering chemistry research*, 52(35):12329–12336, 2013.
- [181] Fredrik Ståhl and Rolf Johansson. Diabetes mellitus modeling and short-term prediction based on blood glucose measurements. *Mathematical biosciences*, 217(2):101–117, 2009.
- [182] Ashenafi Zebene Woldaregay, Eirik Årsand, Taxiarchis Botsis, David Albers, Lena Mamykina, and Gunnar Hartvigsen. Data-driven blood glucose pattern classification and anomalies detection: Machine-learning applications in type 1 diabetes. *Journal of medical Internet research*, 21(5):e11030, 2019.
- [183] Chunhui Zhao, Eyal Dassau, Lois Jovanovič, Howard C Zisser, Francis J Doyle III, and Dale E Seborg. Predicting subcutaneous glucose concentration using a latent-variable-based statistical method for type 1 diabetes mellitus. *Journal of diabetes science and technology*, 6(3):617–633, 2012.
- [184] Chiara Zecchin, Andrea Facchinetti, Giovanni Sparacino, Giuseppe De Nicolao, and Claudio Cobelli. Neural network incorporating meal information improves accuracy of short-time prediction of glucose concentration. *IEEE Transactions on Biomedical Engineering*, 59(6):1550–1560, 2012.
- [185] Carmen Pérez-Gandía, A Facchinetti, G Sparacino, C Cobelli, EJ Gómez, M Rigla, Alberto de Leiva, and ME Hernando. Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring. *Diabetes technology & therapeutics*, 12(1):81–88, 2010.
- [186] Eleni I Georga, José C Príncipe, and Dimitrios I Fotiadis. Short-term prediction of glucose in type 1 diabetes using kernel adaptive filters. *Medical & biological engineering & computing*, 57(1):27–46, 2019.
- [187] Youqing Wang, Xiangwei Wu, and Xue Mo. A novel adaptive-weighted-average framework for blood glucose prediction. *Diabetes technology & therapeutics*, 15(10):792–801, 2013.

- [188] Andrea Facchinetti, Giovanni Sparacino, Emanuele Trifoglio, and Claudio Cobelli. A new index to optimally design and compare continuous glucose monitoring glucose prediction algorithms. *Diabetes technology & therapeutics*, 13(2):111–119, 2011.
- [189] Qingnan Sun, Marko V Jankovic, Lia Bally, and Stavroula G Mougiakakou. Predicting blood glucose with an lstm and bi-lstm based deep neural network. In *2018 14th Symposium on Neural Networks and Applications (NEUREL)*, pages 1–5. IEEE, 2018.
- [190] Elena Daskalaki, Aikaterini Prountzou, Peter Diem, and Stavroula G Mougiakakou. Real-time adaptive models for the personalized prediction of glycemic profile in type 1 diabetes patients. *Diabetes technology & therapeutics*, 14(2):168–174, 2012.
- [191] Scott M Pappada, Brent D Cameron, Paul M Rosman, Raymond E Bourey, Thomas J Papadimos, William Olorunto, and Marilyn J Borst. Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes. *Diabetes technology & therapeutics*, 13(2):135–141, 2011.
- [192] Stavroula G Mougiakakou, Aikaterini Prountzou, Dimitra Iliopoulou, Konstantina S Nikita, Andriani Vazeou, and Christos S Bartsocas. Neural network based glucose-insulin metabolism models for children with type 1 diabetes. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pages 3545–3548. IEEE, 2006.
- [193] Qian Wang, Peter Molenaar, Saurabh Harsh, Kenneth Freeman, Jinyu Xie, Carol Gold, Mike Rovine, and Jan Ulbrecht. Personalized state-space modeling of glucose dynamics for type 1 diabetes using continuously monitored glucose, insulin dose, and meal intake: An extended kalman filter approach. *Journal of diabetes science and technology*, 8(2):331–345, 2014.
- [194] ED Lehmann and T Deutsch. A physiological model of glucose-insulin interaction in type 1 diabetes mellitus. *Journal of biomedical engineering*, 14(3):235–242, 1992.
- [195] Markus Berger and David Rodbard. Computer simulation of plasma insulin and glucose dynamics after subcutaneous insulin injection. *Diabetes care*, 12(10):725–736, 1989.
- [196] Razvan Bunescu, Nigel Struble, Cindy Marling, Jay Shubrook, and Frank Schwartz. Blood glucose level prediction using physiological models and support vector regression. In *2013 12th International*

- Conference on Machine Learning and Applications*, volume 1, pages 135–140. IEEE, 2013.
- [197] Carlos A Verdonk, Robert A Rizza, and John E Gerich. Effects of plasma glucose concentration on glucose utilization and glucose clearance in normal man. *Diabetes*, 30(6):535–537, 1981.
- [198] Konstantia Zarkogianni, Eleni Litsa, Andriani Vazeou, and Konstantina S Nikita. Personalized glucose-insulin metabolism model based on self-organizing maps for patients with type 1 diabetes mellitus. In *13th IEEE International Conference on BioInformatics and BioEngineering*, pages 1–4. IEEE, 2013.
- [199] Chiara Zecchin, Andrea Facchinetti, Giovanni Sparacino, and Claudio Cobelli. Jump neural network for online short-time prediction of blood glucose from continuous monitoring sensors and meal information. *Computer methods and programs in biomedicine*, 113(1):144–152, 2014.
- [200] Konstantia Zarkogianni, Konstantinos Mitsis, Eleni Litsa, M-T Arredondo, G Fico, Alessio Fioravanti, and Konstantina S Nikita. Comparative assessment of glucose prediction models for patients with type 1 diabetes mellitus applying sensors for glucose and physical activity monitoring. *Medical & biological engineering & computing*, 53(12):1333–1343, 2015.
- [201] Keith W Hipel and A Ian McLeod. *Time series modelling of water resources and environmental systems*, volume 45. Elsevier, 1994.
- [202] Peter J Brockwell, Richard A Davis, and Matthew V Calder. *Introduction to time series and forecasting*, volume 2. Springer, 2002.
- [203] Chris Chatfield. *Time-series forecasting*. Chapman and Hall/CRC, 2000.
- [204] GP Nason. *Stationary and non-stationary time series*, pages 129 – 142. The Geological Society, 2006.
- [205] William WS Wei. Time series analysis. In *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2*. 2006.
- [206] David J Thomson. Jackknifing multiple-window spectra. In *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on*, volume 6, pages VI–73. IEEE, 1994.
- [207] J.D. Hamilton and Princeton University Press. *Time Series Analysis*. Princeton University Press, 1994.

- [208] Herman Wold. *A study in the analysis of stationary time series*. PhD thesis, Almqvist & Wiksell, 1938.
- [209] George Udny Yule. On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London Series A*, 226:267–298, 1927.
- [210] Gilbert Thomas Walker. On periodicity in series of related terms. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 131(818):518–532, 1931.
- [211] James Durbin. The fitting of time-series models. *Revue de l'Institut International de Statistique*, pages 233–244, 1960.
- [212] Marcello Pagano. An algorithm for fitting autoregressive schemes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 21(3):274–281, 1972.
- [213] Ruey S Tsay. *Analysis of financial time series*, volume 543. John Wiley & Sons, 2005.
- [214] Maurice H Quenouille. Approximate tests of correlation in time-series 3. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 45, pages 483–484. Cambridge University Press, 1949.
- [215] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.
- [216] Gebhard Kirchgässner and Jürgen Wolters. *Introduction to modern time series analysis*. Springer Science & Business Media, 2007.
- [217] Hirotugu Akaike. A new look at the statistical model identification. In *Selected Papers of Hirotugu Akaike*, pages 215–222. Springer, 1974.
- [218] Gideon Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.
- [219] Chris Chatfield. *The analysis of time series: an introduction*. Chapman and Hall/CRC, 2016.
- [220] George E. P Box and Gwilym M. Jenkins. *Time series analysis : forecasting and control*. San Francisco : Holden-Day, 1970.
- [221] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

- [222] Wayne A Fuller. *Introduction to statistical time series*, volume 428. John Wiley & Sons, 2009.
- [223] David A Dickey and Wayne A Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431, 1979.
- [224] Said E Said and David A Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607, 1984.
- [225] George EP Box and David R Cox. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243, 1964.
- [226] A.C. Harvey. *Time Series Models*. MIT Press, 1993.
- [227] George EP Box and David A Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332):1509–1526, 1970.
- [228] Greta M Ljung and George EP Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.
- [229] Carlos M Jarque and Anil K Bera. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, 6(3):255–259, 1980.
- [230] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.
- [231] H. Theil. *Applied economic forecasting*. Studies in mathematical and managerial economics. North-Holland Pub. Co., 1966.
- [232] Charu C. Aggarwal and Chandan K. Reddy, editors. *Data Clustering: Algorithms and Applications*. CRC Press, 2014.
- [233] G. Gan, C. Ma, and J. Wu. *Data Clustering: Theory, Algorithms, and Applications*. ASA-SIAM Series on Statistics and Applied Probability. Society for Industrial and Applied Mathematics, 2007.
- [234] Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1):43–49, 1978.
- [235] Carmelo Cassisi, Placido Montalto, Marco Aliotta, Andrea Cannata, and Alfredo Pulvirenti. Similarity measures and dimensionality reduction techniques for time series data mining. In *Advances in data mining knowledge discovery and applications*. InTech, 2012.



- [236] TH Cormen, CE Leiserson, and RL Rivest. Introduction to algorithms. themit electrical engineering and computer science series, 1990.
- [237] Padhraic Smyth. Clustering sequences with hidden markov models. In *Advances in neural information processing systems*, pages 648–654, 1997.
- [238] Konstantinos Kalpakis, Dhiral Gada, and Vasundhara Puttagunta. Distance measures for effective clustering of arima time-series. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 273–280. IEEE, 2001.
- [239] Robert John Bennett. Spatial time series. *Pion, London*, 674, 1979.
- [240] Xiaozhe Wang, Kate Smith, and Rob Hyndman. Characteristic-based clustering for time series data. *Data mining and knowledge Discovery*, 13(3):335–364, 2006.
- [241] Subana Shanmuganathan and Sandhya Samarasinghe. *Artificial neural network modelling*, volume 628. Springer, 2016.
- [242] T Warren Liao. Clustering of time series data: a survey. *Pattern recognition*, 38(11):1857–1874, 2005.
- [243] Donald J Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, 1994.
- [244] Jason Lines and Anthony Bagnall. Time series classification with ensembles of elastic distance measures. *Data Mining and Knowledge Discovery*, 29(3):565–592, 2015.
- [245] Ben D Fulcher. Feature-based time-series analysis. In *Feature Engineering for Machine Learning and Data Analytics*, pages 87–116. CRC Press, 2018.
- [246] Alex Nanopoulos, Rob Alcock, and Yannis Manolopoulos. Feature-based classification of time-series data. *International Journal of Computer Research*, 10(3):49–61, 2001.
- [247] Rakesh Agrawal, Christos Faloutsos, and Arun Swami. Efficient similarity search in sequence databases. In *International conference on foundations of data organization and algorithms*, pages 69–84. Springer, 1993.
- [248] Kin-Pong Chan and Wai-Chee Fu. Efficient time series matching by wavelets. In *icde*, page 126. IEEE, 1999.

- [249] Flip Korn, Hosagrahar V Jagadish, and Christos Faloutsos. Efficiently supporting ad hoc queries in large datasets of time sequences. In *Acm Sigmod Record*, volume 26, pages 289–300. ACM, 1997.
- [250] Eamonn Keogh, Kaushik Chakrabarti, Michael Pazzani, and Sharad Mehrotra. Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and information Systems*, 3(3):263–286, 2001.
- [251] Jessica Lin, Eamonn Keogh, Li Wei, and Stefano Lonardi. Experiencing sax: a novel symbolic representation of time series. *Data Mining and knowledge discovery*, 15(2):107–144, 2007.
- [252] Sarab S Sethi, Valerio Zerbi, Nicole Wenderoth, Alex Fornito, and Ben D Fulcher. Structural connectome topology relates to regional bold signal dynamics in the mouse brain. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(4):047405, 2017.
- [253] Cyril Goutte, Peter Toft, Egill Rostrup, Finn Å Nielsen, and Lars Kai Hansen. On clustering fmri time series. *NeuroImage*, 9(3):298–310, 1999.
- [254] David Minnen, Thad Starner, Irfan Essa, and Charles Isbell. Discovering characteristic actions from on-body sensor data. In *Wearable computers, 2006 10th IEEE international symposium on*, pages 11–18. IEEE, 2006.
- [255] Antonello Panuccio, Manuele Bicego, and Vittorio Murino. A hidden markov model-based approach to sequential data clustering. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 734–743. Springer, 2002.
- [256] Nitin Kumar, Venkata Nishanth Lolla, Eamonn Keogh, Stefano Lonardi, Chotirat Ann Ratanamahatana, and Li Wei. Time-series bitmaps: a practical visualization tool for working with large time series databases. In *Proceedings of the 2005 SIAM international conference on data mining*, pages 531–535. SIAM, 2005.
- [257] Marcella Corduas and Domenico Piccolo. Time series clustering and classification by the autoregressive metric. *Computational statistics & data analysis*, 52(4):1860–1872, 2008.
- [258] Anil K Jain and Richard C Dubes. Algorithms for clustering data. 1988.

- [259] Godfrey N Lance and William Thomas Williams. A general theory of classificatory sorting strategies: 1. hierarchical systems. *The computer journal*, 9(4):373–380, 1967.
- [260] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.
- [261] Darshan Sonagara and Soham Badheka. Comparison of basic clustering algorithms. *International Journal of Computer Science and Mobile Computing*, 3(10):58–61, 2014.
- [262] Martin J Bayley, Valerie J Gillet, Peter Willett, John Bradshaw, and Darren VS Green. Computational analysis of molecular diversity for drug discovery. In *Proceedings of the third annual international conference on Computational molecular biology*, pages 321–330. ACM, 1999.
- [263] Miklos Feher and Jonathan M Schmidt. Fuzzy clustering as a means of selecting representative conformers and molecular alignments. *Journal of chemical information and computer sciences*, 43(3):810–818, 2003.
- [264] Joseph C Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. 1973.
- [265] James C Bezdek. Objective function clustering. In *Pattern recognition with fuzzy objective function algorithms*, pages 43–93. Springer, 1981.
- [266] Richard J Hathaway and James C Bezdek. Fuzzy c-means clustering of incomplete data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 31(5):735–744, 2001.
- [267] John K Dixon. Pattern recognition with partly missing data. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(10):617–621, 1979.
- [268] Allan D Gordon. Cluster validation. In *Data science, classification, and related methods*, pages 22–39. Springer, 1998.
- [269] Huapeng Li, Shuqing Zhang, Xiaohui Ding, Ce Zhang, and Patricia Dale. Performance evaluation of cluster validity indices (cvis) on multi/hyperspectral remote sensing datasets. *Remote Sensing*, 8(4):295, 2016.
- [270] Subhash Sharma and Subhash Sharma. Applied multivariate techniques. 1996.
- [271] Maria Halkidi, Yannis Batistakis, and Michalis Vazirgiannis. Clustering validity checking methods: part ii. *ACM Sigmod Record*, 31(3):19–27, 2002.

- [272] Joseph C Dunn. Well-separated clusters and optimal fuzzy partitions. *Journal of cybernetics*, 4(1):95–104, 1974.
- [273] David L Davies and Donald W Bouldin. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2):224–227, 1979.
- [274] James C. Bezdek. Pattern recognition with fuzzy objective function algorithms. In *Advanced Applications in Pattern Recognition*, 1981.
- [275] Xuanli Lisa Xie and Gerardo Beni. A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (8):841–847, 1991.
- [276] M Ramze Rezaee, Boudewijn PF Lelieveldt, and Johan HC Reiber. A new cluster validity index for the fuzzy c-mean. *Pattern recognition letters*, 19(3-4):237–246, 1998.
- [277] Rolf Johansson. *System modeling and identification*, volume 1. Prentice Hall Englewood Cliffs, NJ, 1993.
- [278] Laura Lema-Perez, Rafael Muñoz-Tamayo, Jose Garcia-Tirado, and Hernan Alvarez. On parameter interpretability of phenomenological-based semiphysical models in biology. *Informatics in Medicine Unlocked*, 15:100158, 2019.
- [279] J Doyne Farmer and John J Sidorowich. Predicting chaotic time series. *Physical review letters*, 59(8):845, 1987.
- [280] John J Sidorowich. Modeling of chaotic time series for prediction, interpolation, and smoothing. In *[Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 121–124. IEEE, 1992.
- [281] Jörg Walter, H Riter, and Klaus Schulten. Nonlinear prediction with self-organizing maps. In *1990 IJCNN International Joint Conference on Neural Networks*, pages 589–594. IEEE, 1990.
- [282] Jose C Principe and Ludong Wang. Non-linear time series modeling with self-organization feature maps. In *Proceedings of 1995 IEEE Workshop on Neural Networks for Signal Processing*, pages 11–20. IEEE, 1995.
- [283] Bjarne A Foss, Tor A Johansen, and Aage V Sørensen. Nonlinear predictive control using local models applied to a batch fermentation process. *Control Engineering Practice*, 3(3):389–396, 1995.
- [284] Léon Bottou and Vladimir Vapnik. Local learning algorithms. *Neural computation*, 4(6):888–900, 1992.

- [285] Roderick Murray-Smith and Kenneth Hunt. Local model architectures for nonlinear modelling and control. In *Neural Network Engineering in dynamic control systems*, pages 61–82. Springer, 1995.
- [286] José Luis Díez, José Luis Navarro, and Antonio Sala. A fuzzy clustering algorithm enhancing local model interpretability. *Soft Computing*, 11(10):973–983, 2007.
- [287] José Luis Díez, Antonio Sala, and José Luis Navarro. Fuzzy clustering algorithm for local model control. *IFAC Proceedings Volumes*, 35(1):337–342, 2002.
- [288] Fátima Barceló-Rico, Jorge Bondia, José Luis Díez, and Paolo Rossetti. A multiple local models approach to accuracy improvement in continuous glucose monitoring. *Diabetes technology & therapeutics*, 14(1):74–82, 2012.
- [289] José Luis Díez. *Clustering techniques for identification and control by local models (in Spanish)*. PhD thesis, Tesis Doctoral, Universidad Politécnica de Valencia, Valencia, España, 2003.
- [290] Fátima Barceló Rico et al. *Multimodel Approaches for Plasma Glucose Estimation in Continuous Glucose Monitoring. Development of New Calibration Algorithms*. PhD thesis, 2012.
- [291] Tor Arne Johansen and Roderick Murray-Smith. The operating regime approach to nonlinear modelling and control. *Multiple model approaches to modelling and control*, 1:3–72, 1997.
- [292] Tor A Johansen and Bjarne A Foss. Identification of non-linear system structure and parameters using regime decomposition. *Automatica*, 31(2):321–326, 1995.
- [293] Lotfi A Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on systems, Man, and Cybernetics*, (1):28–44, 1973.
- [294] H Tanaka-S Uegima-K Asai, S Tanaka, and K Uegima. Linear regression analysis with fuzzy model. *IEEE Trans. Systems Man Cybern*, 12:903–907, 1982.
- [295] Robert Babuška. *Fuzzy modeling for control*, volume 12. Springer Science & Business Media, 2012.
- [296] Ebrahim H Mamdani. Application of fuzzy logic to approximate reasoning using linguistic synthesis. In *Proceedings of the sixth international symposium on Multiple-valued logic*, pages 196–202. IEEE Computer Society Press, 1976.

- [297] Witold Pedrycz. An identification algorithm in fuzzy relational systems. *Fuzzy sets and systems*, 13(2):153–167, 1984.
- [298] Soo Yeong Yi and Myung Jin Chung. Identification of fuzzy relational model and its application to control. *Fuzzy Sets and Systems*, 59(1):25–33, 1993.
- [299] Tomohiro Takagi and Michio Sugeno. Fuzzy identification of systems and its applications to modeling and control. In *Readings in Fuzzy Sets for Intelligent Systems*, pages 387–403. Elsevier, 1993.
- [300] Richard J Hathaway and James C Bezdek. Switching regression models and fuzzy clustering. *IEEE Transactions on fuzzy systems*, 1(3):195–204, 1993.
- [301] Dimiter Driankov, Hans Hellendoorn, and Michael Reinfrank. Stability of fuzzy control systems. In *An Introduction to fuzzy control*, pages 245–292. Springer, 1993.
- [302] Tomohiro Takagi and Michio Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE transactions on systems, man, and cybernetics*, (1):116–132, 1985.
- [303] R Babuska. Fuzzy modeling and identification phd dissertation. In *Deft University of Technology*. 1996.
- [304] Iván Contreras, Carmen Quirós, Marga Giménez, Ignacio Conget, and Josep Vehi. Profiling intra-patient type i diabetes behaviors. *Computer methods and programs in biomedicine*, 136:131–141, 2016.
- [305] Iván Contreras, Ignacio Arnaldo, Natalio Krasnogor, and J Ignacio Hidalgo. Blind optimisation problem instance classification via enhanced universal similarity metric. *Memetic Computing*, 6(4):263–276, 2014.
- [306] Iván Contreras, Josep Vehí, Roberto Visentin, and Martina Vettoretti. A hybrid clustering prediction for type 1 diabetes aid: towards decision support systems based upon scenario profile analysis. In *Proceedings of the Second IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies*, pages 64–69. IEEE Press, 2017.
- [307] Stefan Schneider, Ronald J Iannotti, Tonja R Nansel, Denise L Haynie, Bruce Simons-Morton, Douglas O Sobel, Linda Zeitsoff, Loretta Clark, and Leslie P Plotnick. Identification of distinct self-management styles of adolescents with type 1 diabetes. *Diabetes Care*, 30(5):1107–1112, 2007.

- [308] GD Garson. Cluster analysis statnotes: Topics in multivariate analysis 2010, 2010.
- [309] Jennifer M Rohan, Alan Delamater, Jennifer Shroff Pendley, Lawrence Dolan, Grafton Reeves, and Dennis Drotar. Identification of self-management patterns in pediatric type 1 diabetes using cluster analysis. *Pediatric diabetes*, 12(7):611–618, 2011.
- [310] Lyvia Biagi, Josep Antoni Martin-Fernandex, and Josep Vehí. Compositional data analysis of type 1 diabetes data. In *KHD@ IJCAI*, pages 8–12, 2018.
- [311] Giacomo Cappon, Pantelis Georgiou, Pau Herrero, et al. Classification of postprandial glycemic status with application to insulin dosing in type 1 diabetes: An in silico proof-of-concept. *Sensors*, 19(14):3168, 2019.
- [312] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794. ACM, 2016.
- [313] Svend Hylleberg. *Modelling seasonality*. Oxford University Press, 1992.
- [314] A Revert, F Garelli, Jesús Picó, H De Battista, P Rossetti, J Vehi, and J Bondia. Safety Auxiliary Feedback Element for the Artificial Pancreas in Type 1 Diabetes. *IEEE Transactions on Biomedical Engineering*, 60(8):2113–2122, August 2013.
- [315] Sheri R. Colberg, Ronald J. Sigal, Jane E. Yardley, Michael C. Riddell, David W. Dunstan, Paddy C. Dempsey, Edward S. Horton, Kristin Castorino, and Deborah F. Tate. Physical activity/exercise and diabetes: A position statement of the american diabetes association. *Diabetes Care*, 39(11):2065–2079, 2016.
- [316] William L Clarke, Daniel Cox, Linda A Gonder-Frederick, William Carter, and Stephen L Pohl. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes care*, 10(5):622–628, 1987.
- [317] William L Clarke. The original clarke error grid analysis (ega). *Diabetes technology & therapeutics*, 7(5):776–779, 2005.