AALBORG UNIVERSITY

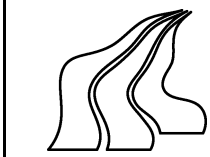
Bayesian Model Discrimination for Glucose-Insulin Homeostasis

by

Kim E. Andersen, Stephen P. Brooks and Malene Højbjerre

May 2004 R-2004-15

DEPARTMENT OF MATHEMATICAL SCIENCES
AALBORG UNIVERSITY
Fredrik Bajers Vej 7 G ▪ DK - 9220 Aalborg Øst ▪ Denmark
Phone: +45 96 35 80 80 ▪ Telefax: +45 98 15 81 29
URL: www.math.auc.dk/research/reports/reports.htm



ISSN 1399–2503 ▪ On-line version ISSN 1601–7811

Bayesian Model Discrimination for Glucose-Insulin Homeostasis

Kim E. Andersen* Stephen P. Brooks‡ Malene Højbjerre*

May 19, 2004

Abstract

In this paper we analyse a set of experimental data on a number of healthy and diabetic patients and discuss a variety of models for describing the physiological processes involved in glucose absorption and insulin secretion within the human body. We adopt a Bayesian approach which facilitates the reformulation of existing deterministic models as stochastic state space models which properly accounts for both measurement and process variability. The analysis is further enhanced by Bayesian model discrimination techniques and model averaged parameter estimation which fully accounts for model as well as parameter uncertainty. Markov chain Monte Carlo methods are used, combining Metropolis Hastings, reversible jump and simulated tempering updates to provide rapidly mixing chains so as to provide robust inference. We demonstrate the methodology for both healthy and type II diabetic populations concluding that whilst both populations are well modelled by a common insulin model, their glucose dynamics differ considerably.

1 Introduction

Quantitative assessment of the integrated glucose and insulin system within the human body is a vital component in the study of diabetes. The intravenous glucose tolerance test (IVGTT) is commonly used to study this system and involves sampling the glucose and insulin concentration in the blood following an intravenous glucose injection. Such studies provide valuable data on the body's response to increased blood sugar levels and can be used to improve the classification, prognosis and therapy of diabetes in human patients (Martin *et al* 1992).

Several models have been proposed in the medical literature to describe the human glucose and insulin systems. However, the most popular is the minimal model proposed for the glucose kinetics in Bergman *et al* (1979) and extended to the insulin kinetics by (Toffolo *et al* 1980). The minimal model couples the insulin and glucose processes by a non-observable latent process representing the effect of the insulin upon the glucose absorption rate. However, the minimal model possesses several drawbacks. First it is an entirely deterministic model and takes no account of measurement error or variability between individual patients. Secondly, some components of the model do not agree with current thinking as to the physiological processes actually involved in glucose absorption and insulin secretion within the human body. As a result, several modifications to the minimal model have been suggested as well as an entirely new model recently suggested by de Gaetano and Arino (2000) in

*Department of Mathematical Sciences, Aalborg University, Denmark
‡Statistical Laboratory, University of Cambridge, U.K.

which the glucose and insulin systems are coupled directly via integration, removing the need for the latent process used within the minimal model.

In this paper we examine data obtained from a study on both healthy and diabetic patients and use Bayesian model discrimination techniques to determine the model which best describes the physiological processes observed in each group. We consider the analysis of data arising from a series of intravenous glucose tolerance tests (IVGTT) in which a dose of glucose is administered intravenously to individual patients, and the subsequent glucose and insulin concentrations within the bloodstream are recorded at pre-specified though irregular times. See Pacini and Bergman (1986) for example. The data analysed in this paper are provided by Professor D. R. Owens, Diabetes Research Unit, University of Wales College of Medicine, Wales, and include 19 healthy young male individuals and 52 Type II diabetic male individuals.

We begin, in Section 2 with a description of the basic models proposed in the literature and extend them to account for both process and system errors by reformulating them as state space models. In Section 3 we construct the likelihood of these models, and in Section 4 we provide details of the Bayesian approach to the modelling used here, together with a simulation study to demonstrate the utility and robustness of the methods we propose. We present our results in Section 5 before providing additional interpretation and more general discussion in Section 6.

2 Data and Models

We begin with a brief description of diabetes and the human glucose-insulin system and how it differs between diabetic and healthy patients, before discussing the various models proposed in the literature for describing this system.

2.1 Diabetes and the Human Glucose-Insulin System

Diabetes is a progressive and incurable disease of the pancreas that causes an abnormally high level of glucose to build up in the blood. The two most common types of diabetes are caused by the pancreas' inability to produce insulin or by the body's inability to respond to raised insulin levels to stimulate glucose absorption. The former is referred to as *type I diabetes* whereas the latter is known as *type II diabetes*. In type II diabetics the pancreas produces greatly reduced quantities of insulin, and the fat, muscle, and liver cells all have a diminished ability to respond to the action of the insulin. With possibly more than 250 million people suffering worldwide from diabetes, most in the form of type II diabetes, the disease has reached epidemic proportions. In addition, recent prognoses say that this number may more than double within the next thirty years and that early detection of diabetes is vital for reducing the risk of obtaining the most severe symptoms such as impaired vision and gangrene in consequence of becoming diabetic. It is therefore important to improve the understanding of the human glucose-insulin system.

The human glucose-insulin system seeks to keep the levels of glucose and insulin within the bloodstream in a state of equilibrium. Both body tissues and the liver absorb glucose from the blood stream at a roughly constant rate p_1 depending upon the current level of glucose compared to a natural baseline, G_b. The rate p_1 then denotes the natural absorption rate of glucose from the bloodstream that occurs independently of the presence of insulin. Similarly insulin is constantly removed from the body in the urine at a roughly constant rate, n depending upon the current insulin level and it's deviation from the natural baseline level, I_b. This insulin elimination occurs in both healthy and diabetic patients.

However, the absorption of glucose from the bloodstream is also affected by the level of insulin and this additional absorption is known as the insulin action. In a healthy patient, ingestion of sugar (or, in our case, an intravenous glucose dose) rapidly elevates the glucose concentration $G(t)$ in the bloodstream initiating the pancreatic β-cells to secrete insulin. The level of insulin $I(t)$ within the bloodstream therefore also increases, triggering the absorption of glucose from the bloodstream into the adipose tissue, liver and muscles at some unobservable rate $X(t)$. $X(t)$ is the remote insulin action and reflects the proportion of glucose removed from the bloodstream as a result of the current insulin level at time t. This absorption process lowers the glucose concentration, affecting a reduction in pancreatic insulin production. This, in turn reduces the rate of glucose absorption and the downward spiral continues until the levels of both glucose and insulin return to normal basal levels. For a healthy individual this entire process normally take less than three hours.

Figure 1 provides plots of the glucose and insulin levels during an IVGTT for both a healthy and a type II diabetic patient. The rapid increase in insulin levels immediately after the administration of the glucose dose in the healthy patient can be clearly seen, as can the subsequent steady decline in both the glucose and insulin levels. Type II diabetic patients, on the other hand, produce only a small amount of insulin in reaction to glucose stimuli, and are not able to use this amount efficiently. Therefore no significant increase in the insulin level is seen after the glucose injection, only an unstable and small reaction is observed in Figure 1. The glucose level therefore rises dramatically and is lowered only very slowly at the constant absorption rate p_1.

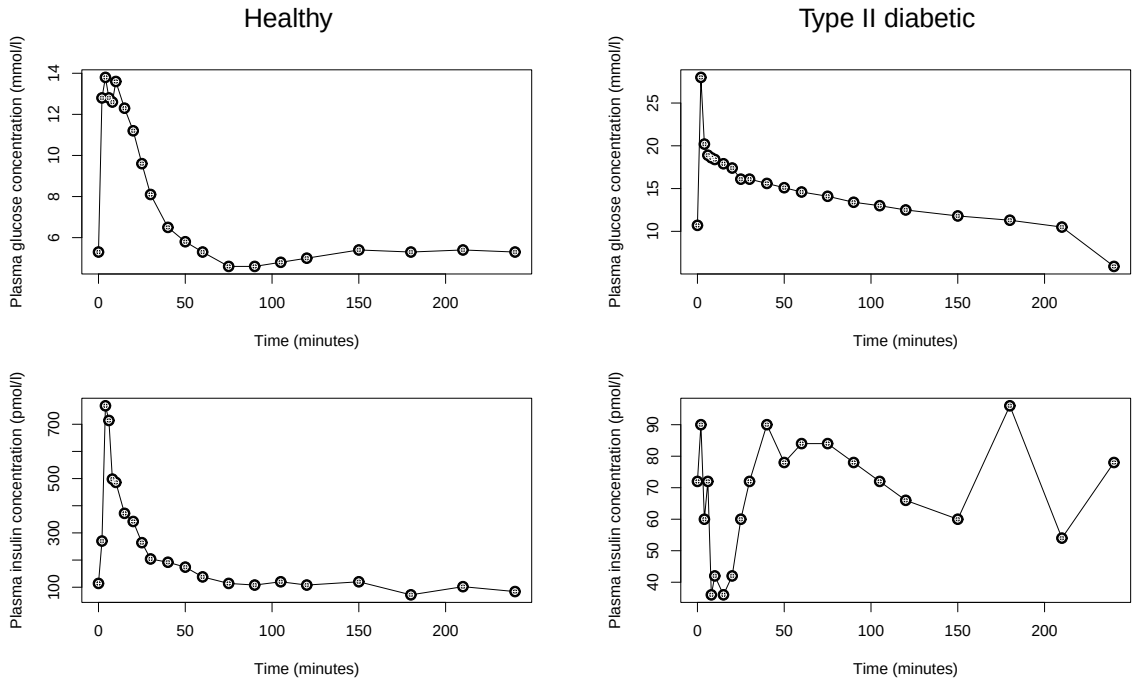


Figure 1: Glucose and insulin concentrations in a healthy person and a type II diabetic recorded over a 240-minute period after an intravenous glucose injection.

2.2 Minimal and Alternative Models

The medical literature provides a range of alternative models to describe the integrated glucose-insulin system. Here, we begin by describing a standard formulation of Bergman's minimal model

before discussing various extensions and finally the entirely different delay-differential model proposed by de Gaetano and Arino (2000).

2.2.1 The Minimal Model

As described earlier, the minimal model is probably the most popular of all of those proposed in the literature due, in part, to the availability of the software package MINMOD for fitting the minimal model to IVGTT data (Pacini and Bergman 1986).

The minimal model is wholly deterministic and can be represented by the compartment model depicted in Figure 2, or by the following set of differential equations

$$\begin{aligned} \dot{G}(t) &= -p_1(G(t) - G_b) - X(t)G(t), & G(0) &= G_0, \\ \dot{X}(t) &= -p_2 X(t) + p_3(I(t) - I_b), & X(0) &= 0, \\ \dot{I}(t) &= -n(I(t) - I_b) + \gamma J_+(G(t) - h)t, & I(0) &= I_0. \end{aligned} \tag{1}$$

Here $J_+(x) = x$ if $x > 0$ and zero otherwise.

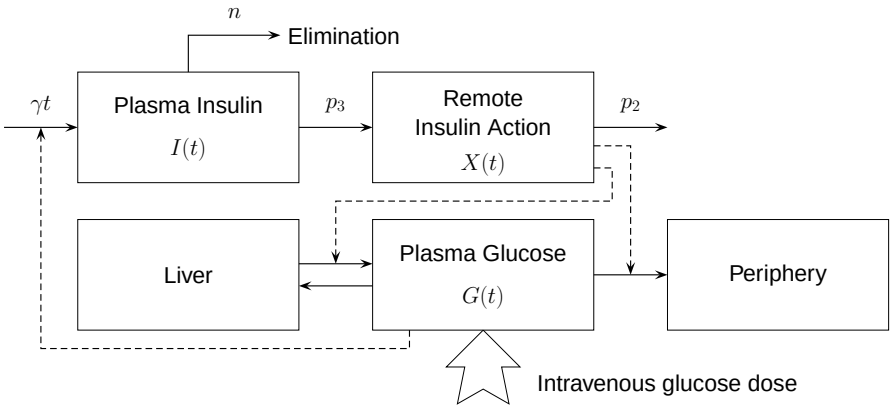


Figure 2: The Minimal Model describing the glucose and insulin system in an IVGTT study.

In this model, p_2 describes the decreasing level of insulin action with time whilst p_3 describes the rate in which insulin action is increased as the level of insulin deviates from the corresponding baseline. Similarly, γ denotes the rate at which insulin is produced as the level of glucose rises above some threshold level, h. In the context of IVGTT data, G_0 denotes the theoretical glucose concentration in plasma extrapolated to the time of glucose injection, i.e. at time $t = 0$, and I_0 denotes the corresponding theoretical insulin concentration. We label the insulin model in the minimal model as I_1 and the combined glucose and insulin action model as G_1.

Three factors, referred to as the metabolic portrait (Pacini and Bergman 1986), play an important role for the characterisation of an individual's glucose disposal: (1) *insulin sensitivity*, defined by $S_I = p_3/p_2$, specifies the insulin's capability to increase the glucose disposal to muscles, liver and tissue, (2) *glucose effectiveness*, defined by $S_G = p_1$, represents the ability of blood glucose to enhance its own disposal at basal insulin level, and (3) *pancreatic responsiveness*, defined jointly by $\varphi_1 = (I_0 - I_b)/[n(G_0 - G_b)]$ and $\varphi_2 = \gamma \times 10^4$, characterises the ability of the pancreas to secrete insulin in response to glucose stimuli. Failure in any of these factors may lead to impaired glucose tolerance, or, if severe, diabetes. The assessment of these three key factors has proven to be very useful in the classification, prognosis and therapy of diabetes (Martin *et al* 1992) and so will be the focus of our study here.

All four parameters can be estimated via the minimal model, and usually this is done using iterative non-linear weighted least squares estimation techniques in which the insulin is treated as known, i.e. the minimal model is regarded as a deterministic model composed only of the equations for G and X. There are many problems with this approach. For example, parameter estimates of insulin sensitivity obtained by the use of MINMOD are often estimated to be close or even equal to zero, especially for type II diabetic individuals, and the corresponding confidence intervals commonly include negative values for these strictly positive parameters. Furthermore φ_1 and φ_2 are not estimated in this approach, since the insulin model is fixed and thus the parameters n and γ are not estimated.

The original derivation of the minimal model assumes that G, X and I constitute a single dynamical system. However traditional analyses typically treat one of these processes as known and then estimate the others conditional on the first, thereby analysing only part of the system at any one time. In addition, the minimal model takes no account of either individual variability or process error. All of these problems will be addressed by our reformulation and analysis of the minimal model described later. However, we first describe several alternative models for the description of the underlying physiological process.

2.2.2 Variations on the Minimal Model

The glucose-insulin model has attracted many mathematical descriptions. Using a combination of Akaike information criteria and arguments based upon the minimisation of residual sums of squares, Bergman *et al* (1979) and Toffolo *et al* (1980) showed that the minimal model dominated a large range of alternative models. Nevertheless, several criticisms of the model remain, namely: (a) that the positive truncation J is physiologically questionable (Bergman *et al* 1981); and (b) that the multiplicative effect of time t in system I_1 suggests that the effect of circulating hyperglycemia on the rate of pancreatic secretion of insulin is proportional to the time elapsed from the glucose stimulus (Toffolo *et al* 1980), which is difficult to justify biologically. We therefore consider three additional variants of the insulin component of the model removing combinations of these two assumptions:

$$\begin{aligned}
\dot{I}(t) &= -n(I(t) - I_b) + \gamma J_+(G(t) - h), \quad I(0) = I_0, \\
\dot{I}(t) &= -n(I(t) - I_b) + \gamma (G(t) - h)t, \qquad I(0) = I_0, \\
\dot{I}(t) &= -n(I(t) - I_b) + \gamma (G(t) - h), \qquad I(0) = I_0.
\end{aligned}$$

These alternative insulin models are labelled I_2, I_3 and I_4 respectively.

2.2.3 The de Gaetano & Arino Model

De Gaetano and Arino (2000) undertake a thorough analysis of the standard minimal model and demonstrate that the solution set is unstable in that both glucose and insulin levels may rise indefinitely as time increases. Whilst it is possible to argue that observations are taken only over a finite time interval and so the model need only approximate reality over a finite time scale, the inherent instability of the system does occasionally lead to highly volatile solutions for particular data sets.

Another drawback with the minimal model is that the solution to the three differential equations that make up the minimal model and its variants is generally very difficult to obtain and traditional model-fitting techniques tend to treat the insulin observations as a fixed forcing function so that only the insulin action and glucose system equations need be solved.

In order to overcome these problems de Gaetano and Arino (2000) propose an alternative two-system model in which the coupling of the insulin and glucose process is made by the use of integration,

removing the need for the non-observable state variable X. As with the minimal model the de Gaetano and Arino (GA) model is fully deterministic, and can be illustrated as the compartment model given in Figure 3.

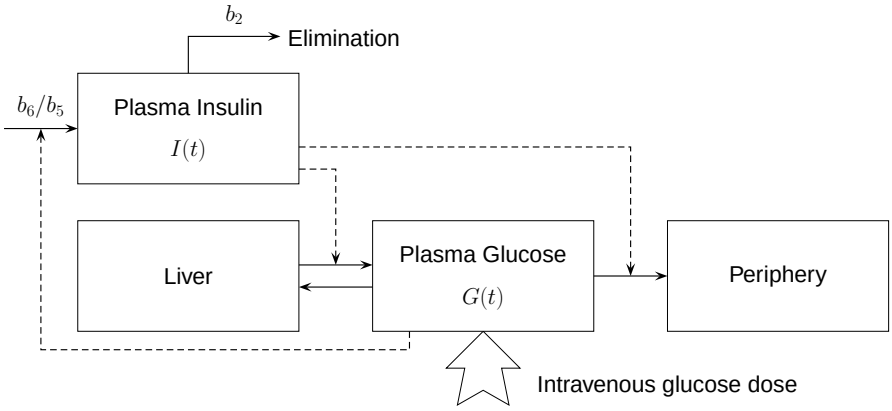


Figure 3: De Gaetano & Arino model describing the glucose and insulin system in an IVGTT study.

The following set of differential equations provide a mathematical description of the system.

$$\dot{G}(t) = -b_1 G(t) - b_4 I(t) G(t) + b_7, \quad G(0) = G_b + b_0,$$
$$\dot{I}(t) = -b_2 I(t) + \frac{b_6}{b_5} \int_{t-b_5}^{t} G(s)\, ds, \quad I(0) = I_b + b_3 b_0,$$

where $G(t) \equiv G_b$ for $t \in [-b_5, 0)$. This model is both easier to fit and has more stable solutions than the minimal model and its variants (Li *et al* 2001; Mukhopadhyay *et al* 2004).

The GA model introduces several new parameters, many of which can be expressed in terms of the parameters under the minimal model. The parameter b_0 is the theoretical increase in glucose concentration over basal level caused by the glucose injection. Parameter b_1 represents the constant glucose absorption rate and is equivalent to the parameter p_1 under the minimal model. Similarly, b_2 denotes the constant elimination rate of insulin and is equivalent to n under the minimal model. Parameter b_3 denotes the increase in insulin concentration per increase in glucose concentration at time zero caused by the injection of glucose, whilst b_4 denotes the glucose absorption rate per insulin concentration and b_5 the length of the time for which glucose concentration influences the current pancreatic insulin secretion. Finally, b_6 denotes the second-phase insulin release rate per average glucose concentration during the last b_5 minutes and b_7 denotes the constant increase in glucose concentration due to constant baseline liver glucose release. We label this glucose model as G_2 and the insulin model as I_5.

2.2.4 Relating the Minimal and GA Models

The GA model is, in fact over-parameterised and de Gaetano and Arino (2000) show that setting $b_7 = G_b(b_1 + b_4 I_b)$ and $b_6 = b_2 I_b / G_b$ provides an adequate description of the underlying physiological processes.

Comparing the GA model with the minimal model is it fairly easy to see that $b_4 = p_3/p_2$, $G_b + b_0 = G_0$ and $I_b + b_3 b_0 = I_0$. Thus, the metabolic portrait under the GA model is given by $S_I = b_4$, $S_G = b_1$, and $\varphi_1 = b_3/b_2$. The second pancreatic responsiveness parameter φ_2 cannot be estimated under the GA model.

3 Likelihood Construction

Having described the data and introduced the basic models, we next describe how these essentially deterministic models can be expressed as stochastic state space models that approximate the corresponding stochastic differential equation models.

We use the same approach as Andersen and Højbjerre (2003), in which an entire set of coupled differential equations is solved by first applying a logarithmic transformation of the system in order to bring the process entities onto the same scale so that they can share a common variance, then discretising and embedding them within a stochastic state space modelling framework allowing for random errors both within the system and measurement processes. This state space formulation then provides a natural framework for likelihood construction and therefore statistical analysis.

3.1 Logarithmic Transformation

We begin by taking a logarithmic transformation of all the glucose and insulin systems. Here, we let $g(t) = \log G(t)$, $x(t) = \log X(t)$ and $i(t) = \log I(t)$, so that $\dot{g}(t) = \dot{G}(t)/G(t)$, $\dot{x}(t) = \dot{X}(t)/X(t)$ and $\dot{i}(t) = \dot{I}(t)/I(t)$. Adopting this transformation and reparameterising in terms of the metabolic portrait parameters S_I, S_G, φ_1 and φ_2 as well as the GA model parameters b_0 and b_3, the minimal model becomes

$$\begin{aligned}
\dot{g}(t) &= -S_G(1 - G_b\, e^{-g(t)}) - e^{x(t)}, & g(0) &= \log(G_b + b_0),\\
\dot{x}(t) &= -p_2(1 - S_I(e^{i(t)} - I_b)e^{-x(t)}), & x(0) &\to -\infty,\\
\dot{i}(t) &= -\frac{b_3}{\varphi_1}(1 - e^{-i(t)}I_b) + 10^{-4}e^{-i(t)}\varphi_2 J_+(e^{g(t)} - h)t, & i(0) &= \log(I_b + b_3 b_0).
\end{aligned} \tag{2}$$

Similarly, the three alternative insulin models derived from the minimal model become

$$\begin{aligned}
\dot{i}(t) &= -\frac{b_3}{\varphi_1}(1 - e^{-i(t)}I_b) + 10^{-4}e^{-i(t)}\varphi_2 J_+(e^{g(t)} - h), & i(0) &= \log(I_b + b_3 b_0),\\
\dot{i}(t) &= -\frac{b_3}{\varphi_1}(1 - e^{-i(t)}I_b) + 10^{-4}e^{-i(t)}\varphi_2 (e^{g(t)} - h)t, & i(0) &= \log(I_b + b_3 b_0),\\
\dot{i}(t) &= -\frac{b_3}{\varphi_1}(1 - e^{-i(t)}I_b) + 10^{-4}e^{-i(t)}\varphi_2 (e^{g(t)} - h), & i(0) &= \log(I_b + b_3 b_0),
\end{aligned} \tag{3}$$

and the GA model becomes

$$\begin{aligned}
\dot{g}(t) &= -S_G(1 - G_b\, e^{-g(t)}) - S_I(e^{i(t)} - I_b G_b e^{-g(t)}), & g(0) &= \log(G_b + b_0),\\
\dot{i}(t) &= -\frac{b_3}{\varphi_1}\left[1 - \frac{I_b}{G_b b_5}e^{-i(t)}\int_{t-b_5}^{t} e^{g(s)}\, ds\right], & i(0) &= \log(I_b + b_3 b_0).
\end{aligned} \tag{4}$$

Thus, the parameter vector for any model obtained by combining either glucose process from the minimal model with any of the four insulin process $I_1, \ldots, I_4$ contains the parameters $(S_G, S_I, \varphi_1, b_3, b_0, G_b, I_b, \varphi_2, p_2, h)$, whereas the two models adopting the GA insulin process contain parameters $(S_G, S_I, \varphi_1, b_3, b_0, G_b, I_b, b_5)$. Later these parameter vectors will be extended to also include the precisions of the error processes on the underlying system and the measurements.

3.2 Model Discretisation and Stochasticity

We discretise the process in order to obtain an approximation to the continuous time process described by the different combination of differential equations given in the previous section.

We begin by dividing the sample period into suitably small time intervals in which the end points, $\Lambda = \{t_1, t_2, \ldots, t_{|\Lambda|}\}$, of the time intervals need not necessarily be equidistant, but are chosen so that the observation times all lie within Λ. We introduce a more convenient notation for $g(t_k)$, $x(t_k)$ and $i(t_k)$ by using t_k as a subscript and adding s as superscript indicating that they represent the underlying system process, e.g. $g(t_k) = g^s_{t_k}$. Discretising the processes and adding noise with variance dependent on the discretisation intervals, we derive the underlying system process for e.g. the log-transformed minimal model as

$$\begin{aligned} g^s_{t_k} &= f^g_{t_{k-1}} + \epsilon^{g^s}, \\ x^s_{t_k} &= f^x_{t_{k-1}} + \epsilon^{x^s}, \\ i^s_{t_k} &= f^i_{t_{k-1}} + \epsilon^{i^s}, \end{aligned}$$

where

$$\begin{aligned} f^g_{t_{k-1}} &= g_{t_{k-1}} - (t_k - t_{k-1})\big(S_G(1 - G_b\, e^{-g_{t_{k-1}}}) + e^{x_{t_{k-1}}}\big), \\ f^x_{t_{k-1}} &= x_{t_{k-1}} - (t_k - t_{k-1})p_2\big(1 - S_I(e^{i_{t_{k-1}}} - I_b)e^{-x_{t_{k-1}}}\big), \\ f^i_{t_{k-1}} &= i_{t_{k-1}} - (t_k - t_{k-1})\Big(\frac{b_3}{\varphi_1}(1 - e^{-i_{t_{k-1}}} I_b) - 10^{-4} e^{-i_{t_{k-1}}} \varphi_2 J_+ (e^{g_{t_{k-1}}} - h) t_{k-1}\Big), \end{aligned}$$

and where ϵ^{g^s}, ϵ^{x^s} and ϵ^{i^s} each follows a normal distribution with zero mean and common variance $\nu^{-1}(t_k - t_{k-1})$.

Thus, the corresponding stochastic differential equations become a discrete-time stochastic process where

$$\begin{aligned} g^s_{t_k} \,|\, g^s_{t_{k-1}}, x^s_{t_{k-1}}, \nu &\sim \mathcal{N}(f^g_{t_{k-1}}, \nu^{-1}(t_k - t_{k-1})), \\ x^s_{t_k} \,|\, x^s_{t_{k-1}}, i^s_{t_{k-1}}, \nu &\sim \mathcal{N}(f^x_{t_{k-1}}, \nu^{-1}(t_k - t_{k-1})), \\ i^s_{t_k} \,|\, i^s_{t_{k-1}}, g^s_{t_{k-1}}, \nu &\sim \mathcal{N}(f^i_{t_{k-1}}, \nu^{-1}(t_k - t_{k-1})). \end{aligned} \tag{5}$$

We can derive similar process for each of the remaining models obtained by combining the two glucose models with any of the five insulin models. The corresponding mean values for these process are provided in Appendix A, as the variances remain unchanged.

Note, that the larger the value of $|\Lambda|$, the smaller the time intervals in the discretisation and the more accurate the approximation of the system processes. Thus, the time intervals indicate a level of coarseness of the solution. In practice, at resolution level τ, we divide the interval $[t_k, t_{k+1}]$ into $\kappa_{k\tau}$ sections so that $\kappa_{k\tau}$ quantifies the coarseness and τ indexes the coarseness level. As an example, we might set $\kappa_{k\tau} = \tau$, so that each time period is divided into an equal number parts. We return to the definition of κ in the next section. Of course, for large values of $|\Lambda|$ the likelihood evaluations are extremely computational expensive.

The true underlying system described is not directly observable. We have instead a series of observations from which we wish to estimate the true underlying process. If we denote the observed log-values of the glucose and insulin levels at time t_k by $g^o_{t_k}$ and $i^o_{t_k}$ respectively, then observations are made at times $t_k \in \mathcal{T} \subseteq \Lambda$. Thus, the discretisation above is sufficiently fine so that observations occur on a subset of the discrete time points in Λ. We model the measurement error on g^o_t and i^o_t by random white noise processes with precisions ν_{g^o} and ν_{i^o}, respectively. Consequently the distributional assumptions for $g^o_{t_k}$ and $i^o_{t_k}$ are

$$\begin{aligned} g^o_{t_k} \,|\, g^s_{t_k}, \nu^{-1}_{g^o} &\sim \mathcal{N}(g^s_{t_k}, \nu_{g^o}), \\ i^o_{t_k} \,|\, i^s_{t_k}, \nu^{-1}_{i^o} &\sim \mathcal{N}(i^s_{t_k}, \nu_{i^o}). \end{aligned} \tag{6}$$

3.3 Population Modelling

From the sections above, we have two distinct glucose models, G_1 and G_2, and five different insulin models, $I_1, \ldots, I_5$. We combine these to obtain 10 different models to choose from when describing the glucose and insulin system. We use m as a model index, i.e. $m = 1, \ldots, 10$, where the former five models contain the glucose model G_1 and insulin model I_1 to I_5, and the latter five models contain the glucose model G_2 and insulin models I_1 to I_5.

These models describe the system for a specific individual but can be further extended to consider a population of individuals as follows. Suppose our population comprises L individuals and let the vector of system process values for individual j in model m be given by

$$\boldsymbol{\Phi}^s_{jm} = \begin{cases} \{g^s_{jmt_k}, x^s_{jmt_k}, i^s_{jmt_k}\}_{t_k \in \Lambda}, & \text{for } m = 1, \ldots, 5, \\ \{g^s_{jmt_k}, i^s_{jmt_k}\}_{t_k \in \Lambda}, & \text{for } m = 6, \ldots, 10. \end{cases}$$

The corresponding data vector of observations is denoted by $\boldsymbol{\Phi}^o_j = \{g^o_{jt_k}, i^o_{jt_k}\}_{t_k \in \mathcal{T}}$. Using the distributional assumptions in (5) and (6) the densities associated with the individual system and observation processes are simply products of univariate normal densities, so that

$$p_m(\boldsymbol{\Phi}^s_{jm} \,|\, \boldsymbol{\theta}_{jm}) \propto \begin{cases} \nu_j^{|\Lambda|} \exp(-V_m(\boldsymbol{\Phi}^s_{jm}, \boldsymbol{\theta}_{jm})), & \text{for } m = 1, \ldots, 5, \\ \nu_j^{3|\Lambda|/2} \exp(-V_m(\boldsymbol{\Phi}^s_{jm}, \boldsymbol{\theta}_{jm})), & \text{for } m = 6, \ldots, 10, \end{cases}$$

$$p_m(\boldsymbol{\Phi}^o_j \,|\, \boldsymbol{\theta}_{jm}, \boldsymbol{\Phi}^s_{jm}) \propto (\nu_{g^o_j} \nu_{i^o_j})^{|\mathcal{T}|/2} \exp(-W(\boldsymbol{\Phi}^o_j, \boldsymbol{\Phi}^s_{jm}, \boldsymbol{\theta}_{jm})),$$

where $\boldsymbol{\theta}_{jm}$ denotes the vector of parameters for individual j in model m (as stated in Section 3.1 but extended to include the individual precisions $\nu_j, \nu_{g^o_j}, \nu_{i^o_j}$) and corresponding variance functions are given by

$$V_m(\boldsymbol{\Phi}^s_{jm}, \boldsymbol{\theta}_{jm}) = \begin{cases} \frac{1}{2}\nu_j \sum_{t_k \in \Lambda} (g^s_{jmt_k} - f^g_{jmt_k})^2 + (x^s_{jmt_k} - f^x_{jmt_k})^2 + (i^s_{jmt_k} - f^i_{jmt_k})^2, & m = 1, \ldots, 5, \\ \frac{1}{2}\nu_j \sum_{t_k \in \Lambda} (g^s_{jmt_k} - f^g_{jmt_k})^2 + (i^s_{jmt_k} - f^i_{jmt_k})^2, & m = 6, \ldots, 10, \end{cases}$$

$$W(\boldsymbol{\Phi}^o_j, \boldsymbol{\Phi}^s_{jm}, \boldsymbol{\theta}_{jm}) = \frac{1}{2} \sum_{t_k \in \mathcal{T}} \nu_{g^o_j} (g^o_{jt_k} - g^s_{jmt_k})^2 + \nu_{i^o_j} (i^o_{jt_k} - i^s_{jmt_k})^2.$$

Thus, we assume that each individual has their own individual system process and associated parameters. However, we further assume that the values of the model parameters for each individual are drawn from some common population distribution. Essentially, we assume that the $\boldsymbol{\theta}_{jm}$ are random effects. Here, we assume that the individual system parameters in $\boldsymbol{\theta}_{jm}$ are independent and log-normally distributed, except for the precisions which we assume to be gamma distributed. This means that $\boldsymbol{\theta}_{jm}$ belongs to a common population distribution $p_m(\boldsymbol{\theta}_{jm} \,|\, \boldsymbol{\Psi}_m)$ which is the product of log-normal and gamma distributions and with population parameters $\boldsymbol{\Psi}_m$ consisting of the means and precisions for the log-normal distributed systems parameters, e.g. μ_{S_G} and τ_{S_G}, and the scale and shape parameters for the gamma-distributed precisions, e.g. α_ν and β_ν.

The likelihood for the whole population in model m can then be written as

$$L(\boldsymbol{\Psi}_m, \boldsymbol{\theta}_m, \boldsymbol{\Phi}^s_m \,|\, \boldsymbol{\Phi}^o) = \prod_{j=1}^{L} p_m(\boldsymbol{\Phi}^o_j \,|\, \boldsymbol{\theta}_{jm}, \boldsymbol{\Phi}^s_{jm}) p_m(\boldsymbol{\Phi}^s_{jm} \,|\, \boldsymbol{\theta}_{jm}) p_m(\boldsymbol{\theta}_{jm} \,|\, \boldsymbol{\Psi}_m),$$

where $\boldsymbol{\Phi}^s_m = (\boldsymbol{\Phi}^s_{1m}, \boldsymbol{\Phi}^s_{2m}, \ldots, \boldsymbol{\Phi}^s_{Lm})$ is the set of all L individual system processes in model m, $\boldsymbol{\Phi}^o = (\boldsymbol{\Phi}^o_1, \boldsymbol{\Phi}^o_2, \ldots, \boldsymbol{\Phi}^o_L)$ is the set of all individual observations and $\boldsymbol{\theta}_m = (\boldsymbol{\theta}_{1m}, \boldsymbol{\theta}_{2m}, \ldots, \boldsymbol{\theta}_{Lm})$ denotes the set of all individual system parameters in a given model m.

Now that we have a likelihood, we can undertake our Bayesian analysis as described in the next section.

4 Bayesian Analysis

Here, we undertake a Bayesian analysis of the data, both to facilitate the model discrimination problem and to ensure efficient propagation of the information in the data throughout the model. We begin by giving a brief overview of the Bayesian approach used here.

4.1 The Bayesian Approach

The Bayesian approach involves constructing a posterior distribution for the model parameters as a product of the joint probability distribution of the data (essentially the likelihood) with prior distributions representing our prior beliefs about the parameters before any data was observed.

Dropping the model dependence here for notational convenience, we have data $\mathbf{\Phi}^o$ and wish to estimate the unobserved process $\mathbf{\Phi}^s$, the individual parameter values $\boldsymbol{\theta}_j$ and the population parameters, $\mathbf{\Psi}$. The corresponding posterior distribution for these unknowns is therefore

$$\pi(\mathbf{\Psi}, \boldsymbol{\theta}, \mathbf{\Phi}^s \mid \mathbf{\Phi}^o) \propto \prod_{j=1}^{L} p(\mathbf{\Phi}_j^o \mid \boldsymbol{\theta}_j, \mathbf{\Phi}_j^s) p(\mathbf{\Phi}_j^s \mid \boldsymbol{\theta}_j) p(\boldsymbol{\theta}_j \mid \mathbf{\Psi}) p(\mathbf{\Psi})$$

for some appropriate (hyper)prior distribution $p(\mathbf{\Psi})$ (see Section 4.4 for discussion on priors).

In order to obtain inference about the parameters of interest we use Markov chain Monte Carlo (MCMC) methods (Brooks 1999; Gilks *et al* 1996) to obtain the corresponding posterior means and variances. Implementational details of the MCMC algorithm used here are provided in Appendix B. To ensure adequate performance, the MCMC scheme is tuned via an initial pilot simulation. For the population parameters (which, as described in Appendix B, are updated individually) we adapt the proposal scales by running the simulation for an initial N iterations and then calculating the mean acceptance ratio for the updates for each parameter in the model. For any parameter with a mean acceptance ratio less than 0.2, we halve the current proposal variance. For any parameter with a mean acceptance ratio greater than 0.5, we multiply the current proposal variance by 1.5. This process is continued until all mean acceptance ratios lie within $(0.2, 0.5)$, see Gelman *et al* (1996). For the individual parameters (which we update together as a block in the main simulation) we conduct an initial pilot simulation in which the parameters are updated individually and follow the process above to find suitable proposal scales. We then conduct a second simulation in which the parameters are updated together and rescale all of the proposal variances together following the scheme above in order to get a suitable acceptance rate for the full joint update described in Appendix B. For each of these pilot simulations, a value of $N = 10\,000$ appears to work well.

4.2 Model Uncertainty

As well as parameter estimation, we are interested in determining which combination of glucose and insulin models best describes the observed data. In order to address this question, we extend the posterior distribution above to include both parameter and model uncertainty, so that if $\boldsymbol{\theta}_{jm}$ and $\mathbf{\Psi}_m$ denote the individual and population parameters present in model m, then the joint posterior

distribution over both model and parameter space is given by

$$\pi(\boldsymbol{\Psi}_m, \boldsymbol{\theta}_m, \boldsymbol{\Phi}^s_m, m \,|\, \boldsymbol{\Phi}^o) \propto \prod_{j=1}^{L} p_m(\boldsymbol{\Phi}^o_j \,|\, \boldsymbol{\theta}_{jm}, \boldsymbol{\Phi}^s_{jm}) p_m(\boldsymbol{\Phi}^s_{jm} \,|\, \boldsymbol{\theta}_{jm}) p_m(\boldsymbol{\theta}_{jm} \,|\, \boldsymbol{\Psi}_m) p(\boldsymbol{\Psi}_m \,|\, m) p(m).$$

Here $p(m)$ denotes the prior model probability for model m and p_m the corresponding probability distribution under model m.

Posterior model probabilities are then obtained by integrating the joint posterior over the corresponding parameter space. These posterior model probabilities can then be used either to discriminate between competing models or to provide model-averaged predictive inference incorporating both parameter and model uncertainty. In particular, they can be used to provide model-averaged inference with regard to parameters such as S_I which retain a constant interpretation across all models. These posterior model probabilities are obtained by augmenting the MCMC algorithm discussed above to include reversible jump (RJ)MCMC (Green 1995) transitions for moves between models. Details of the RJMCMC transitions used in this analysis are provided in Appendix B.

The implementation of the RJMCMC scheme can be rather problematic because of the need to specify both a map between the parameter spaces for different models and a corresponding proposal distribution. For problems with only a small number of competing models the between-model transitions can be pilot-tuned beforehand (Richardson and Green 1997). However, for more general modelling problems more complex automated schemes are required (see Brooks *et al* 2003; Green 2003).

These transition problems can be eased somewhat by periodically "relaxing" the algorithm so that the accept-reject step becomes more lenient, allowing a wider variety of proposals to be accepted. This can be achieved through the introduction of a suitable tempering scheme.

4.3 Improving Mixing

The MCMC scheme described above can exhibit poor mixing properties with particularly low acceptance rates for between-model transition. In order to improve the mixing rate and, in particular, the ability for the simulation to move between models, we introduce a simulated tempering scheme (Marinari and Parisi 1992; Geyer and Thompson 1995) which involves further augmenting the posterior distribution above to depend upon an arbitrary temperature parameter. Conceptually, at high temperatures we want the corresponding posterior to be reasonably flat so that movement around the joint model and parameter space is unrestricted. We also require a temperature corresponding to the posterior of interest, the so-called cold distribution as it is from this distribution that inference will be obtained.

In our context, a very natural tempering scheme is obtained by equating temperature with the resolution level τ described in Section 3.2. Thus, our augmented posterior distribution incorporating resolution level becomes

$$\begin{aligned}\pi_\tau(\boldsymbol{\Psi}_m, \boldsymbol{\theta}_m, \boldsymbol{\Phi}^s_m, m \,|\, \boldsymbol{\Phi}^o) \propto \prod_{j=1}^{L} & p_{m,\tau}(\boldsymbol{\Phi}^o_j \,|\, \boldsymbol{\theta}_{jm}, \boldsymbol{\Phi}^s_{jm}) p_{m,\tau}(\boldsymbol{\Phi}^s_{jm} \,|\, \boldsymbol{\theta}_{jm}) \\ & \times p_m(\boldsymbol{\theta}_{jm} \,|\, \boldsymbol{\Psi}_m) p(\boldsymbol{\Psi}_m \,|\, m) p(\tau \,|\, m) p(m),\end{aligned}$$

where $p_{m,\tau}$ denotes the corresponding probability distribution associated with model m at resolution level τ. Here, we take $\kappa_{k\tau} = 2^{a_k}/\tau - 1$, where a_k is chosen sufficiently large so that all observations in $\boldsymbol{\Phi}^o$ are used for approximating the joint probability distribution. In order to account for the perturbation of the glucose-insulin system at time $t = 0$, we let a_k decrease gradually over time as to allow

for better approximation at the beginning of the experiment than at the end, i.e. we choose $a_k = 5$ for $t_{k+1} \leq 2$, $a_k = 4$ for $2 < t_{k+1} \leq 10$ and $a_k = 3$ otherwise. Here, we take $T = 3$ temperatures.

Since $p(m \,|\, \tau = 1)$ is pre-determined as our model prior (assuming that $\tau = 1$ corresponds to the cold distribution) and since $p(m \,|\, \tau = 1) \propto p(\tau = 1 \,|\, m)p(m)$, we fix $p(\tau = 1 \,|\, m) = 1$ for all m. The remainder of the prior weights can be determined so as to maximise the efficiency of the resulting algorithm. In practice, we use a pilot-tuning scheme to choose values for $p(\tau \,|\, m)$ for $\tau \neq 1$ which ensures that transitions between temperatures can occur with relative ease. Pilot-tuning can be a time-consuming process, so we adopt the following automated procedure. For fixed model m, let $c_\tau = p(\tau \,|\, m)$. Then, given a simulated tempering scheme with fixed model and temperatures $\tau = 1, \ldots, T$ we begin by setting $c_\tau = 1$ for all $\tau = 1, \ldots, T$. We fix $c_1 = 1$ to ensure a unique solution and, without loss of generality, restrict attention to tempering schemes under which transitions may only be made between neighbouring temperatures. We begin our simulated tempering scheme in some arbitrary starting position and after each temperature transition, we update the prior weights as follows. Suppose we currently have weights $c_1, \ldots, c_T$ and that we have just proposed a move from temperature τ to $\tau + 1$, say. Whether or not we accept this move we now divide all weights $c_{\tau+1}, \ldots, c_T$ by the acceptance ratio for the proposed move. Thus, if the move attracts a low acceptance probability, the normalising constant for the proposed model is increased. Similarly, since we have learnt only about the $\tau \to \tau + 1$ transition from this proposed move, all of the normalisation constants to the right of τ are increased to preserve their relative size. Similarly, if we propose to move from τ to $\tau - 1$, we would *multiply* the constants $c_\tau, \ldots, c_T$ by the corresponding acceptance ratio. This procedure is repeated until the weights settle to roughly constant values and, once the weights have been determined, they can then be used as inputs for the main simulation.

One advantage of this approach is that as the resolution level decreases (with increasing temperature) so does computational expense. Thus, transitions within the hotter temperatures not only enable more rapid movement (due to increasing coarseness of the likelihood) but are also computationally cheaper than transitions in the cold distribution. However, one problem with this approach is that though the prior weights can be tuned so as to allow adequate mixing between temperatures, the decrease in resolution may not on its own be sufficient to achieve satisfactorily rapid mixing between models in the hottest temperatures. To increase mixing between models we further adapt the posterior distribution so that as the temperature increases, not only does the resolution decrease, but also the affect of the likelihood decreases, so that the posterior looks increasingly like the prior. This is most easily achieved by taking

$$\pi_\tau(\boldsymbol{\Psi}_m, \boldsymbol{\theta}_m, \boldsymbol{\Phi}^s_m, m \,|\, \boldsymbol{\Phi}^o) \propto \prod_{j=1}^{L} \Big(p_{m,\tau}(\boldsymbol{\Phi}^o_j \,|\, \boldsymbol{\theta}_{jm}, \boldsymbol{\Phi}^s_{jm}) p_{m,\tau}(\boldsymbol{\Phi}^s_{jm} \,|\, \boldsymbol{\theta}_{jm}) \Big)^{s(\tau)} \times p_m(\boldsymbol{\theta}_{jm} \,|\, \boldsymbol{\Psi}_m) p(\boldsymbol{\Psi}_m \,|\, m) p(\tau \,|\, m) p(m),$$

where $s(\tau) = 2^{-(\tau-1)n}$ for $n > 0$. Clearly, s decreases as τ increases so the influence of the likelihood terms on the posterior similarly decreases. In addition, when $\tau = 1$, $s = 1$ so that the cold distribution remains the distribution of interest.

Having described the algorithm to be used to undertake the analysis, we next describe the priors to be used before undertaking a simulation study to examine the algorithm's performance.

4.4 Priors

The only parameters requiring prior distributions are the population parameters $\boldsymbol{\Psi}_m$. We assume normal priors for all mean parameters and gamma priors for the corresponding precisions. These priors are adapted from the results of previous studies (e.g. Pacini and Bergman 1986) and from

information elicited from collaborators at Novo Nordisk A/S. However, for μ_{G_b} and μ_{I_b} the priors are based on observations of the pre-injection levels of the glucose and insulin concentrations. The MINMOD programme fixes these base levels of glucose and insulin, but the parameter estimation there is very sensitive to these fixed values. Therefore we impose fairly informative prior distributions for μ_{G_b} and μ_{I_b} based upon observations of each patient's glucose and insulin levels taken 30 minutes and again 15 minutes before the glucose injection. These observations are obviously excluded from the likelihood terms used in the analysis. Finally for the population parameters corresponding to the individual precisions, we assume vague Gamma priors.

In terms of models we have no preference *a priori* for one model over another, and so we assume equal prior probabilities across all models.

4.5 Simulation Study

In order to assess the performance of our methodology both for parameter estimation and for model discrimination, we conduct a brief study in which simulated data is analysed and compared with the truth. Here, we construct an artificial population distribution $p(\boldsymbol{\Psi}_m \,|\, m)$ from which simulated IVGTT data can be drawn under a fixed model m. The MCMC simulation algorithm is then applied to these data to assess the method's performance.

We simulated four different sets of data consisting of 25 subjects from the same population distribution under the four different models $m = 1, 5, 6,$ and 10. The four chains were run for 3 million iterations each and exhibited good estimation properties. In particular, the algorithm was easily able to distinguish the true model for each of the four data sets, with posterior model probabilities of more that 90% for the true model. See Table 1 for details. Kass and Raftery (1995) discuss the interpretation of the ratio of the posterior to prior odds which, in this case as we have flat priors, is simply the ratio of two posterior model probabilities. They conclude that a ratio greater than 3 suggests some evidence whereas a ratio of greater than 20 represents very strong evidence in favour of the most probable model in favour of the other. Each of the Bayes factors corresponding to the posterior model probabilities in Table 1 suggests strong evidence in favour of the correct model over the next best alternative. The posterior population and individual parameter estimates are also extremely accurate, with posterior means close to the true values and high posterior precision. We therefore conclude that our method provides a robust and accurate analysis of IVGTT data.

	Posterior model probability									
True model	1	2	3	4	5	6	7	8	9	10
1	**0.94**	0.02	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.05	**0.95**	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	**0.91**	0.06	0.02	0.01	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	**0.97**

Table 1: The achieved cold posterior model probabilities for four simulations under known 'true' models.

In order to examine the influence of the population size on the model discrimination we conduct a final simulation study. Here, we subsample at random three groups of subjects from the 25 simulated subject from model $m = 10$. These groups comprise 10, 15 and 20 subjects, respectively. Three additional Markov chains were run and essentially all chains exhibited similar behaviour implying so that the models visited during the cold distribution were almost identical. Parameter estimates were also consistent and accurate, though the larger group had smaller credible intervals from which we may conclude, as expected, that the inclusion of more data provides better estimates of population and individual parameters.

5 Results

It is worth noting here, that the minimal model is based upon the assumption of a single well-mixed pool of glucose (Pacini and Bergman 1986), which requires the injected glucose load to be fully circulated within the venous system before glucose measurements can be used for the actual model fitting. It is in general believed that this mixing phase lasts approximately 10 minutes and consequently the MINMOD software package zero-weights all glucose observations taken at times $t < 8$ minutes. However, exploratory pilot simulation studies indicate that with a logarithmic transformation of the data the observations taken at $t = 2, 4, 6$ minutes exhibit similar behaviour around the fitted mean values to those taken at $t \geq 8$ minutes. We expect these glucose observations to contain important information which may be lost if not included in our simulations. The observations taken at $t = 0$ must be discarded as to allow for adequate estimates of G_0 and I_0. This applies to our approach as well as the MINMOD package.

For adequate model discrimination of the two populations, we follow the procedure outlined in Section 4 noting that we used a small degree of fine tuning to obtain the tempering parameter n. It was found that suitable tempering schemes for both populations was obtained for $n = 1.5$. Our MCMC simulation is run for a total of 5 million iterations during which we update the temperature every 50 iterations and the model every 500.

Output from the two chains is shown in Figure 4, from which we see that the chains appear to move fairly easy between resolution levels. Shown also are the models visited at each resolution level τ. It is clear that between-model mixing is excellent at resolution level $\tau = 4$. Note that, as expected, the stationary distributions become more and more centred on only a few models as the resolution level is decreased. For the healthy population we obtain models 2 and 3 as the most probable models, whereas for the type II diabetic population the most probable models are 8 and 10. See Table 2 for the corresponding posterior model probabilities. In both cases the support for the *a posteriori* most probable model is significant.

		Posterior model probability									
Population	Resolution	1	2	3	4	5	6	7	8	9	10
Healthy	1	0.00	0.21	**0.76**	0.03	0.00	0.00	0.00	0.00	0.00	0.00
	2	0.10	0.15	0.23	0.20	0.08	0.02	0.07	0.08	0.05	0.01
	4	0.11	0.13	0.12	0.10	0.10	0.16	0.07	0.05	0.09	0.06
Diabetic	1	0.00	0.00	0.07	0.00	0.01	0.00	0.00	**0.75**	0.00	0.17
	2	0.05	0.09	0.15	0.07	0.08	0.04	0.13	0.20	0.08	0.11
	4	0.13	0.05	0.11	0.10	0.07	0.08	0.14	0.13	0.12	0.09

Table 2: The achieved posterior model probabilities under the three degrees of resolution.

Clearly, the variants of Bergman's minimal model dominates all others for the healthy data set as there is no posterior support for the GA model. Overall, we would conclude that the minimal model without the positive truncation is the most appropriate model for the healthy data though, with approximately 25% of the posterior mass on alternative models, model averaged predictive estimates may be useful as they would properly reflect this element of uncertainty. For the type II diabetic population, we see that the same insulin model as for the healthy population is strongly identified in the posterior with a posterior model probability of 82% (summing over models 3 and 8). However, here we find a large degree of evidence for the use of the glucose part of the GA model (92%) as opposed to the minimal glucose model observed for the healthy patients. Comparing the two populations we deduce that insulin model I_3 is the best to describe both populations, however for the glucose model we need to distinguish between the population, since the GA model is the best for the type II diabetic population, whereas for the healthy population the minimal model is the best.

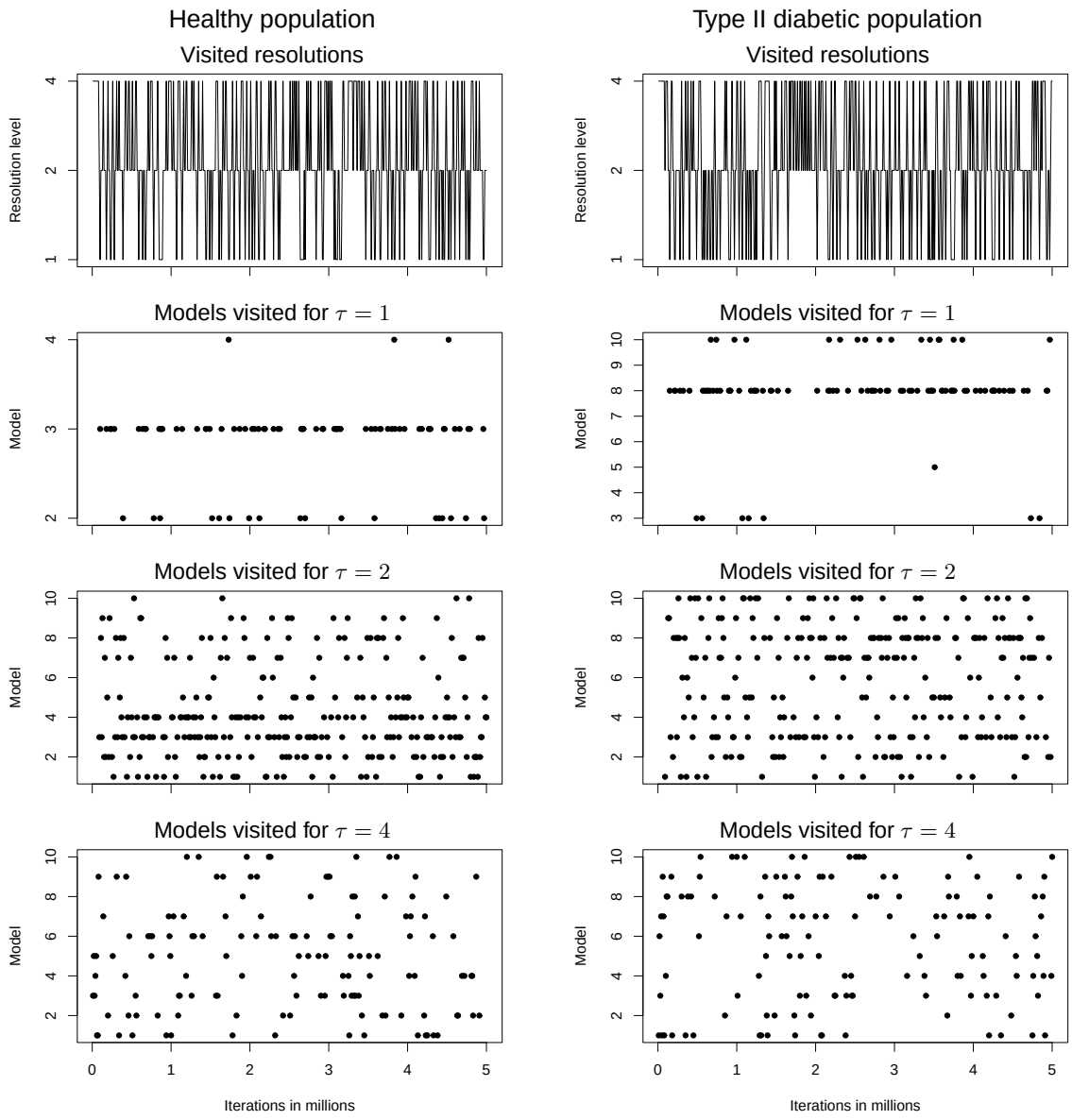


Figure 4: Simulation output for the healthy and diabetic populations showing the trace plots for the resolution and models within-resolution parameters over time.

Table 3 provides the posterior means and corresponding credible intervals of the population parameters for the two populations under the two most probable models identified, together with model-averaged values. For the healthy population, each of these parameters retain a consistent interpretation across the three models identified and we can see that most parameters vary only little between models. However, for the type II diabetic population we obtain quite different estimates for the baseline insulin level between the two top models, for example, with correspondingly large credible intervals.

Healthy	Model $m = 3$ (76%)	Model $m = 2$ (21%)	Model averaged
S_G	$0.025_{(0.020,0.031)}$	$0.025_{(0.020,0.032)}$	$0.026_{(0.019,0.032)}$
$S_I \cdot 10^5$	$1.638_{(0.785,2.387)}$	$1.746_{(0.583,2.699)}$	$1.693_{(0.607,2.761)}$
φ_1	$323.775_{(301.937,342.026)}$	$322.570_{(299.497,348.975)}$	$322.160_{(295.953,349.681)}$
φ_2	$540.488_{(451.745,625.243)}$	$538.327_{(449.116,610.311)}$	$533.957_{(447.655,620.240)}$
G_b	$4.351_{(3.069,5.259)}$	$4.275_{(3.164,5.010)}$	$4.375_{(3.407,5.289)}$
I_b	$53.983_{(48.600,59.019)}$	$54.446_{(47.362,60.195)}$	$54.593_{(47.476,61.107)}$
b_0	$4.531_{(4.402,4.641)}$	$4.541_{(4.406,4.682)}$	$4.539_{(4.402,4.694)}$
b_3	$51.229_{(46.415,56.132)}$	$50.444_{(46.526,54.894)}$	$50.723_{(46.803,54.531)}$
p_2	$0.020_{(0.018,0.022)}$	$0.020_{(0.017,0.022)}$	$0.020_{(0.017,0.022)}$
h	$6.941_{(4.142,9.748)}$	$6.752_{(4.046,9.984)}$	$6.875_{(3.782,10.205)}$
ν	$4165.656_{(3232.793,5101.553)}$	$4102.646_{(3200.843,5180.017)}$	$4143.755_{(3112.828,5253.806)}$
Diabetic	**Model $m = 8$ (75%)**	**Model $m = 10$ (17%)**	**Model averaged**
S_G	$0.008_{(0.005,0.011)}$	$0.007_{(0.005,0.009)}$	$0.007_{(0.005,0.010)}$
$S_I \cdot 10^5$	$0.106_{(0.041,0.179)}$	$0.105_{(0.043,0.170)}$	$0.102_{(0.038,0.177)}$
φ_1	$157.383_{(139.250,175.218)}$	$157.814_{(142.135,171.744)}$	$158.970_{(139.552,175.266)}$
φ_2	$776.508_{(535.876,954.114)}$	–	$754.391_{(537.810,958.524)}$
G_b	$9.994_{(7.028,11.897)}$	$10.250_{(8.171,11.974)}$	$10.128_{(7.013,12.436)}$
I_b	$64.860_{(1.639,331.458)}$	$118.951_{(4.911,390.844)}$	$77.614_{(2.079,351.639)}$
b_0	$20.656_{(19.402,22.895)}$	$20.851_{(19.511,22.776)}$	$20.624_{(19.397,22.573)}$
b_3	$2.966_{(2.410,3.730)}$	$2.997_{(2.650,3.438)}$	$2.978_{(2.392,3.520)}$
b_5	–	$14.636_{(11.363,18.155)}$	$14.642_{(10.812,18.200)}$
p_2	–	–	$0.007_{(0.003,0.011)}$
h	$0.020_{(0.010,0.027)}$	–	$0.019_{(0.008,0.027)}$
ν	$2760.839_{(1852.464,3431.459)}$	$2817.343_{(2333.624,3246.538)}$	$2723.238_{(1670.877,3435.933)}$

Table 3: Posterior mean and 95% credible intervals (in subscript) for the population parameters in the most probable models in the cold distribution. Given is also the model-averaged posterior mean and a model-averaged 95% credible intervals for the same parameters.

Possibly one of the most considerable consequences of the analysis, is the estimation of the healthy population mean glucose process, together with a measure of our associated uncertainty. A graphical representation of the true underlying model-averaged glucose and insulin processes are provided in Figure 5. Note how the diabetic blood glucose reaction to glucose stimuli is more pronounced than for the healthy population, which may be explained by the healthy subjects being able to dispose of glucose instantaneously. We also see, for the diabetics, that the insulin rises immediately and decreases very slowly as the pancreas remains producing insulin in order to dispose of the glucose.

Normal curves such as these can be used to monitor a particular patients' reaction to the IVGTT test to determine whether or not the patient is in the risk group for developing type II diabetes. Attaining a reliable estimate of the true underlying glucose disposal process facilitates a more robust classification procedure which is further improved by the provision of the corresponding error bounds. It is on the basis of such plots that the reaction to treatment, for example, is judged and so obtaining reliable estimates is of paramount importance.

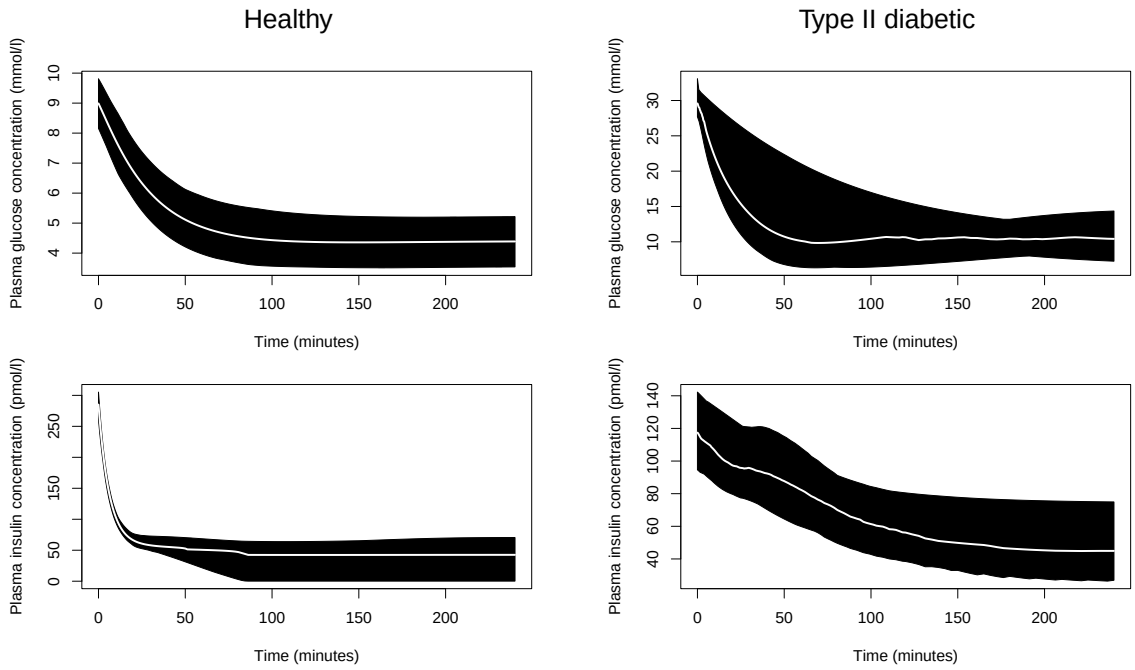


Figure 5: Posterior mean (white line) and 95% credible intervals superimposed in black for the glucose and insulin levels for the healthy and the type II diabetic populations.

6 Discussion

In this paper we consider the problem of discriminating between several proposed variations of the minimal models in addition to the new model proposed by de Gaetano and Arino (2000). Thus we obtain a total of ten different models subject for discrimination. The models have been logarithmically transformed and recast as state space models treating each of them as unified systems in a Bayesian setting and allowing for incorporation of noise on both the glucose and insulin concentrations, together with noise on the differential equations. The models we use here are therefore stochastic versions of those traditionally adopted in the literature and we use Bayesian model-fitting techniques to deal with this stochasticity.

For the healthy population we see that the model describing the data best is model 3 combining the minimal glucose model G_1 and the insulin model I_3 which has no positive reflection. This is, in fact, the model suggested by Bergman *et al* (1981). However, for the type II diabetic population we find that model 8 describes the data the best, and combines the same insulin model as for the healthy patients with the glucose model proposed by de Gaetano and Arino (2000). This supports the criticism by Pillonetto *et al* (2002) that the minimal model does not adequately describe the glucose-insulin system for type II diabetics because of the inherently low response in insulin concentration observed for these patients.

It is worth noting that though the conclusions drawn from the analyses presented here may well apply to the original deterministic models, we have fitted stochastic versions of those original models in this paper. This means also that comparisons with metabolic portraits estimated from our stochastic version of the minimal model should be compared to estimates obtained via MINMOD with care. However, as demonstrated in Andersen and Højbjerre (2003) the Bayesian approach tends to provide far more precise parameter estimates that those obtained via MINMOD and are more robust to the instabilities inherent in the underlying model. Further, Cobelli *et al* (1999) showed that estimates of

the metabolic portrait obtained by MINMOD are typically biased compared to estimates obtained by expensive and time consuming clamp studies, and that a Bayesian approach to the minimal model improves the accuracy of the estimates obtained. This is partly caused by MINMOD incorrectly estimating very low insulin sensitivity indices. It would be interesting to apply the approach presented in this paper on subjects that have undergone both an IVGTT study as well as a clamp study to further investigate the accuracy of our estimates, though the simulation studies do already appear to suggest that accuracy is very high.

Acknowledgements

The authors gratefully acknowledge Dr. Owens, Diabetes Research Unit, University of Wales College of Medicine, Wales, for providing the data. The authors also gratefully acknowledge the support of Novo Nordisk A/S for partially funding the work of KEA and MH. The work of SPB was funded by the UK Engineering and Physical Sciences research Council under grant number AF/000537.

References

Andersen, K. E. and M. Højbjerre (2003), A Population based Bayesian Approach to the Minimal Model of Glucose and Insulin Homeostasis. Technical Report R-2003-25, Department of Mathematical Sciences, Aalborg University

Bergman, R. N., Y. Z. Ider, C. R. Bowden and C. Cobelli (1979), Quantitative estimation of insulin sensitivity. *American Journal of Physiology* **236**, E667 – E677

Bergman, R. N., L. S. Phillips and C. Cobelli (1981), Physiologic Evaluation of Factors Controlling Glucose Tolerance in Man. *Journal of Clinical Investigation* **68**, 1456 – 1467

Brooks, S. P. (1999), Bayesian Analysis of Animal Abundance Data via MCMC. In J. M. Bernardo, A. F. M. Smith, A. P. Dawid and J. O. Berger (eds.), *Bayesian Statistics 6*, pp. 723–731, Oxford University Press

Brooks, S. P., P. Giudici and G. O. Roberts (2003), Efficient Construction of Reversible Jump Proposal Distributions (with discussion). *Journal of the Royal Statistical Society, Series B* **65**, 3–55

Cobelli, C., A. Caumo and M. Omenetto (1999), Minimal model S_G overestimation and S_I underestimation: improved accuracy by a Bayesian two-compartment model. *American Journal of Physiology* **277**, E481 – 8

de Gaetano, A. and O. Arino (2000), Mathematical modelling of the intravenous glucose tolerance test. *Journal of Mathematical Biology* **40**, 136 – 168

Gelman, A., G. O. Roberts and W. R. Gilks (1996), Efficient Metropolis Jumping Rules. In J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds.), *Bayesian Statistics 5*, pp. 599–608, New York: Oxford University Press

Geyer, C. J. and E. A. Thompson (1995), Annealing Markov Chain Monte Carlo with Applications to Ancestral Inference. *Journal of the American Statistical Association* **90**, 909–920

Gilks, W. R., S. Richardson and D. J. Spiegelhalter (1996), *Markov Chain Monte Carlo in Practice*. Chapman and Hall

Green, P. J. (1995), Reversible Jump MCMC Computation and Bayesian Model determination. *Biometrika* **82**, 711–732

Green, P. J. (2003), Trans-dimensional Markov chain Monte Carlo. In P. J. Green, N. L. Hjort and S. Richardson (eds.), *Highly Structured Stochastic Systems Volume*, pp. 179 – 198, Oxford University Press

Kass, R. E. and A. E. Raftery (1995), Bayes Factors. *Journal of the American Statistical Association* **90**, 773–795

Li, J., Y. Kuang and B. Li (2001), Analysis of IVGTT glucose-insulin interaction models with time delay. *Discrete and Continous Dynamical Systems – Series B* **1**, 103 – 124

Marinari, E. and G. Parisi (1992), Simulated Tempering: A New Monte Carlo Scheme. *Europhysics letters* **19**, 451–458

Martin, B. C., J. H. Warram, A. S. Krolewski, R. N. Bergman, J. S. Soeldner and C. R. Kahn (1992), Role of glucose and insulin resistance in development of type 2 diabetes mellitus: results of a 25-year follow-up study. *The Lancet* **340**, 925 – 929

Mukhopadhyay, A., A. De Gaetano and A. Arino (2004), Modelling the intra-venous glucose tolerance test: a global study for a single-distributed-delay model. *Discrete and Continous Dynamical Systems – Series B* **4**, 407 – 417

Pacini, G. and R. N. Bergman (1986), MINMOD: a computer program to calculate insulin sensitivity and pancreatic responsivity from the frequently sampled intravenous glucose tolerance test. *Computer Methods and Programs in Biomedicine* **23**, 113 – 122

Pillonetto, G., G. Sparacino, P. Magni, R. Bellazzi and C. Cobelli (2002), Minimal model $S_I = 0$ problem in NIDDM subjects: nonzero Bayesian estimates with credible intervals. *American Journal of Physiology - Endocrinology and Metabolism* **282**, E565 – E573

Richardson, S. and P. J. Green (1997), On Bayesian Analysis of Mixtures with an Unknown Number of Components. *Journal of the Royal Statistical Society, Series B* **59**, 731–792

Toffolo, G., D. R.N. Bergman, C. R. Finegood, Bowden and C. Cobelli (1980), Quantitative estimation of beta cell sensitivity to glucose in the intact organism: a minimal model of insulin kinetics in the dog. *Diabetes* **29**, 979–990

A System Process Means

The system means in the different models are given below. Note that the integral in the GA model is approximated here by a trapezoidal quadrature rule.

Glucose model G_1 and insulin model I_1 to I_5:

$$
\begin{aligned}
G_1 : f^g_{t_{k-1}} &= g_{t_{k-1}} - (t_k - t_{k-1})\big(S_G(1 - G_b\, e^{-g_{t_{k-1}}}) + e^{x_{t_{k-1}}}\big),\\
\quad f^x_{t_{k-1}} &= x_{t_{k-1}} - (t_k - t_{k-1})p_2\big(1 - S_I(e^{i_{t_{k-1}}} - I_b)e^{-x_{t_{k-1}}}\big),\\
I_1 : f^i_{t_{k-1}} &= i_{t_{k-1}} - (t_k - t_{k-1})\Big(\frac{b_3}{\varphi_1}(1 - e^{-i_{t_{k-1}}} I_b) - 10^{-4} e^{-i_{t_{k-1}}}\varphi_2 J_+(e^{g_{t_{k-1}}} - h)t_{k-1}\Big),\\
I_2 : f^i_{t_{k-1}} &= i_{t_{k-1}} - (t_k - t_{k-1})\Big(\frac{b_3}{\varphi_1}(1 - e^{-i_{t_{k-1}}} I_b) - 10^{-4} e^{-i_{t_{k-1}}}\varphi_2 J_+(e^{g_{t_{k-1}}} - h)\Big),\\
I_3 : f^i_{t_{k-1}} &= i_{t_{k-1}} - (t_k - t_{k-1})\Big(\frac{b_3}{\varphi_1}(1 - e^{-i_{t_{k-1}}} I_b) - 10^{-4} e^{-i_{t_{k-1}}}\varphi_2 (e^{g_{t_{k-1}}} - h)t_{k-1}\Big),\\
I_4 : f^i_{t_{k-1}} &= i_{t_{k-1}} - (t_k - t_{k-1})\Big(\frac{b_3}{\varphi_1}(1 - e^{-i_{t_{k-1}}} I_b) - 10^{-4} e^{-i_{t_{k-1}}}\varphi_2 (e^{g_{t_{k-1}}} - h)\Big),\\
I_5 : f^i_{t_{k-1}} &= i_{t_{k-1}} - (t_k - t_{k-1})\Big(-\frac{b_3}{\varphi_1}\big(1 - \frac{I_b}{G_b b_5} e^{-i_{t_{k-1}}} \sum_{t_j \in \Lambda \cap [t_k - b_5, t_k]} (t_j - t_{j-1})(e^{g(t_j)} + e^{g(t_{j-1})})/2\big)\Big)
\end{aligned}
$$

Glucose model G_2 and insulin model I_1 to I_5:

$$
\begin{aligned}
G_2 : f^g_{t_{k-1}} &= g_{t_{k-1}} - (t_k - t_{k-1})\big(S_G(1 - G_b\, e^{-g_{t_{k-1}}}) - S_I\big(e^{i_{t_{k-1}}} - I_b G_b e^{-g_{t_{k-1}}}\big),\\
I_1 : f^i_{t_{k-1}} &= i_{t_{k-1}} - (t_k - t_{k-1})\Big(\frac{b_3}{\varphi_1}(1 - e^{-i_{t_{k-1}}} I_b) - 10^{-4} e^{-i_{t_{k-1}}}\varphi_2 J_+(e^{g_{t_{k-1}}} - h)t_{k-1}\Big),\\
I_2 : f^i_{t_{k-1}} &= i_{t_{k-1}} - (t_k - t_{k-1})\Big(\frac{b_3}{\varphi_1}(1 - e^{-i_{t_{k-1}}} I_b) - 10^{-4} e^{-i_{t_{k-1}}}\varphi_2 J_+(e^{g_{t_{k-1}}} - h)\Big),\\
I_3 : f^i_{t_{k-1}} &= i_{t_{k-1}} - (t_k - t_{k-1})\Big(\frac{b_3}{\varphi_1}(1 - e^{-i_{t_{k-1}}} I_b) - 10^{-4} e^{-i_{t_{k-1}}}\varphi_2 (e^{g_{t_{k-1}}} - h)t_{k-1}\Big),\\
I_4 : f^i_{t_{k-1}} &= i_{t_{k-1}} - (t_k - t_{k-1})\Big(\frac{b_3}{\varphi_1}(1 - e^{-i_{t_{k-1}}} I_b) - 10^{-4} e^{-i_{t_{k-1}}}\varphi_2 (e^{g_{t_{k-1}}} - h)\Big),\\
I_5 : f^i_{t_{k-1}} &= i_{t_{k-1}} - (t_k - t_{k-1})\Big(-\frac{b_3}{\varphi_1}\big(1 - \frac{I_b}{G_b b_5} e^{-i_{t_{k-1}}} \sum_{t_j \in \Lambda \cap [t_k - b_5, t_k]} (t_j - t_{j-1})(e^{g(t_j)} + e^{g(t_{j-1})})/2\big)\Big)
\end{aligned}
$$

B (RJ)MCMC Updates

Here we provide additional details on the MCMC transitions used to explore the posterior described in the text.

B.1 Parameter Updates

The parameter updates occur in two distinct steps. First we update the population parameters using a series of univariate Metropolis Hasting updates. Then, due to the high correlations between the individual parameters and system process, these are updated together using a Metropolis Hastings step for each individual j in turn.

The posterior conditional distributions for the population parameters all have standard form except for those corresponding to the three precision parameters. Thus, Gibbs moves are used to update these, using the corresponding normal or gamma conditional distributions, whilst random walk Metropolis Hastings updates are required to the population parameters associated with the precisions.

High posterior correlations between the elements of $\boldsymbol{\theta}_{jm}$ means that updating the entire vector in a single step is likely to be most efficient. Similarly, the high dependence between $\boldsymbol{\Phi}^s_{jm}$ and $\boldsymbol{\theta}_{jm}$ suggests blocking these two vectors together within a single Metropolis Hastings update. An additional advantage of this approach is that all of these parameters are updated at a cost of only one likelihood evaluation. Here, we use a random walk Metropolis proposal for $\boldsymbol{\theta}_{jm}$ which is uniformly distributed within a hyper-rectangle centred on the current position. The length of the proposal interval for each parameter is determined separately via an initial pilot-tuning simulation, as discussed in Section 4.1. As a proposal for the $\boldsymbol{\Phi}^s_{jm}$ we simply sample a new value $\boldsymbol{\Phi}^s_{jm}{}'$ from the model $p_{m,\tau}(\boldsymbol{\Phi}^s_{jm} \,|\, \boldsymbol{\theta}'_{jm})$ where $\boldsymbol{\theta}'_{jm}$ denotes the proposed new value for the vector of individual parameters. The proposed jump from $(\boldsymbol{\Phi}^s_{jm}, \boldsymbol{\theta}_{jm})$ to $(\boldsymbol{\Phi}^s_{jm}{}', \boldsymbol{\theta}'_{jm})$ is subsequently accepted with probability $\alpha = \min(1, A)$, where

$$A = \frac{\left[p_{m,\tau}(\boldsymbol{\Phi}^o_j \,|\, \boldsymbol{\theta}'_{jm}, \boldsymbol{\Phi}^s_{jm}{}')\right]^{s(\tau)} \left[p_{m,\tau}(\boldsymbol{\Phi}^s_{jm}{}' | \boldsymbol{\theta}'_{jm})\right]^{s(\tau)-1} p(\boldsymbol{\theta}'_{jm}|\boldsymbol{\Psi}_m)}{\left[p_{m,\tau}(\boldsymbol{\Phi}^o_j \,|\, \boldsymbol{\theta}_{jm}, \boldsymbol{\Phi}^s_{jm})\right]^{s(\tau)} \left[p_{m,\tau}(\boldsymbol{\Phi}^s_{jm} | \boldsymbol{\theta}_{jm})\right]^{s(\tau)-1} p(\boldsymbol{\theta}_{jm}|\boldsymbol{\Psi}_m)}.$$

B.2 Model Updates

Jumps between some models (e.g., updating the insulin model from I_1 to $I_t, t = 2, 3, 4$, say) do not involve changing the number of parameters. For moves of this kind, we simply compare the posterior distribution under the current and proposed new model with the current parameter values and adopt the usual Metropolis Hastings acceptance ratio. This works reasonably well in practice, since the majority of parameters that persist in all models change very little between models.

For moves that involve adding or deleting a parameter (e.g., when we update the glucose model) we require a reversible jump MCMC update which we implement as follows. First, we select a new model to which we propose to jump, this we do by picking one of the remaining models with equal probability. We then use the prior corresponding to the new model to generate values for any population parameters which exist in the proposed new model but not the current one. Formally, suppose $\tilde{\boldsymbol{\Psi}}$ denotes the set of population parameters common to all models and $\tilde{\boldsymbol{\Psi}}_m$ denote those population parameters which belong to model m but are not common to all other models, so that $\boldsymbol{\Psi}_m = (\tilde{\boldsymbol{\Psi}}, \tilde{\boldsymbol{\Psi}}_m)$. Thus, supposing that we propose to model m', we generate $\tilde{\boldsymbol{\Psi}}'_{m'}$ from its corresponding prior.

Next, for any individual parameters that exist in the new but not the current model, $\tilde{\boldsymbol{\theta}}_{jm'}$, we generate new values from the corresponding conditional distribution $p_{m'}(\tilde{\boldsymbol{\theta}}'_{jm'} | \tilde{\boldsymbol{\Psi}}'_{m'})$. Finally, we generate an entirely new system vector $\boldsymbol{\Phi}^s_{jm'}{}'$ from the conditional distribution $p_{m',\tau}(\boldsymbol{\Phi}^s_{jm'}{}' \,|\, (\tilde{\boldsymbol{\theta}}_j, \tilde{\boldsymbol{\theta}}'_{jm'})$, with obvious notation. This proposal is then accepted with probability $\alpha = \min(1, A)$, where

$$A = \frac{p(\tau \,|\, m')p(m')}{p(\tau \,|\, m)p(m)} \prod_{j=1}^{L} \left(\frac{p_{m',\tau}(\boldsymbol{\Phi}^0_j \,|\, [\tilde{\boldsymbol{\theta}}_j, \tilde{\boldsymbol{\theta}}'_{jm'}], \boldsymbol{\Phi}^s_{jm'}{}')}{p_{m,\tau}(\boldsymbol{\Phi}^0_j \,|\, [\tilde{\boldsymbol{\theta}}_j, \tilde{\boldsymbol{\theta}}_{jm}], \boldsymbol{\Phi}^s_{jm})} \right)^{s(\tau)} \left(\frac{p_{m',\tau}(\boldsymbol{\Phi}^s_{jm'}{}' \,|\, [\tilde{\boldsymbol{\theta}}_j, \tilde{\boldsymbol{\theta}}'_{jm'}])}{p_{m,\tau}(\boldsymbol{\Phi}^s_{jm} \,|\, [\tilde{\boldsymbol{\theta}}_j, \tilde{\boldsymbol{\theta}}_{jm}])} \right)^{s(\tau)-1}.$$

B.3 Temperature Updates

Suppose the Markov chain is in model m at temperature τ, and a new temperature τ' is proposed with equal probability, then the corresponding acceptance probability is given by $\alpha = \min(1, A)$, where

$$A = \prod_{j=1}^{L} \left(\frac{p_{m,\tau'}(\boldsymbol{\Phi}^o_j \,|\, \boldsymbol{\theta}_{jm}, \boldsymbol{\Phi}^s_{jm}) p_{m,\tau'}(\boldsymbol{\Phi}^s_{jm} \,|\, \boldsymbol{\theta}_{jm})}{p_{m,\tau}(\boldsymbol{\Phi}^o_j \,|\, \boldsymbol{\theta}_{jm}, \boldsymbol{\Phi}^s_{jm}) p_{m,\tau}(\boldsymbol{\Phi}^s_{jm} \,|\, \boldsymbol{\theta}_{jm})} \right)^{s(\tau')} \frac{p(\tau'|m)}{p(\tau \,|\, m)}.$$