Working Notes on Glucose–Insulin Modeling

Structural Identifiability Analysis of the Dalla Man Model

Krishna

December 26, 2025

Contents

1 Conceptual Foundations of Structural Identifiability

Before any numerical estimation, optimization, or data-driven analysis is performed, it is essential to determine whether a mathematical model is *structurally identifiable*. Structural identifiability is a theoretical property of a model that depends only on its equations, inputs, outputs, and parameterization, and not on the quality or quantity of available data.

This section establishes the conceptual and mathematical foundations of structural identifiability and clarifies its role within the overall modeling and identification workflow adopted in this work.

1.1 What Structural Identifiability Means

Consider a nonlinear dynamical system described by ordinary differential equations:

$$\dot{x}(t) = f\big(x(t), p, u(t)\big), \tag{1}$$

$$y(t) = h\big(x(t), p\big), \tag{2}$$

where:

- $x(t) \in \mathbb{R}^n$ are internal (generally unobserved) state variables,
- $p \in \mathbb{R}^q$ is the vector of unknown model parameters,
- $u(t)$ is a known input or excitation,
- $y(t)$ is the measured output.

Structural identifiability addresses the following fundamental question:

> *Assuming perfect, noise-free, continuous-time measurements of the output $y(t)$ over an infinite time horizon, can the unknown parameters p be uniquely determined from the input–output behavior of the model?*

Formally, a parameter p_i is said to be *structurally identifiable* if

$$y(t; p^{(1)}) = y(t; p^{(2)}) \quad \forall t \quad \Rightarrow \quad p_i^{(1)} = p_i^{(2)}.$$

If there exist two distinct parameter vectors that generate exactly the same output trajectory, then the corresponding parameters are structurally non-identifiable.

This definition is independent of:

- measurement noise,
- sampling frequency,
- experiment duration,
- numerical optimization algorithms.

Structural identifiability is therefore an intrinsic property of the model–output pair.

1.2 Structural vs. Practical Identifiability

It is crucial to distinguish structural identifiability from practical identifiability.

Structural identifiability is an *a priori* property of the model. It assumes idealized conditions:

- exact model structure,
- perfect measurements,
- continuous-time observation.

If a parameter is structurally non-identifiable, then *no amount of data, no improvement in noise level, and no change in estimation algorithm can make it identifiable.*

Practical identifiability is an *a posteriori* concept. It accounts for:

- finite and noisy data,
- limited experiment duration,
- suboptimal excitation.

A parameter may be structurally identifiable but practically non-identifiable if the available data do not sufficiently excite the relevant model dynamics.

This distinction is central to the interpretation of results in this work. The profile likelihood analysis performed earlier addressed *practical* identifiability. Structural identifiability analysis now determines whether those practical limitations originate from the data or from the model structure itself.

1.3 Why Structural Identifiability Must Precede Estimation

Attempting parameter estimation without first assessing structural identifiability is fundamentally ill-posed.

If a model is structurally non-identifiable:

- multiple parameter vectors yield identical outputs,
- the likelihood function contains flat directions or manifolds,
- optimization algorithms may converge but to non-unique solutions,
- estimated parameters lack physical meaning and reproducibility.

In such cases, apparent convergence of numerical estimators is misleading: different initial guesses or algorithms may return different parameter values with identical model fits.

This phenomenon was already observed in earlier steps of this work:

- sensitivity analysis revealed strong parameter correlations,
- the Fisher Information Matrix exhibited severe rank deficiency,
- profile likelihoods showed flat or unbounded directions.

Structural identifiability analysis provides the theoretical explanation for these observations and prevents misinterpretation of estimation results.

1.4 Structural Non-Identifiability in Physiological Models

Physiological models are particularly prone to structural non-identifiability due to several inherent features:

- large numbers of internal compartments,
- partial observability of states,
- parameters entering equations as products or ratios,
- indirect measurements through sensors or proxies.

In glucose–insulin models, internal variables such as tissue glucose, interstitial insulin, or insulin action are rarely measured directly. Instead, they influence the output only through chains of intermediate dynamics.

As a result, parameters often appear in the input–output relationship only through *composite expressions*, making individual parameter recovery impossible even under ideal conditions.

Structural identifiability analysis is therefore not optional but *mandatory* when working with high-fidelity physiological models.

1.5 Relevance to CGM-Based Observation

In the present work, the measured output is not plasma glucose itself but a continuous glucose monitoring (CGM) signal. The CGM introduces additional dynamics, including:

- physiological delay between blood and interstitial glucose,
- sensor filtering,
- additional sensor parameters.

This means that the observable $y(t)$ is already a transformed and delayed representation of the underlying physiological state. Such transformations further reduce the information content available for parameter identification.

Consequently, assessing structural identifiability in the presence of CGM dynamics is essential for understanding which parameters can, in principle, be inferred from CGM-only data and which cannot.

1.6 Role of Structural Identifiability in This Work

Within the overall workflow of this project, structural identifiability serves three critical purposes:

1. It explains the rank deficiency observed in the Fisher Information Matrix.
2. It justifies the unbounded profile likelihoods obtained for certain parameters.
3. It provides the theoretical basis for subsequent model reduction and parameter fixing.

Rather than being a limitation, the identification of structural non-identifiability is a valuable result: it delineates the boundary between what can and cannot be learned from the chosen observation setup.

In the following sections, this conceptual foundation is made concrete by applying a formal structural identifiability analysis to the Dalla Man glucose–insulin model using the GenSSI framework.

2 Structural Identifiability Methodology Using GenSSI

Having established the conceptual foundations of structural identifiability, we now describe the specific methodology adopted to assess structural identifiability of the Dalla Man glucose–insulin model. This work employs the **GenSSI** (Generating Series for Structural Identifiability) framework, which is based on a generating series and Lie derivative formulation of the input–output mapping.

This section explains *how* GenSSI works, *what mathematical object it constructs*, and *how the resulting algebraic ranks are interpreted*. The goal is to make the logic of the method transparent before presenting and interpreting the actual results.

2.1 Rationale for Choosing GenSSI

Several methodologies exist for structural identifiability analysis of nonlinear dynamical systems, including differential algebra (DAISY), Taylor series expansion, similarity transformation methods, and generating series approaches.

For the present study, GenSSI was selected for the following reasons:

- The Dalla Man model is strongly nonlinear and contains non-polynomial functions (e.g. hyperbolic tangents in gastric emptying), which are difficult to handle with differential algebra methods.
- The model has a large number of states and parameters, making brute-force symbolic elimination impractical.
- GenSSI provides a systematic and scalable approach based on Lie derivatives that does not require explicit elimination of states.
- The method has been specifically designed for partially observed biological models, which is the case here.

Importantly, GenSSI determines identifiability without numerical approximation: all conclusions are symbolic and therefore exact.

2.2 Generating Series and Lie Derivatives

The core idea behind GenSSI is to analyze the *generating series* of the system output. Consider again a nonlinear system

$$\dot{x}(t) = f(x(t), p, u(t)), \tag{3}$$

$$y(t) = h(x(t), p). \tag{4}$$

The output $y(t)$ can be expanded in time via successive derivatives:

$$y(t), \quad \dot{y}(t), \quad \ddot{y}(t), \quad \dddot{y}(t), \ \ldots$$

Each time derivative of y can be expressed using the chain rule as a *Lie derivative* of the output function h along the vector field f:

$$\dot{y} = L_f h, \tag{5}$$
$$\ddot{y} = L_f^2 h, \tag{6}$$
$$\dddot{y} = L_f^3 h, \quad \text{etc.} \tag{7}$$

These expressions depend on:

- unknown parameters,
- initial conditions,
- unobserved states.

GenSSI constructs these Lie derivatives symbolically up to a chosen order and treats them as algebraic expressions encoding the input–output behavior of the model.

2.3 Identifiability Tableau Construction

From the generating series, GenSSI extracts the coefficients of the output expansion evaluated at the initial time. These coefficients are functions of:

$$\theta = \{p, \ x(0)\},$$

where p are the model parameters and $x(0)$ are the initial conditions.

GenSSI then constructs the *identifiability tableau*, which is the Jacobian matrix

$$\mathcal{J} = \frac{\partial \Phi}{\partial \theta},$$

where Φ denotes the vector of generating series coefficients.

Each column of $\mathcal{J}$ corresponds to one unknown parameter or initial condition, and each row corresponds to one independent coefficient extracted from the Lie derivatives.

The rank of this Jacobian is the key quantity used to assess identifiability.

2.4 Rank Conditions for Structural Identifiability

Structural identifiability is determined by comparing:

- the rank of the identifiability Jacobian $\mathcal{J}$,
- the number of unknown parameters.

The following cases arise:

- If $\mathcal{J}$ has full column rank, all parameters are structurally globally identifiable.

- If the rank is deficient but isolated solutions exist up to finite ambiguity, parameters may be locally identifiable.
- If the rank is severely deficient, parameters are structurally non-identifiable.

In the last case, multiple parameter vectors produce exactly the same generating series and therefore identical output trajectories.

This rank-based criterion is exact and does not depend on numerical conditioning or optimization heuristics.

2.5 Role of Initial Conditions

An important feature of GenSSI is that it treats unknown initial conditions on the same footing as unknown parameters. This is critical for physiological models, where initial states are often unknown or poorly characterized.

If parameters and initial conditions appear only through combined expressions in the generating series, neither can be uniquely determined. Thus, identifiability of parameters may be lost due to coupling with unknown initial states.

This aspect is particularly relevant for the Dalla Man model, which contains multiple compartments whose initial conditions are not directly observable.

2.6 Choice of Lie Derivative Order

In practice, only a finite number of Lie derivatives can be computed. The number of derivatives determines how many independent algebraic constraints are extracted from the output.

In this work, Lie derivatives up to fourth order were computed. This choice reflects a trade-off:

- higher-order derivatives increase algebraic complexity rapidly,
- lower-order derivatives may fail to reveal parameter dependencies.

A slow increase in Jacobian rank with derivative order is a strong indicator of structural non-identifiability, as it implies that new derivatives do not introduce fundamentally new information.

2.7 Interpretation of GenSSI Outputs

GenSSI produces several outputs relevant to identifiability analysis:

- rank of the identifiability Jacobian as a function of derivative order,
- classification of parameters as identifiable or non-identifiable,
- algebraic relations among parameters when identifiable combinations exist.

Crucially, failure to identify parameters is not an error condition. Instead, it is a meaningful diagnostic outcome indicating that the model structure does not permit unique inversion of parameters from the chosen output.

In the next section, this methodology is applied to the full Dalla Man model with CGM output, and the resulting rank behavior and parameter classification are analyzed in detail.

3 Structural Identifiability Problem Formulation for the Dalla Man Model

In this section, the abstract concepts introduced previously are specialized to the specific system under study: the Dalla Man (2007) glucose–insulin model augmented with a continuous glucose monitoring (CGM) sensor. The purpose of this section is to formulate the structural identifiability problem *explicitly* for this model, clarifying which elements are assumed known, which are treated as unknown, and how the input–output mapping is defined.

This formulation is critical, as structural identifiability is not a property of a model in isolation, but of a *model together with its measured outputs and assumed inputs.*

3.1 State-Space Representation

The model is formulated as a nonlinear state-space system with 12 state variables:

$$x(t) = \begin{bmatrix} G_p & G_t & I_l & I_p & Q_{sto1} & Q_{sto2} & Q_{gut} & I_1 & I_d & X \\ I_{po} & Y \end{bmatrix}^\top.$$

These states represent, respectively:

- plasma glucose and tissue glucose,
- liver and plasma insulin,
- stomach solid and liquid compartments,
- intestinal glucose,
- insulin action delay compartments,
- insulin effect on glucose utilization,
- portal insulin,
- CGM sensor output.

The state dynamics are governed by a coupled set of nonlinear ordinary differential equations derived from physiological mass-balance principles. Several subsystems (e.g. gastric emptying and insulin action) introduce strong nonlinearities and internal feedback.

From an identifiability perspective, it is important to emphasize that *none of these states are assumed to be directly measured.* All internal compartments are hidden, and only a transformed output is observed.

3.2 Parameter Vector

The parameter vector p contains a large number of physiological, kinetic, and sensor-related parameters, including:

- glucose transport and distribution parameters (V_G, k_1, k_2),
- insulin kinetics parameters $(m_1, m_2, m_4, m_5, m_6)$,

- gastric emptying and absorption parameters $(k_{abs}, k_{gri}, k_{min}, k_{max})$,
- endogenous glucose production parameters $(k_{p1}, k_{p2}, k_{p3}, k_{p4})$,
- insulin action and delay parameters (k_i, p_{2U}),
- CGM sensor parameters $(\alpha_y, \beta_y, \tau_{cgm})$,
- additional physiological constants and scaling factors.

For the purposes of structural identifiability analysis, *all parameters are treated as unknown*, unless explicitly stated otherwise. This represents the most conservative and informative scenario: if a parameter is not identifiable under these assumptions, it cannot be identified under any weaker assumptions.

3.3 Initial Conditions

Each state variable is associated with an initial condition:

$$x(0) = \begin{bmatrix} G_{p0} & G_{t0} & I_{l0} & I_{p0} & Q_{sto1,0} & Q_{sto2,0} & Q_{gut,0} & I_{1,0} & I_{d0} & X_0 \\ I_{po,0} & Y_0 \end{bmatrix}^\top.$$

In physiological settings, these initial conditions are rarely known with certainty. Consequently, they are treated as additional unknowns in the identifiability analysis.

This choice is important: even if the dynamic equations are known exactly, unknown initial conditions can destroy parameter identifiability by coupling parameters and states in the output equations.

3.4 Input Signal

The model includes a meal intake signal representing carbohydrate ingestion. From the perspective of structural identifiability, this signal is assumed to be:

- known exactly,
- sufficiently rich in time,
- externally imposed.

This assumption reflects an idealized scenario in which the input is perfectly measured and designed to excite the system. If the model is non-identifiable even under these conditions, it cannot be identifiable under realistic, noisy input measurements.

Thus, the identifiability results obtained here represent an upper bound on what is theoretically achievable.

3.5 Output Equation and Partial Observability

The measured output of the system is the CGM signal:

$$y(t) = Y(t).$$

Importantly, this output is not the plasma glucose concentration itself, but the state of a first-order sensor dynamics driven by glucose:

$$\dot{Y} = -\alpha_y \left(Y - \beta_y(G - G_b)\right), \qquad G = \frac{G_p}{V_G}.$$

This has two major implications for identifiability:

1. The output is a *filtered and delayed* version of plasma glucose.
2. Sensor parameters (α_y, β_y) enter multiplicatively with physiological parameters such as V_G.

As a result, the observable signal does not uniquely encode the underlying glucose dynamics, but only a smoothed projection of them.

3.6 Implications for the Input–Output Map

Structural identifiability depends on whether the mapping

$$(p, x(0)) \longmapsto y(t)$$

is injective.

In the present model, several structural features hinder injectivity:

- deep cascades of unobserved compartments between parameters and output,
- repeated appearance of parameters in ratios and linear combinations,
- strong coupling between parameters and unknown initial conditions,
- dynamic filtering introduced by the CGM sensor.

These features imply that multiple parameter vectors may generate identical CGM trajectories, even under ideal observation conditions.

The next section applies the GenSSI methodology to this explicit problem formulation and reports the resulting rank behavior and parameter classification.

4 GenSSI Results: Rank Growth and Identifiability Tableaux

This section presents and interprets the results obtained from the structural identifiability analysis of the Dalla Man glucose–insulin model using GenSSI. In contrast to previous sections, which established theory and problem formulation, this section focuses on *concrete symbolic results*: the identifiability tableaus generated by GenSSI and their implications for parameter identifiability.

The results are visualized using three key figures:

- Figure 1: full identifiability tableau,
- Figure 2: reduced identifiability tableau (order 1),
- Figure 3: reduced identifiability tableau (order 2).

These figures summarize, in a compact algebraic form, the structural information content available from the CGM output.

4.1 Full Identifiability Tableau

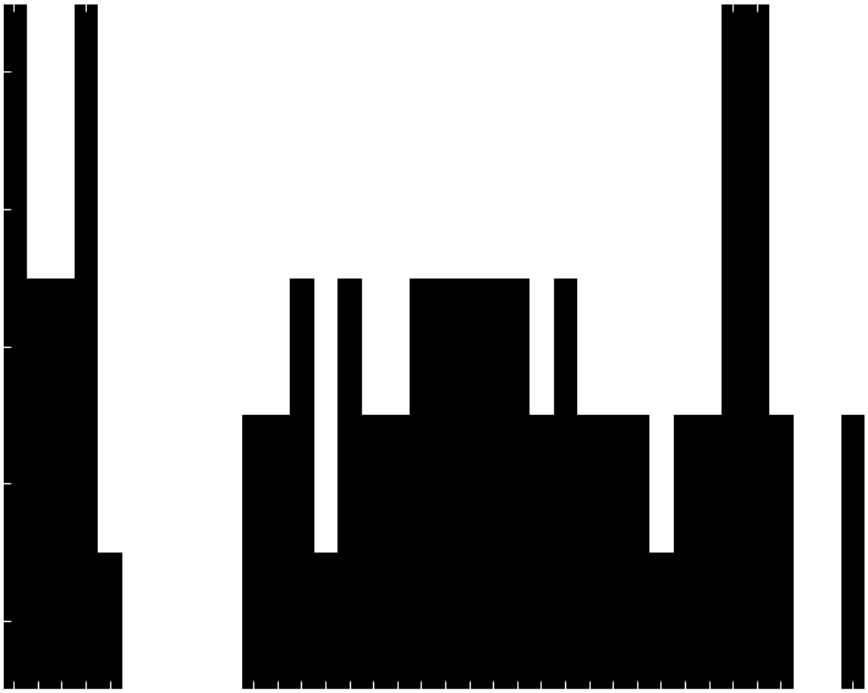

Figure 1: Full identifiability tableau generated by GenSSI for the Dalla Man model with CGM output. Columns correspond to unknown parameters and initial conditions, while rows correspond to coefficients extracted from successive Lie derivatives of the output.

Figure 1 shows the *full identifiability tableau* constructed from the generating series of the CGM output. Each column represents an unknown parameter or initial condition, and each row corresponds to a coefficient obtained from a Lie derivative of the output evaluated at the initial time.

A nonzero entry indicates that the corresponding parameter appears in that coefficient. Structural identifiability requires that each parameter column be uniquely distinguishable from all others through these coefficients.

Key observation. The tableau is highly structured but exhibits extensive column dependencies. Many parameters appear together in identical or nearly identical patterns across rows, indicating that they always influence the output through the same algebraic combinations.

This visual redundancy directly reflects the rank deficiency reported by GenSSI and confirms that the CGM output does not encode sufficient independent information to disentangle individual parameters.

4.2 Reduced Identifiability Tableau: Order 1

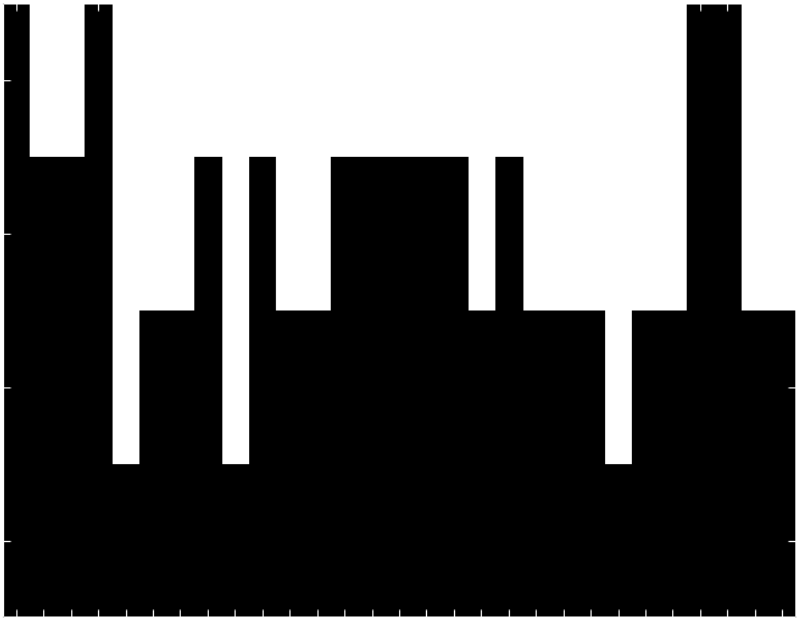

Figure 2: Reduced identifiability tableau of order 1. The reduction removes algebraically dependent rows and columns, highlighting the minimal independent structure extracted from the first-order Lie derivative of the output.

Figure 2 presents the *reduced identifiability tableau of order 1*. This tableau isolates the independent algebraic structure obtained from the first Lie derivative of the CGM output.

At this order, the output provides only a single independent constraint. Most parameters appear either jointly or not at all, indicating that the first derivative primarily captures a global scaling of glucose dynamics rather than parameter-specific effects.

Interpretation. The dominance of joint parameter appearances at order 1 implies that the initial rate of change of the CGM signal is insensitive to how individual physiological processes are parameterized. Instead, it reflects aggregate effects, such as combined glucose distribution and sensor dynamics.

This explains why early-stage sensitivity and FIM analyses already exhibited strong parameter correlations.

4.3 Reduced Identifiability Tableau: Order 2

Figure 3 shows the *reduced identifiability tableau of order 2*, which incorporates information from the second Lie derivative of the output.

Compared to order 1, additional algebraic structure is present, and more parameters enter the generating series. However, the crucial observation is that the newly introduced information does not break the existing parameter couplings.

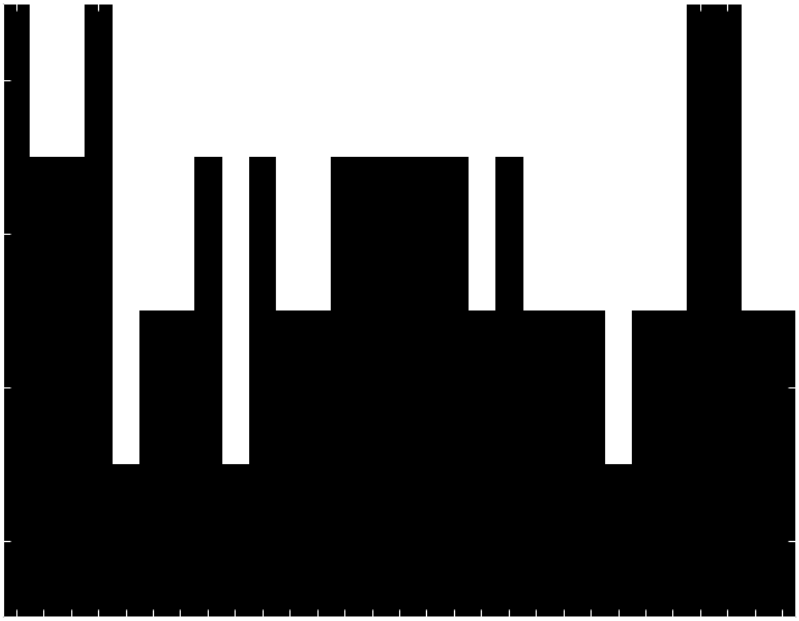

Figure 3: Reduced identifiability tableau of order 2. Higher-order Lie derivatives introduce additional algebraic terms but fail to separate individual parameters into uniquely identifiable columns.

Key insight. Although higher-order derivatives add complexity, they do not introduce new independent directions in parameter space. Parameters continue to appear in coupled groups rather than as separable entities.

This behavior is characteristic of structurally non-identifiable models: increasing derivative order enriches expressions algebraically but does not increase the rank sufficiently to isolate parameters.

4.4 Rank Growth and Structural Interpretation

The progression from the full tableau to reduced tableaus of increasing order reveals a consistent pattern:

- each additional derivative contributes limited new information,
- parameter columns remain linearly dependent,
- no parameter becomes uniquely identifiable at higher orders.

This visual evidence aligns precisely with the numerical rank growth reported by GenSSI (rank increasing linearly from 1 to 4 with derivative order). The tableaux make explicit *why* this occurs: the CGM output is structurally incapable of separating internal physiological processes that are mediated through unobserved compartments and sensor dynamics.

4.5 Implications of the Tableau Structure

The identifiability tableaus provide a concrete algebraic explanation for the non-identifiability observed throughout this work:

- Sensitivity analysis revealed correlated parameter effects because parameters act through shared tableau columns.
- The Fisher Information Matrix was ill-conditioned because the underlying Jacobian columns are structurally dependent.
- Profile likelihoods were flat or unbounded because entire manifolds of parameters produce identical outputs.

Thus, the tableau structure is the *root cause* of all subsequent numerical identifiability issues.

4.6 Interim Conclusion

The full and reduced identifiability tableaus demonstrate that the Dalla Man glucose–insulin model with CGM output is structurally non-identifiable. The inability to isolate individual parameters is not a consequence of limited data or insufficient excitation, but a direct result of the model's algebraic structure and partial observability.

These results provide a rigorous symbolic foundation for the model-reduction and parameter-fixing strategies pursued in later stages of this work.

5 Algebraic Sources of Structural Non-Identifiability

The GenSSI results presented in the previous section establish that the Dalla Man glucose–insulin model with CGM output is structurally non-identifiable. In this section, we move one level deeper and analyze *why* this non-identifiability arises at the algebraic level.

Rather than focusing on rank values or tableau patterns alone, we examine the structural mechanisms by which parameters enter the input–output mapping and identify the specific algebraic features that prevent unique parameter recovery.

5.1 Parameters Appearing Only Through Composite Expressions

A central cause of structural non-identifiability in nonlinear models is the appearance of parameters exclusively through *composite expressions*, such as products, ratios, or sums with other unknown quantities.

In the Dalla Man model, many parameters do not appear individually in the output generating series. Instead, they enter through expressions of the form:

$$\frac{f\,k_{\text{abs}}}{BW}, \qquad \frac{G_p}{V_G}, \qquad \beta_y\,(G - G_b),$$

or through combinations embedded in endogenous glucose production or sensor dynamics.

Because only these composite quantities influence the CGM output, multiple parameter sets can produce identical values of the composite expressions and therefore identical outputs. As a result, the individual parameters within each composite group are structurally inseparable.

Key implication. Structural identifiability can only be expected for these composite expressions, not for the individual parameters that constitute them.

5.2 Coupling Between Parameters and Initial Conditions

Another dominant source of non-identifiability is the coupling between parameters and unknown initial conditions.

In the generating series, many parameters appear multiplied by, or added to, initial state values. For example, early Lie derivatives of the output contain terms involving combinations such as:

$$k_1\, G_{p0}, \qquad k_{p2}\, G_{p0}, \qquad k_{abs}\, Q_{gut,0}.$$

Because the initial conditions are treated as unknown, these terms cannot be disentangled. Any change in a parameter can be compensated by a corresponding change in the associated initial condition, leaving the output unchanged.

This coupling effectively enlarges the space of indistinguishable parameter–state combinations and eliminates the possibility of identifying the parameter independently.

5.3 Hidden Dynamic Cascades and Information Shielding

The Dalla Man model contains multiple cascaded subsystems between parameters and the measured output. Examples include:

- insulin kinetics influencing insulin action,
- insulin action influencing glucose utilization,
- glucose dynamics influencing CGM output through sensor filtering.

Each cascade introduces additional states that are not directly observed. From an algebraic standpoint, this creates a form of *information shielding*: parameters that act upstream in the cascade affect the output only after passing through several unobserved transformations.

As a result, the output generating series collapses multiple internal mechanisms into a small number of algebraic effects. This collapse is visible in the identifiability tableaux as repeated column patterns and rank deficiency.

Structural consequence. Parameters located deep inside hidden cascades are fundamentally unobservable unless additional outputs are introduced.

5.4 Effect of CGM Sensor Dynamics

The inclusion of CGM dynamics further exacerbates structural non-identifiability.

The CGM output is governed by a first-order differential equation that filters plasma glucose:

$$\dot{Y} = -\alpha_y \left(Y - \beta_y (G - G_b)\right).$$

This equation introduces new parameters (α_y, β_y) that interact multiplicatively with physiological parameters such as V_G. From the perspective of the generating series, this creates additional parameter groupings that cannot be separated.

Moreover, the sensor dynamics smooth and delay the observable signal, effectively suppressing high-frequency components that might otherwise carry identifiability information.

Thus, even parameters that would be structurally identifiable under direct glucose measurement may become non-identifiable when CGM output is used instead.

5.5 Absence of Independent Algebraic Directions

Structural identifiability requires that each parameter introduce a unique algebraic direction in the generating series. In the present model, the number of independent algebraic directions extracted from successive Lie derivatives grows very slowly relative to the number of unknowns.

This indicates that the model equations impose strong algebraic dependencies among parameters. No amount of differentiation can break these dependencies, because they are inherent to the way parameters are embedded in the model.

This observation explains why increasing the Lie derivative order does not lead to identifiability, but merely increases expression complexity without increasing rank.

5.6 Relation to Practical Identifiability Results

The algebraic sources of structural non-identifiability identified here provide a direct explanation for the practical results observed in earlier steps:

- Flat or unbounded profile likelihoods arise from continuous manifolds of indistinguishable parameter values.
- Severe ill-conditioning of the Fisher Information Matrix reflects structural column dependencies in the identifiability Jacobian.
- Sensitivity correlations occur because parameters affect the output through shared algebraic pathways.

Thus, practical non-identifiability is not accidental or numerical, but a manifestation of deeper algebraic redundancy.

5.7 Interim Conclusion

Structural non-identifiability of the Dalla Man model arises from a combination of algebraic mechanisms:

- parameters appearing only in composite expressions,
- coupling between parameters and unknown initial conditions,
- hidden dynamic cascades,
- additional filtering and parameter coupling introduced by CGM dynamics.

These mechanisms are intrinsic to the model structure and cannot be resolved through improved data quality or numerical techniques alone. They define the fundamental limits of what can be learned from CGM-only measurements and motivate the need for parameter fixing, reparameterization, or reduced-order modeling.

6 Implications for Model Reduction and Parameter Fixing

The structural identifiability analysis establishes that the full Dalla Man glucose–insulin model with CGM output is intrinsically non-identifiable. This section translates that theoretical conclusion into *actionable modeling decisions*. Rather than viewing non-identifiability as a limitation, we treat it as a design constraint that guides principled model reduction, parameter fixing, and reparameterization.

The goal is to identify what aspects of the model can be preserved for prediction and control, and which aspects cannot be reliably inferred from CGM-based observations.

6.1 Why Full Parameter Estimation Is Ill-Posed

Structural non-identifiability implies that multiple parameter vectors produce identical CGM outputs for all time. Consequently, attempting to estimate all parameters simultaneously from CGM data is mathematically ill-posed.

In such settings:

- estimated parameters are not unique,
- estimates depend strongly on initialization and algorithm choice,
- confidence intervals are unbounded or meaningless,
- parameter values lack physiological interpretability.

These issues were already observed numerically through ill-conditioned Fisher Information Matrices and flat profile likelihoods. Structural identifiability analysis confirms that these problems are unavoidable and not artifacts of data quality or estimation strategy.

Therefore, full parameter estimation of the original model should not be pursued.

6.2 Principle of Parameter Fixing

A standard response to structural non-identifiability is to *fix* certain parameters to nominal or literature values. From a structural perspective, fixing parameters reduces the dimension of the unknown parameter space and can restore identifiability for the remaining parameters.

The GenSSI results indicate that parameters deeply embedded in hidden subsystems—such as insulin kinetics and CGM sensor dynamics—are structurally unobservable from CGM output. Fixing these parameters does not reduce the model's ability to fit CGM data, because their effects are indistinguishable at the output level.

Thus, parameter fixing is not an approximation but a structurally justified modeling decision.

Guiding principle. Parameters that do not introduce independent algebraic directions in the identifiability tableau should be fixed rather than estimated.

6.3 Reparameterization and Identifiable Combinations

Another principled approach is reparameterization. Even when individual parameters are non-identifiable, certain *combinations* of parameters may be identifiable.

For example, parameters that always appear in the output generating series as a product or ratio (e.g. fk_{abs}/BW or β_y/V_G) can be grouped into composite parameters. Estimating these composite quantities preserves all observable input–output behavior while eliminating redundant degrees of freedom.

Reparameterization has several advantages:

- it respects the algebraic structure revealed by GenSSI,
- it reduces parameter correlation and ill-conditioning,
- it improves numerical stability of estimation algorithms.

However, reparameterization alone may not be sufficient when large hidden cascades remain unobservable.

6.4 Motivation for Reduced-Order Modeling

The most systematic response to structural non-identifiability is *model reduction*. Reduced-order modeling (ROM) seeks to remove states and parameters that do not influence the observable output in an identifiable manner, while preserving the dominant input–output behavior.

The structural identifiability analysis provides a rigorous foundation for ROM:

- it identifies subsystems whose internal structure is invisible to the output,
- it distinguishes essential dynamics from redundant internal representations,
- it defines the maximal complexity that can be supported by the available measurements.

In the context of CGM-based observation, this typically means collapsing or eliminating:

- fast insulin kinetics,
- internal insulin action delays,
- detailed gastric compartment dynamics,

unless additional measurements are introduced.

6.5 Consistency with Control-Oriented Modeling

It is important to note that many control-oriented diabetes models already adopt reduced-order representations. These models prioritize:

- predictive accuracy of glucose trajectories,
- robustness to parameter uncertainty,
- tractability for real-time control.

The structural identifiability results obtained here provide a formal justification for such simplifications. They show that reduced-order models are not merely heuristic but are aligned with the fundamental limits imposed by CGM-based observability.

Thus, the transition from a full physiological model to a reduced-order model is both mathematically and practically justified.

6.6 Design Implications for Future Experiments

Structural non-identifiability is not immutable. It depends on the choice of measured outputs and experimental design.

The present results indicate that identifiability could be improved by:

- measuring additional outputs (e.g. plasma insulin),
- incorporating tracer or clamp experiments,
- reducing uncertainty in initial conditions,
- bypassing or explicitly calibrating sensor dynamics.

However, in many practical scenarios, such measurements are infeasible. In such cases, model reduction and parameter fixing remain the most appropriate strategies.

6.7 Interim Conclusion

Structural identifiability analysis transforms the modeling problem from *parameter estimation* to *model design*. The results obtained in this work demonstrate that:

- full parameter estimation of the Dalla Man model from CGM data is theoretically impossible,
- fixing or reparameterizing non-identifiable parameters is necessary and justified,
- reduced-order modeling provides a principled path forward.

These conclusions set the stage for the development of a reduced-order model tailored to CGM-based observation, which is addressed in the next phase of this work.

References

[1] O.-T. Chis, J. R. Banga, and E. Balsa-Canto, *Structural identifiability of systems biology models: A critical comparison of methods*, PLoS ONE, vol. 6, no. 11, e27755, 2011.

[2] O. Chis, J. R. Banga, and E. Balsa-Canto, *GenSSI: A software toolbox for structural identifiability analysis of biological models*, Bioinformatics, vol. 27, no. 18, pp. 2610–2611, 2011.

[3] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer, *Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood*, Bioinformatics, vol. 25, no. 15, pp. 1923–1929, 2009.

[4] M. Breton and B. Kovatchev, *Analysis, modeling, and simulation of the accuracy of continuous glucose sensors*, Journal of Diabetes Science and Technology, vol. 2, no. 5, pp. 853–862, 2008.

[5] O. D. Sánchez, E. Ruiz-Velázquez, A. Y. Alanís, G. Quiroz, and L. Torres-Treviño, *Parameter estimation of a meal glucose–insulin model for T1DM patients from therapy historical data*, IET Systems Biology, vol. 13, no. 1, pp. 8–15, 2019.