A Bayesian Framework to Identify Type 1 Diabetes Physiological Models Using Easily Accessible Patient Data

Giacomo Cappon, *Member, IEEE*, Andrea Facchinetti, Giovanni Sparacino, and Simone Del Favero

***Abstract*—Mathematical physiological models of type 1 diabetes (T1D) glucose-insulin dynamics have been of great help in designing and preliminary assessing new algorithm for glucose control. Derivation of models at the individual level is however difficult because of identifiability issues. Recently, fitting these models against data of real patients with T1D has been made possible by both the use of Bayesian estimation techniques and the availability of individual datasets including plasma glucose and insulin concentration samples gathered in clinical protocols. The aim of this work is to make a step further and develop a methodology able to estimate the parameters of T1D physiological models using easily accessible data only, i.e. continuous glucose monitoring (CGM) sensor, carbohydrate intakes (CHO), and exogenous insulin infusion (I) data. The methodology is tested on synthetic data of 100 patients generated by a composite model of glucose-insulin dynamics. To solve identifiability problems, a Bayesian approach numerically implemented by Markov Chain Monte Carlo (MCMC) has been used to obtain point estimates and confidence intervals of model unknown parameters exploiting a priori knowledge available from the literature. Results show goodness of model fit and acceptable precision of parameter estimates. The methodology is also successful in reconstructing of "non-accessible" glucose-insulin fluxes, i.e. glucose rate of appearance and plasma insulin. These preliminary results encourage further development of this framework and its assessment in more challenging setups.**

I. Introduction

Type 1 diabetes (T1D) is a chronical metabolic condition in which secretion of endogenous insulin by the pancreatic beta-cells is totally absent. As a result, patients with T1D are required to inject exogenous insulin and constantly monitor their blood glucose (BG) concentration in order to keep it within the normal range [1]. To help patients in managing such a burdensome therapy, in the last few years, new technologies, e.g. insulin pumps for continuous subcutaneous insulin infusion (CSII), and minimally invasive continuous glucose monitoring sensors (CGM) have been introduced. These technologies also opened new possibilities in the development of novel algorithms to improve patients' quality of life [2][3]. In this context, simulation models of the glucose-insulin system can help to preliminary design and test new methodologies for BG control in a safe and cost efficient environment [4]. Average models are commonly used. Another approach is to use a population of virtual subjects collectively covering the inter-subject variability seen in the real population. A more interesting perspective would be fitting individual models against data of real patients with T1D in order to mimic their specific physiologies and obtain their virtual "clones" usable in *in silico* simulations aimed at designing new, personalized, algorithms for T1D management.

In the literature, many mathematical models of different complexities have been proposed [5]-[7]. Unfortunately, the simpler one is too simplistic for our purposes, the more complicated one is affected by identifiability issues. Identifiability problems can be solved by introducing some prior knowledge on their distributions, adopting a Bayesian approach where point parameter estimates are obtained from the posterior density function by either analytical (e.g. Maximum A Posteriori (MAP)) [8] or numerical integration (e.g. Markov Chain Monte Carlo (MCMC)) [9]. However, in both [8] and [9], rich datasets of frequent plasma glucose and insulin concentration obtained in a clinical setting have been used. This clearly limits the applicability of these methodologies.

In this work, by using only easily accessible data, i.e. CGM sensor, carbohydrate intakes, and CSII data; a Bayesian framework is proposed to identify by MCMC a physiological model of T1D glucose-insulin dynamics.

Here, an ad-hoc model of glucose-insulin dynamics has been proposed and used to test the methodology on 100 virtual subjects. In particular, model parameters of each subject are supposed to be known and, for each of them, the model has been used to simulate synthetic data and evaluate the methodology performance. Specifically, the identification accuracy of the algorithm has been evaluated by assessing whether the MCMC is able to (re-)identify the same model used to generate the data.

Obtained results show good model fit and accurate parameter estimates encouraging further development of the methodology.

II. Proposed Mathematical of Glucose-Insulin Regulation in T1D

The proposed model consists of four subsystems describing subcutaneous insulin absorption, oral glucose absorption, glucose-insulin kinetics, and CGM sensor error respectively. It has been built in order to be simple, i.e. to improve identifiability, but still allowing to capture the patient specific glucose-insulin dynamics. The model takes, as inputs, the exogenous insulin infusion and meal carbohydrate intake

This work was partially supported by "Ministero dell'Istruzione, Università e Ricerca" (Italian Ministry of Education, Universities and Research) through the project "Learn4AP: Patient-Specific Models for an Adaptive, Fault-Tolerant Artificial Pancreas" (initiative "SIR: Scientific Independence of young Researchers", project ID: RBSI14JYM2).

G. Cappon, A. Facchinetti, G. Sparacino, and S. Del Favero are with Department of Information Enginnering (DEI), University of Padova, Padova, Italy (e-mail: {cappongi, delfave, facchine, gianni}@dei.unipd.it).

978-1-5386-1311-5/19/$31.00 ©2019 IEEE

recordings, providing, as an output, the glucose concentration. In the following, for each subsystem, model equations, unknown parameters, and a priori distributions used for their identification, are defined.

A. Glucose-Insulin Kinetics Subsystem

The glucose-insulin kinetics subsystem describes the effect of the insulin action and the glucose rate of appearance on plasma glucose dynamics through time [5]. Model equations are:

$$\begin{cases} \dot{G}(t) = -(S_G + X(t))G(t) + S_G G_b + \frac{Ra(t)}{V_G} \\ \dot{X}(t) = -p_2[X(t) - \frac{SI}{VI}(I_p(t) - I_{pb})] \end{cases} \quad (1)$$

where G (mg/dL) is the glucose concentration in the plasma; X (min^{-1}) is the insulin action on glucose disposal and production; S_G (min^{-1}) is the glucose effectiveness; G_b (mg/dL) is the basal glucose concentration; Ra (mg/kg/min) is the glucose rate of appearance; V_G (dL/kg) is the volume of glucose distribution; p_2 (min^{-1}) is the rate constant of insulin action dynamics; SI (mL/μU·min) is the insulin sensitivity; I_p (mU/kg) is the plasma insulin; I_{pb} (mU/kg) is the basal insulin infusion. V_G has been fixed to population value, i.e. 0.126 L/kg. Thus, unknown model parameters of glucose-insulin kinetics subsystem, i.e. $\boldsymbol{\theta_{glu}}$, are $[S_G, G_b, p_2, SI]$. A priori distributions of model parameters have been obtained from the literature [10].

B. Subcutaneous Insulin Absorption Subsystem

The subcutaneous insulin absorption subsystem is a three compartment model that uses the exogenous insulin infusion recordings, as an input, to describe fast-acting insulin analogues diffusion dynamics through the subcutis to the plasma [11]. It is defined as:

$$\begin{cases} \dot{I_{sc1}}(t) = -(k_{a1} + k_d)I_{sc1}(t) + I(t-\tau) \\ \dot{I_{sc2}}(t) = k_d I_{sc1}(t) - k_{a2} I_{sc2}(t) \\ \dot{I_p}(t) = k_{a1} I_{sc1}(t) + k_{a2} I_{sc2}(t) - k_e I_p(t) \end{cases} \quad (2)$$

where I_{sc1} (mU/kg) and I_{sc2} (mU/kg) represent the insulin in a non-monomeric and monomeric state, respectively; k_{a1} (min^{-1}) and k_{a2} (min^{-1}) are the rate constant of subcutaneous insulin absorption from the first and the second compartment, respectively, to the plasma; k_e (min^{-1}) is the fractional clearance rate; V_I (L/kg) is the volume of insulin distribution; I (mU/kg/min) is the rate of exogenous insulin infusion at time t; τ (min^{-1}) defines the delay in the appearance of insulin in the first compartment. V_I and τ have been fixed to population values, i.e. 1.7 dL/kg and 8 min, respectively. Additionally, k_e has also been fixed to population value, since being the static gain of subsystem (2), it made SI to vary freely making it impossible to identify its value. As such, model unknown parameters of subcutaneous insulin absorption subsystem are $\boldsymbol{\theta_{ins}} = [k_{a1}, k_{a2}, k_d]$. A priori distributions of model parameters have been obtained from the literature [11].

C. Oral Glucose Absorption Subsystem

The oral glucose absorption subsystem is a three compartment model that, by taking the meal carbohydrate intakes as an input, describes the glucose transit through the stomach and upper small intestine [12] as follows:

$$\begin{cases} \dot{Q_{sto1}}(t) = -k_{gri} Q_{sto1}(t) + CHO(t) \\ \dot{Q_{sto2}}(t) = k_{gri} Q_{sto1}(t) - k_{empt} Q_{sto2}(t) \\ \dot{Q_{gut}}(t) = k_{empt} Q_{sto2}(t) - k_{abs} Q_{gut}(t) \end{cases} \quad (3)$$

where Q_{sto1} (mg/kg) and Q_{sto2} (mg) are the glucose amount in the stomach in a solid and liquid state, respectively; Q_{gut} (mg) is the glucose concentration in the intestine; k_{gri} (min^{-1}) is the rate constant of grinding; k_{empt} (min^{-1}) is the rate constant of gastric empting; k_{abs} (min^{-1}) is the rate constant of intestinal absorption; CHO (g/min) is the ingested glucose absorption rate. Using (3), it is possible to estimate Ra as:

$$Ra(t) = f \cdot k_{abs} Q_{gut} / BW \quad (4)$$

where f (dimensionless) is the fraction of the intestinal absorption that actually appears in the plasma; and BW (kg) is the body weight. f has been fixed to 0.9 and k_{gri} has been set equal to k_{empt}. Unknown model parameter of oral glucose absorption subsystem are $\boldsymbol{\theta_{oral}} = [k_{gri}, k_{abs}]$. A priori distributions of model parameters have been obtained from [12].

D. CGM Sensor Error Subsystem

In order to simulate CGM sensor error, the following simplified model has been defined to describe measurement data:

$$CGM(t) = IG(t) + n(t) \quad (5)$$

where n (mg/dL) is the measurement error, assumed to be normally distributed, with zero-mean and constant coefficient of variation (CV_n); and IG (mg/dL) is the interstitial glucose concentration whose evolution through time has been modeled as a one compartment model describing the plasma-to-interstitial glucose dynamics as follows:

$$\dot{IG}(t) = -\frac{1}{\alpha}(IG(t) - G(t)) \quad (6)$$

where α (min) represents the delay between the plasma and interstitial glucose compartments (fixed to 7).

Here, the unknown model parameter is $\boldsymbol{\theta_{meas}} = [CV_n]$. No a priori distribution of CV_n is available from the literature. As such, CV_n has been supposed to be uniformly distributed between [0, 1].

III. Bayesian Estimation of Unknown Model Parameters via MCMC

Let us rewrite the model as:

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, t, \boldsymbol{\theta}) \\ \mathbf{y} = \mathbf{g}(\mathbf{x}, \mathbf{u}, t, \boldsymbol{\theta}) \end{cases} \quad (7)$$

where $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ are the state and output functions, respectively, i.e. two possibly non-linear functions of the state vector $\mathbf{x} \in \mathbb{R}^{n_x} := [G, X, I_{sc1}, I_{sc2}, I_p, Q_{sto1}, Q_{sto2}, Q_{gut}, IG]$, the input vector $\mathbf{u} \in \mathbb{R}^{n_u} := [CHO, I]$, the time t; and the set of unknown parameters that need to be identified $\boldsymbol{\theta} := [\boldsymbol{\theta_{glu}},$

$\boldsymbol{\theta}_{ins}$, $\boldsymbol{\theta}_{oral}$, $\boldsymbol{\theta}_{meas}$]. Here, $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ are defined by (1), (2), (3), (4), (6), and (5) respectively.

As mentioned in the introduction, in order to identify the model parameters, a Bayesian approach has been adopted. Bayesian inference estimates the unknown parameter vector $\boldsymbol{\theta}$ by exploiting the available a priori information on $\boldsymbol{\theta}$, specified by the probability density function $p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$. Then, knowledge on measurement data $Y:=\{CGM(t_k) \mid t_k = k\cdot T_s, k = 1,\ldots,m\}$ allows to obtain the a posteriori density function using the Bayes theorem:

$$p_{\boldsymbol{\theta}|Y}(\boldsymbol{\theta}|Y)=\frac{p_{Y|\boldsymbol{\theta}}(Y|\boldsymbol{\theta})p_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{\int p_{Y|\boldsymbol{\theta}}(Y|\boldsymbol{\theta})p_{\boldsymbol{\theta}}(\boldsymbol{\theta})d\boldsymbol{\theta}} \quad (8)$$

where $p_{Y|\boldsymbol{\theta}}(Y|\boldsymbol{\theta})$ is the likelihood function, i.e. the probability of observing a certain measurement data set given the parameter vector $\boldsymbol{\theta}$.

The aim being identifying (7), a point estimate of $\boldsymbol{\theta}$ can be obtained choosing a proper Bayesian estimator. Here, the posterior mean has been used (known to be the minimum variance estimate of $\boldsymbol{\theta}$):

$$\hat{\boldsymbol{\theta}}=E[\boldsymbol{\theta}|Y]=\int \boldsymbol{\theta} p_{\boldsymbol{\theta}|Y}(\boldsymbol{\theta}|Y)\, d\boldsymbol{\theta} \quad (9)$$

Unfortunately, (9) is, in general, analytically intractable. However, this limitation can be surpassed using a MCMC approach [13]. Specifically, it consists of two main steps: first a Markov chain is built to converge in distribution to the underlying target distribution $p_{\boldsymbol{\theta}|Y}(\boldsymbol{\theta}|Y)$; second, after obtaining the posterior distribution in a sampled form, a Monte Carlo integration is performed to compute (9), thus obtaining a point estimate of $\boldsymbol{\theta}$. Details of MCMC implementation are described in Appendix A.

IV. Test Framework

A. Simulated dataset

Model (1)-(6) has been used to generate data of 100 virtual subjects by simulation over 2 week. For each simulation, different combinations of model inputs, i.e. CHO and I; have been generated.

In particular, each virtual subject is assigned to a specific set of model parameters, which are supposed to be known, and randomly generated from their prior distributions found in the literature. Therefore, the model used to simulate the data and the model we identified via the MCMC strategy described in Section III, perfectly match.

Model equations have been solved numerically by first order Euler approximation with fixed integration step (T_s) equal to 1 minute.

For each subject, data have been split in two sets: the first 3 days of data have been used to identify the model, the remaining 11 to evaluate the goodness of model fit.

B. Performance metrics

Stressing that the true value of the parameters are known for each subject, absolute relative deviation (ARD), of each parameter from its true value has been evaluated:

$$ARD_i=100\frac{|\theta_i-\hat{\theta}_i|}{\theta_i}\ (\%) \quad (11)$$

where θ_i and $\hat{\theta}_i$ denote the i-th element of vector $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$, respectively.

In addition, we quantify goodness of model prediction vs. real data. To do that, the root mean squared error (RMSE) has been computed for each subject as:

$$RMSE=\sqrt{\sum_{k=1}^{m}\left(IG(t_k)-\widehat{IG}(t_k)\right)^2}\ (mg/dL) \quad (10)$$

where $\widehat{IG}$ is the interstitial glucose concentration obtained by simulation from the identified model. In the following, obtained ARD and RMSE distributions have been reported as median [interquantile range].

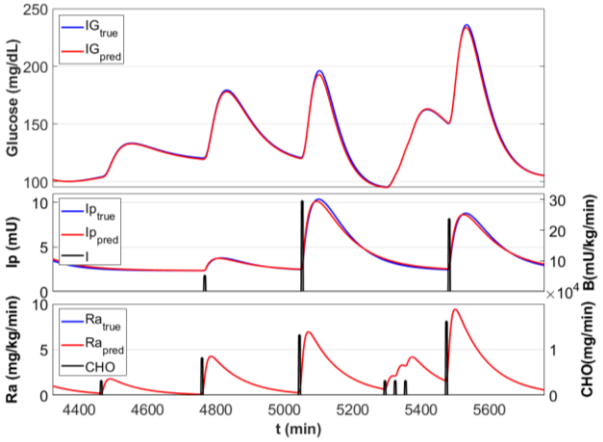

Figure 1. Model fit to interstitial glucose, plasma insulin, and glucose rate of appearance in a representative subject. Blue lines denote measured glucose concentration. Real and predicted Ip and Ra are shown in red and blue, respectively. Black lines represent model inputs.

V. Results

Fig. 1 shows model predictions (IG_{pred}) vs. true IG (IG_{true}) in a representative subject during the fourth day of simulation. Moreover, the predicted Ip (Ip_{pred}) and Ra (Ra_{pred}) traces are also shown and compared to the respective real traces, i.e. Ip_{true} and Ra_{true}. Results, show good model fit to IG. In particular, good RMSE is achieved (6.59 mg/dL). Moreover, thanks to the accuracy in estimating the parameters of (1) and (2), almost identical Ra and I_p traces have been reconstructed. This result was not the aim of the algorithm, but corroborates the validity of the proposed methodology.

Good fit is achieved at the population level as well. In details, obtained RMSE is 6.44 mg/dL [5.51-7.40] (median [interquartile range]). Table I reports the obtained identification performance in terms of ARD. Of note, higher ARD has been obtained for the parameters related to theinsulin subsystem. This is expected since no direct access to (2) is given to the identification procedure.

VI. Conclusion

In this work, a Bayesian strategy to identify a T1D physiological model using easily accessible data, i.e. glucose sensor data, CHO intakes, and exogenous insulin recordings,

TABLE I. ACCURACY OF PARAMETER ESTIMATES

	SG	G_b	SI	p_2	k_{a1}	k_{a2}	k_d	k_{empt}	k_{abs}	CV_n
ARD	**1.22** [0.68-2.44]	**0.30** [0.14-0.53]	**1.31** [0.59-2.67]	**27.14** [15.09-38.38]	**22.82** [12.19-33.52]	**13.97** [6.99-27.76]	**12.91** [6.91-24.55]	**2.14** [1.22-5.70]	**1.20** [0.51-2.25]	**1.71** [1.04-2.89]

Accuracy of parameter estimates obtained in the synthetic population. Results are reported as median [interquartile range]

has been proposed. To preliminary study the potentiality of this approach in solving the problem at hand, the assessment of the methodology on synthetic data has been performed. In particular, accurate parameter estimates have been obtained across the synthetic population allowing to achieve good model fit. Moreover, beside fitting glucose data, the methodology was successfully used to estimate the plasma insulin and glucose rate of appearance traces and allowed identifying the glucose sensor error model. This allowed to exploit this technique not only to obtain accurate and reliable glucose predictions, but also to estimate "not accessible" fluxes of glucose-insulin physiology.

Preliminary results are encouraging but further work is needed. First, "model mismatch" should be introduced. Indeed, validation of this technique to fit data simulated by a different model is needed in order to assess the algorithm performance. Then, this methodology will be exploited to devise new model-based glucose prediction techniques. From this perspective, research on this topic is currently undergoing in our laboratory and is achieving promising results. Finally, the methodology will be evaluated on real data, to assess model performance in a more realistic scenario, where multiple sources of errors and uncertainties affect the quality of data and inevitably deteriorates parameter estimation.

APPENDIX A

MCMC has been implemented through single component Metropolis-Hastings algorithm [13]. The algorithm consists in dividing $\boldsymbol{\theta}$ into p partitions $\boldsymbol{\theta}:=[\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_h]$. Any iteration of the algorithm is divided into h steps and each step consists in updating the c-th component of $\boldsymbol{\theta}$ by mean of a proposal density function $q_c(\cdot|\cdot)$ as described in Algorithm 1:

Algorithm 1 Single Component Metropolis-Hastings

Inizialize $\boldsymbol{\theta}_0$, set i = 0
Repeat{
  for c = 1,…,h {
    Set $\boldsymbol{\theta}_{i,-c}=[\boldsymbol{\theta}_{i+1,1},\ldots,\boldsymbol{\theta}_{i+1,c-1},\boldsymbol{\theta}_{i,c+1},\ldots,\boldsymbol{\theta}_{i,h}]$
    Sample a point ϕ_c from $q_c(\cdot|\cdot)$
    Set $\alpha = \min\left(1, \frac{\pi(\boldsymbol{\theta}_{i+1,c}|\boldsymbol{\theta}_{i,-c})q_c(\boldsymbol{\theta}_{i,c}|\boldsymbol{\theta}_{i+1,c},\boldsymbol{\theta}_{i,-c})}{\pi(\boldsymbol{\theta}_{i,c}|\boldsymbol{\theta}_{i,-c})q_c(\boldsymbol{\theta}_{i+1,c}|\boldsymbol{\theta}_{i,c},\boldsymbol{\theta}_{i,-c})}\right)$
    Set, with probability α, $\boldsymbol{\theta}_{i+1,c} = \phi_c$
    Otherwise set $\boldsymbol{\theta}_{i+1,c} = \boldsymbol{\theta}_{i,c}$
  }
  i = i + 1
}

where $\pi(\boldsymbol{\theta})$ is the target distribution, i.e. $\propto p_{\boldsymbol{\theta}|Y}(\boldsymbol{\theta}|Y)$.

Specifically, $\boldsymbol{\theta}$ has been divided into four components, namely $\boldsymbol{\theta}_1$:=[S_G, G_b, SI], $\boldsymbol{\theta}_2$:=[p_2, k_{a1}, k_{a2}, k_d], $\boldsymbol{\theta}_3$:=[k_{empt}, k_{abs}], $\boldsymbol{\theta}_4$:=[CV_v]. This scheme has been chosen since it improved MCMC mixing. In particular, SI and p_2 have been assigned to two different components in order to break the correlation these parameters have, known to be critical from the literature [10]. As far as the MCMC chain convergence is concerned, the Raftery-Lewis criterion [13] has been used to compute the number of iterations necessary to ensure the chain to represent the posterior distribution of interest.

REFERENCES

[1] American Diabetes Association, "Standards of medical care in diabetes," *Diabetes Care*, vol. 41, no. S1, pp. S1-S159, Dec. 2017.

[2] Cappon et al., "Wearable continuous glucose monitoring sensors: a revolution in diabetes treatment," *Electronics*, vol. 6, no. 3, pp. 65, Sep. 2017.

[3] Vettoretti et al., "Continuous glucose monitoring: current use in diabetes management and possible future applications," *J. Diabetes Sci. Technol.*, vol. 12, no. 5, pp. 1064-1071, Sep. 2018.

[4] B. P. Kovatchev et al., "In silico preclinical trials: A proof of concept in closed-loop control in type 1 diabetes," *J. Diabetes Sci. Technol.*, vol. 3, no. 2, pp. 44-55, Jan. 2009.

[5] R. N. Bergman et al., "Quantitative estimation of insulin sensitivity," *Am. J. Physiol.*, vol. 236, no. 6, pp. e667-e677, Jun. 1979.

[6] R. Hovorka et al., "Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes," *Physiol. Meas.*, vol. 25, no. 4, pp. 905-920, Aug. 2004.

[7] C. Dalla Man et al., "Meal simulation model of the glucose-insulin system," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 10, pp. 1740-1749, Oct. 2007.

[8] R. Visentin et al., "One-day Bayesian cloning of type 1 diabetes subjects: Toward a single-day UVa/Padova type 1 diabetes simulator," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 11, pp. 2416-2424, Nov. 2016.

[9] A. Haidar et al., "Stochastic virtual population of subjects with type 1 diabetes for the assessment of closed-loop glucose controllers," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 12, pp. 3524-3533, Dec. 2013.

[10] G. Pillonetto et al., "Numerical non-identifiability regions of the minimal model of glucose kinetics: superiority of Bayesian estimation," *Math. Biosci.*, vol. 184, no. 1, pp. 53-67, Jul. 2003.

[11] M. Schiavon et al., "Modeling subcutaneous absorption of fast-acting insulin in type 1 diabetes," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 2079-2086, Sep. 2018.

[12] C. Dalla Man et al., "A system model of oral glucose absorption: Validation on gold standard data," vol. 53, no. 12, pp. 2472-2478, Dec. 2006.

[13] W. R. Gilks et al., "Markov Chain Monte Carlo in practice," *Chapman and Hall*, London, 1996.