
UNIVERSITY OF PADUA

DEPARTMENT OF INFORMATION ENGINEERING

Masters degree in Bioengineering

# Modeling the metabolic trajectory of patients with type 1 diabetes using digital twins and Markov Models

**Supervisor:** Prof. Giacomo Cappon

**Candidate:** Francesco Izzi

**Student ID:** 2063527

**Co-supervisor:** Dott. Luca Cossu

**Academic Year:** 2025–2026

**Graduation date:** 20 aprile 2026

# Abstract

Digital twins represent a promising tool for conducting in-silico clinical trials in type 1 diabetes (T1D), enabling researchers to safely evaluate therapeutic strategies without exposing patients to risk and substantially reducing time and cost. However, current digital twin implementations typically generate daily simulations that are independent of each other, lacking a model of how patient-specific physiological parameters evolve over multiple days. Understanding the temporal trajectory of these parameters is essential for developing multi-day simulators that more accurately reflect metabolic adaptation, stability, or deterioration over time.

This thesis aims to model the inter-day metabolic trajectory of patients with type 1 diabetes by analyzing parameter sets extracted from digital twin simulations. A data set comprised approximately 100 patients, each characterized by multiple simulated days. For every patientday instance, a set of physiological and pharmacokinetic parameters was availableincluding basal glucose (Gb), glucose effectiveness (SG), subcutaneous and gastric absorption rates (ka2, kd, kempt), insulin sensitivities in different compartments (SI_B, SI_L, SI_D), and absorption/shape parameters for multiple insulin formulations (kabs_*, beta_*).

To investigate whether patient days exhibit distinctive and recurrent metabolic states, these parameters were grouped into nine biologically motivated combinations (combos), each reflecting a specific physiological mechanism or subsystem. For each combination, clustering techniques were applied to identify latent daily metabolic patterns. The resulting cluster sequences, representing the series of metabolic states visited by each digital twin on days, were subsequently modeled using first-order Markov models. The transition matrices derived from these models quantify the probability of moving from one metabolic state to another, thus providing a representation of the patient's multi-day metabolic dynamics.

The results show that several combos yield well-separated clusters, indicating the presence of distinct daily metabolic regimes throughout the simulated population. Markov models derived from these clusters reveal non-uniform, patient-specific transition structures, with certain metabolic states displaying strong persistence while others occur transiently. These findings suggest that the digital twin dynamics encode meaningful temporal patterns that can be captured and formalized probabilistically.

Overall, this work demonstrates that the metabolic evolution of digital twin patients can be effectively modeled as a Markov process. The proposed framework lays the groundwork for

the construction of multi-day metabolic simulators, which have the potential to accelerate the development and evaluation of therapeutic strategies, improve the realism of in-silico trials, and ultimately support personalized decision-making in the management of type 1 diabetes.

# Summary

I digital twin rappresentano uno strumento promettente per la conduzione di studi clinici in silico nel diabete di tipo 1 (T1D), consentendo ai ricercatori di valutare in modo sicuro diverse strategie terapeutiche senza esporre i pazienti a rischi e riducendo in modo significativo tempi e costi. Tuttavia, le attuali implementazioni di digital twin generano tipicamente simulazioni giornaliere indipendenti tra loro, senza modellare levoluzione nel tempo dei parametri fisiologici specifici del paziente su più giorni consecutivi. La comprensione della traiettoria temporale di tali parametri è essenziale per lo sviluppo di simulatori multi-giornalieri in grado di rappresentare in modo più realistico i processi di adattamento, stabilità o deterioramento metabolico nel tempo.

Questa tesi si propone di modellare la traiettoria metabolica inter-giornaliera di pazienti affetti da diabete di tipo I mediante lanalisi di insiemi di parametri estratti da simulazioni basate su digital twin. Il dataset considerato comprende circa 100 pazienti, ciascuno caratterizzato da più giorni simulati. Per ogni coppia pazientegiorno è disponibile un insieme di parametri fisiologici e farmacocinetici, tra cui la glicemia basale (Gb), lefficacia del glucosio (SG), i tassi di assorbimento sottocutaneo e gastrico (ka2, kd, kempt), le sensibilità insuliniche in diversi compartimenti ($SI_B$, $SI_L$, $SI_D$) e i parametri di assorbimento e di forma associati a diverse formulazioni insuliniche ($kabs_*$, $\beta_*$).

Per indagare se le giornate dei pazienti presentino stati metabolici distinti e ricorrenti, tali parametri sono stati raggruppati in nove combinazioni (combos) motivate biologicamente, ciascuna rappresentativa di uno specifico meccanismo fisiologico o sottosistema. Per ogni combinazione sono state applicate tecniche di clustering al fine di identificare pattern metabolici latenti a livello giornaliero. Le sequenze di cluster risultanti, che rappresentano la successione degli stati metabolici attraversati da ciascun digital twin nel tempo, sono state successivamente modellate mediante modelli di Markov del primo ordine. Le matrici di transizione derivate da tali modelli quantificano la probabilità di transizione da uno stato metabolico a un altro, fornendo così una rappresentazione dinamica dellevoluzione metabolica multi-giornaliera del paziente.

I risultati mostrano che diverse combinazioni producono cluster ben separati, indicando la presenza di regimi metabolici giornalieri distinti allinterno della popolazione simulata. I modelli di Markov costruiti a partire da tali cluster rivelano strutture di transizione non uniformi e specifiche per paziente, con alcuni stati metabolici caratterizzati da elevata persistenza e altri di natura più transitoria. Questi risultati suggeriscono che le dinamiche

dei digital twin incorporano pattern temporali significativi, che possono essere catturati e formalizzati in modo probabilistico.
Nel complesso, questo lavoro dimostra che levoluzione metabolica dei pazienti digital twin può essere efficacemente modellata come un processo di Markov. Il framework proposto pone le basi per la costruzione di simulatori metabolici multi-giornalieri, che hanno il potenziale di accelerare lo sviluppo e la valutazione di strategie terapeutiche, migliorare il realismo degli studi in silico e, in ultima analisi, supportare processi decisionali personalizzati nella gestione del diabete di tipo 1.

# Ringraziamenti

Gentili lettori,
Desidero esprimere la mia più sincera gratitudine al Prof. Giacomo Cappon, relatore di questa tesi, per la disponibilità, la competenza e il costante supporto fornito durante tutto il percorso di lavoro. Un ringraziamento va anche al Dott. Luca Cossu per i preziosi suggerimenti e per il confronto sempre costruttivo. Ringrazio la mia famiglia, che mi ha sostenuto in ogni fase di questo percorso, permettendomi di arrivare fin qui. Senza il loro appoggio e la loro fiducia, nulla di tutto questo sarebbe stato possibile. Un grazie speciale va agli amici e ai colleghi, con cui ho condiviso momenti di studio, difficoltà e soddisfazioni. Il loro supporto, dentro e fuori dalluniversità, ha reso questo percorso più leggero e significativo. Infine, ringrazio tutte le persone che, anche solo con un incoraggiamento o una parola al momento giusto, hanno contribuito a questo traguardo

# Indice

# Elenco delle figure

# Elenco delle tabelle

# Capitolo 1

# Introduction

## 1.1 Type 1 Diabetes and Standard Therapy

Type 1 diabetes mellitus (T1D) is a chronic autoimmune disease resulting from the selective and progressive destruction of pancreatic $\beta$-cells, the endocrine cells responsible for producing and secreting insulin. This autoimmune process is primarily mediated by autoreactive T lymphocytes and is influenced by a complex interplay of genetic susceptibility, environmental triggers, and dysregulation of immune tolerance [1]. Although T1D is often associated with childhood and adolescence, it can develop at any age as a result of the interaction between genetic predisposition, immune dysregulation, and environmental factors. Cases presenting in adulthood are commonly referred to as latent autoimmune diabetes in adults (LADA), a form characterized by a slower decline in $\beta$-cell function, but ultimately leading to insulin dependence.

The autoimmune response in T1D is mediated primarily by autoreactive T lymphocytes. Cytotoxic $CD8^+$ T cells recognize $\beta$-cell antigens and directly induce apoptosis, while helper $CD4^+$ T cells amplify inflammatory responses within the pancreatic islets [1]. B cells also contribute to disease pathogenesis through the production of autoantibodies directed against specific $\beta$-cell antigens, which represent a hallmark of T1D. The autoantibodies detected the most frequently include those targeting glutamic acid decarboxylase (GAD65), insulin, protein tyrosine phosphatase IA-2, and the zinc transporter ZnT8. The presence of two or more of these autoantibodies is a strong predictor of progression toward explicit diabetes.

Genetic predisposition also plays an significant role, particularly genes within the human leukocyte antigen (HLA) complex. Variants such as HLA-DR3, HLA-DR4, and HLA-DQ8 confer the highest risk (Noble & Valdes, 2011). Additional non-HLA genesincluding PTPN22, CTLA4, and IL2RAfurther modulate immune tolerance and pancreatic autoimmunity. However, genetics alone does not fully explain the onset of the disease: the concordance rate among identical twins is approximately 50%, underscoring the importance of environmental factors.

The main environmental factors associated with the development of T1D include viral

infections, particularly enteroviruses, alterations of the intestinal microbiota, dietary patterns during the first years of life, vitamin D deficiency, and psychosocial stress factors. These factors do not act alone, but can interact with individual genetic susceptibility, thus contributing to the initiation or acceleration of the autoimmune process that ultimately leads to the destruction of pancreatic $\beta$-cells. Therefore, within the broader framework of the etiopathogenesis of T1D, environmental influences represent a crucial link between genetic predisposition and the clinical manifestation of the disease.

These elements do not act in isolation, but can interact with individual genetic susceptibility, contributing to triggering or accelerating the autoimmune process that leads to the destruction of pancreatic $\beta$-cells. When placed within the broader framework of T1D etiopathogenesis, environmental factors therefore represent an important link between genetic predisposition and the clinical manifestation of the disease.

Insulin is an essential anabolic hormone that regulates glucose uptake, suppresses hepatic glucose production, and regulates lipid metabolism and protein synthesis. Its absolute deficiency results in a cascade of metabolic abnormalities: impaired peripheral glucose uptake, excessive hepatic gluconeogenesis, increased lipolysis, and elevated ketone production [18]. Clinically, this translates into persistent hyperglycaemia, polyuria, polydipsia, weight loss, fatigue, and, in severe cases, diabetic ketoacidosis (DKA) a life-threatening emergency. Long-term, inadequate glycaemic control markedly increases the risk of microvascular complications such as retinopathy, nephropathy, and peripheral neuropathy, as well as macrovascular complications including coronary artery disease and stroke. The landmark DCCT study demonstrated that intensive glucose control dramatically reduces the incidence and progression of these complications, establishing strict metabolic management as a cornerstone of T1D care [10].

Given the near-complete loss of endogenous insulin production, exogenous insulin replacement is the only effective therapy for T1D. Standard clinical practice relies on two main approaches: Multiple Daily Injections (MDI) and Continuous Subcutaneous Insulin injection (CSII). MDI therapy combines one or two daily injections of long-acting basal insulin with rapid-acting insulin administered before meals or as corrective doses. This regimen requires continued patient participation for carbohydrate counting, dose adjustment, and frequent glucose monitoring.

The introduction of CSII, commonly known as insulin pump therapy, represented a major advancement in the management of diabetes. By continuously delivering rapid-acting insulin through programmable basal rates and bolus doses, CSII allows for more flexible and physiologically adaptive insulin delivery. Evidence from randomized clinical trials and meta-analyzes shows that CSII can reduce glycemic variability, improve HbA1c, and lower the risk of severe hypoglycemia compared to MDI in selected patient populations [14].

Glucose monitoring technologies have evolved significantly in the past two decades. While traditional fingerstick glucometers provide discrete measurements, Continuous Glucose Monitoring (CGM) systems offer real-time, dynamic glucose readings every few minutes.

These devices detect hyperglycaemic and hypoglycaemic excursions that might otherwise remain unnoticed and introduce novel metrics for assessing glycaemic control. Among these, the *Time in Range* (TIR), defined as the percentage of time spent with glucose values between 70180 mg/dL, has emerged as a clinically meaningful index strongly associated with reduced risk of complications [3, 4]. CGM data also enable the quantification of glycaemic variability, time above range, time below range, and glycaemic stability metrics that offer a more detailed picture of metabolic control than HbA1c alone.

Integration of CGM with insulin pumps led to the development of Hybrid Closed-Loop (HCL) systems, often referred to as artificial pancreatic systems. These systems employ control algorithms to automatically adjust basal insulin delivery in response to real-time glucose measurements. Although users still administer mealtime boluses, automated modulation of basal rates substantially increases TIR and reduces both nocturnal and daytime hypoglycemia. Meta-analyzes show that HCL systems significantly improve glycemic outcomes and quality of life for individuals with DM1 [5]. Despite these technological improvements, day-to-day management remains demanding: patients must continuously adapt their therapy to account for variable eating patterns, physical activity, stress, illness, and hormonal fluctuations.

The clinical complexity and burden of self-management underscore the growing interest in computational models and digital health technologies. Physiological glucoseinsulin models, decision-support algorithms and, more recently, *digital twins* aim to assist clinicians and patients by enabling personalized therapy optimization and predicting responses to therapeutic adjustments. A digital twin for T1D is a computational model tailored to the characteristics of an individual patient, capable of reproducing their metabolic dynamics under different conditions. These models offer a safe and controlled environment for conducting *in* silico experiments, testing algorithms, or exploring long-term therapeutic strategies without exposing the patient to risks.

In this context, the growing availability of high-resolution longitudinal data and advanced computational tools has motivated the use of in situ experimentation as a complement to traditional clinical studies. In-silico clinical trials allow for systematic evaluation of therapeutic strategies, decision-support algorithms, and automated insulin delivery systems under controlled and reproducible conditions, without exposing patients to risk. However, most current simulation-based approaches remain limited to short time horizons and focus primarily on single-day dynamics, implicitly assuming that patient-specific physiological characteristics remain constant over time.

Therefore, a solid understanding of the pathophysiology of T1D and the principles of standard therapy is essential to contextualize the use of digital twins and computational methods. The increasing availability of patient-specific physiological parameters, combined with advancements in simulation and data-driven modeling, opens new possibilities for analyzing metabolic trajectories over multiple days, predicting long-term outcomes, and developing more realistic in-silico clinical trials. This conceptual framework provides the

motivation for the present thesis, which focuses on modeling multi-day metabolic evolution in T1D using clustering and Markov modeling applied to digital twinderived parameters.

## 1.2 Advanced Therapies and Decision-Support Algorithms

Treatment of TD has evolved dramatically in the last two decades, driven by technological innovations that aim to reduce the cognitive and practical burden placed on patients. Managing T1D requires continuous decision-making throughout the day: estimating carbohydrate intake, adjusting insulin doses, interpreting glucose trends, accounting for stress, physical activity, hormonal fluctuations, and unpredictable biological variability. This dynamic and high-risk environment makes DD1 one of the most complex chronic diseases to self-manage (American Diabetes Association, 2023).

For this reason, significant research and industrial efforts have focused on developing therapeutic technologies and algorithmic support systems capable of automating or assisting the numerous tasks typically performed manually. This chapter provides an overview of the most advanced therapeutic systems available todayincluding insulin pumps, continuous glucose monitoring (CGM), sensor-augmented pumps, hybrid closed-loop systems, and digital health toolsexplaining their functioning, clinical benefits, and limitations. The discussion culminates in a critical analysis of why the development of automated systems requires simulation environments and digital twins.

### 1.2.1 Insulin Pump Therapy and Sensor Augmentation

Insulin pump therapy, formally known as *Continuous Subcutaneous Insulin Infusion* (CSII), represents one of the most significant technological advances in the treatment of DM1. Unlike multiple daily injections (MDI), which rely on intermittent and manually administered doses of insulin, CSII provides a continuous, programmable, and fine-grained infusion of rapid-acting insulin throughout the day. This mode of delivery is designed to approximate the physiological secretion of insulin by pancreatic $\beta$-cells and reduces the cognitive and practical burden associated with daily diabetes management [14].

An insulin pump is a compact electromechanical device composed of several integrated components that operate together to ensure accurate and flexible insulin delivery. These typically include a reservoir containing rapid-acting insulin, usually with a capacity ranging from approximately 150 to 300 units, a miniature motor-driven pumping mechanism that advances a plunger or piston with high precision, and a microprocessor-based control system capable of executing programmable delivery profiles. The device also incorporates a user interface, which may consist of physical buttons or a touchscreen, as well as an infusion set composed of thin tubing connected to a subcutaneous cannula that delivers insulin to the subcutaneous tissue.

Insulin delivery through CSII is organized into two complementary components: a basal rate and bolus doses. Basal insulin is administered continuously in small micro-pulses to cover background insulin requirements and can be programmed with high temporal resolution, allowing different hourly rates throughout the day to accommodate circadian

variations such as the dawn phenomenon. Bolus doses are delivered on demand to cover meals or correct hyperglycemia and can be administered using different delivery modes, including standard, extended, or dual-wave boluses. In many devices, bolus administration is supported by built-in calculators that integrate carbohydrate estimates, current glucose values, and insulin-on-board (IOB). Compared to MDI, CSII offers increased therapeutic flexibility, reduces glycemic variability, and is associated with a lower risk of severe hypoglycemia, particularly in individuals with unstable glucose profiles [14].

The effectiveness of pump therapy increases substantially when paired with *Continuous Glucose Monitoring* (CGM). A CGM sensor consists of a thin subcutaneous filament coated with glucose oxidase and an electrochemical electrode that measures the current generated by glucose oxidation reactions. This signal is processed and transmitted every 1–5 minutes, providing real-time estimates of interstitial glucose levels. Modern CGM systems achieve high accuracy, often with a mean absolute relative difference (MARD) below 10% % [3]. In addition to absolute glucose values, CGM devices report trend information, predictive alerts for hypoglycaemia and hyperglycaemia, and overnight glucose behavior–capabilities unavailable with fingerstick meters.

Integration of CGM with insulin pumps is known as *sensor-augmented pump therapy*. This combined system marks the first major step toward automation in diabetes care. By providing continuous glucose feedback, sensor-augmented pumps introduce safety-driven control features such as:

- **Low-Glucose Suspend (LGS)**, which automatically stops insulin infusion when the sensor detects hypoglycemia;
- **Predictive Low-Glucose suspension (PLGS)**, which uses CGM trend analysis to forecast imminent hypoglycemia and suspends insulin pre-emptively [6].

These functionalities reduce the frequency and severity of hypoglycaemic episodes, particularly during the night, and improve both safety and quality of life without requiring continuous manual intervention from the patient.

From an engineering perspective, the combined use of CSII and CGM transforms diabetes management into a real-time control problem. The human glucose–insulin system can be viewed as the *plant*, meals and physical activity act as *disturbances*, the insulin pump serves as the *actuator*, and the CGM provides the *sensor*. Under MDI therapy, this system operates in an open-loop configuration. CSII alone remains open-loop but provides finer actuation. Sensor-augmented pumping introduces feedback information, creating a sensorized open-loop system that forms the conceptual and technological foundation for hybrid closed-loop control architectures.

In summary, insulin pump therapy combined with CGM constitutes a crucial intermediate stage between traditional manual treatment and fully automated insulin delivery. It improves clinical outcomes, improves patient safety, and reduces the cognitive burden of self-management of DM1. Most importantly, sensor-augmented pumps provide the building blocks –continuous sensing, fine-grained actuation, and safety-oriented automation—that

underpin the development of advanced closed-loop and artificial pancreas systems.

### 1.2.2 Hybrid closed-loop systems and automated insulin delivery

Hybrid Closed-Loop systems (HCL), also known as automated insulin delivery systems (AID) or artificial pancreas technologies, represent a major advance in the technological evolution of T1D. These systems integrate three key components: Continuous Subcutaneous Insulin Infusion (CSII), Continuous Glucose Monitoring (CGM), and model-based control algorithmsto create a semi-automated feedback-control architecture capable of adjusting insulin delivery in real time. Although users must still administer mealtime boluses, the system autonomously modulates basal insulin throughout the day, bridging the gap between manual therapy and fully automated insulin regulation [11].

At the core of an HCL system is a control algorithm that receives glucose estimates from the CGM every few minutes and determines the insulin infusion rate needed to maintain glucose levels within a desired target range. This transforms diabetes management into a closed-loop control problem, where the CGM provides continuous feedback, the insulin pump acts as the actuator, and the human glucose–insulin system constitutes the plant. Depending on the specific commercial or research system, different control strategies are implemented, including:

- **Proportional–Integral–Derivative (PID) control**, which adjusts insulin delivery based on real-time deviations from the target glucose level, as well as their rate and cumulative magnitude;
- **Model Predictive Control (MPC)**, which uses physiological models to predict future glucose trajectories and optimizes insulin delivery over a prediction horizon;
- **Adaptive and data-driven algorithms** that update control parameters based on individual responses to insulin, meals, and physical activity.

HCL algorithms typically operate on a 5-minute control cycle, during which the controller continuously updates insulin delivery based on the most recent physiological information. At each iteration, the system acquires the latest glucose measurement and trend information from the CGM and uses internal models or data-driven methods to predict future glucose levels over a short-term horizon. These predictions are then compared with the predefined target range, allowing the controller to determine whether an adjustment of insulin delivery is required. Based on this comparison, the algorithm computes an appropriate modification of the basal insulin infusion, which may involve an increase, a decrease, or a temporary suspension of delivery, and subsequently commands the insulin pump to implement the updated basal rate.

This closed-loop feedback ensures that insulin delivery is continuously adapted to changing physiological demands, such as stress, hormonal fluctuations, and unannounced meals or snacks.

Clinical evidence consistently demonstrates substantial benefits of HCL systems in diverse populations. Major trials show improvements in *Time in Range* (TIR), reductions in

the time spent in hypoglycemia, and improved nocturnal glycemic stability compared to standard pump therapy or sensor-augmented pumps [11]. Users also report reduced therapy-related burden, improved sleep quality, and increased confidence in day-to-day diabetes management.

Despite these advances, current systems remain "hybrid" rather than fully automated, as they continue to require active user participation in several aspects of therapy management. In particular, users must announce meals and manually administer bolus insulin, adjust dosing strategies for meals with a high carbohydrate or fat content, and respond to alerts or alarms related to sensor inaccuracies, infusion set failures, or pump malfunctions. These limitations stem primarily from the pharmacokinetics of subcutaneous insulin delivery: even rapid-acting insulin analogs exhibit an onset of action on the order of 10–20 minutes and reach peak activity after approximately 60–90 minutes, which is substantially slower than the near-instantaneous endogenous insulin secretion into the portal vein. As a consequence, current control algorithms cannot fully compensate for rapid postprandial glucose excursions, thereby necessitating user-initiated bolus administration.

Fully closed-loop systems – capable of autonomously managing both basal and bolus insulin delivery—represent the next frontier in diabetes technology. However, their development faces several challenges, including the following:

- **delayed insulin absorption**, which limits the responsiveness of the controller;
- **inter- and intra-patient variability** in insulin sensitivity, gastric emptying, and daily routines;
- **CGM noise and time lag**, which complicate real-time control decisions;
- **the need for robust safety layers** to prevent insulin stacking or over-delivery in unpredictable conditions.

To overcome these barriers, research is exploring alternative strategies such as ultra-rapid insulin formulations, dual-hormone systems delivering both insulin and glucagon, and machine learning approaches that adaptively model patient-specific behaviors. Although promising, these innovations must be validated through extensive tests to ensure safety in a wide range of real-world conditions.

In general, hybrid closed-loop systems represent a critical milestone on the path toward full automation in T1D therapy. By integrating continuous sensing, intelligent control algorithms, and precise insulin actuation, they significantly improve metabolic outcomes and quality of life while reducing the day-to-day burden of disease management. At the same time, their limitations highlight the need for sophisticated simulation environments capable of capturing physiological variability and edge-case scenarios, an essential step toward developing safe and reliable next-generation artificial pancreas systems.

### 1.2.3 Decision-support systems and digital health tools

Parallel to the development of automated insulin delivery systems, a broad ecosystem of decision-support systems (DSS) and digital health tools has emerged to help patients and

clinicians with the daily management of type 1 diabetes. Although hybrid closed-loop systems aim to automate insulin delivery, decision-support tools are designed primarily to support, structure, and augment human decision-making rather than to replace it entirely. These systems exploit the growing availability of continuous glucose monitoring (CGM) data, detailed insulin dosing histories, meal logs, and contextual information (such as physical activity, stress, or intercurrent illness) to generate personalized recommendations and clinically meaningful insights [2].

From a functional perspective, diabetes decision-support systems can be broadly grouped into several categories:

- **Patient-facing tools**, such as smartphone applications and on-device features that help people calculate insulin doses, interpret glucose trends, or plan meals and physical activity;
- **Clinician-facing platforms**, typically web-based dashboards that aggregate longitudinal CGM, pump, and self-reported data into summaries and visualizations that support therapeutic adjustments during clinic visits or telemedicine encounters;
- **Integrated DSS within medical devices**, such as smart bolus calculators and pattern-recognition modules embedded directly in insulin pumps or CGM receivers.

A prototypical example of a patient-facing decision-support tool is the *bolus calculator*. These algorithms estimate the appropriate mealtime or correction bolus by combining carbohydrate estimates, current glucose levels, active insulin (insulin-on-board), and individualized parameters such as the insulin-to-carbohydrate ratio and the correction factor. By systematizing calculations that would otherwise be performed mentally or with ad hoc heuristics, bolus calculators reduce cognitive load and help prevent systematic under- or over-dosing. More advanced tools incorporate features such as meal type (e.g., high-fat or high-protein meals), planned physical activity, or specific glycemic targets (for example, before exercise or sleep).

Another important class of decision-support tools is represented by *pattern recognition systems*. These algorithms analyze CGM traces and insulin dosing histories over days or weeks to identify recurring glycemic patterns, such as consistent nocturnal hypoglycemia, post-breakfast hyperglycemia, or late-evening glucose rises. The insights derived from this analysis can be translated into clinically actionable results, including suggested adjustments to base insulin profiles, recommendations to modify insulin-to-carbohydrate ratios for specific meals, and targeted alerts that prompt patients or clinicians to review particular time periods of interest. Such systems are particularly valuable in clinical practice, where limited consultation time often makes it impractical to manually inspect and interpret large volumes of CGM data.

Digital health platforms extend these ideas by integrating multiple data streams–including CGM, pump or injection records, physical activity from wearable sensors, and patient-reported outcomes–into unified cloud-based environments. These platforms enable remote monitoring, asynchronous review of data, and telemedicine consultations, thereby facilita-

ting more frequent but less intrusive adjustments to therapy. For clinicians, they provide aggregated views at different time scales (e.g., daily profiles, weekly summaries, monthly statistics) and derived metrics such as Time in Range (TIR), glycemic variability indices, and hypoglycemic events counts, which can guide evidence-based decisions.

In recent years, machine learning (ML) and data-driven approaches have been increasingly explored for decision support in diabetes management. These methods leverage information from CGM data, insulin dosing histories, and meal records to perform a variety of tasks, including the prediction of short-term glucose trajectories, the estimation of time-varying individual insulin sensitivity, and the classification of days into recurring behavioral or metabolic patterns. In addition, ML-based systems have been proposed to provide personalized recommendations for insulin dosing strategies or lifestyle changes. Despite their potential, the integration of ML algorithms into regulated medical devices and routine clinical workflows remains challenging. Key issues include limited model interpretability, sensitivity to sensor noise and missing data, potential biases in the training datasets, and the need for rigorous validation across diverse patient populations and real-world conditions.

Despite their potential, all decision-support systems – from simple bolus calculators to complex ML-based platforms – must satisfy strict clinical safety requirements. Algorithms are expected to perform reliably not only under typical conditions but also in unusual physiological states, such as illness, hormonal changes, or extreme exercise, and in rare edge-case scenarios that can carry a high risk of severe hypoglycemia or ketoacidosis. Validating such systems directly in real-world settings is difficult and expensive and raises ethical concerns, especially when testing unproven strategies or exploring worst-case conditions.

These limitations highlight the importance of complementary *in* situ evaluation frameworks. Physiologically grounded simulators and digital twins can generate large and diverse virtual patient populations, reproduce a wide spectrum of behavioral and metabolic scenarios, and allow decision-support algorithms to be stress-tested under controlled, yet realistic conditions. In this context, digital health tools and decision-support systems do not simply consume data generated by digital twins; they can also be developed, calibrated, and refined using synthetic data before being deployed to support real patients. This interplay between advanced decision-support algorithms and high-fidelity simulation environments represents a key enabler for the safe and efficient translation of digital technologies into clinical practice.

### 1.2.4 Clinical validation and safety considerations for algorithm-based therapies

The increasing adoption of algorithm-driven technologies in T1D including sensor-augmented pumps, hybrid closed-loop systems, and advanced decision-support toolshas introduced new challenges in clinical validation and safety assurance. Unlike traditional insulin

therapies, which are based primarily on user-driven decisions, algorithm-based systems continuously interpret data streams, generate dosing recommendations, or autonomously adjust insulin delivery. As a result, their safety depends not only on the accuracy of the underlying physiological models, but also on their robustness to noisy data, unexpected real-life conditions, and individual variability.

Validating such systems is inherently complex. Clinical trials for algorithm-based diabetes therapies must demonstrate that the system operates safely in a wide spectrum of physiological and behavioral scenarios. These include variations in insulin sensitivity, unpredictable meal timing and composition, physical activity, stress, illness, and sensor artifacts. However, real-world studies cannot afford to explore all relevant conditions ethically or economically. Rare but dangerous situationssuch as rapid glucose drops after intense exercise, sensor failure during sleep, or delayed meal absorptionare difficult to reproduce in a controlled clinical environment, but are critical to testing system safety. Trial participants typically follow structured protocols and receive close monitoring, conditions that differ substantially from everyday life.

Another major challenge arises from the variability between patients and within the patient. Even within the same individual, insulin needs vary between days due to hormonal cycles, stress, circadian rhythms, and unpredictable behavioral factors. For algorithm-based therapies, this variability creates a vast landscape of potential operating conditions. Ensuring that an algorithm generalizes safely to all of them requires extensive testing that cannot be achieved solely through human-subject studies. In addition, regulatory approval demands evidence of safety in edge-case scenariosrare events that are unlikely to occur during the limited duration of a clinical trial but may still pose significant risks if the algorithm does not handle them appropriately.

Technical factors further complicate the validation. CGM sensors introduce measurement noise, calibration drift, and physiological time lags between interstitial and blood glucose. Insulin pumps may experience occlusions, infusion-set failures, or variable absorption kinetics. Algorithms must remain stable and safe under these imperfections, even when the input data is unreliable or delayed. Similarly, algorithm updates, software revisions, and firmware modifications require repeated rounds of validation, generating substantial regulatory burden.

In light of these challenges, traditional clinical trials, though essential, are insufficient to guaranty a comprehensive safety evaluation of algorithm-based therapies. The enormous space of physiological states, behavioral patterns, and device-related disturbances cannot be fully explored in vivo. This limitation motivates the adoption of complementary testing strategies that allow researchers and developers to explore a broader range of scenarios than those attainable with human participants alone.

These considerations naturally lead to the need for controlled, scalable, and ethically safe evaluation environments. *In-silico* simulation platforms, physiologically grounded virtual patients, and digital twin technologies offer a powerful means to investigate algorithm

behavior under diverse and extreme conditions, enabling systematic stress-testing that would be impractical or unsafe in the real world. The following section examines how simulation environments address these validation challenges and why they have become indispensable tools for the development of advanced diabetes algorithms.

### 1.2.5 Need for simulation environments for algorithm development

As diabetes technologies become increasingly algorithm-driven, their development requires testing in a wide and heterogeneous range of physiological and behavioral conditions. However, Real-world studies can only capture a fraction of this variability and ethically cannot explore extreme or unsafe scenarios. To overcome these limitations, the field has progressively adopted *in-silico* simulation environments, which provide scalable, controllable, and reproducible platforms for evaluating insulin delivery algorithms under various circumstances.

Physiologically based simulators of glucoseinsulin dynamics enable thousands of virtual experiments to be performed rapidly and without risk to patients. These tools support early-stage algorithm design, edge case stress-testing, and systematic comparison of therapeutic strategies. Among the most advanced simulation paradigms is the concept of a *digital twin*: a personalized virtual patient whose metabolic behavior reflects that of a specific individual. Digital twins can capture patient-specific characteristics, offer a safe environment to testing unvalidated control strategies, and generate synthetic datasets that would be difficult or impossible to obtain clinically.

Despite their potential, most existing digital twin frameworks operate on a single-day timescale and do not account for how physiological parameters evolve over days. As a result, they cannot reproduce long-term metabolic adaptation, variability, or progression. This limitation motivates the present work, which focuses on modeling multi-day metabolic evolution in type 1 diabetes and providing a framework capable of capturing inter-day physiological dynamics for next-generation digital twin simulations.

## 1.3 Digital Twins in Type 1 Diabetes

Digital twins (DTs) are increasingly recognized as a key enabling technology for precision medicine, as they allow the creation of patient-specific computational models that can be continuously updated using real-world data and interrogated through simulation. In the context of type 1 diabetes (T1D), DTs are particularly attractive because disease management already relies on a tightly integrated technological ecosystem comprising continuous glucose monitoring, insulin delivery devices, and algorithmic decision-support tools. This combination of sensing, actuation, and data availability provides a natural foundation for the development of virtual patient replicas capable of reproducing glucose–insulin dynamics and supporting the safe and cost-effective exploration of alternative therapeutic or behavioral scenarios [12, 19].

In recent years, a growing body of literature has proposed DT frameworks specifically tailored to T1D. These systems aim to emulate how an individual patient responds to meals, insulin administration, physical activity, stress, and other metabolic disturbances by personalizing parameterized physiological models using CGM data, insulin delivery records, and contextual information [16, 17]. Once instantiated, such digital twins can be queried through simulation to perform short-term glucose prediction, evaluate therapy adjustments, support clinical decision-making, and test automated insulin delivery strategies. Importantly, DT-based simulators have also enabled large-scale in-silico clinical trials, reducing the cost, duration, and ethical burden associated with traditional human studies [20].

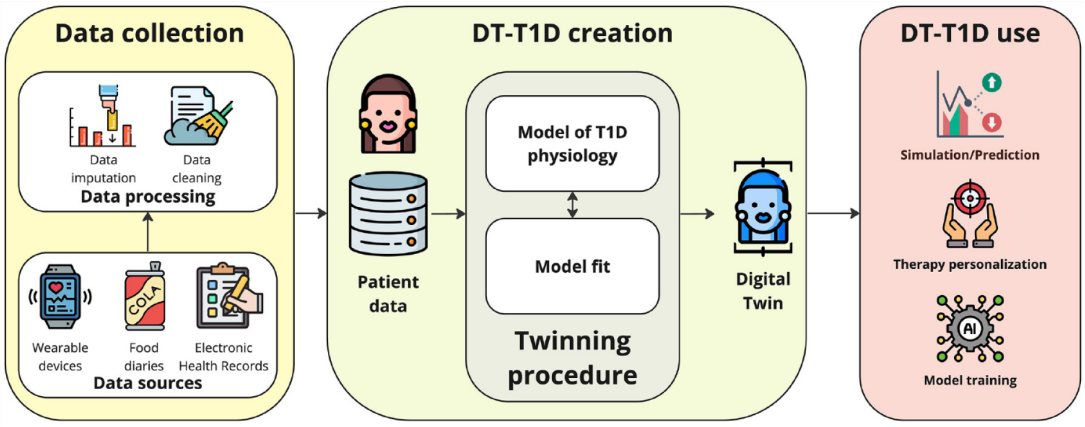


Figura 1.1: Schematic workflow of a digital twin for T1D, including multi-source data collection, personalization of a patient-specific model, simulation of alternative scenarios, and decision support.

From a methodological standpoint, most DTs for T1D are based on mechanistic models formulated as systems of ordinary differential equations that describe glucose–insulin dynamics at varying levels of detail. These models are typically personalized through a *twinning procedure*, in which physiological parameters such as insulin sensitivity, glucose effectiveness, and absorption rates are estimated from patient data. High-fidelity simulators, including those derived from the UVA/Padova framework, offer detailed multi-compartment

representations and are widely used for research-grade simulations and regulatory evaluation of automated insulin delivery systems [20]. At the same time, more recent approaches integrate data-driven components with mechanistic backbones to improve predictive performance under free-living conditions and to capture variability induced by factors such as stress, sleep, or physical activity [16].

Despite these advances, a fundamental limitation of current DT implementations concerns their treatment of time. In the majority studies, digital twins are calibrated using data from a single day or a short time window, and the resulting physiological parameters are assumed to remain constant during subsequent simulations. As a consequence, even when multiple days are simulated, each day is effectively treated as an independent realization of the same underlying model. This assumption contrasts with real-world T1D behavior, where insulin sensitivity, absorption dynamics, hormonal responses, and behavioral patterns can vary substantially from day to day [12]. Although existing DTs are highly effective in reproducing intra-day glucose fluctuations, they generally do not capture the inter-day evolution of metabolic states.

An additional and closely related challenge concerns the validation of T1D digital twins. In principle, a digital twin should be considered clinically meaningful only if it can generalize beyond the specific data used for its personalization and reproduce patient responses under previously unseen conditions. However, in practice, most validation strategies remain limited to short-term horizons or single-day settings. Many studies evaluate DT performance by comparing simulated and measured glucose trajectories over the same day used for calibration, or by testing the model on synthetic data generated from physiological simulators. Although these approaches are useful for assessing internal consistency, they provide limited insight into the ability of a digital twin to capture longer-term metabolic behavior.

Only a small number of works explicitly assess cross-day or longitudinal generalization, for example, by calibrating a digital twin on one day and evaluating its performance on subsequent days. This limitation is particularly relevant in T1D, where physiological parameters and behavioral factors exhibit substantial variability between days. As a result, current validation practices implicitly reinforce the assumption of static patient characteristics and do not test whether digital twins can reliably track or predict transitions between different metabolic regimes over time.

This limitation has direct implications for the design and interpretation of in-silico clinical trials. Current DT-based simulations are well suited to assess acute outcomes, such as postprandial control or overnight stability, but are less informative when addressing questions that inherently span longer time horizons. Examples include the robustness of insulin delivery algorithms to evolving physiological conditions, the persistence of glycemic control over weeks, or the cumulative impact of behavioral changes. Therefore, simply extending simulations over multiple consecutive days without taking into account changes in patient-specific parameters may lead to unrealistic or overly optimistic predictions [19].

In this context, modeling inter-day metabolic evolution represents an open and largely underexplored challenge in T1D digital twins. Rather than treating daily simulations as isolated instances, an alternative perspective is to summarize each day through a latent metabolic state and to explicitly model how patients transition between such states over time. This thesis adopts this point of view by analyzing daily parameter sets extracted from digital twin simulations, identifying recurrent metabolic regimes, and representing their temporal evolution through probabilistic models. By doing so, the proposed framework provides a principled foundation for multi-day digital twin simulations and addresses a critical gap in the current state of the art.

## 1.4 In-silico Clinical Trials and the Need for Modelling Temporal Evolution

Digital twins have emerged as a powerful tool for supporting research, development, and decision-making in type 1 diabetes. By allowing simulation of patient-specific metabolic responses under controlled and reproducible conditions, DTs allow researchers and clinicians to explore alternative insulin strategies, behavioral scenarios, and control algorithms without exposing patients to risk. One of the most prominent applications of this paradigm is the use of digital twins to conduct in-silico clinical trials, in which virtual patient cohorts are employed to evaluate therapies and technologies in a scalable and cost-effective manner.

In silico clinical trials represent a transformative complement to traditional human studies. By replacing or augmenting in vivo experimentation with virtual experiments, they allow the systematic evaluation of thousands of scenarios involving meals, insulin strategies, physical activity, or closed-loop algorithms, while significantly reducing cost, duration, and ethical constraints. Regulatory agencies have already endorsed simulation-based assessment for the early-stage evaluation of automated insulin delivery systems, further highlighting the relevance of this approach.

Despite this progress, most current in-silico studies operate on relatively short time horizons, typically ranging from a few hours to a few days. These settings are well suited for assessing acute responses such as postprandial control, correction boluses, or overnight stability, but they are insufficient when addressing questions that intrinsically involve longer time scales. Examples include the long-term robustness of insulin delivery algorithms, the cumulative impact of behavioral changes, or the stability of metabolic control over weeks or months.

A key limitation underlying this gap is that many simulation frameworks implicitly assume that each simulated day is independent. Even when longer simulations are performed, the physiological parameters of the digital twinsuch as insulin sensitivity, gastric emptying, or subcutaneous absorptionare often kept fixed. As a consequence, these models capture intra-day variability driven by meals, insulin, activity, and sensor noise, but do not reflect the inter-day physiological fluctuations that are well documented in individuals with DM1. In real patients, metabolic behavior varies substantially from day to day as a result of stress, illness, hormonal changes, exercise patterns, circadian rhythms, behavioral choices, and adherence.

This mismatch between modeling assumptions and real-world variability creates a critical limitation in the ability of in-silico trials to serve as realistic substitutes for long-term clinical studies. To rigorously evaluate an insulin delivery algorithm or a decision-support tool, it is necessary to assess not only its performance within a single day but also its ability to adapt to evolving metabolic conditions over time. Patients may transition between stable control regimes, periods of elevated variability, or phases characterized by recurrent hypoglycemia or hyperglycemia, and capture of these transitions is essential to assess long-term safety and effectiveness.

Extending physiological simulations naï on many consecutive days is not sufficient, as it is computationally demanding, conceptually limited when parameters are assumed to remain static, and unable to capture higher-level metabolic patterns. A more principled approach requires summarizing the metabolic state of each day and explicitly modeling how these states evolve over time. This motivates the development of higher-level representationssuch as clusters of physiologically similar daysand stochastic models that describe the transitions between them.

In this work, we address this challenge by modeling the inter-day evolution of digital twin parameters in a compact yet expressive form. By analyzing how virtual patients transition between different daily metabolic states, we lay the foundation for multi-day in-silico simulations that better approximate long-term metabolic trajectories. This approach bridges the gap between short-term virtual experiments and realistic long-term virtual trials, providing a more reliable framework for the evaluation of future algorithmic and therapeutic innovations in type 1 diabetes.

## 1.5 Objectives and structure of the thesis

This thesis addresses the problem of modeling the inter-day metabolic evolution of individuals with T1D using digital twin simulations. Although existing approaches primarily focus on short-term or single-day dynamics, the central research objective of this work is to investigate whether recurrent daily metabolic states can be identified from digital twin–derived parameters and whether the temporal evolution between such states can be described in a principled, interpretable, and informative way.

The analysis is conducted on real-world clinical data collected from individuals with T1D and made available through the Tidepool Data Donation Project, which provides high-resolution longitudinal records of glucose measurements and insulin therapy.

To this end, the thesis focuses on the characterization of daily metabolic behavior and the representation of its evolution over consecutive days. Individual daily simulations are summarized through latent metabolic states that capture salient features of patient-specific physiology. These states are then used to describe how patients transition between different metabolic regimes over time, enabling the analysis of persistence, variability, and structural patterns in metabolic dynamics at a multi-day time scale.

A further objective of the thesis is to assess the extent to which the proposed framework supports meaningful long-term simulation and analysis. In particular, the work investigates whether the resulting representations allow for the generation of plausible multi-day metabolic trajectories and whether they capture heterogeneity in temporal behavior across different individuals. This evaluation is essential to determine whether the proposed approach can serve as a realistic abstraction of longer-term metabolic evolution rather than a mere extension of single-day simulations.

Where informative, the identified metabolic states are also related to clinically relevant outcome metrics such as Time in Range. This additional analysis is intended to support the interpretation of states not only in terms of abstract model parameters, but also in relation to clinically meaningful indicators of glycemic control, thereby facilitating a more intuitive connection between computational representations and real-world outcomes.

The structure of the thesis reflects this progression from problem formulation to methodological development and analysis. Following the present introductory chapter, which outlines the clinical background, the role of digital twins, and the motivation to model multi-day metabolic evolution, Chapter 2 presents the methodological framework underlying the proposed approach. Chapter 3 reports the main findings obtained from the analysis of digital twin simulations and illustrates the identified metabolic states and their temporal transitions. Chapter 4 discusses these results in light of the physiology and clinical management of type 1 diabetes, highlighting both strengths and limitations of the approach. Finally, Chapter 5 summarizes the main contributions of the work and outlines potential directions for future research.

# Capitolo 2

# Methods

## 2.1 ReplayBG: A Digital Twin Framework for Personalized Simulation in Type 1 Diabetes

ReplayBG is a digital twin framework for type 1 diabetes (T1D) that enables personalized simulation of glucose–insulin dynamics from real-world patient data. Using continuous glucose monitoring (CGM) measurements together with insulin delivery records (basal and bolus) and carbohydrate intake events, ReplayBG instantiates a patient-specific physiological model whose parameters are identified from the available observations. The framework is described in Cappon et al. [8] and has been further extended with a complementary data-augmentation approach in Cappon et al. [7]. As summarized in Figure 2.1, ReplayBG is organized around two core phases. In the *twinning* phase, the physiological model is personalized to the individual by estimating a set of parameters that reproduce the observed glucose response to insulin and meals, while accounting for uncertainty arising from sensor noise, reporting inaccuracies, and day-to-day variability. In the *replay* phase, the personalized model is simulated under modified inputs to generate counterfactual glucose trajectories corresponding to alternative therapeutic scenarios, such as changes in bolus dosing, basal profiles, meal timing or carbohydrate amounts. This replay capability supports the systematic evaluation of treatment strategies and control policies in a controlled in-silico setting. In addition to point estimates, ReplayBG provides uncertainty-aware simulation outputs in the form of credible intervals, enabling a probabilistic characterization of the predicted glucose response under the specified scenario.

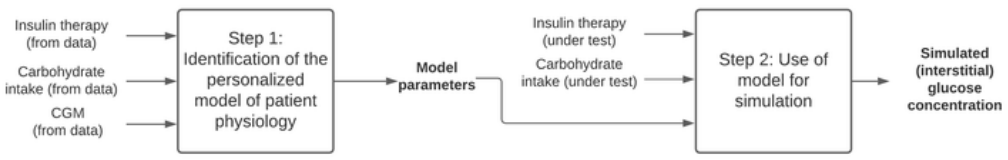


Figura 2.1: Block diagram of the ReplayBG modeling workflow, showing data collection, parameter identification, simulation, and decision-support steps.

### 2.1.1 Physiological Model Structure

The physiological core of ReplayBG is a mechanistic compartmental model describing glucose–insulin dynamics through the interaction of insulin absorption, oral glucose appearance, and glucose–insulin regulation. A schematic overview of the model is shown in Figure 2.2, which highlights the main subsystems and the flow of information between them. Exogenous inputs in the form of insulin delivery and carbohydrate intake are processed by dedicated absorption subsystems and integrated within a regulatory module that generates simulated glucose trajectories consistent with continuous glucose monitoring (CGM) observations.

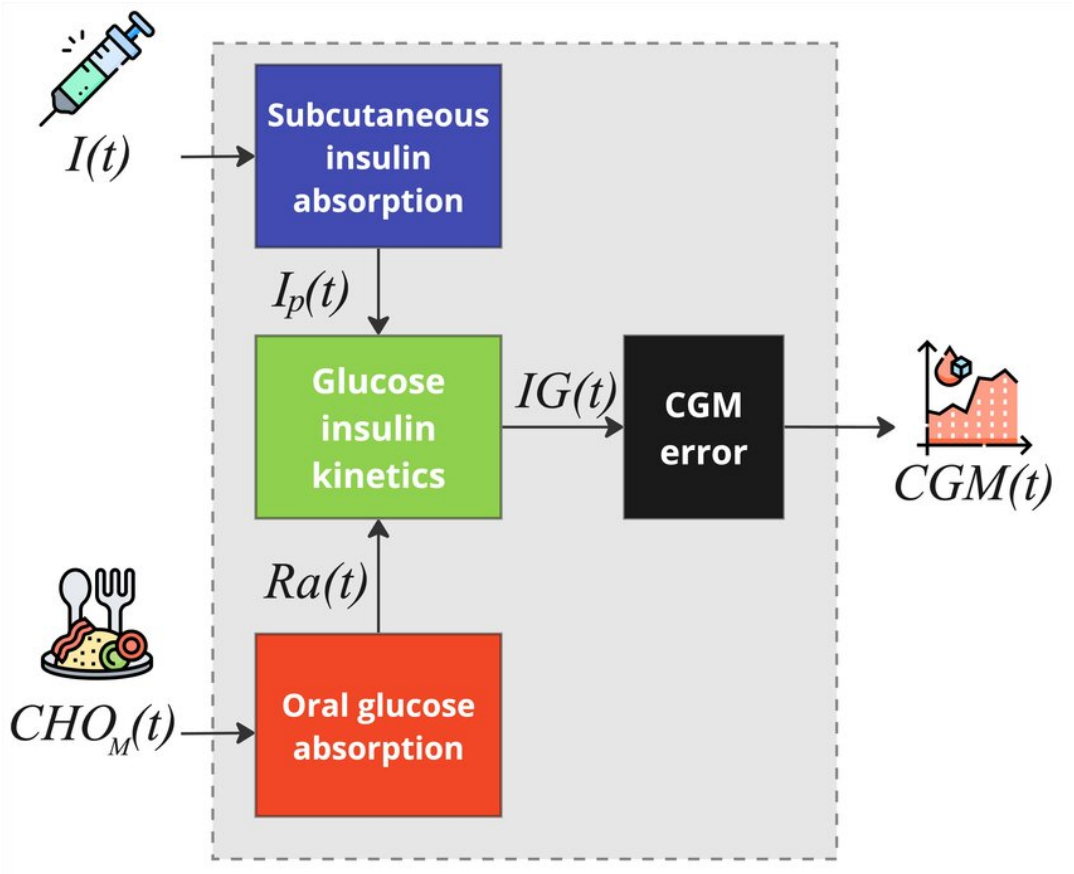


Figura 2.2: ReplayBG physiological compartmental model. Insulin delivery and carbohydrate intake are processed through dedicated absorption subsystems and combined within the glucose–insulin regulatory dynamics, followed by a CGM observation model.

The first block of the model represents subcutaneous insulin absorption. Insulin delivered by basal infusion or bolus administration enters a subcutaneous depot and undergoes a sequence of dissociation and absorption processes before reaching the plasma compartment. This delayed and smoothed pharmacokinetic behavior is modeled through a

three-compartment structure described by

$$\begin{cases} \dot{I}_{\mathrm{sc1}}(t) = -k_d\, I_{\mathrm{sc1}}(t) + \dfrac{I(t-\beta)}{V_I}, \\ \dot{I}_{\mathrm{sc2}}(t) = k_d\, I_{\mathrm{sc1}}(t) - k_{a2}\, I_{\mathrm{sc2}}(t), \\ \dot{I}_p(t) = k_{a2}\, I_{\mathrm{sc2}}(t) - k_e\, I_p(t), \end{cases} \tag{2.1}$$

where $I_p(t)$ denotes the plasma insulin concentration. The parameters governing dissociation, absorption, and clearance determine the timing and magnitude of insulin action and constitute key descriptors of patient-specific pharmacokinetics.

Oral glucose absorption is modeled through a mechanistic representation of the gastrointestinal tract that converts carbohydrate intake into a glucose appearance rate in plasma. In the single-meal formulation, gastric emptying and intestinal absorption are described by a stomach–gut structure,

$$\begin{cases} \dot{Q}_{\mathrm{sto1}}(t) = -k_{\mathrm{empt}} Q_{\mathrm{sto1}}(t) + \mathrm{CHO}(t), \\ \dot{Q}_{\mathrm{sto2}}(t) = k_{\mathrm{empt}} Q_{\mathrm{sto1}}(t) - k_{\mathrm{empt}} Q_{\mathrm{sto2}}(t), \\ \dot{Q}_{\mathrm{gut}}(t) = k_{\mathrm{empt}} Q_{\mathrm{sto2}}(t) - k_{\mathrm{abs}} Q_{\mathrm{gut}}(t), \end{cases} \tag{2.2}$$

with the corresponding glucose appearance rate given by

$$Ra(t) = f\, k_{\mathrm{abs}} Q_{\mathrm{gut}}(t).$$

To account for multiple meals, snacks, and hypoglycemia treatments occurring within the same day, this formulation is extended by superposing multiple absorption pathways,

$$Ra(t) = f \sum_{M} k_{\mathrm{abs},M} Q_{\mathrm{gut},M}(t), \tag{2.3}$$

where each type of meal $M$ is characterized by its own absorption dynamics. The resulting glucose appearance rate drives postprandial glucose excursions in the regulatory subsystem.

The glucose–insulin regulatory block integrates the effects of plasma insulin and glucose appearance to determine plasma glucose dynamics. This component extends the classical minimal model by incorporating additional states and nonlinearities to improve accuracy across a wide glycemic range. The system is described by

$$\begin{cases} \dot{G}(t) = -\Big[S_G + \rho(G)X(t)\Big]G(t) + S_G G_b + \dfrac{Ra(t)}{V_G}, \\ \dot{X}(t) = -p_2 \Big[X(t) - S_I\Big(I_p(t) - I_{pb}\Big)\Big], \\ \dot{IG}(t) = -\dfrac{1}{\alpha}\Big[IG(t) - G(t)\Big], \end{cases} \tag{2.4}$$

where $G(t)$ denotes plasma glucose, $X(t)$ represents delayed insulin action, and $IG(t)$ models interstitial glucose as a physiological proxy for CGM measurements. The parameters

governing glucose effectiveness, insulin sensitivity, and temporal delays summarize key regulatory properties of the individual metabolic system.

Together, the insulin absorption, oral glucose absorption, and glucose–insulin regulatory subsystems form a unified dynamical system that maps insulin delivery and carbohydrate intake into simulated glucose trajectories. The complete physiological model can be compactly expressed as

$$\dot{x}_{\text{phy}}(t) = f_{\text{phy}}(x_{\text{phy}}(t), u_{\text{phy}}(t), \theta_{\text{phy}}) \,,$$

where the state vector $x_{\text{phy}}$ collects insulin, glucose, gastrointestinal and CGM-related compartments, the input vector $u_{\text{phy}}$ comprises insulin and meal events, and the parameter vector $\theta_{\text{phy}}$ includes physiological and pharmacokinetic quantities governing absorption, regulation, and detection.

In ReplayBG, these parameters are estimated during the twinning phase for each simulated day and constitute a compact description of the patients metabolic state. In the present work, the daily parameter sets extracted from this model represent the primary objects of analysis and provide the basis for the inter-day modeling framework introduced in the following sections.

### 2.1.2 Physiological and Pharmacokinetic Parameters Extracted from ReplayBG

The twinning procedure implemented in ReplayBG produces, for each simulated day and for each individual, a set of physiological and pharmacokinetic parameters that characterize the inferred metabolic behavior of the patient during that day. These parameters summarize the dynamics of insulin absorption, oral glucose appearance, and glucose–insulin regulation described in the previous section, and provide a compact and physiologically interpretable representation of the daily metabolic state.

In the present work, each patient–day instance is therefore represented by a vector of parameters estimated from real-world data. Rather than analyzing raw glucose time series directly, the proposed framework treats these parameter vectors as high-level descriptors of daily metabolism, allowing comparison between days and individuals and supporting the identification of recurrent metabolic regimes.

The first group of parameters describes the basal regulation of glucose and the insulin-independent glucose dynamics. The basal glucose concentration $G_b$ represents the equilibrium level around which daily glucose fluctuations occur, while the glucose effectiveness $S_G$ quantifies the ability of glucose to promote its own disposal independently of insulin action. Variations in these parameters reflect differences in baseline metabolic control and endogenous glucose regulation between patients and days.

Parameters that describe insulin sensitivity and insulin action play a central role. ReplayBG allows insulin sensitivity to vary in different time windows throughout the day, resulting

in distinct parameters $S_{I,B}$, $S_{I,L}$, and $S_{I,D}$, corresponding to the morning, daytime, and evening periods. These parameters capture circadian and behavioral effects on insulin response and are among the most informative quantities to characterize day-to-day metabolic variability. Additional parameters such as the insulin action rate constant $p_2$ influence the temporal dynamics of insulin effects, determining how rapidly insulin-mediated glucose removal responds to changes in plasma insulin.

The pharmacokinetic aspects of insulin delivery are summarized by parameters associated with the subcutaneous absorption subsystem. The dissociation rate $k_d$ governs the conversion of insulin from a non-monomeric subcutaneous depot to an absorbable form, while the absorption rate $k_{a2}$ controls the transfer of insulin to the plasma compartment. Together, these parameters strongly influence the timing and magnitude of insulin action and play a key role in shaping the postprandial glucose dynamics. Additional parameters such as plasma insulin clearance rate $k_e$ and infusion delay $\beta$ further modulate insulin availability and contribute to variability between days and between individuals.

Oral glucose absorption is characterized by parameters that describe gastric emptying and intestinal uptake. The gastric emptying rate $k_{\text{empt}}$ regulates the speed with which ingested carbohydrates leave the stomach and become available for absorption, while intestinal absorption rates $k_{\text{abs},M}$ describe glucose transfer from the gut to plasma for different types of meals. Among these, $k_{\text{empt}}$ is particularly relevant, as gastric emptying is one of the most variable physiological processes that affect postprandial glucose excursions. The Day-to-day fluctuations in this parameter capture the influence of meal composition, stress, physical activity, and other contextual factors that are not explicitly modeled.

Additional parameters describe distribution volumes and observation dynamics. Insulin and glucose distribution volumes, $V_I$ and $V_G$, contribute to inter-individual variability in concentration dynamics, while the plasma–interstitial delay constant $\alpha$ governs the relationship between plasma glucose and CGM measurements. Although these parameters tend to be more stable within individuals, they influence the overall shape and timing of glucose trajectories and are therefore retained in the parameter set.

In general, the parameters extracted from ReplayBG provide a structured and physiologically meaningful summary of daily metabolic behavior. In the following sections, these daily parameter vectors are used as features to identify recurrent metabolic states through clustering and to model inter-day transitions between such states using Markov models.

For completeness, Table 2.1 summarizes all physiological and pharmacokinetic parameters estimated by ReplayBG and used in the present analysis, together with their physiological interpretation.

The personalization of the ReplayBG digital twin is performed using a Bayesian parameter identification procedure, which estimates a posterior distribution over physiological and pharmacokinetic parameters using real-world CGM, insulin and meal data. This probabilistic formulation explicitly accounts for the uncertainty arising from sensor noise, variability in insulin absorption, inaccuracies in carbohydrate reporting, and intra-day

Tabella 2.1: Physiological and pharmacokinetic parameters estimated by ReplayBG for each patient–day instance.

| **Parameter** | **Subsystem** | **Physiological meaning** |
|---|---|---|
| $G_b$ | Glucose regulation | Basal glucose concentration |
| $S_G$ | Glucose regulation | Glucose effectiveness (insulin-independent disposal) |
| $S_{I,B}$ | Insulin action | Morning insulin sensitivity |
| $S_{I,L}$ | Insulin action | Daytime insulin sensitivity |
| $S_{I,D}$ | Insulin action | Evening/night insulin sensitivity |
| $p_2$ | Insulin action | Insulin action rate constant |
| $k_d$ | Insulin absorption | Insulin dissociation rate from subcutaneous depot |
| $k_{a2}$ | Insulin absorption | Subcutaneous insulin absorption rate |
| $k_e$ | Insulin absorption | Plasma insulin clearance rate |
| $\beta$ | Insulin delivery | Insulin infusion delay |
| $k_{\text{empt}}$ | Glucose absorption | Gastric emptying rate |
| $k_{\text{abs},M}$ | Glucose absorption | Intestinal absorption rate per meal type |
| $V_I$ | Distribution | Insulin distribution volume |
| $V_G$ | Distribution | Glucose distribution volume |
| $\alpha$ | Observation model | Plasma–interstitial glucose delay (CGM) |

physiological fluctuations. While full posterior inference is obtained via Markov Chain Monte Carlo sampling, ReplayBG also supports faster calibration through Maximum A Posteriori estimation when uncertainty quantification is not required.

Once personalized, the digital twin can be used to simulate glucose dynamics under modified therapeutic or behavioral conditions. By altering insulin delivery profiles, meal characteristics, or control actions, ReplayBG generates counterfactual glucose trajectories together with uncertainty bounds, enabling systematic in-silico exploration of alternative treatment strategies. This replay capability supports both clinical interpretation and the development and testing of algorithmic solutions in a controlled and risk-free setting.

The ReplayBG framework has been validated in previous work on both synthetic and real-world datasets, demonstrating accurate reproduction of glucose responses under a range of perturbations and clinically relevant scenarios. In the present thesis, ReplayBG is used as a parameter-generating engine: instead of focusing on individual glucose trajectories, the analysis leverages the daily parameter sets estimated by the digital twin as compact descriptors of metabolic behavior. These descriptors form the basis for the inter-day modeling framework introduced in the following sections, where recurrent metabolic states and their temporal evolution are investigated.

## 2.2 Clustering and Definition of Parameter Combos

This section describes the methodology adopted to identify recurrent daily metabolic states from digital twin simulations. The approach consists of three main steps: (i) preprocessing and representation of daily parameter sets, (ii) reduction of dimensionality through the definition of physiologically motivated parameter combinations (combos), and (iii) unsupervised clustering of daily observations to obtain discrete metabolic states. These states form the basis for the temporal analysis developed in the subsequent Markov modeling framework.

### 2.2.1 Data processing and daily representation

Each day simulated is summarized by the set of physiological and pharmacokinetic parameters estimated through the ReplayBG digital twin. These parameters describe the latent metabolic state of the patient for a given day and constitute the input for the clustering analysis. Before clustering, a preprocessing pipeline is applied to ensure numerical stability, physiological plausibility, and comparability between parameters.
Occasional missing values may arise due to convergence issues in parameter estimation or incomplete data segments. Missing entries are handled either through population-based imputation, using the median value of the parameter across all available days, or through forward/backward filling when the parameter is expected to vary slowly across consecutive days. Parameters falling outside physiologically plausible ranges are excluded or clipped based on domain knowledge, in order to prevent biologically unrealistic values from biasing the clustering results.
Since the clustering algorithm relies on distance-based similarity measures, it is essential that all parameters contribute comparably to the distance computation. Therefore, each parameter $p_k$ is standardized in the data set according to

$$\tilde{p}_k = \frac{p_k - \mu_k}{\sigma_k},$$

where $\mu_k$ and $\sigma_k$ denote the mean and standard deviation of the parameter $k$ calculated on all days. This transformation removes scale effects and prevents parameters with larger numerical ranges from dominating the clustering.
After preprocessing and standardization, each day $d$ is represented as a feature vector

$$\mathbf{v}_d = [\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_K] \in \mathbb{R}^K,$$

where $K$ is the number of parameters included in the combination of selected parameters. This representation compresses the full intra-day glucose dynamics into a compact and interpretable description of the underlying metabolic configuration estimated for that day. By representing each day as a single point in parameter space, the analysis shifts from time-series data to a state-based description, enabling the identification of recurrent daily

metabolic states and supporting subsequent temporal modeling across days.

### 2.2.2 Parameter dimensionality and definition of combos

The complete set of ReplayBG parameters is high-dimensional and includes quantities associated with different physiological subsystems, such as glucose regulation, insulin absorption, gastric emptying, and circadian insulin sensitivity. Clustering the full parameter space directly would reduce interpretability, increase sensitivity to noise, and exacerbate issues related to scale heterogeneity and the curse of dimensionality.

For this reason, parameters are grouped into physiologically motivated subsets, termed *combos*. Each combo is designed to isolate a specific metabolic mechanism or a controlled interaction between mechanisms, allowing clustering to focus on coherent physiological processes. Clustering is performed independently for each combo, enabling the identification of recurrent daily metabolic states associated with distinct aspects of glucose regulation.

In the following, the nine parameter combos used in this work are described together with their physiological interpretation.

**Combo 1: Core glucose–insulin regulatory dynamics**

$$\{Gb,\ SG,\ ka2,\ kd,\ kempt\}$$

This combination includes parameters that directly shape baseline glucose regulation and the immediate response to insulin and meals. Basal glucose ($Gb$) and glucose efficiency ($SG$) characterize insulin-independent glucose absorption, while $ka2$ and $kd$ describe the speed of subcutaneous insulin absorption. The gastric emptying rate ($kempt$) governs the appearance of glucose after meals. Together, these parameters determine the overall temporal profile of postprandial glucose excursions.

**Combo 2: Circadian insulin sensitivity**

$$\{SI_B,\ SI_L,\ SI_D\}$$

This set captures variations in insulin sensitivity during the day, reflecting circadian and behavioral influences on insulin action. By isolating morning, midday, and evening sensitivity, this combo allows identification of recurrent states driven by daily rhythms rather than acute events.

**Combo 3: Insulin absorption shape parameters**

$$\{kabs_B,\ \beta_B,\ kabs_L,\ \beta_L,\ kabs_S,\ \beta_S,\ kabs_D,\ \beta_D\}$$

This combo focuses on the shape and timing of insulin absorption associated with different meals. The $kabs_*$ parameters control the speed of absorption, while the corresponding $\beta_*$ parameters influence its temporal profile. This combination highlights variability in insulin pharmacokinetics between meals and days.

**Combo 4: Insulin absorption rates only**

$$\{kabs_B,\ kabs_L,\ kabs_S,\ kabs_D\}$$

This reduced version of Combo 3 isolates absorption rates while ignoring shape parameters. It allows a review of whether clustering is primarily driven by differences in insulin uptake speed rather than absorption profile.

**Combo 5: Full physiological parameter set**

$$\{Gb,\ SG,\ ka2,\ kd,\ kempt,\ SI_B,\ SI_L,\ SI_D,\ kabs_B,\ kabs_L,\ kabs_S,\ kabs_D\}$$

This combo combines glucose regulation, insulin absorption, gastric emptying, and circadian insulin sensitivity into a single comprehensive representation. Capture global metabolic behavior and allow for identification of states emerging from interactions across multiple subsystems.

**Combo 6: Glucose regulation, insulin sensitivity, and absorption shape**

$$\{Gb,\ kd,\ kempt,\ SI_B,\ SI_L,\ SI_D,\ \beta_B,\ \beta_L,\ \beta_S,\ \beta_D\}$$

This set emphasizes interactions between baseline glucose dynamics, insulin sensitivity that changes over time, and insulin absorption shape. It is designed to highlight metabolic states driven by the combined effects of variability in meal absorption and circadian insulin responsiveness.

**Combo 7: Extended full parameter set**

$$\{Gb,\ SG,\ ka2,\ kd,\ kempt,\ SI_B,\ SI_L,\ SI_D,\ kabs_B,\ \beta_B,\ kabs_L,\ \beta_L,\ kabs_S,\ \beta_S,\ kabs_D,\ \beta_D\}$$

This is the most comprehensive combination, including all the main physiological and pharmacokinetic parameters estimated by ReplayBG. It captures nearly the entire metabolic model and enables clustering based on fully integrated glucose–insulin dynamics.

**Combo 8: Glucose regulation and insulin sensitivity**

$$\{Gb,\ SG,\ ka2,\ kd,\ kempt,\ SI_B,\ SI_L,\ SI_D\}$$

By excluding insulin absorption shape parameters, this combo focuses on intrinsic physiological regulation of glycemia, highlighting states driven by insulin effectiveness and sensitivity rather than pharmacokinetic variability.

**Combo 9: Insulin absorption and insulin sensitivity**

$$\{kabs_B,\, kabs_L,\, kabs_S,\, kabs_D,\, SI_B,\, SI_L,\, SI_D\}$$

This final combination isolates the interaction between insulin pharmacokinetics and circadian insulin response. It is particularly suited to identify states associated with variability in insulin action timing and magnitude across the day.
In general, the design of these combos balances physiological interpretability with numerical robustness. Smaller, mechanism-specific combos facilitate clear interpretation of cluster structure, while larger combos allow exploration of more complex, multi-mechanism metabolic patterns. Each combo is clustered independently, generating distinct state sequences that are subsequently used for temporal analysis and Markov modeling.

### 2.2.3 Clustering methodology

For each parameter combo, daily feature vectors are clustered using the K-means algorithm. The K-means partition the data set into $C$ clusters by minimizing the sum within the cluster of squared Euclidean distances:

$$J = \sum_{i=1}^{N} \sum_{c=1}^{C} z_{ic} \left\| \mathbf{x}^{(i)} - \boldsymbol{\mu}_c \right\|_2^2,$$

where $\boldsymbol{\mu}_c$ denotes the centroid of the cluster $c$.
The algorithm alternates between the assignment and centroid update steps until convergence and is initialized multiple times to mitigate sensitivity to local minima.
The number of clusters $C$ is selected using a combination of the elbow method and the silhouette score. The silhouette score for a given day $i$ is defined as

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))},$$

where $a(i)$ is the average distance to points in the same cluster and $b(i)$ is the minimum average distance to points in other clusters. Values close to 1 indicate well-separated clusters. The final choice of $C$ maximizes the average silhouette score while preserving interpretability.

### 2.2.4 Cluster centroids and super-patient representation

Each cluster is represented by its centroid in parameter space, which can be interpreted as a *superpatient* describing the average metabolic configuration of all days assigned to

that cluster. The super-patient provides a compact and interpretable summary of the physiological characteristics associated with a given metabolic state.

This representation is particularly useful for subsequent Markov modeling, as transitions between clusters correspond to transitions between super-patients, enabling the study of inter-day metabolic evolution at a higher level of abstraction.

### 2.2.5 Clinical characterization of clusters

Although clustering is performed exclusively on model-derived physiological parameters, each identified cluster is subsequently characterized using clinically relevant glycemic metrics computed from the corresponding CGM data. This post-hoc analysis is essential to relate the abstract metabolic states emerging from the clustering to clinically interpretable indicators of glycemic control.

In particular, three standard metrics derived from continuous glucose monitoring are considered: Time in Range (TIR), Time Above Range (TAR), and Time Below Range (TBR). These metrics quantify the proportion of time during which glucose concentrations fall within, above, or below clinically defined target ranges, respectively, and are widely used in both clinical practice and research to assess glycemic control.

Let $G(t)$ denote the CGM glucose signal sampled on a given day, and let $T$ be the total observation time. The Time in Range is defined as

$$\mathrm{TIR} = \frac{1}{T} \int_0^T \mathbb{I}\Big(70 \leq G(t) \leq 180\Big)\, dt,$$

where $\mathbb{I}(\cdot)$ is the indicator function. TIR represents the fraction of time spent within the recommended target range of 70–180 mg/dL and is strongly associated with a reduced risk of long-term complications.

Time Above Range quantifies exposure to hyperglycemia and is defined as

$$\mathrm{TAR} = \frac{1}{T} \int_0^T \mathbb{I}\Big(G(t) > 180\Big)\, dt,$$

while Time Below Range captures hypoglycemic exposure and is defined as

$$\mathrm{TBR} = \frac{1}{T} \int_0^T \mathbb{I}\Big(G(t) < 70\Big)\, dt.$$

These two metrics provide complementary information, as elevated TAR is associated with chronic hyperglycemia and an increased risk of microvascular complications, while elevated TBR reflects the risk of hypoglycemia, which is particularly relevant for patient safety.

For each cluster, TIR, TAR, and TBR are computed for all days assigned to that cluster and summarized using descriptive statistics (e.g. mean and dispersion). This allows clusters to be interpreted in clinically meaningful terms. For example, clusters characterized by high TIR and low TBR can correspond to stable and well-controlled metabolic states,

whereas clusters with elevated TAR or TBR can reflect states associated with impaired insulin sensitivity, delayed absorption, or increased variability.

Importantly, this clinical characterization does not influence the clustering itself, but serves to contextualize and validate the identified metabolic states. By linking parameter-based clusters to CGM-derived outcomes, the analysis bridges the gap between physiological modeling and real-world clinical relevance, facilitating interpretation of cluster transitions in the subsequent Markov modeling framework.

## 2.3 Markov Modeling of Multi-Day Metabolic Dynamics

Once each simulated day is assigned to a discrete metabolic state through clustering, it becomes possible to study how these states evolve over time using probabilistic temporal models. In this work, the inter-day evolution of metabolic states is modeled using discrete-time, first-order Markov chains. This framework provides a compact, interpretable, and patient-specific representation of how individuals transition between different metabolic regimes across days, and forms the basis for generating synthetic multi-day digital twin trajectories.

### 2.3.1 From clustered days to discrete-time Markov chains

For a given parameter combo and patient, clustering produces a sequence of discrete labels

$$X_1, X_2, \ldots, X_T \in \{0, 1, \ldots, K-1\},$$

where $X_t$ denotes the metabolic state assigned to day $t$ and $K$ is the total number of clusters identified for this combination. This sequence represents a temporal realization of the patient's daily metabolic evolution.

Two different constructions of the state sequence are considered. In the *all-days* setting, the full ordered sequence of available days is retained, regardless of possible gaps in the calendar. In the *adjacent-days* setting, only transitions between strictly consecutive days are considered, discarding transitions separated by longer temporal gaps. The latter formulation isolates true day-to-day dynamics, while the former captures longer-term trends across the observation period.

### 2.3.2 Markov property and transition matrix

The modeling assumption underlying this approach is the first-order Markov property:

$$\mathbb{P}(X_{t+1} = j \mid X_t = i, X_{t-1}, \ldots, X_1) = \mathbb{P}(X_{t+1} = j \mid X_t = i), \tag{2.5}$$

which states that the probability of transitioning to the next-day metabolic state depends only on the current state and not on the full history.

The one-step transition probabilities are collected in a $K \times K$ transition matrix

$$\mathbf{P} = [p_{ij}], \qquad p_{ij} = \mathbb{P}(X_{t+1} = j \mid X_t = i),$$

with the constraints

$$p_{ij} \geq 0, \qquad \sum_{j=0}^{K-1} p_{ij} = 1.$$

A separate transition matrix is estimated for each patient, each combination of parameters, and for both sequence constructions described above.
The adoption of a first-order Markov chain implies several modeling assumptions that are appropriate for the scope of this work. First, it is assumed that the metabolic state the next-day depends only on the current state and not on the full history of the previous days. This choice reflects a balance between model simplicity, interpretability, and the limited length of available patient trajectories.
Second, the transition process is assumed to be time-homogeneous, meaning that the transition probabilities encoded in the matrix $\mathbf{P}$ are constant over the observation window. While physiological processes may evolve over longer time scales, this assumption is reasonable within the temporal resolution and duration considered in this study.
Finally, metabolic states are treated as discrete and mutually exclusive, with each day assigned to a single cluster. This discretization enables a compact representation of inter-day dynamics and facilitates probabilistic modeling, at the cost of neglecting within-day variability, which is instead captured at the parameter estimation stage.

### 2.3.3 Estimation of transition probabilities

Given an observed sequence of states,

$$x_1, x_2, \ldots, x_T,$$

the number of observed transitions from state $i$ to state $j$ is defined as

$$N_{ij} = \#\{t \in \{1, \ldots, T-1\} : x_t = i,\ x_{t+1} = j\}.$$

The collection of all $N_{ij}$ forms the transition count matrix $\mathbf{N}$. Under the assumption of a first-order Markov process, the maximum likelihood estimate of the transition probabilities is obtained by row-wise normalization.

$$\hat{p}_{ij} = \frac{N_{ij}}{\sum_{m=0}^{K-1} N_{im}}.$$

Rows corresponding to states with no observed outgoing transitions are handled by appropriate regularization or exclusion from subsequent analysis.

### 2.3.4 Next-day prediction and multi-day simulation

Once the transition matrix $\mathbf{P}$ has been estimated, it can be used both for prediction and simulation. Given the current metabolic state $X_t = i$, the distribution of the next-day state is given by the $i$-th row of $\mathbf{P}$:

$$\mathbb{P}(X_{t+1} = j) = p_{ij}.$$

Repeated sampling from these transition probabilities allows the generation of synthetic state trajectories

$$X_0, X_1, \ldots, X_T,$$

which represent hypothetical multi-day metabolic evolutions consistent with the dynamics learned from data.

These simulated state sequences are subsequently mapped to representative digital twin parameters by associating each state with its corresponding cluster centroid (super-patient). In this way, the Markov model enables the generation of coherent multi-day digital twin simulations without explicitly integrating the underlying physiological model over long horizons.

### 2.3.5 Train–test split and chi-square comparison

To assess the robustness and temporal consistency of the estimated Markov dynamics, a train–test validation procedure is applied independently for each patient and parameter combo. The observed sequence of daily metabolic states is divided into two segments: a training segment comprising the first 80% of available days and a test segment containing the remaining 20%.

Separate transition count matrices and transition probability matrices are estimated from the two segments:

$$\mathbf{N}^{(\text{train})}, \quad \mathbf{N}^{(\text{test})}, \qquad \mathbf{P}^{(\text{train})}, \quad \mathbf{P}^{(\text{test})}.$$

This separation allows for evaluation of whether the transition structure learned from the training data generalizes to unseen observations.

To quantitatively compare the dynamics observed in the training and test segments, a row-wise chi-square goodness-of-fit test is performed on the transition counts. For each state $i$, the transition counts observed in the test segment, denoted by $N_{i\cdot}^{(\text{test})}$, are compared to the expected counts derived from the training distribution. Specifically, the expected number of transitions from state $i$ to state $j$ in the test segment is computed as

$$E_{ij} = N_{ij}^{(\text{train})} \cdot \frac{\sum_m N_{im}^{(\text{test})}}{\sum_m N_{im}^{(\text{train})}},$$

which corresponds to re-scaling the training transition frequencies to match the total number of outgoing transitions observed in the test data.

A chi-square statistic and the corresponding $p$-value are computed for each state by comparing observed and expected counts across outgoing transitions. These values are then aggregated by reporting their mean across all states,

$$\chi^2_{\text{mean}}, \qquad p_{\text{mean}}.$$

High average $p$-values indicate that the transition patterns observed in the test segment

are statistically consistent with those learned during training, supporting the stability and generalizability of the estimated Markov model. In contrast, low $p$-values suggest temporal variability in the transition dynamics or insufficient data to reliably estimate transition probabilities for certain states.

### 2.3.6 Model validation: implied timescales and Chapman–Kolmogorov test

Beyond assessing consistency between training and test data, it is essential to validate whether the evolution of daily metabolic states can be reasonably approximated as a Markov process at the chosen temporal resolution. In particular, a key question is whether the clustering-based state representation satisfies the Markov property for transitions observed at a given time lag. To address this issue, validation tools from the Markov State Model (MSM) literature are employed, namely the analysis of implied timescales and the Chapman–Kolmogorov (CK) test.

**Lag time and Markovianity.** A central concept in MSM-based modeling is the *lag time* $\tau$, which represents the temporal spacing between successive observations used to estimate the probabilities of transition. In the present context, $\tau$ is measured in days and corresponds to the number of days that separate consecutive states in the Markov chain. The choice of $\tau$ is critical. If $\tau$ is too short, the system may retain the memory of previous states beyond the immediately preceding one, violating the Markov assumption. In such cases, the observed transitions are influenced by unresolved fast dynamics or incomplete relaxation within clusters. Conversely, if $\tau$ is sufficiently long, these fast processes are effectively averaged out, and the dynamics between states can be approximated as memoryless. However, excessively large values of $\tau$ reduce the number of usable transitions and degrade statistical reliability. Therefore, model validation aims to identify a temporal regime in which the Markov approximation is both valid and data-efficient.

**Implied timescales.** The implied timescale analysis provides a quantitative criterion to assess Markovianity as a function of the lag time. For a given $\tau$, a transition matrix $\mathbf{P}(\tau)$ is estimated and its eigenvalues $\lambda_i(\tau)$ are calculated. Each non-trivial eigenvalue corresponds to a relaxation process in the system, and the associated implied timescale is defined as [15, 9]

$$t_i(\tau) = -\frac{\tau}{\ln \lambda_i(\tau)}.$$

Intuitively, the implied timescale represents the characteristic time over which a slow dynamical process is correlated. If the dynamics are Markovian at lag time $\tau$, the implied timescales should be intrinsic properties of the system and therefore remain approximately constant when $\tau$ is varied. In contrast, a strong dependence of $t_i(\tau)$ on $\tau$ indicates residual memory effects and insufficient state separation.

In practice, implied timescales are computed for increasing values of $\tau$ (e.g. $\tau = 1, 5, 10, 20, 30$ days). The presence of a plateau region, where the time scales exhibit limited variation with $\tau$, supports the validity of the Markov approximation at those temporal resolutions. The early values of $\tau$ that show significant drift are interpreted to belong to a non-Markovian regime and are effectively discarded. This procedure allows for the identification of a minimal lag time beyond which the clustered states behave as approximately memoryless.

**Chapman–Kolmogorov test.** While implied timescales assess Markovianity in the spectral domain, the Chapman–Kolmogorov test provides complementary validation in the time domain. For a Markov process, the transition matrix must satisfy the Chapman–Kolmogorov consistency condition:

$$\mathbf{P}(k\tau) \approx \mathbf{P}(\tau)^k,$$

implying that the probability of transitioning between states over $k$ lag steps should be reproducible by successive applications of the one-step transition matrix [13, 15].
In the CK test, multi-step transition probabilities are estimated directly from the data for several multiples of $\tau$ and compared against the corresponding predictions obtained from $\mathbf{P}(\tau)^k$. Good agreement between empirical and predicted transition probabilities indicates that the Markov chain provides a self-consistent description of the observed dynamics at the chosen lag time.
Discrepancies between the two suggest either insufficient sampling, inadequate state discretization, or the presence of memory effects not captured by a first-order model. As with implied timescales, the CK test is typically satisfied only beyond a certain minimal lag time, reinforcing the notion that early transitions may need to be excluded to obtain a reliable Markovian representation.

**Role in the present work.** Together, implied timescale analysis and the Chapman–Kolmogorov test provide a principled validation of the Markov modeling framework adopted in this thesis. They ensure that the discrete metabolic states identified through clustering evolve in a manner that is consistent with the assumptions of a first-order, time-homogeneous Markov process at the selected temporal resolution. This validation step is crucial to justify the use of the estimated transition matrices for next-day prediction, long-term simulation, and analysis of inter-day metabolic dynamics.

# Capitolo 3

# Results

## 3.1 Dataset overview and preprocessing outcome

This section provides an overview of the parameters derived from digital twins used in the analysis and summarizes their statistical properties prior to clustering. The goal is to characterize the variability and scale of the parameters and to motivate the preprocessing steps applied before unsupervised learning.

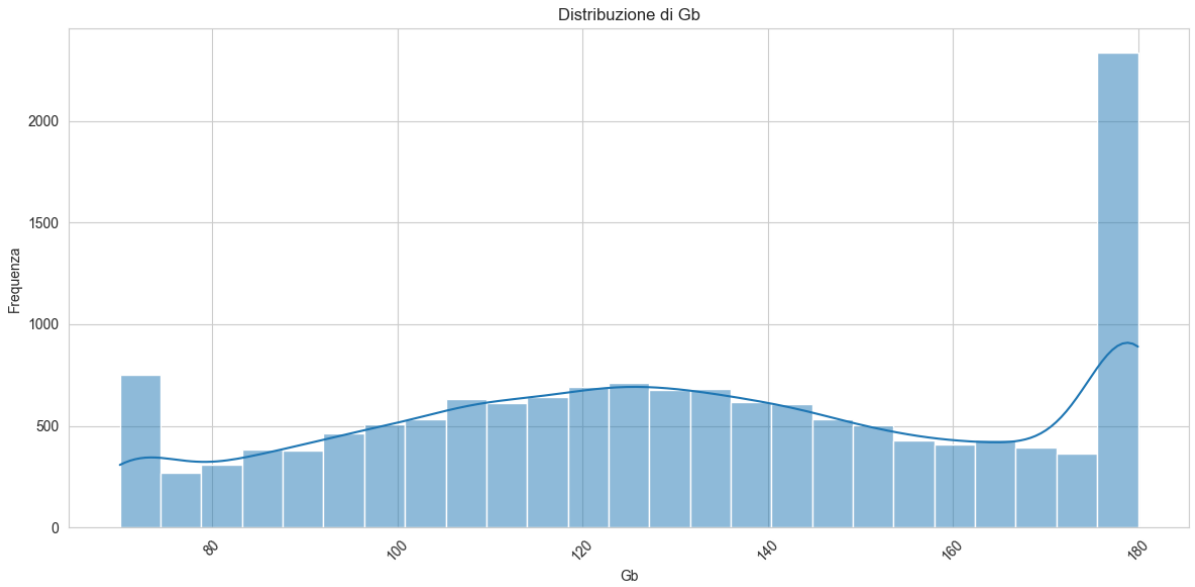


Figura 3.1: Marginal distribution of basal glucose concentration ($Gb$) across all simulated patient–day instances. The distribution summarizes inter-individual and inter-day variability in baseline glycaemic conditions captured by the digital twin fitting process.

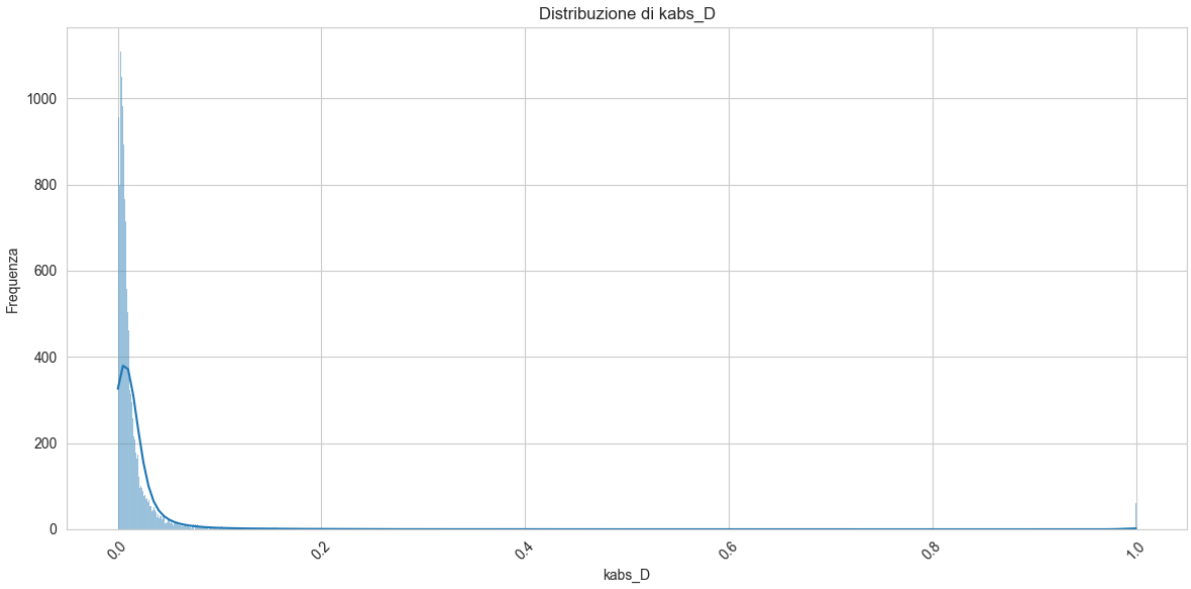


Figura 3.2: Marginal distribution of the insulin absorption rate parameter ($kabs_S$) across all simulated patient–day instances. The strong right-skewness and long tail reflect substantial inter-individual and inter-day variability in subcutaneous insulin absorption kinetics.

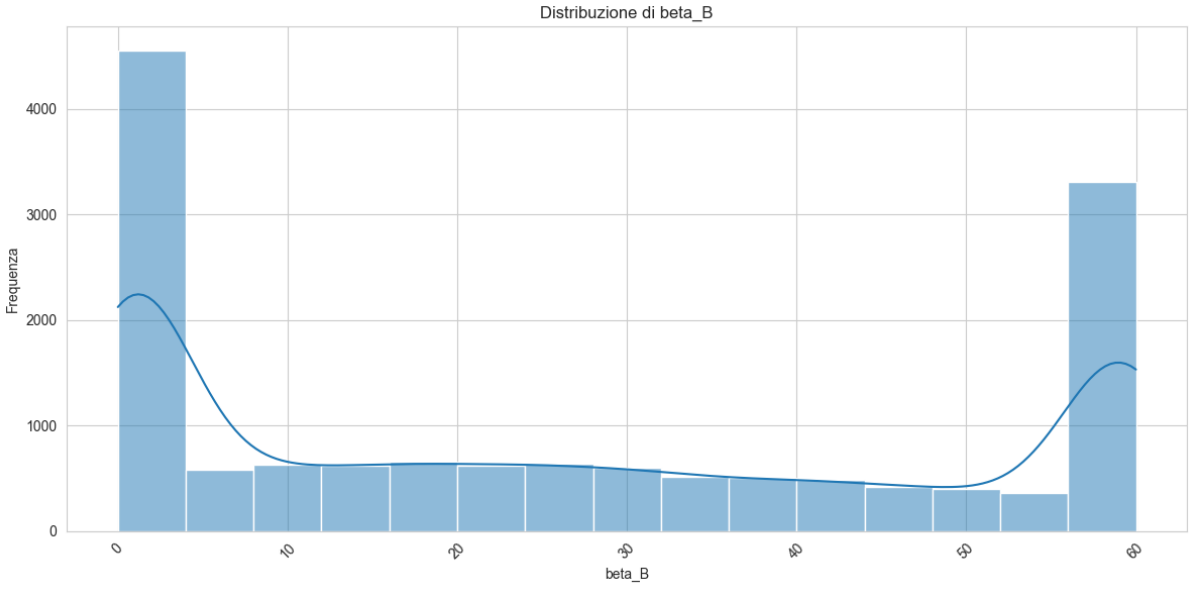


Figura 3.3: Marginal distribution of the insulin absorption shape parameter ($\beta_B$) across all simulated patient–day instances. The multimodal distribution suggests the presence of distinct kinetic regimes governing the temporal profile of insulin absorption.

## 3.2 Clustering results

For each combo, the number of clusters k is not fixed a priori but is chosen based on silhouette analysis and visualized in line plots (see Fig. 3.4), allowing for an explicit inspection of the trade-off between model complexity and separation quality.

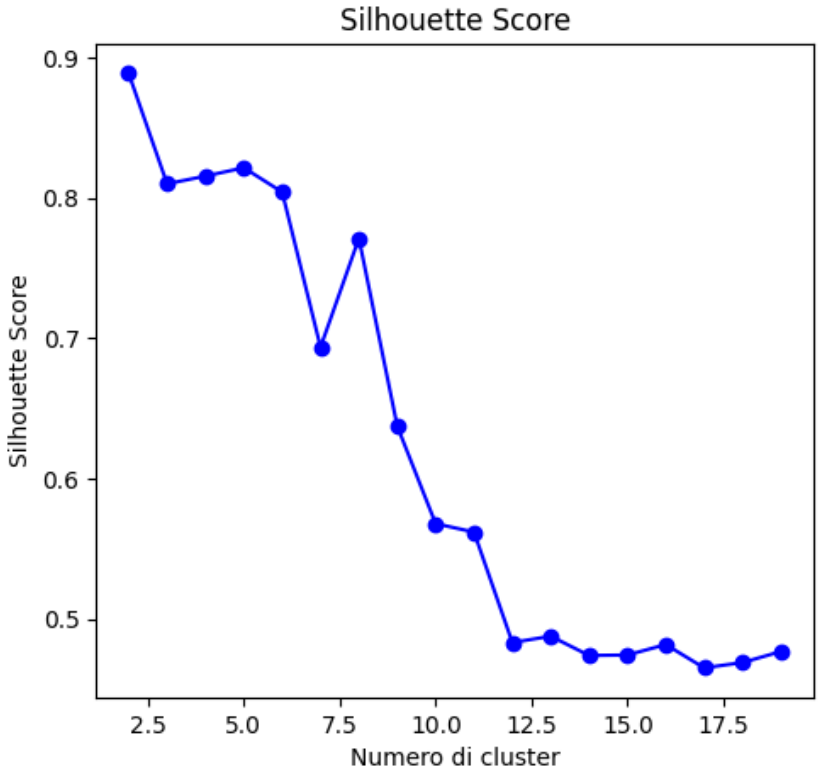


Figura 3.4: Example of silhouette score as a function of the number of clusters for a given combo. The selected $k$ corresponds to a relatively high silhouette score and acceptable within-cluster compactness, while avoiding very small clusters.

The clustering results obtained for each parameter combination are reported in Table 3.1.

Tabella 3.1: Summary of clustering configurations for each parameter combo.

| **Combo** | **Parameters in the combo** | **# Clusters** | **Silhouette mean** |
|---|---|---|---|
| 1 | $\{Gb, SG, ka2, kd, kempt\}$ | 5 | 0.336 |
| 2 | $\{SI_B, SI_L, SI_D\}$ | 3 | 0.857 |
| 3 | $\{kabs_B, \beta_B, kabs_L, \beta_L, kabs_S, \beta_S, kabs_D, \beta_D\}$ | 2 | 0.80 |
| 4 | $\{kabs_B, kabs_L, kabs_S, kabs_D\}$ | 2 | 0.92 |
| 5 | $\{Gb, SG, ka2, kd, kempt, SI_B, SI_L, SI_D, kabs_B, kabs_L, kabs_S, kabs_D\}$ | 14 | 0.245 |
| 6 | $\{Gb, kd, kempt, SI_B, SI_L, SI_D, \beta_B, \beta_L, \beta_S, \beta_D\}$ | 7 | 0.134 |
| 7 | $\{Gb, SG, ka2, kd, kempt, SI_B, SI_L, SI_D, kabs_B, \beta_B, kabs_L, \beta_L, kabs_S, \beta_S, kabs_D, \beta_D\}$ | 6 | 0.245 |
| 8 | $\{Gb, SG, ka2, kd, kempt, SI_B, SI_L, SI_D\}$ | 8 | 0.27 |
| 9 | $\{kabs_B, kabs_L, kabs_S, kabs_D, SI_B, SI_L, SI_D\}$ | 3 | 0.81 |

To provide an intuitive visualization of cluster separation in a low-dimensional space, a two-dimensional PCA projection is reported for selected parameter combinations. Each point represents a patient–day instance, colored according to the assigned cluster label.

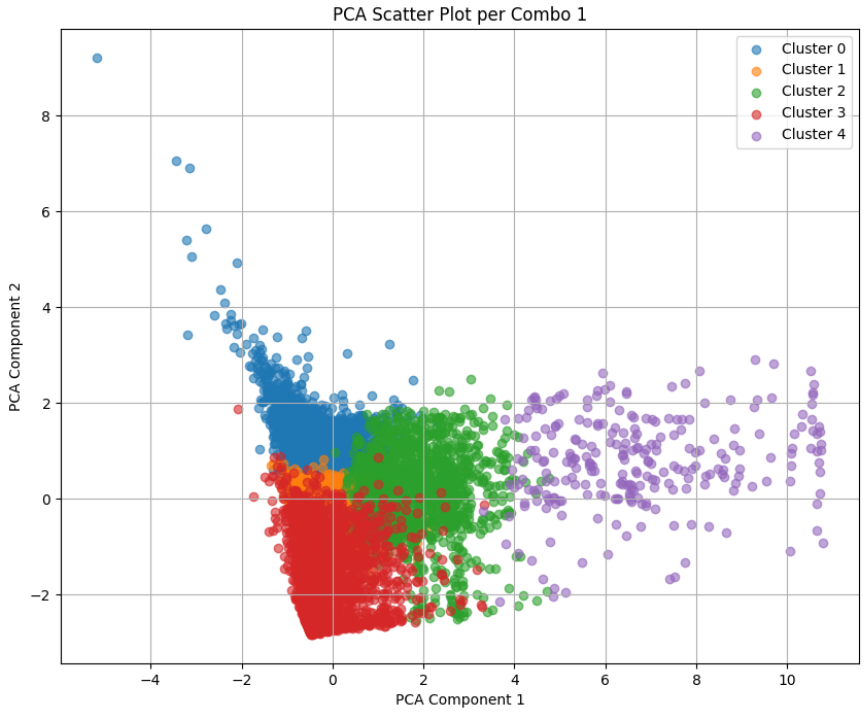


Figura 3.5: PCA projection of clustered days for Combo 1 ({Gb, SG, ka2, kd, kempt}).

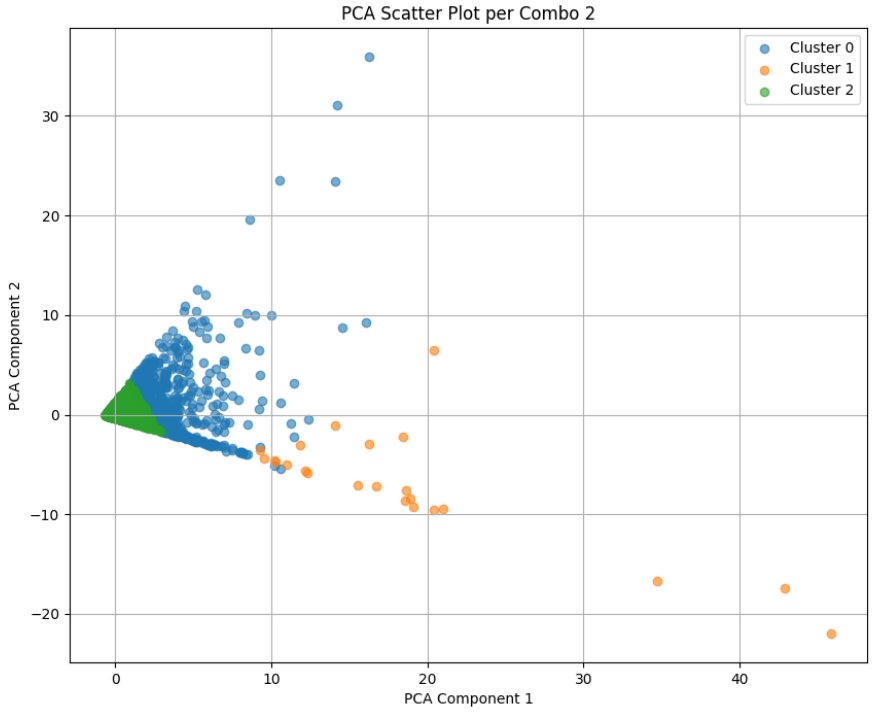


Figura 3.6: PCA projection of clustered days for Combo 2 ({SI_B, SI_L, SI_D}).

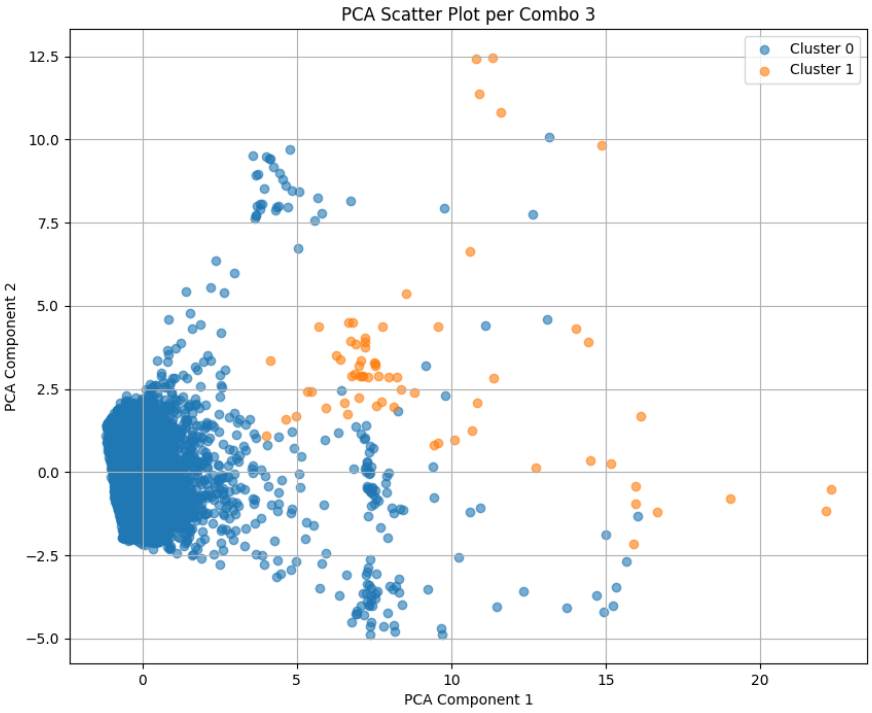


Figura 3.7: PCA projection of clustered days for Combo 3 ({kabs_B, $\beta_B$, kabs_L, $\beta_L$, kabs_S, $\beta_S$, kabs_D, $\beta_D$}).

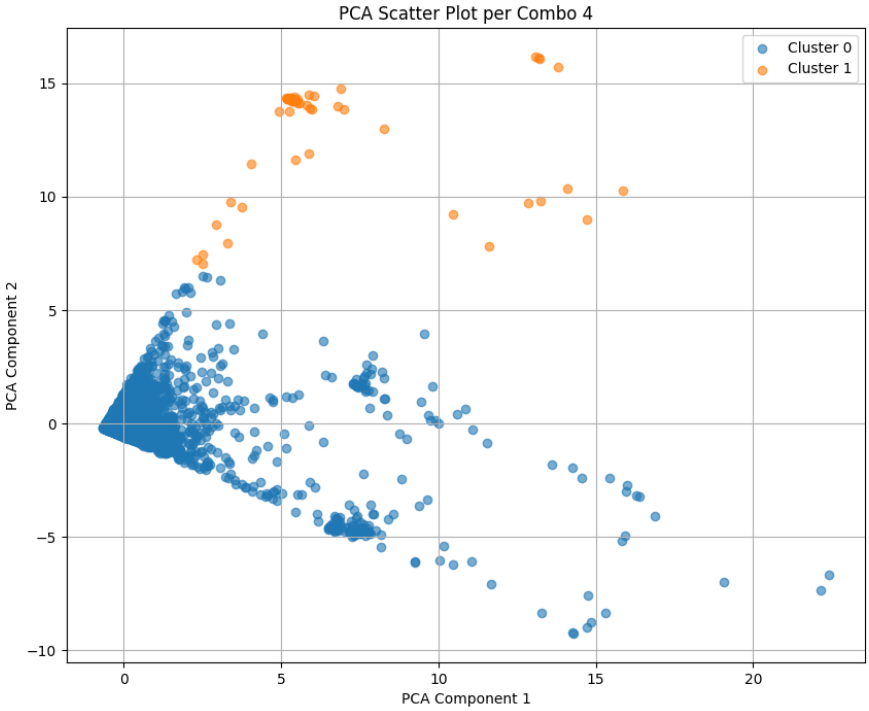


Figura 3.8: PCA projection of clustered days for Combo 4 ({kabs_B, kabs_L, kabs_S, kabs_D}).

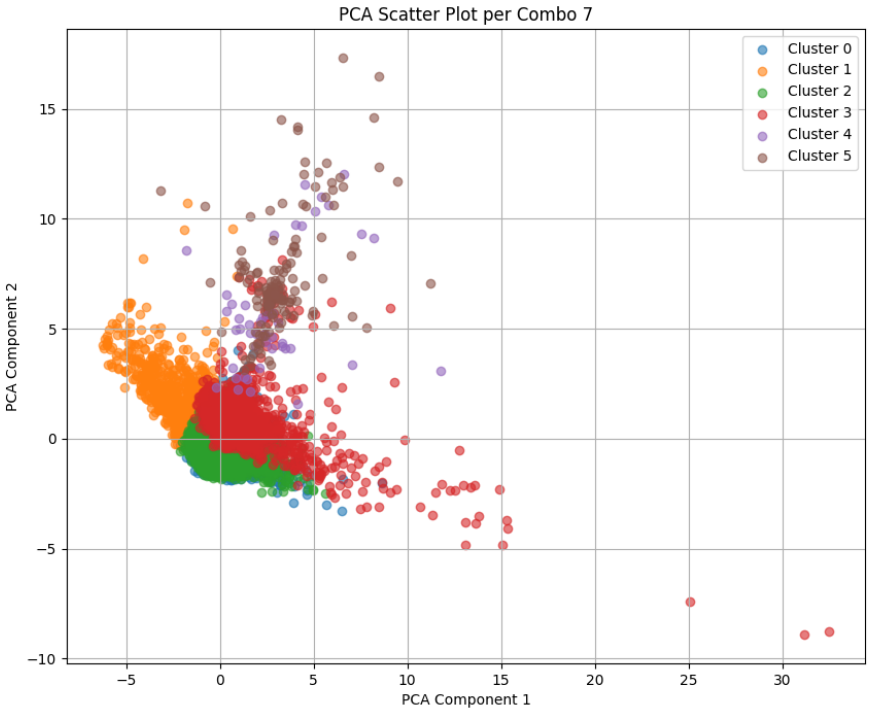


Figura 3.9: PCA projection of clustered days for Combo 7 (full parameter set including absorption shape parameters).

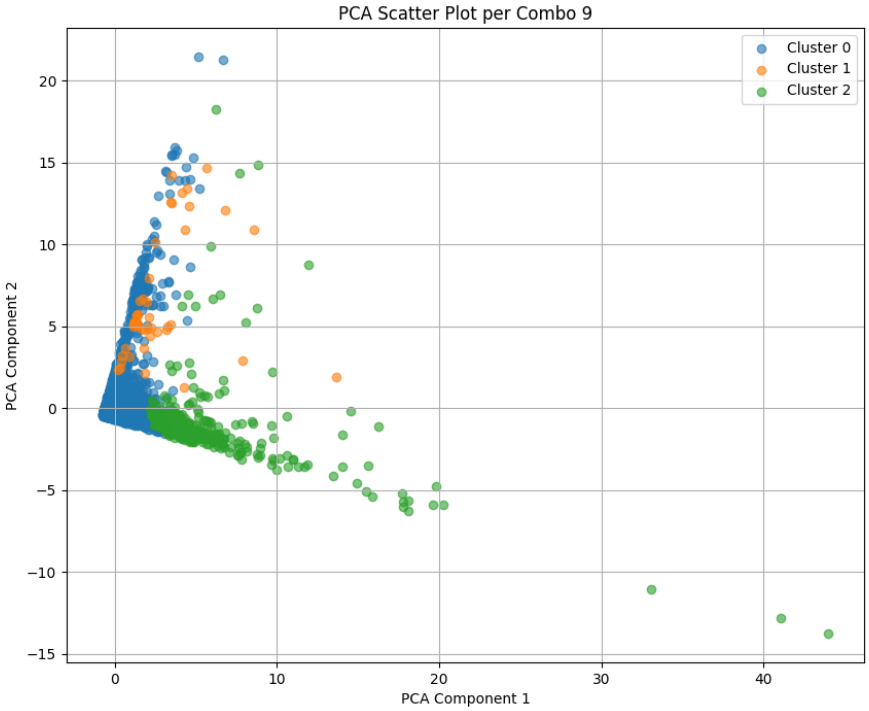


Figura 3.10: PCA projection of clustered days for Combo 9 ({kabs_B, kabs_L, kabs_S, kabs_D, SI_B, SI_L, SI_D}).

For low-dimensional combinations, pairwise scatter plots are additionally reported to directly visualize the separation between clusters in the original parameter space.

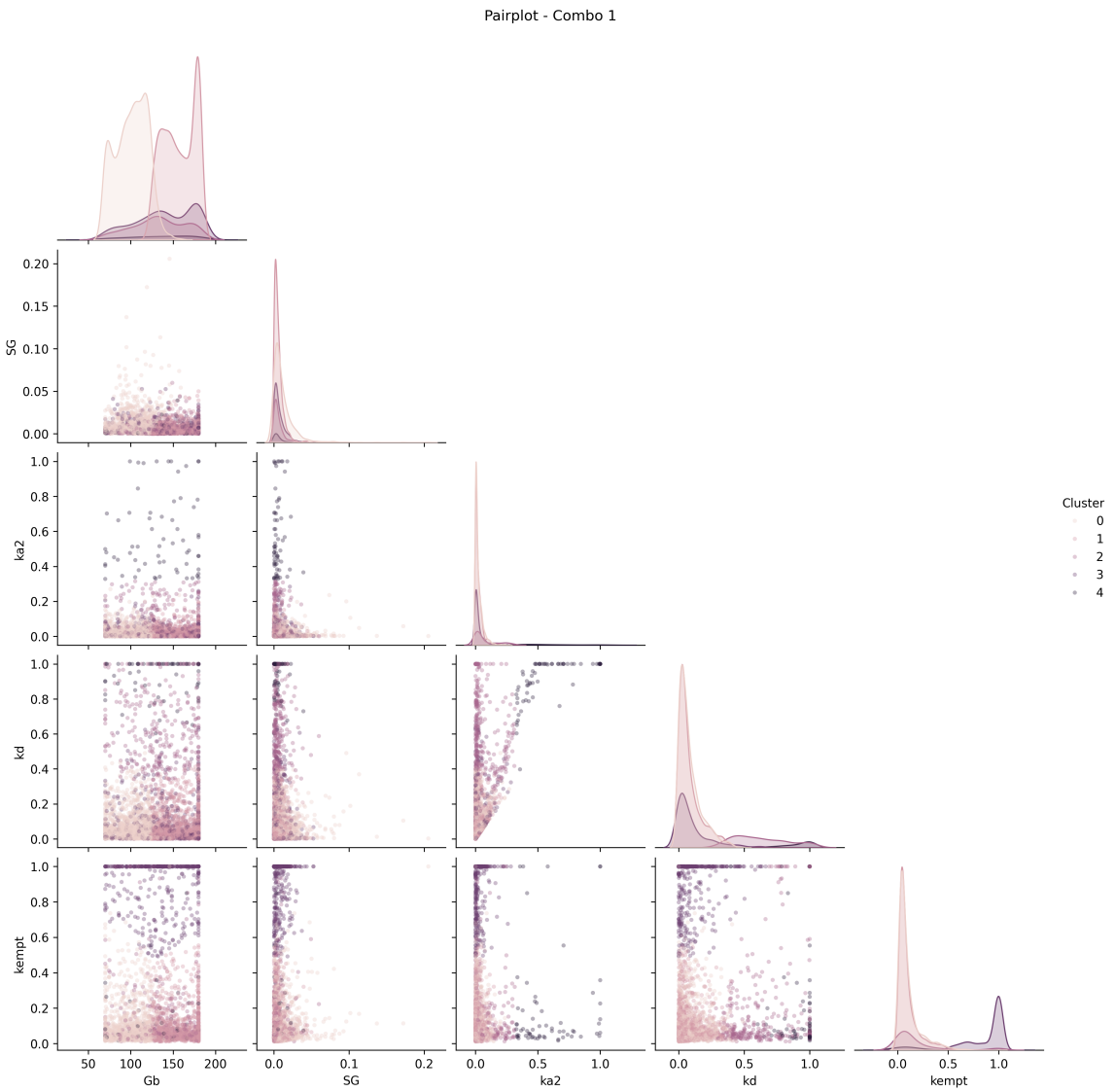


Figura 3.11: Pairwise scatter plot for Combo 1 ({Gb, SG, ka2, kd, kempt}), colored by cluster membership.

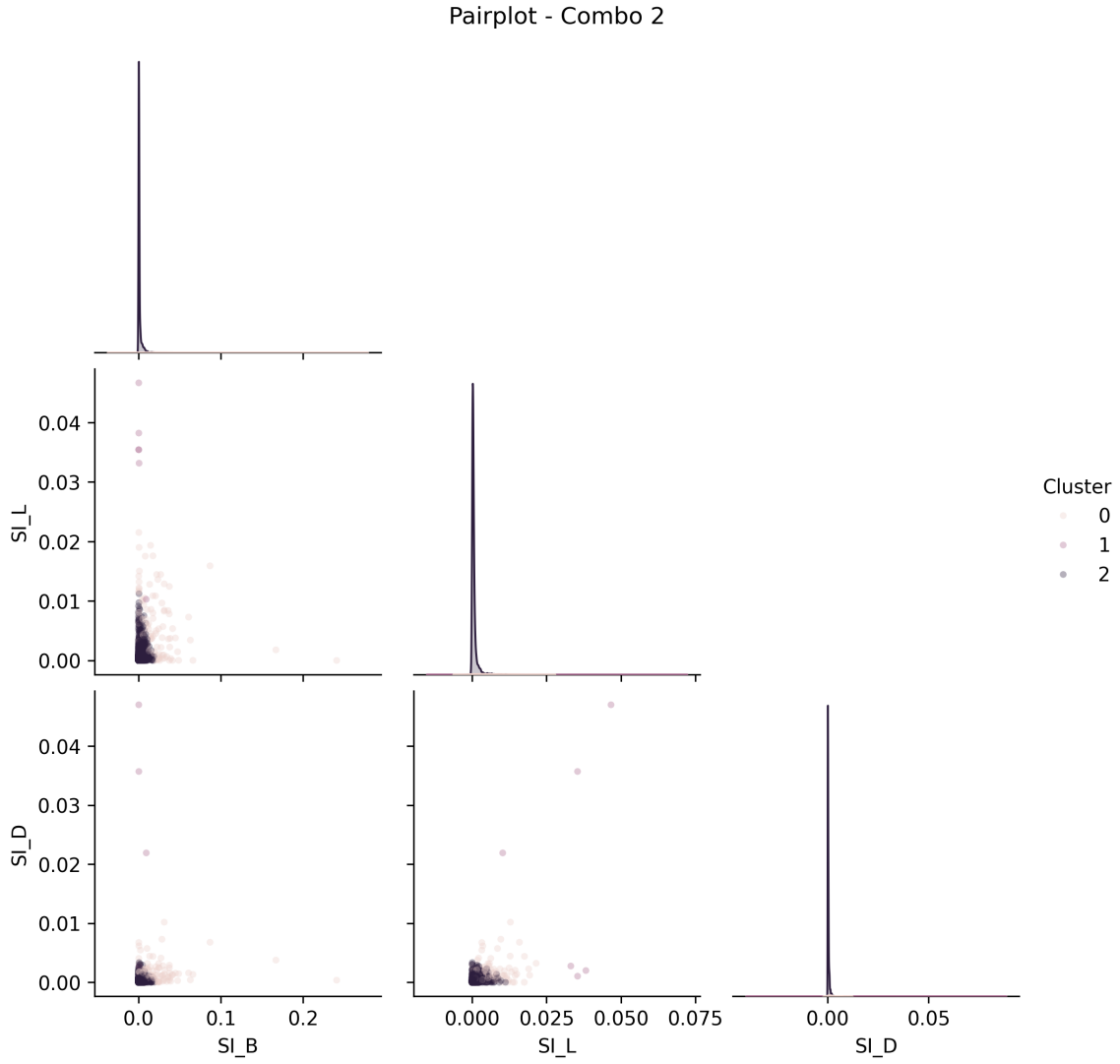


Figura 3.12: Pairwise scatter plot for Combo 2 ({SI_B, SI_L, SI_D}), colored by cluster membership.

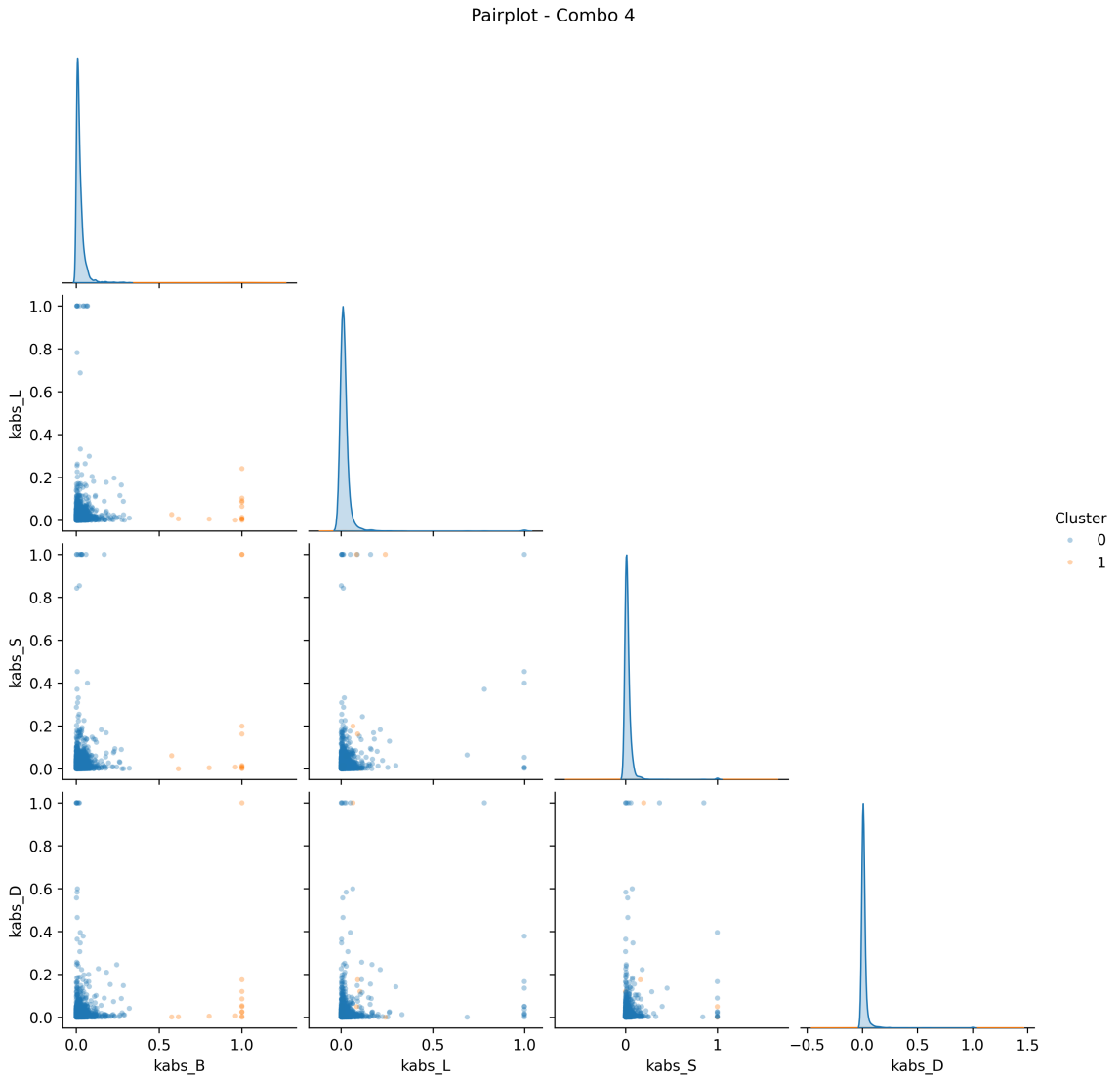


Figura 3.13: Pairwise scatter plot for Combo 4 ({kabs_B, kabs_L, kabs_S, kabs_D}), colored by cluster membership.

## 3.3 Characterization of clusters

Although clustering was performed exclusively on the ReplayBG parameter space, each identified cluster was subsequently characterized using clinically meaningful CGM-derived metrics. In particular, for each patient day instance assigned to a given cluster, we computed TIR, TAR, and TBR. This post-hoc characterization allows the clusters to be interpreted not only as latent parameter regimes, but also in relation to glycemic control quality and risk profiles.

Tabella 3.2: Clinical characterization of clusters for Combo 1 using TIR, TAR and TBR.

| Cluster | Days | TIR (%) | TAR (%) | TBR (%) |
|---|---|---|---|---|
| 0 | 4007 | $84.54 \pm 17.99$ (91.09) | $12.25 \pm 18.75$ (3.16) | $3.21 \pm 5.15$ (0.98) |
| 1 | 3494 | $72.06 \pm 22.95$ (77.99) | $26.69 \pm 23.59$ (20.69) | $1.24 \pm 2.45$ (0.00) |
| 2 | 973 | $75.75 \pm 21.77$ (83.62) | $22.24 \pm 22.81$ (14.24) | $2.01 \pm 3.43$ (0.29) |
| 3 | 1423 | $67.27 \pm 25.91$ (72.99) | $31.01 \pm 27.08$ (25.00) | $1.72 \pm 3.95$ (0.00) |
| 4 | 215 | $60.02 \pm 26.86$ (62.07) | $38.43 \pm 28.15$ (37.36) | $1.55 \pm 2.89$ (0.00) |

Tabella 3.3: Clinical characterization of clusters for Combo 2 using TIR, TAR and TBR.

| Cluster | Days | TIR (%) | TAR (%) | TBR (%) |
|---|---|---|---|---|
| 0 | 324 | $60.71 \pm 20.26$ (61.92) | $38.16 \pm 21.01$ (37.36) | $1.13 \pm 2.78$ (0.00) |
| 1 | 15 | $55.18 \pm 14.24$ (49.71) | $39.86 \pm 20.24$ (47.68) | $4.97 \pm 8.01$ (0.00) |
| 2 | 9773 | $76.99 \pm 22.60$ (84.79) | $20.81 \pm 23.62$ (11.65) | $2.20 \pm 4.13$ (0.00) |

Tabella 3.4: Clinical characterization of clusters for Combo 3 using TIR, TAR and TBR.

| Cluster | Days | TIR (%) | TAR (%) | TBR (%) |
|---|---|---|---|---|
| 0 | 10059 | $76.55 \pm 22.62$ (84.42) | $21.28 \pm 23.65$ (12.36) | $2.17 \pm 4.10$ (0.00) |
| 1 | 53 | $54.14 \pm 28.20$ (52.87) | $44.32 \pm 29.81$ (47.13) | $1.54 \pm 4.20$ (0.00) |

For space reasons, detailed visualizations are reported only for the most informative combinations (Combos 1, 2, 3, 4 and 8), while full tabular summaries are provided for all combos.

Tabella 3.5: Clinical characterization of clusters for Combo 4 using TIR, TAR and TBR.

| Cluster | Days | TIR (%) | TAR (%) | TBR (%) |
|---|---|---|---|---|
| 0 | 10080 | $76.50 \pm 22.67$ (84.27) | $21.33 \pm 23.70$ (12.36) | $2.17 \pm 4.10$ (0.00) |
| 1 | 32 | $56.10 \pm 27.26$ (50.28) | $42.33 \pm 27.98$ (43.62) | $1.57 \pm 3.73$ (0.00) |

Tabella 3.6: Clinical characterization of clusters for Combo 8 using TIR, TAR and TBR.

| Cluster | Days | TIR (%) | TAR (%) | TBR (%) |
|---|---|---|---|---|
| 0 | 429 | $92.86 \pm 9.41$ (96.84) | $3.68 \pm 8.25$ (0.00) | $3.46 \pm 6.23$ (0.62) |
| 1 | 3120 | $72.79 \pm 22.67$ (78.92) | $25.95 \pm 23.30$ (19.70) | $1.26 \pm 2.46$ (0.00) |
| 2 | 925 | $75.34 \pm 22.10$ (83.38) | $22.70 \pm 23.12$ (14.44) | $1.96 \pm 3.38$ (0.29) |
| 3 | 215 | $60.02 \pm 26.86$ (62.07) | $38.43 \pm 28.15$ (37.36) | $1.55 \pm 2.89$ (0.00) |
| 4 | 202 | $57.49 \pm 19.00$ (59.08) | $41.51 \pm 19.68$ (40.47) | $1.00 \pm 2.42$ (0.00) |
| 5 | 3 | $74.48 \pm 1.34$ (74.85) | $7.50 \pm 3.26$ (9.36) | $18.02 \pm 4.57$ (15.79) |
| 6 | 1331 | $67.39 \pm 26.02$ (73.56) | $30.92 \pm 27.20$ (24.71) | $1.69 \pm 3.84$ (0.00) |
| 7 | 3887 | $82.80 \pm 19.24$ (89.94) | $14.15 \pm 20.17$ (4.69) | $3.05 \pm 4.91$ (0.87) |

## 3.4 Markov transition matrix

The Markov modeling procedure was applied to all patients in the data set and to all combinations of parameters. For clarity and readability, the results are illustrated for a representative subject, while aggregate validation metrics are reported across the full cohort. The representative subject was selected as the patient with the longest available observation window, i.e. the highest number of simulated days, ensuring sufficient data to estimate transition probabilities with reasonable statistical support and to illustrate multi-day metabolic dynamics in a stable and informative manner.

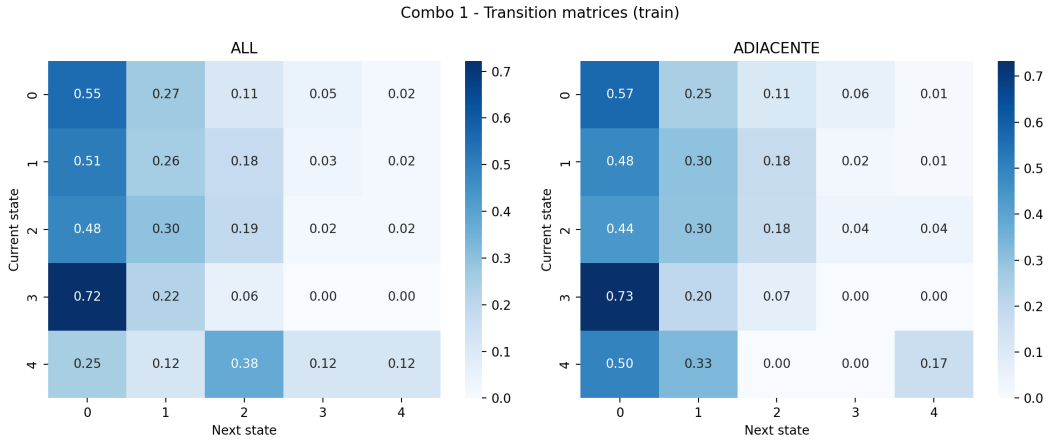


Figura 3.14: Estimated Markov transition matrices for the representative patient and Combo 1. The left matrix is computed using all consecutive and non-consecutive days, while the right matrix is obtained considering only consecutive (adjacent) day transitions.

To assess the consistency of the estimated transition dynamics over time, we compare the transition matrices obtained from the training and test segments using a chi-square goodness-of-fit test.

Tabella 3.7: Chi-square test comparing train and test transition counts for the *all* transition definition (Combo 1).

| Cluster | Train transitions | Test transitions | $\chi^2$ | $p$-value |
|---|---|---|---|---|
| 0 | 244 | 61 | 5.64 | 0.227 |
| 1 | 121 | 30 | 7.82 | 0.099 |
| 2 | 64 | 18 | 4.01 | 0.405 |
| 3 | 18 | 4 | – | – |

Tabella 3.8: Chi-square test comparing train and test transition counts for the *adjacent* transition definition (Combo 1).

| Cluster | Train transitions | Test transitions | $\chi^2$ | $p$-value |
|---|---|---|---|---|
| 0 | 197 | 49 | 9.20 | 0.056 |
| 1 | 99 | 22 | 3.99 | 0.407 |
| 2 | 50 | 17 | 5.98 | 0.201 |
| 3 | 15 | 4 | – | – |

After presenting the results obtained with the set of core physiological parameters (Combo 1), we now report the Markov modeling results for Combo 8, which explicitly extend the core dynamics to include intraday variations in insulin sensitivity.

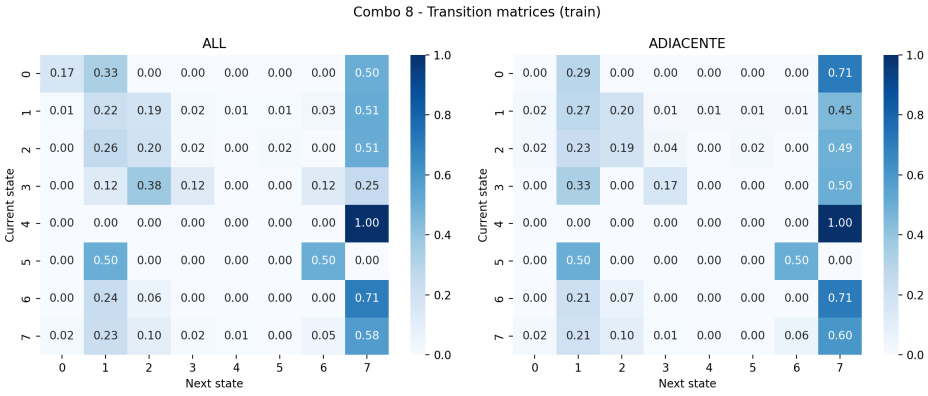


Figura 3.15: Estimated Markov transition matrices for Combo 8 and the representative patient, obtained using all transitions (left) and only adjacent-day transitions (right).

Similarly to Combo 1, the chi-square analysis is also reported for Combo 8 to assess the consistency between training and the transition dynamics of the test.

Tabella 3.9: Chi-square comparison between training and test transition distributions for Combo 8 using all transitions.

| State | Train transitions | Test transitions | $\chi^2$ | $p$-value |
|---|---|---|---|---|
| 0 | 6 | 2 | – | – |
| 1 | 105 | 29 | 12.04 | 0.099 |
| 2 | 61 | 18 | – | – |
| 6 | 17 | 4 | – | – |
| 7 | 253 | 60 | – | – |

Tabella 3.10: Chi-square comparison between training and test transition distributions for Combo 8 using only adjacent-day transitions.

| State | Train transitions | Test transitions | $\chi^2$ | $p$-value |
|---|---|---|---|---|
| 1 | 84 | 21 | 9.31 | 0.231 |
| 2 | 47 | 17 | – | – |
| 6 | 14 | 4 | – | – |
| 7 | 204 | 50 | – | – |

## 3.5 Validation of Markov models

The results of Markov chain validation are reported in the following, using implied timescale analysis and the ChapmanKolmogorov test to assess the internal consistency of the inferred stochastic dynamics under the adopted modeling assumptions.

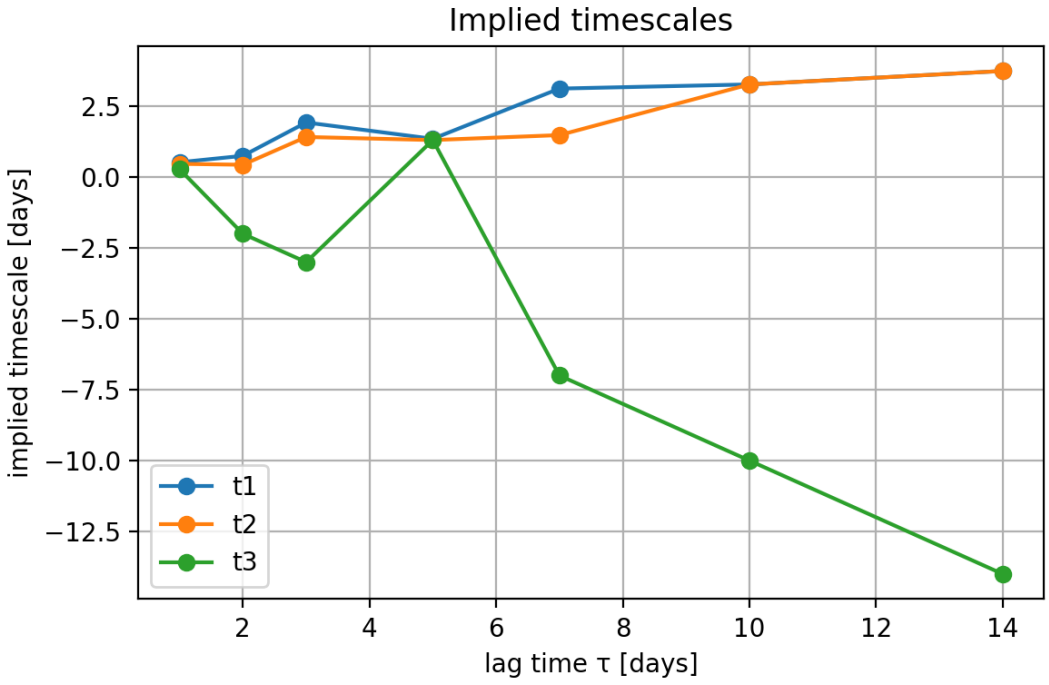


Figura 3.16: Implied timescales obtained for different lag times for Combo 1, shown for *adiacente* transition definition, within the analyzed observation window.

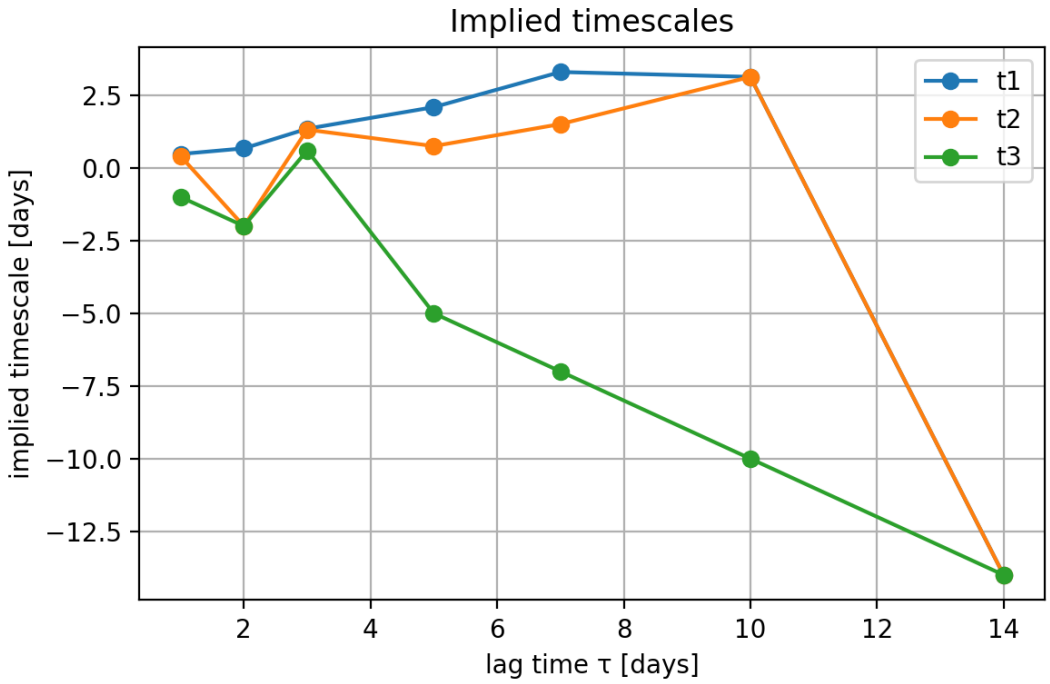


Figura 3.17: Implied timescales obtained for different lag times for Combo 1, shown for *all* transition definition, within the analyzed observation window.

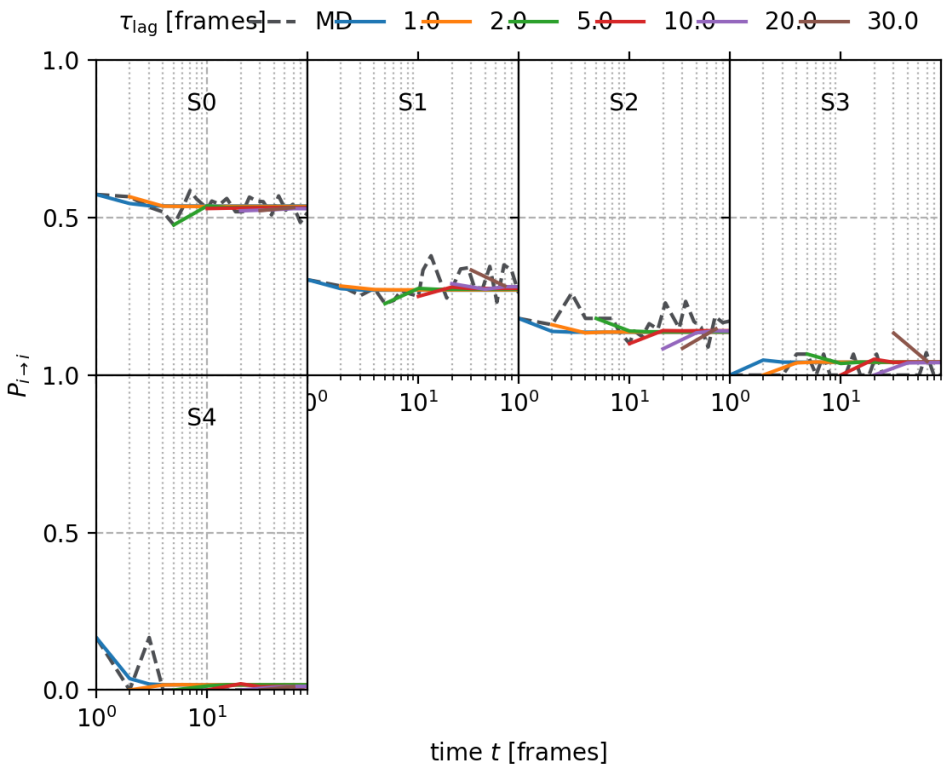


Figura 3.18: Chapman–Kolmogorov test for the Markov model of Combo 1 for adiacente transitions definition, comparing empirical and model-predicted multi-step transition probabilities.

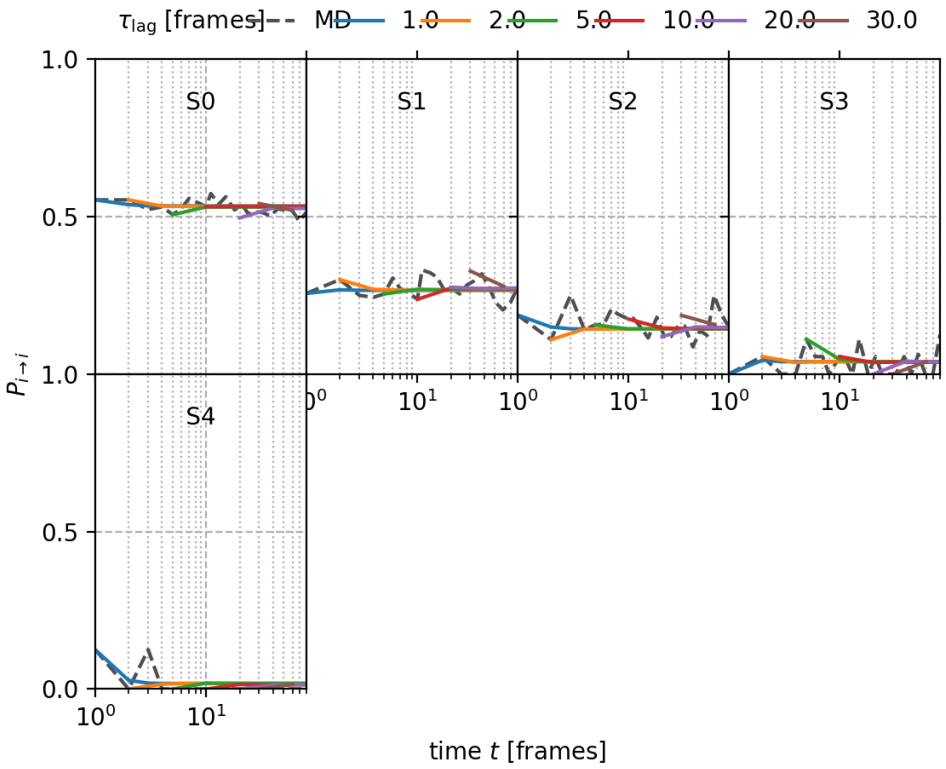


Figura 3.19: Chapman–Kolmogorov test for the Markov model of Combo 1 for all transition definition, comparing empirical and model-predicted multi-step transition probabilities.

The validation analysis is now extended to Combo 8, in order to verify whether the Markovian assumption and the inferred temporal dynamics remain consistent when focusing on glucose regulation and insulin sensitivity parameters.

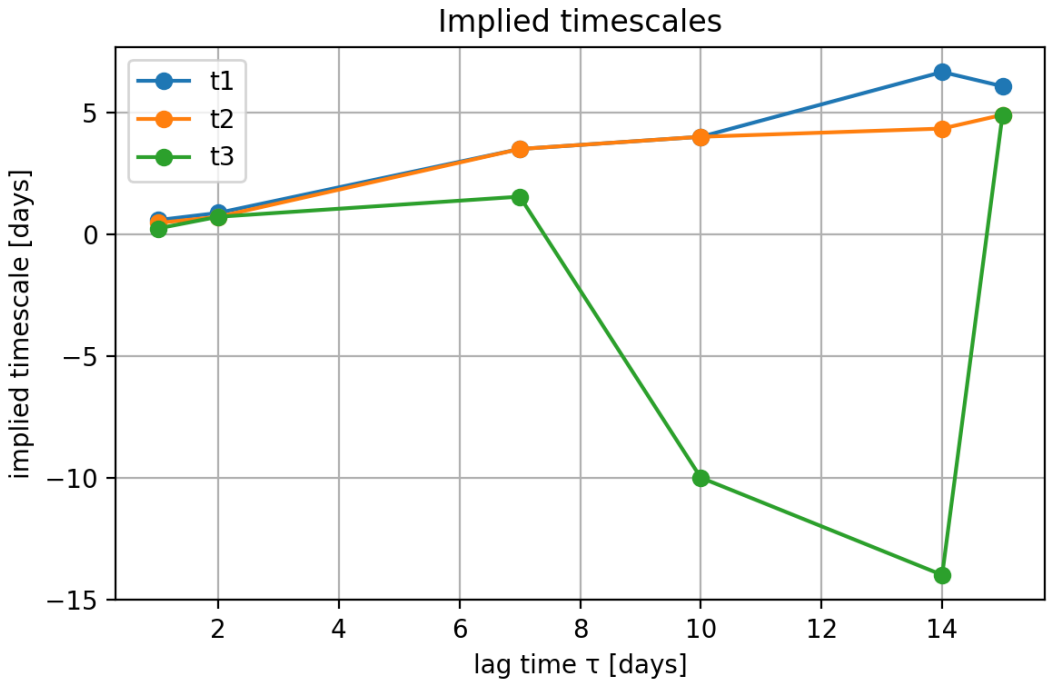


Figura 3.20: Implied timescales obtained for different lag times for Combo 8, shown for *adiacente* transition definition, within the analyzed observation window.

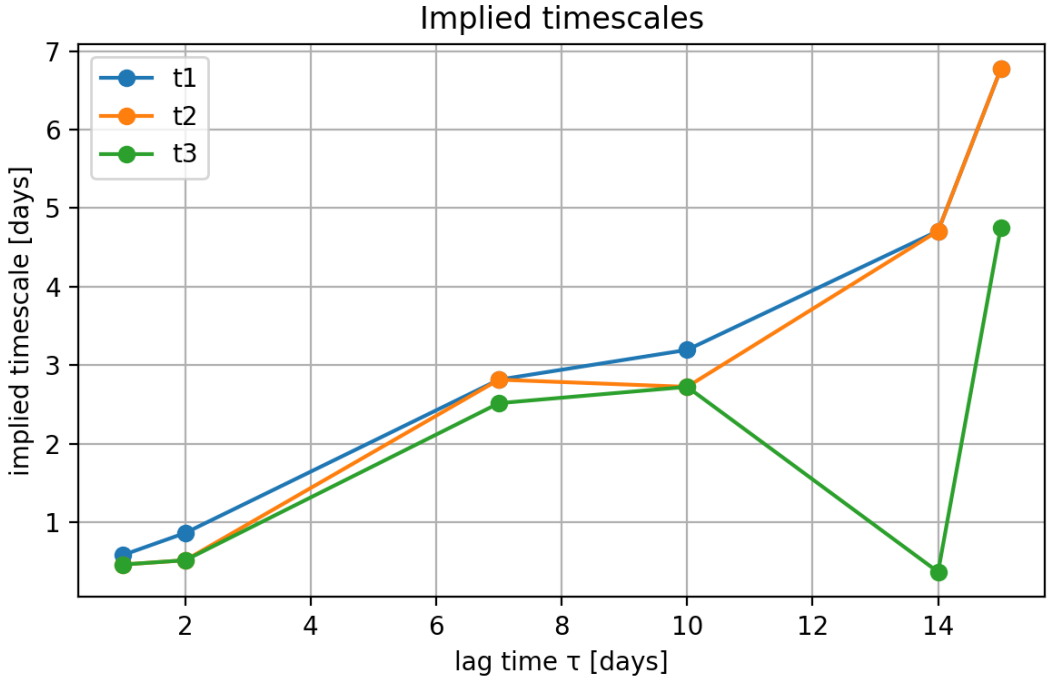


Figura 3.21: Implied timescales obtained for different lag times for Combo 8, shown for *all* transition definition, within the analyzed observation window.

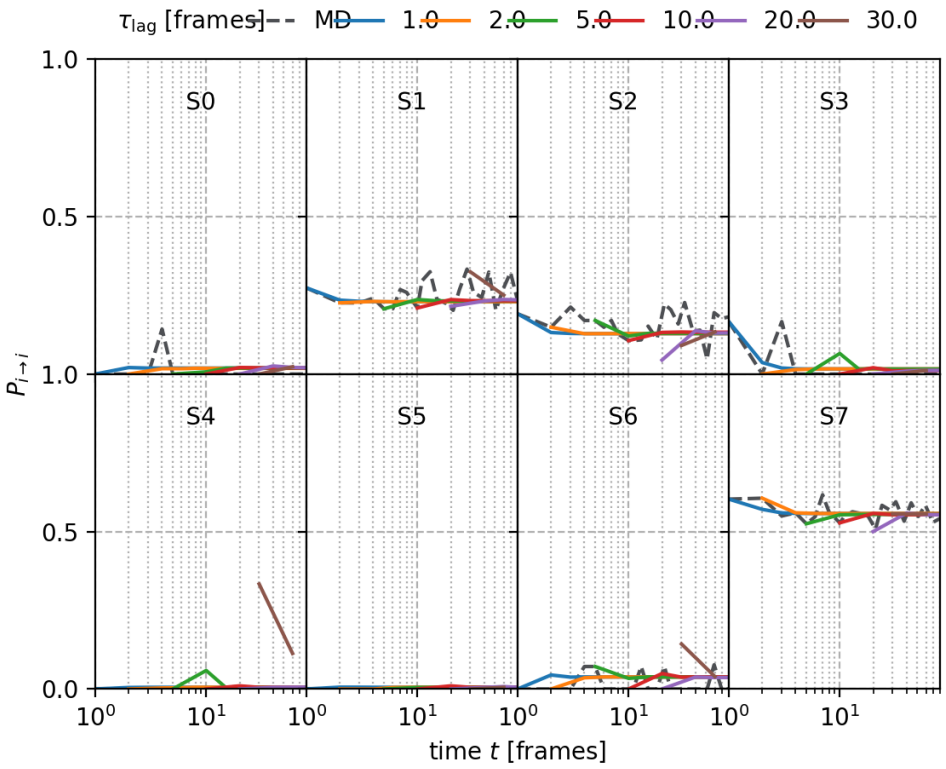


Figura 3.22: Chapman–Kolmogorov test for the Markov model of Combo 8 for *adiacente* transition definition, comparing empirical and model-predicted multi-step transition probabilities.

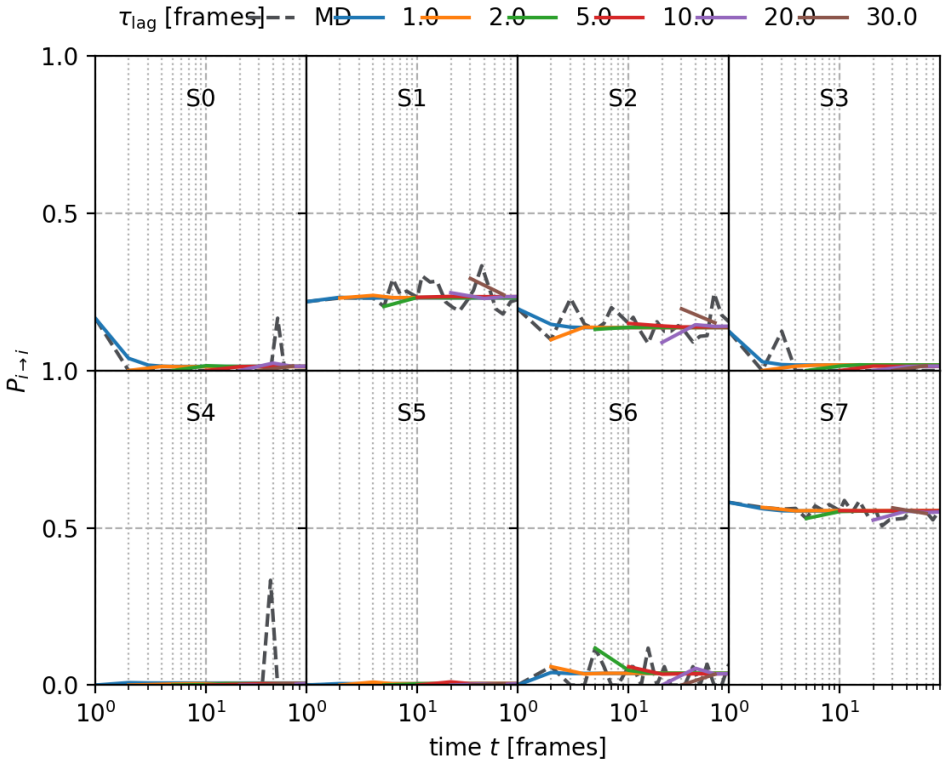


Figura 3.23: Chapman–Kolmogorov test for the Markov model of Combo 8 for *all* transition definition, comparing empirical and model-predicted multi-step transition probabilities.

# Capitolo 4

# Discussion

This chapter discusses the results presented in Chapter 3, with the goal of interpreting the clustering outcomes, their clinical characterization, and their implications for modeling multi-day metabolic dynamics. Particular emphasis is placed on understanding how parameter-based clusters relate to glycemic control and on assessing the physiological meaning of the identified metabolic states.

## 4.1 Discussion of clustering results

Table 3.1 summarizes the clustering configurations obtained for each parameter combination, reporting the number of clusters selected through the silhouette analysis and the corresponding mean silhouette score. These results highlight how different physiological subsystems encoded in the ReplayBG parameters exhibit distinct clustering behavior, both in terms of separability and structural complexity.

Parameter combinations composed of a limited number of closely related parameters tend to yield a small number of clusters with comparatively high silhouette values. This behavior is observed for Combo 4, which includes only absorption rate parameters ($kabs_B, kabs_L, kabs_S, kabs_D$), and for Combo 3, which additionally incorporates the corresponding shape parameters $\beta$. In both cases, clustering results in two dominant states with high silhouette scores (0.92 and 0.80, respectively), suggesting a clear separation between absorption-related regimes. This indicates that, at the daily scale and within the analyzed dataset, absorption dynamics can be described using a limited number of recurrent states.

A similar behavior is observed for Combo 2, which focuses exclusively on insulin sensitivity parameters ($SI_B, SI_L, SI_D$). The three resulting clusters exhibit a relatively high silhouette score (0.857), suggesting that daily variations in insulin sensitivity tend to organize into distinguishable regimes in this dataset. This finding supports the role of insulin sensitivity as a key driver of inter-day variability in glucose regulation.

In contrast, parameter combinations that integrate multiple physiological subsystems exhibit lower silhouette scores and require a larger number of clusters. Combo 1, which

combines basal glucose, glucose effectiveness, insulin kinetics, and gastric emptying parameters, results in five clusters with a moderate silhouette score (0.336). This reflects increased heterogeneity due to the interaction of multiple mechanisms, leading to partially overlapping but still meaningful metabolic states.
The most complex combinations (Combos 5, 6 and 7), which jointly include glucose dynamics, insulin sensitivity, and absorption-related parameters, require a larger number of clusters and show lower silhouette values. This may reflect the intrinsic complexity of the underlying physiology: when multiple subsystems interact, variability can manifest along more continuous dimensions, reducing sharp separability and increasing the presence of intermediate or less frequent regimes.
In general, these results demonstrate that the proposed combo-based clustering strategy provides a flexible trade-off between interpretability and expressiveness. Simpler combos isolate dominant physiological mechanisms, while more comprehensive combos capture richer metabolic structures at the cost of reduced cluster separation.

## 4.2 Clinical interpretation of clusters through glycemic metrics

Although clustering was performed exclusively in the ReplayBG parameter space, each identified cluster was subsequently characterized using clinically meaningful CGM-derived metrics, namely Time in Range (TIR), Time Above Range (TAR), and Time Below Range (TBR). This post-hoc characterization enables the interpretation of parameter-based clusters in terms of glycemic control quality and risk profile.

### 4.2.1 Combo 1: Core glucose dynamics and gastric emptying

For Combo 1 (Table 3.2), the five identified groups show a coherent trend toward reduced glycemic control. Cluster 0, which contains the largest number of days, is characterized by high TIR (median above 90%) and low TAR, corresponding to well-controlled conditions. Across the remaining clusters, TIR tends to decrease, while TAR tends to increase, with Cluster 4 showing the least favorable profile (median TIR around 62% and TAR close to 40%).
In all clusters, TBR remains relatively low, indicating that the differences are primarily driven by hyperglycemic exposure rather than hypoglycemia. This pattern is consistent with variations in parameters related to gastric emptying and glucose turnover, which predominantly affect postprandial glucose excursions.

### 4.2.2 Combo 2: Insulin sensitivity regimes

The combination 2 (Table 3.3) reveals three distinct states of insulin sensitivity with markedly different clinical outcomes. The dominant group (group 2) is associated with

a high TIR and a relatively low TAR, indicating an effective insulin action and stable glycemic control. In contrast, smaller clusters (groups 0 and 1) exhibit substantially lower TIR and higher TAR, reflecting reduced insulin sensitivity and an increased risk of hyperglycemia.
The sharp separation between these states aligns with the high silhouette score observed for this combo and reinforces the interpretation of insulin sensitivity as a major determinant of daily metabolic state.

### 4.2.3 Combo 3 and Combo 4: Gastric absorption dynamics

For Combo 3 and Combo 4 (Tables 3.4 and 3.5), clustering essentially separates a predominant absorption regime from a rare alternative state associated with poor glycemic control. The smaller clusters, although representing a limited number of days, consistently exhibit markedly reduced TIR and elevated TAR.
These results highlight the clinical relevance of absorption-related parameters: even infrequent deviations in gastric absorption dynamics can lead to substantial deterioration in glycemic control. At the same time, the strong imbalance in the size of the clusters reflects the relative stability of the absorption behavior on most days.

### 4.2.4 Combo 8: Integrated glucose dynamics and insulin sensitivity

Combo 8 (Table 3.6) provides a more nuanced representation of metabolic variability by combining glucose dynamics and insulin sensitivity parameters. The eight identified groups span a wide spectrum of glycemic outcomes, ranging from extremely well-controlled states (group 0, with median TIR close to 97%) to poorly controlled regimes dominated by hyperglycemia (groups 3 and 4).
In particular, Combo 8 also reveals a small group characterized by elevated TBR (group 5), indicating increased hypoglycemic exposure. Although this group contains very few days, its emergence is clinically significant, as it suggests the existence of specific parameter regimes associated primarily with hypoglycemia rather than hyperglycemia. This information would be difficult to extract from glucose time series alone and highlights the added value of parameter-based representations.

## 4.3 Synthesis and implications

Across the combinations of parameters analyzed, a coherent relationship can be observed between the clusters identified in the ReplayBG parameter space and the clinically interpretable metabolic states. Differences in TIR, TAR, and TBR between clusters suggest that the estimated digital twin parameters capture relevant information related to glycemic control, although glucose data were not explicitly used during the clustering procedure.

Insulin sensitivity and gastric absorption parameters tend to act as important discriminants of metabolic state, while their interaction with glucose dynamics is associated with more complex and gradual transitions between regimes. Overall, these results provide a reasonable physiological and clinical basis for the subsequent temporal modeling of cluster transitions and motivate the use of Markov chains as a compact representation of multi-day metabolic evolution.

## 4.4 Discussion of cluster characterization

Although the clustering procedure was performed exclusively in the ReplayBG parameter space, the post-hoc characterization of clusters using CGM-derived metrics provides a crucial link between abstract physiological parameters and clinically meaningful outcomes. In particular, the use of Time in Range (TIR), Time Above Range (TAR), and Time Below Range (TBR) allows each cluster to be interpreted as a distinct glycemic control profile rather than as a purely mathematical construct.

Across all combinations of parameters analyzed, a consistent and robust correspondence emerges between clusters defined by the parameters and clinically relevant glucose control patterns. This observation supports the central hypothesis of this work: daily digital twin parameters encode sufficient information to identify meaningful metabolic states.

### 4.4.1 Progressive glycemic stratification in Combo 1

Characterization of clusters obtained from Combo 1 reveals a clear and interpretable stratification of the quality of the glycemic control. As shown in Table 3.2, clusters are ordered along a continuum ranging from optimal control to progressively worsening hyperglycemic exposure.

Cluster 0 represents the best-controlled metabolic state, characterized by a high median TIR (above 90%) and minimal TAR. This cluster includes the highest number of days, suggesting that it corresponds to the most frequent and stable physiological regime. Moving toward higher cluster indices, a monotonic decrease in TIR is observed, accompanied by a systematic increase in TAR. Cluster 4, in particular, exhibits a median TIR close to 60% and a TAR approaching 40%, clearly identifying a high-risk hyperglycemic state.

Interestingly, TBR remains consistently low across all clusters. This indicates that for this combination of parameters, the dominant source of variability in metabolic states is hyperglycemia rather than hypoglycemia. From a physiological perspective, this is consistent with the role of parameters such as gastric emptying rate and insulin kinetics, which mainly influence postprandial glucose excursions and prolonged hyperglycemic exposure.

### 4.4.2 Insulin sensitivity-driven regimes in Combo 2

Combo 2, which isolates insulin sensitivity parameters, exhibits one of the clearest clinical separations among clusters. As reported in Table 3.3, the dominant cluster (Cluster 2) is associated with high TIR and relatively low TAR, reflecting effective insulin action and stable glycemic control.

In contrast, the remaining clusters show a marked reduction in TIR and a corresponding increase in TAR. These clusters represent days characterized by impaired insulin sensitivity, leading to prolonged hyperglycemia despite similar therapeutic input. The sharp clinical separation between these states is consistent with the high silhouette score obtained during clustering and confirms insulin sensitivity as a primary driver of day-to-day metabolic variability.

The limited size of the poorest-performing cluster further suggests that severe insulin resistance is a relatively rare but clinically significant condition, which is consistently captured by the digital twin parameters.

### 4.4.3 Absorption-related states in Combos 3 and 4

The Combos 3 and 4 focus on gastric absorption dynamics, including or excluding the associated shape parameters. In both cases, the characterization of the cluster reveals a strongly imbalanced structure, with a dominant cluster corresponding to satisfactory glycemic control and a much smaller cluster associated with markedly worse outcomes.

As shown in Tables 3.4 and 3.5, the minor clusters are characterized by substantially reduced TIR and elevated TAR, indicating that deviations in absorption dynamics, although rare, can have a pronounced negative impact on glycemic control. This highlights the sensitivity of glucose regulation to gastric emptying and carbohydrate absorption, even in the absence of major changes in insulin sensitivity.

The rarity of these clusters suggests that absorption-related disturbances are episodic rather than persistent, but their distinct clinical signature underscores the importance of modeling absorption parameters explicitly when analyzing daily metabolic states.

### 4.4.4 Rich metabolic heterogeneity in Combo 8

The combination 8 integrates glucose dynamics and insulin sensitivity parameters, resulting in a more complex and diverse set of metabolic states. As reported in Table 3.6, the eight identified groups span a wide spectrum of glycemic profiles, ranging from exceptionally well-controlled days to clearly pathological regimes.

Cluster 0 represents an optimal control state, with a median TIR close to 97% and negligible TAR. At the opposite end of the spectrum, Clusters 3 and 4 are characterized by low TIR and high TAR, indicating sustained hyperglycemia. Intermediate clusters capture varying degrees of control degradation, reflecting the combined influence of glucose kinetics and insulin sensitivity.

Of particular interest is Cluster 5, which, despite its small size, exhibits a markedly elevated TBR. This group represents a hypoglycemia-prone metabolic state, distinct from the predominantly hyperglycemic patterns observed in other groups. Its identification demonstrates that the proposed framework is capable of isolating clinically relevant but infrequent risk states, which are often difficult to detect using glucose time series alone.

### 4.4.5 General considerations

Across all parameter combinations, cluster characterization consistently shows that differences in digital twin parameters translate into systematic and clinically interpretable differences in glycemic outcomes. Clusters are not arbitrary partitions of the parameter space, but rather correspond to distinct metabolic regimes associated with specific risk profiles.

The relative stability of TBR in most clusters, in contrast to the wide variability in TAR, suggests that hyperglycemia is the dominant axis of inter-day metabolic variation in the analyzed cohort. However, the emergence of hypoglycemia-dominated clusters in richer parameter combinations highlights the importance of preserving model complexity when the goal is a comprehensive risk assessment.

Overall, these results validate the use of parameter-based clustering as a powerful abstraction layer between high-frequency CGM data and long-term metabolic modeling. The identified clusters provide a clinically grounded state space that is well suited for subsequent temporal modeling using Markov chains.

## 4.5 Markov model as a multi-day simulator

The Markov modeling framework introduced in this work represents the final abstraction layer linking daily digital twin realizations into coherent multi-day trajectories. Based on estimated daily parameter clusters, each cluster is interpreted as a discrete metabolic state, while the estimated transition matrices describe how patients move between these states over time. This section discusses the results obtained from the Markov analysis, focusing on the structure of the inferred transition dynamics, their statistical robustness, and their suitability for multi-day simulation within a digital twin context.

### 4.5.1 Structure of the inferred transition dynamics

The estimated transition matrices reveal a structured, non-uniform dynamics across days. For both transition definitions (*all* and *adiacente*), several states exhibit a tendency toward self-transition, suggesting that once a patient enters a given metabolic state there is often a non-negligible probability of remaining in that state on the following day.
This behavior is particularly evident in groups associated with favorable glycemic control, as identified during the clinical characterization phase. This persistence reflects physiologically plausible dynamics: key determinants of glucose regulation, such as insulin sensitivity or baseline metabolic efficiency, typically evolve over days rather than fluctuating abruptly. The Markov model captures this inertia at the level of the discretized states, reproducing patterns of persistence observed in the clustered sequences.
At the same time, certain clusters display lower self-transition probabilities and more distributed outgoing transitions. These states tend to act as transient or intermediate regimes, from which the system can evolve toward multiple alternative states. This pattern is especially visible in parameter combinations that integrate multiple physiological subsystems, where the metabolic state results from the interaction of several mechanisms. In these cases, the Markov structure suggests that some daily configurations represent unstable equilibria rather than long-term regimes.

### 4.5.2 Effect of transition definition: *all* vs. *adiacente*

A key methodological contribution of this work is the explicit comparison between two transition definitions. The *all* mode includes transitions across non-consecutive days, while the *adiacent* mode restricts the analysis to strictly consecutive days.
The resulting transition matrices show that while the qualitative structure of the dynamics is preserved across the two definitions, the *adiacente* matrices tend to emphasize short-term persistence and local transitions. In contrast, *all* matrices incorporate longer-range transitions that implicitly encode slower temporal processes, such as gradual adaptation or delayed responses to unobserved changes.
This comparison highlights an important modeling trade-off. The *adiacente* definition provides a more faithful representation of true day-to-day dynamics at the cost of reduced

data availability. The *all* definition increases statistical robustness but blends short-term and longer-term effects. The fact that both approaches yield coherent and interpretable transition structures strengthens the validity of the Markov abstraction.

### 4.5.3 Statistical robustness of transition estimates

The stability of the inferred transition dynamics was quantitatively assessed by comparing the transition counts estimated in the training and test segments using chi-square statistics. For most states with sufficient data support, the resulting p-values indicate that there is no statistically significant deviation between training and test distributions.

This result has two important implications. First, it suggests that the estimated transition probabilities are not driven by random fluctuations or overfitting to a specific temporal segment. Second, it supports the assumption of time-homogeneous dynamics over the analyzed windows, which is a key requirement for the validity of the Markov framework.

States for which the chi-square statistic could not be computed typically correspond to clusters with very low occupancy. Rather than indicating a modeling failure, this reflects a data limitation: rare metabolic states do not provide enough transitions to reliably estimate their dynamics. Importantly, these states are already identified as marginal during clustering and characterization, and their limited representation is consistent throughout the analysis stages.

### 4.5.4 Validation of the Markov assumption

Beyond train–test consistency, the Markov assumption itself was evaluated using tools commonly adopted in the Markov State Model literature, namely implied timescale analysis and the Chapman–Kolmogorov (CK) test.

The implied timescale analysis was used to assess whether the estimated dynamics is compatible with a Markovian description at the chosen temporal resolution. In particular, attention was paid to the behavior of the dominant timescales as a function of the lag time. Across the combinations analyzed, the leading timescales exhibit limited variation over a range of intermediate lag values, rather than strong drift or divergence. This behavior suggests that, beyond the very short lags affected by statistical noise, the system can be reasonably approximated as Markovian on the daily scale.

It is important to note that the implied timescales are not interpreted here in absolute quantitative terms but rather as a qualitative diagnostic of model consistency. Deviations observed at very short lag times or for poorly populated states are expected, given the limited number of transitions available and the stochastic nature of the underlying process.

The Chapman–Kolmogorov tests further support this interpretation. For the dominant and sufficiently populated states, the multi-step transition probabilities predicted by the one-step transition matrices are in reasonable agreement with those estimated directly

from the data. Discrepancies are observed primarily for rare states, where limited sampling restricts the reliability of empirical estimates.

Taken together, these validation results indicate that while the inter-day metabolic dynamics is influenced by unobserved factors and exhibit inherent variability, the first-order Markov approximation provides a coherent and empirically supported description of the dominant temporal structure present in the clustered sequences.

### 4.5.5 Interpretation and limitations of predictive performance

The predictive experiments based on cross-validation confirm that the Markov models possess non-trivial forecasting capability. The probability of correctly predicting the next-day cluster based solely on the current state consistently exceeds simple baselines, such as random guessing or majority-class prediction.

At the same time, the prediction accuracy remains moderate rather than high. This outcome is expected and should not be interpreted as a weakness of the framework. Daily glucose regulation is influenced by many exogenous factorssuch as meals, physical activity, stress, and therapy adjustmentsthat are not explicitly modeled here. Consequently, a purely state-based stochastic model cannot be expected to deterministically predict daily evolution.

Instead, the Markov chain should be interpreted as a generative model: it produces statistically plausible trajectories that reproduce the observed structure and variability of the data, rather than exact forecasts. This distinction is fundamental when the goal is simulation rather than short-term prediction.

### 4.5.6 Markov chain as a bridge toward multi-day digital twins

The primary value of the Markov model lies in its role within a layered digital twin architecture. ReplayBG provides detailed, physiologically grounded simulations at the intra-day scale but does not explicitly model how patient-specific parameters evolve across days. The Markov chain introduced in this work fills this gap by operating at a higher temporal scale.

In the proposed framework, the Markov model generates sequences of metabolic states over days. Each state is associated with representative parameter sets derived from the clustering phase, which can be used to reconstruct day-specific digital twin configurations. These configurations then drive ReplayBG simulations, producing detailed glucose trajectories at high temporal resolution.

By separating intra-day physiology from inter-day stochastic dynamics, this approach enables scalable and interpretable multi-day simulations. Importantly, the Markov layer does not replace physiological modeling, but complements it by providing a principled mechanism to connect daily digital twin realizations into coherent long-term scenarios.

### 4.5.7 Implications for in-silico experimentation

The results obtained in this work demonstrate that a relatively simple probabilistic model, when built on physiologically meaningful state definitions, is sufficient to capture key aspects of long-term metabolic dynamics. This opens the door to future in-silico experimentation, such as exploring how modifications in transition probabilities could alter the long-term distribution of metabolic states.

Although such applications are beyond the scope of this study, the proposed framework establishes a solid methodological foundation. By combining parameter-based clustering with validated Markov dynamics, it becomes possible to move from isolated daily simulations toward integrated, multi-day digital twin trajectories that reflect both physiological realism and observed temporal structure.

# Capitolo 5

# Conclusions and Future Work

## 5.1 Summary of the work

The objective of this thesis was to explore whether the information extracted from a physiological digital twin can be used to describe and model the multi-day metabolic evolution of individuals with type 1 diabetes. In particular, the work focused on parameters identified through ReplayBG daily model fitting and investigated whether such parameters contain sufficient structure to support a compact representation of inter-day metabolic dynamics.

To address this question, the thesis combined two methodological components: an unsupervised clustering of daily parameter sets to identify recurrent metabolic configurations and a probabilistic Markov framework to model transitions between these configurations across days. The goal was not to achieve optimal prediction performance, but rather to assess the feasibility, interpretability, and internal consistency of such an approach.

## 5.2 Main findings

Clustering analysis revealed that ReplayBG-derived parameters can be organized into a finite number of recurrent states, whose structure depends strongly on the specific subset of parameters considered. Simpler combinations, focusing on a single physiological subsystem, such as insulin sensitivity or gastric absorption, resulted in a small number of well-separated clusters. More comprehensive parameter sets, which combined multiple subsystems, produced a larger number of clusters with reduced separability, reflecting the increased complexity and gradual variability of the underlying physiology.

Importantly, although clustering was performed without using glucose outcomes, the identified clusters exhibited systematic differences in clinically relevant CGM-derived metrics such as Time in Range, Time Above range and Time Below Range. These differences suggest that the digital twin parameters encode information that is meaningfully related to glycemic control, even when considered independently of glucose time series.

The post-hoc characterization also highlighted that most inter-day variability in the analyzed cohort is driven by hyperglycemic exposure, while hypoglycemia-related patterns appear less frequent and emerge mainly in richer parameter combinations. This observation is consistent with the distribution of the available data and with the physiological role of the modeled parameters.

## 5.3 Insights from Markov modeling

The Markov modeling results indicate that transitions between daily metabolic states exhibit a non-random temporal structure. Several clusters showed high self-transition probabilities, suggesting the presence of relatively stable metabolic regimes that persist for multiple days. Other clusters appeared to be more transient, acting as intermediate configurations between more persistent states.

Statistical validation procedures, including train–test chi-square comparisons, implied timescale analysis, and Chapman–Kolmogorov tests, provided partial support for the adequacy of a first-order, time-homogeneous Markov approximation at the daily scale. Although not all states could be robustly validated due to limited data availability, the overall behavior of the models was consistent with the expected stochastic nature of the day-to-day metabolic evolution.

It is important to emphasize that the predictive performance of the Markov models remains moderate. This is not unexpected, given the absence of explicit information on exogenous factors such as meals, physical activity, stress, or therapy adjustments. In this context, the Markov framework should be interpreted primarily as a generative and descriptive tool, rather than as a deterministic predictor of future glucose control.

## 5.4 Contribution to multi-day digital twin modeling

A key contribution of this work is the demonstration that parameter-based clustering combined with probabilistic state transitions can serve as a viable abstraction layer for linking independent daily digital twin simulations. Within this framework, the Markov chain operates at a higher temporal scale, describing how daily physiological configurations evolve over time, while ReplayBG provides detailed intra-day glucose dynamics conditioned on those configurations.

Although exploratory, this layered approach offers a potential pathway toward multi-day digital twin simulations that are both interpretable and computationally tractable. Rather than recalibrating the digital twin independently for each day, the proposed framework introduces a mechanism to model temporal coherence across days in a principled manner.

## 5.5 Limitations

Several limitations of the present work should be acknowledged. First, the analysis is based on a limited dataset in terms of both the number of patients and the length of observation windows. Although sufficient for methodological exploration, this limits the generalizability of the findings.
Second, all results depend on the accuracy and assumptions of the ReplayBG parameter identification process. Any modeling simplifications or fitting errors propagate directly into the clustering and Markov analyzes, and the conclusions should therefore be interpreted as conditional on the validity of the underlying digital twin.
Third, the use of K-means clustering imposes structural assumptions on the parameter space that may not fully capture its complexity. In addition, the absence of behavioral and clinical covariates restricts the ability of the models to explain sudden or non-physiological changes in glycemic control.
Finally, the Markov models rely on first-order and time-homogeneous assumptions, which limit their ability to represent long-term memory effects or non-stationary dynamics driven by sustained interventions or disease progression.

## 5.6 Future research directions

Future work could extend this framework in several directions. From a methodological perspective, alternative clustering approaches and richer feature sets, including CGM-derived variability metrics and therapy-related variables, could be explored to improve interpretability and robustness.
On the temporal modeling side, higher-order or non-homogeneous Markov models, as well as hidden-state formulations, could be investigated to capture longer-term dependencies and evolving dynamics. Integrating the proposed framework with full multi-day ReplayBG simulations would enable more comprehensive in-silico experiments and allow the assessment of long-term intervention scenarios.
Ultimately, validation on real-world longitudinal datasets will be essential to assess the clinical relevance and practical applicability of the approach.

# Concluding remarks

In summary, this thesis provides an exploratory investigation into the use of digital twin parameters to model multi-day metabolic dynamics in type 1 diabetes. Although the results are preliminary and subject to important limitations, they suggest that physiologically grounded parameters contain structured information that can be organized into interpretable states and linked over time using probabilistic models.
Rather than offering a definitive solution, this work aims to contribute a methodological perspective and a foundation for future research on evolving digital twins. With further

data, refinement, and validation, the proposed framework could support more realistic and informative long-term simulations of metabolic behavior.

# Bibliografia


1. Atkinson MA, Eisenbarth GS e Michels AW. Type 1 diabetes mellitus: etiopathogenesis and clinical characteristics. Lancet 2014; 383:69–82
2. Barnard KD, Lloyd CE et al. Evaluation of structured online diabetes self-management education: a randomized controlled trial. Diabetes Research and Clinical Practice 2015; 110:e1–e4
3. Battelino T e authors other. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus. Diabetes Care 2019; 42:1593–603
4. Beck RW, Bergenstal RM et al. Time in range: a new parameter to evaluate blood glucose control for diabetes. Diabetes Care 2019; 42:2343–8
5. Bekiari E et al. Artificial pancreas systems for people with type 1 diabetes: a systematic review and meta-analysis. BMJ 2018; 361:k1310
6. Bergenstal RM, Klonoff DC et al. Safety of a hybrid closed-loop system with predictive low-glucose management. JAMA 2013; 310:1240–7
7. Cappon G et al. Data Augmentation via Digital Twins for Type 1 Diabetes: A Physiological Framework to Expand CGM Datasets. Preprint / Submitted Manuscript 2024
8. Cappon G, Vettoretti M, Sparacino G, Favero SD e Facchinetti A. ReplayBG: A Digital Twin-Based Methodology to Identify a Personalized Model From Type 1 Diabetes Data and Simulate Glucose Concentrations to Assess Alternative Therapies. IEEE Transactions on Biomedical Engineering 2023; 70:3227–38. DOI: 10.1109/TBME.2023.3286856
9. Chodera JD e Noé F. Markov State Models of Biomolecular Conformational Dynamics. Current Opinion in Structural Biology 2014; 25:135–44. DOI: 10.1016/j.sbi.2014.04.002
10. Control TD e Group CTR. 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 1993; 329:977–86
11. Kovatchev B. The artificial pancreas in 2019: state of the art. Diabetes Care 2019; 42:815–22

12. Li Y et al. Digital Twins and Artificial Intelligence in Metabolic Disease Research. Metabolism 2025; 158:156245

13. Noé F. Probabilistic Markov Models for Dynamical Systems in Equilibrium. The Journal of Chemical Physics 2008; 128:244103. DOI: 10.1063/1.2916718

14. Pickup JC e Sutton AJ. Insulin pump therapy in type 1 diabetes: meta-analysis of randomised controlled trials and observational studies. BMJ 2012; 344:e381

15. Prinz JH, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schütte C e Noé F. Markov Models of Molecular Kinetics: Generation and Validation. The Journal of Chemical Physics 2011; 134:174105. DOI: 10.1063/1.3565032

16. Roquemen-Echeverri D et al. Hybrid Digital Twin Models for Personalized Glucose Prediction in Type 1 Diabetes. Computer Methods and Programs in Biomedicine 2025; 245:107890

17. Sarani Rad M et al. Medical Digital Twin: Technical Principles and Applications. Journal of Clinical Medicine 2024; 14:324

18. Skyler JS. Type 1 diabetes mellitus: pathogenesis and treatment. Endocrinology and Metabolism Clinics 2017; 46:867–79

19. Wang J e Zhao L. Digital Twin Approaches for Personalized Diabetes Management: A Systematic Review. IEEE Journal of Biomedical and Health Informatics 2025

20. Wong S et al. Artificial Intelligence and Digital Twins in Healthcare: Opportunities and Challenges. The Lancet Digital Health 2025