Model predictive control of glucose concentration in type I diabetic patients: An in silico trial

L. Magni a,b, D.M. Raimondo a, C. Dalla Man b, G. De Nicolao a, B. Kovatchev c, C. Cobelli b

ABSTRACT

In this paper, the feedback control of glucose concentration in type I diabetic patients using subcutaneous insulin delivery and subcutaneous continuous glucose monitoring is considered. A recently developed in silico model of glucose metabolism is employed to generate virtual patients on which control algorithms can be validated against interindividual variability. An in silico trial consisting of 100 patients is used to assess the performances of a linear output feedback and a nonlinear state-feedback model predictive controller, designed on the basis of the in silico model. More than satisfactory results are obtained in the great majority of virtual patients. The experiments highlight the crucial role of the anticipative feedforward action driven by the meal announcement information. Preliminary results indicate that further improvements may be achieved by means of a nonlinear model predictive control scheme.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Diabetes (medically known as diabetes mellitus) is the name given to disorders in which the body has trouble regulating its blood glucose, or blood sugar, levels. There are two major types of diabetes: type I diabetes and type II diabetes. Type I diabetes, also called juvenile diabetes or insulin-dependent diabetes, occurs when the body’s immune system attacks and destroys beta cells of the pancreas. Beta cells normally produce insulin, a hormone that regulates, together with glucagon (another hormone produced by alpha cells of the pancreas), the blood glucose concentration in the body, i.e. while insulin lowers it, glucagon increases it. The food intake in the body results in an increase in the glucose concentration. Insulin enables most of the body’s cells to take up glucose from the blood and use it as energy. When the beta cells are destroyed, since no insulin can be produced, glucose remains in the blood causing serious damage to several organs. For this reason, people with type I diabetes must take insulin in order to stay alive. This means undergoing multiple injections daily (normally coinciding with the meal times), and testing blood sugar by pricking their fingers for blood several times a day [8]. This method is just an open-loop control with intermittently monitored glucose levels resulting in intermittently administered insulin doses that are manually administered by a patient using a written algorithm. The current open-loop therapy can be contrasted to glucose management by a closed-loop system controlling blood glucose in diabetic patients also known as an artificial pancreas.

The automated control of normoglycemia in subjects with type I diabetes has been subject of extensive research since the 1970s [1,7]. However, the first devices, e.g. the Biostator™ [4], which used intravenous (i.v.) BG sampling and i.v. insulin and glucose delivery, were cumbersome and unsuited for outpatient use. A minimally invasive closed-loop system using subcutaneous (s.c.) continuous glucose monitoring and s.c. insulin pump delivery was needed. To date, several s.c. = s.c. systems, have been tested [7,9]. From a control viewpoint, the main challenges are time delays, constraints, meal disturbances, and nonlinear dynamics. Most of the control schemes proposed in literature are based on either PID (proportional integral derivative) or MPC (model predictive control) control laws, see e.g. [16,7,6]. Due to the presence of significant delays in the glucose metabolism model, a well-designed feedforward action able to consider in advance a meal announcement signal is essential in order to guarantee satisfactory performances. Within MPC, feedforward and feedback control actions are jointly designed and constraints are taken into account in a very natural way. The recent advancements in nonlinear MPC, see e.g. [15], and the development of a new generation of models of glucose metabolism make this control technique even more promising.
The need of guaranteeing parameter identifiability from easily measurable experimental data motivated the development of parsimonious models of glucose metabolism (see e.g. the classical minimal model [2]). By contrast, in silico large-scale simulation models aim to describe the glucose–insulin system in as much detail as possible. The major limitation of most in silico models is that they were validated on few subjects and only by means of plasma concentration measurements. Recently, a new-generation in silico model has been developed taking advantage of the availability of a unique-meal data set of 204 nondiabetic individuals [5]. The subjects underwent a triple tracer meal protocol, making it possible to obtain in a virtually model-independent fashion the time course of all the relevant glucose and insulin fluxes during a meal. Thus, by using a “concentration and flux” portrait, it was possible to model the glucose–insulin system by resorting to a subsystem forcing function strategy which minimizes structural uncertainties in modeling the various unit processes.

In this paper, the above model is used to randomly generate virtual diabetic patients. Suitable modifications (see [12]) are introduced to reflect the absence of endogenous insulin secretion, while the parameter variability in the original data set is used to model interindividual variability of the diabetic population. Validating algorithms on a whole set of different patients (in silico trial) is the only realistic way of addressing robustness of the artificial pancreas in the face of interindividual variability so as to maximize success chances in the subsequent clinical trials conducted on real patients.

In the present work, the in silico model is used to design a linear output feedback MPC scheme which is then tested on an in silico trial consisting of 100 synthetic type I diabetic subjects “followed” for a day, receiving dinner and breakfast. The overall results, as measured by several established performance indices, are more than satisfactory for the great majority of virtual patients. An advantage of the meal announcement information. Moreover, the feedforward action that takes into account the time course of all the relevant glucose and insulin fluxes during a meal. Thus, by using a “concentration and flux” portrait, it was possible to model the glucose–insulin system by resorting to a subsystem forcing function strategy which minimizes structural uncertainties in modeling the various unit processes.

2. Virtual patients

In order to synthesize and test the controller, we used the meal glucose–insulin model [5,12] summarized in the following subsections.

2.1. Glucose intestinal absorption

Glucose intestinal absorption is modeled by a three-compartment model

\[
\begin{align*}
Q_{101}(t) &= -k_{in}Q_{101}(t) + d(t) \\
Q_{102}(t) &= -k_{empt}(t)Q_{102}(t) + k_{empt}(t)Q_{101}(t) + d(t) \\
Q_{11}(t) &= Q_{101}(t) + Q_{102}(t) \\
R(t) &= \frac{fQ_{11}Q_{11}(t)}{BW}
\end{align*}
\]

where \(Q_{101}\) (mg) is the amount of glucose in the stomach (solid, \(Q_{101}\), and liquid phase, \(Q_{12}\)), \(Q_{11}\) (mg) is the glucose mass in the intestine, \(k_{in}\) is the rate of grinding, \(k_{empt}\) is the rate constant of intestinal absorption, \(f\) is the fraction of intestinal absorption which actually appears in plasma, \(d\) (mg/min) is the amount of ingested glucose, \(BW\) (kg) is the body weight, \(R(t)\) (mg/kg/min) is the glucose rate of appearance in plasma and \(k_{empt}\) is the rate constant of gastric emptying which is a time-varying nonlinear function of \(Q_{100}\)

\[
k_{empt}(t, Q_{100}(t)) = k_{\text{max}} + \frac{k_{\text{max}} - k_{\text{min}}}{2} \left[\tan h(\alpha(Q_{100}(t) - bD(t))) - \tan h(\beta(Q_{100}(t) - dD(t)))\right]
\]

where

\[
\alpha = \frac{5}{2D(t)(1 - b)} \quad \beta = \frac{5}{2D(t)d} \quad D(t) = \int_{t_i}^{t_f} d(t) \, dt
\]

with \(t_i\) and \(t_f\), respectively, start and end time of the last meal, \(b, d, k_{\text{max}}\) and \(k_{\text{min}}\) model parameters.

2.2. Glucose subsystem

A two-compartment model is used to describe glucose kinetics

\[
\begin{align*}
G_p(t) &= -k_1 G_p(t) + k_2 G_t(t) + EGP(t) + Ra(t) - U_a - E(t) \\
G_t(t) &= k_1 G_p(t) - k_2 G_t(t) - U_d(t) \\
G(t) &= \frac{G_p(t)}{V_G}
\end{align*}
\]

where \(V_G\) (dl/kg) is the distribution volume of glucose, \(G\) (mg/dl) is the glycemia, \(G_p\) (mg/kg) and \(G_t\) (mg/kg) are glucose in plasma and rapidly equilibrating tissues, and in slowly equilibrating tissues, respectively, \(EGP\) is the endogenous glucose production (mg/kg/min), \(E\) (mg/kg/min) is the renal excretion, \(U_a\) and \(U_d\) are the insulin-independent and -dependent glucose utilizations, respectively (mg/kg/min), and \(k_1\) and \(k_2\) are the rate parameters. The insulin-independent glucose utilizations \(U_d\) is assumed constant. Basal steady state, i.e. constant glycemia \(G_b\) (mg/dl), is characterized by the following equations:

\[
\begin{align*}
-k_1 G_{pb} + k_2 G_{rb} + EGP_b + Ra_{rb} - U_a - E_b &= 0 \\
k_1 G_{pb} - k_2 G_{rb} - U_d_{rb} &= 0
\end{align*}
\]

so that, noting that at basal equilibrium \(Ra_{rb} = 0\), \(EGP_b = U_d + U_d_{rb} + E_b\)

2.3. Glucose renal excretion

Renal extraction represents the glucose flow which is eliminated by the kidney, when glycemia exceeds a certain threshold \(k_{c2}\)

\[
E(t) = \max \{0, k_{c1}(G_p(t) - k_{c2})\}
\]

The parameter \(k_{c1}\) (1/min) represents renal glomerular filtration rate.

2.4. Endogenous glucose production

\(EGP\) comes from the liver, where a glucose reserve exists (glycogen). \(EGP\) is inhibited by high levels of glucose and insulin

\[
EGP(t) = \max \{0, EGP_b - k_{p2}(G_p(t) - G_{pb}) - k_{p3}(I_d(t) - I_b)\}
\]

where \(k_{p2}\) and \(k_{p3}\) are model parameters and \(I_d\) (pmol/l) is a delayed insulin signal, coming from the following dynamic system:

\[
\begin{align*}
I_1(t) &= k_l I_1(t) - k_d I_1(t) \\
I_2(t) &= k_l I_2(t) - k_d I_2(t)
\end{align*}
\]

where \(I\) (pmol/l) is plasma insulin concentration or insulinemia and \(k_l\) (1/min) is a model parameter.
2.5. Glucose utilization

Glucose utilization is made up of two components: the insulin-independent one \(U_{id}\), which represents the glucose uptake by the brain and erythrocytes, and the insulin-dependent component \(U_{id}\), which depends nonlinearly on glucose in the tissues

\[
U_{id}(t) = \frac{V_m(X(t)) G_{tb}(t)}{K_m + G_{tb}(t)}
\]

where \(V_m\) (1/min) is a linear function of interstitial fluid insulin \(X\) (pmol/l)

\[
V_m(X(t)) = V_{m0} + V_mX(t)
\]

which depends from insulinemia in the following way:

\[
X(t) = -p_{zb}X(t) + p_{zb}(I(t) - I_0)
\]

where \(K_m, V_{m0}, V_{mx}\) are model parameters, \(I_0\) (pmol/l) is the basal insulin level and \(p_{zb}\) (1/min) is called rate of insulin action on peripheral glucose. Considering the basal steady state, one has

\[
U_{idb} = \frac{G_{tb}}{K_m + G_{tb}}
\]

and

\[
G_{tb} = U_{id} - EGPb + k_1 G_{pb} + E_0
\]

\[
V_{m0} = \frac{(EGPb - U_{id} - E_0)K_m + G_{tb})}{G_{tb}}
\]

2.6. Insulin subsystem

Insulin flow \(s\), coming from the subcutaneous compartments, enters the bloodstream and is degraded in the liver and in the periphery:

\[
\begin{bmatrix}
\dot{I}_p(t) \\
\dot{I}(t) \\
I(t)
\end{bmatrix} = \begin{bmatrix}
-(m_2 + m_3)I_p(t) + m_1I(t) + s(t) \\
-(m_1 + m_2)I(t) + m_3I_p(t) \\
I(t)/V_I
\end{bmatrix}
\]

where \(V_I\) (l/kg) is the distribution volume of insulin and \(m_1, m_2, m_3, m_4\) (1/min) are model parameters.

At basal one has

\[
\begin{bmatrix}
0 \\
0 \\
\dot{I_0}
\end{bmatrix} = \begin{bmatrix}
-(m_2 + m_4)I_{pb} + m_1 I_0 + s_0 \\
-(m_1 + m_3)I_0 + m_2 I_{pb} \\
\dot{I_0}/V_I
\end{bmatrix}
\]

so that

\[
\begin{bmatrix}
I_0 \\
\dot{I_0}
\end{bmatrix} = \begin{bmatrix}
m_2 \\
m_3 + m_4
\end{bmatrix} I_{pb} + \begin{bmatrix}
m_1 \\
m_2
\end{bmatrix} I_0
\]

\[
s_0 = (m_2 + m_4)I_{pb} - m_1 I_0
\]

where \(I_{pb} = I_0 V_I\).

2.7. s.c. insulin subsystem

In diabetic patients, insulin is usually administered by subcutaneous injection. Insulin takes some time to reach the circulatory apparatus, unlike in the healthy subject in which the pancreas secretes directly into the portal vein. This delay is modeled here with two compartments, \(S_1\) and \(S_2\) (pmol/kg) (which is a variation of the model described in [17]), which represent, respectively, polymeric and monomeric insulin in the subcutaneous tissue

\[
\begin{bmatrix}
\dot{S}_1(t) \\
\dot{S}_2(t) \\
s(t)
\end{bmatrix} = \begin{bmatrix}
-(k_{a1} + k_{d2})S_1(t) + u(t) \\
k_4S_1(t) - k_{a2}S_2(t) \\
k_2S_1(t) + k_{a2}S_2(t)
\end{bmatrix}
\]

where \(u(t)\) (pmol/kg/min) represents injected insulin flow, \(k_d\) is called degradation constant, \(k_{a1}\) and \(k_{a2}\) are absorption constants.

At basal one has

\[
\begin{bmatrix}
0 \\
0 \\
k_2 S_0
\end{bmatrix} = \begin{bmatrix}
-(k_{a1} + k_{d2})S_1 + u_0 \\
k_4 S_0 - k_{a2} S_2 \\
k_2 S_1 + k_{a2} S_2
\end{bmatrix}
\]

and solving the system

\[
\begin{bmatrix}
S_1 \\
S_2 \\
u_0
\end{bmatrix} = \begin{bmatrix}
k_2 \\
k_4 \\
k_2
\end{bmatrix} S_0 + \begin{bmatrix}
k_1 \\
k_4 \\
k_2
\end{bmatrix} t_0
\]

The quantity \(u_0\) (pmol/min) represents insulin infusion to maintain diabetic patient at basal steady state.

2.8. s.c. glucose subsystem

Subcutaneous glucose \(G_{sb}\) (mg/dl) is, at steady state, highly correlated with plasma glucose; dynamically, instead, it follows the changes in plasma glucose with some delay. This dynamic was modeled with a system of first order

\[
\dot{G}_{sb}(t) = -k_{sb} G_{sb}(t) + k_{sb} G(t)
\]

2.9. Virtual patient generation

In order to obtain parameter joint distributions in type 1 diabetes, the parameter identified in 204 subjects in health were used as starting point [5]. Some modification was needed to realistically describe a type 1 diabetic subject: basal glucose concentration was assumed to be on average 50 mg/dl higher than in subjects in health; steady-state insulin concentration (due to an external insulin pump) was assumed to be on average four times higher than in subjects in health; basal endogenous glucose production was assumed to be 35% higher than in subjects in health, and steady-state insulin clearance was assumed to be one third lower than in subjects in health; parameters relating to insulin action on both glucose production and utilization were assumed to be one-third lower than in subjects in health. For all these parameters/variables the same inter-subject variability found in subjects in health was maintained. The parameters were assumed to be log-normal, distributed to guarantee that they were always positive, so the covariance matrix \((26 \times 26)\) was calculated using the log-transformed parameters. Then, 100 subjects were generated using the joint distribution, i.e. 100 realizations of the log-transformed parameter vector were randomly extracted from the multivariate normal distribution with mean equal to mean of the log-transformed parameters and \(26 \times 26\) covariance matrix. Finally, the parameters in the 100 in silico subjects were obtained by antitransformation.

3. Metrics to assess control performance

In evaluating the performance of a control algorithm one has to remember that its basic function is to mimic as best as possible the feature of the beta cells, which is to maintain glucose levels between 80 and 120 mg/dl in the face of disturbances such as meals or physical activity. A good algorithm should be able to maintain blood sugar low enough, as this reduces the long-term complications related to diabetes, but must also avoid even isolated hypoglycemic episodes. To be useful, a metric has to take into account these features. The HbA1c cannot be considered in this context because in the first closed-loop clinical trial patients are controlled only for a short period.
3.1. Blood Glucose Index (BGI)

Blood Glucose Index is a metric proposed by Kovatchev et al. [10], to evaluate the clinical risk related to a particular glycemic value

\[
BGI(\cdot) = 10(g[\ln a(\cdot) - b])^2
\]  

(2)

where \(a, b\) and \(g\) are fixed equal to 1.084, 5.3811 and 1.509, respectively. The BGI function is asymmetric, because hypoglycaemia is considered more dangerous than hyperglycaemia (see Fig. 1). Starting from the BGI, two synthetic indices for sequences of \(n\) measurements of blood glucose can be defined.

3.1.1. LBGI (low BGI)

Measures hypoglycemic risk

\[
LBGI = \frac{10}{n} \sum_{i=1}^{n} r_l^2(G_i)
\]

where \(r_l(\cdot) = \min(0, g[\ln a(\cdot) - \ln a(G_0)])\).

3.1.2. HBGI (high BGI)

Measures hyperglycemic risk

\[
HBGI = \frac{10}{n} \sum_{i=1}^{n} r_h^2(G_i)
\]

where \(r_h(\cdot) = \max(0, g[\ln a(\cdot) - \ln a(G_0)])\).

LBGI index captures the tendency of the algorithm to overshoot the target and eventually trigger hypoglycaemia. Directly linked with LBGI, HBGI captures the tendency of the algorithm to stay above the target range.

3.2. Control-Variability Grid Analysis (CVGA)

The Control-Variability Grid Analysis (CVGA) [13] is a graphical representation of min/max glucose values in a population of patients either real or virtual. The CVGA provides a simultaneous assessment of the quality of glycemic regulation in all patients. As such, it has the potential to play an important role in the tuning of closed-loop glucose control algorithms and also in the comparison of their performances. Assuming that for each subject a time series of measured blood glucose (BG) values over a specified time period (e.g. 1 day) is available, the CVGA is obtained as follows: for each subject, a point is plotted whose \(X\)-coordinate is the minimum BG and whose \(Y\)-coordinate is the maximum BG within the considered time period (see Fig. 2). Note that the \(X\)-axis is reversed as it goes from 110 (left) to 50 (right) so that optimal regulation is located in lower left corner. The appearance of the overall plot is a cloud of points located in various regions of the plane. Different regions on the plane can be associated with different qualities of glycemic regulation. In order to classify subjects into categories, nine rectangular zones are defined as follows:

- **Zone A** (\(G_{\text{max}} < 180, G_{\text{min}} > 90\)): Accurate control.
- **Zone B** high (\(180 < G_{\text{max}} < 300, G_{\text{min}} > 90\)): Benign trend towards hyperglycaemia.
- **Zone B** low (\(G_{\text{max}} < 180, 70 < G_{\text{min}} < 90\)): Benign trend towards hypoglycaemia.
- **Zone C** high (\(G_{\text{max}} > 300, G_{\text{min}} > 90\)): overCorrection of hyperglycaemia.
- **Zone C** low (\(G_{\text{max}} < 180, G_{\text{min}} < 70\)): overCorrection of hypoglycaemia.
- **Zone D** high (\(G_{\text{max}} > 300, 70 < G_{\text{min}} < 90\)): Failure to deal with hyperglycaemia.
- **Zone D** low (\(180 < G_{\text{max}} < 300, G_{\text{min}} < 70\)): Failure to deal with hypoglycaemia.
- **Zone E** (\(G_{\text{max}} > 300, G_{\text{min}} < 70\)): Erroneous control.

4. Model predictive control

The glucose metabolism model can be rewritten in the following compact way:

\[
\begin{aligned}
x(t) &= f(t, x(t), u(t), d(t)) \\
y(t) &= G_M(t)
\end{aligned}
\]

where \(x = [Q_{\text{ins}}, Q_{\text{gut}}, G_p, G_s, I_p, I_d, S_1, S_2, G_M]\), and \(f(\cdot, \cdot, \cdot, \cdot)\) is derived from the model equations reported in Section 2. The system is subject to the following constraints:

\[
\begin{aligned}
x_{\text{min}} &\leq x \leq x_{\text{max}} \\
u_{\text{min}} &\leq u \leq u_{\text{max}}
\end{aligned}
\]

(3)

(4)

where \(x_{\text{min}}, x_{\text{max}}, u_{\text{min}}\) and \(u_{\text{max}}\) denote lower and upper bounds on the state and input, respectively. Typically, they represent limits on the glucose concentration and on the insulin delivery rate. In the
following, it is assumed that meal announcement is available, i.e. the disturbance signal (meal) is known in advance.

The MPC control law ([15,3,11]) is based on the solution of a Finite Horizon Optimal Control Problem (FHOCP), where a cost function $J(x, u)$ is minimized with respect to the input $u$ subject to the state dynamics of a model of the system, and to state- and input-constraints. Letting $u^*$ be the solution of the FHOCP, according to the Receding Horizon paradigm, the feedback control law $u = K_{MPC}(x)$ is obtained by applying to the system only the first part of the optimal solution. In this way, a closed-loop control strategy is obtained solving an open-loop optimization problem.

MPC control laws can be formulated for both discrete- and continuous-time systems. In this paper, a discrete-time linear MPC (LMPC) is derived from an input–output linearized approximation of the full model. Moreover, a state-feedback nonlinear MPC (NMPC) scheme is derived that exploits the full nonlinear model.

4.1. Unconstrained linear model predictive control

In order to obtain a time-invariant model, the rate constant $k_{mean}$ has been fixed to $k_{mean}(t, Q_0) = k_{mean} = (k_{max} + k_{min})/2$. The system is linearized around the basal steady-state point of each patient and discretized with sampling time $T_s$ yielding

$$\begin{align*}
\dot{x}(k+1) &= A_0 x(k) + B_0 u(k) + B_0 d(k) \\
y(k) &= C x(k)
\end{align*}$$

After a model order reduction step the following transfer function representation is obtained

$$Y(z) = \frac{N_0(z)}{D(z)} U(z) + \frac{N_0(z)}{D(z)} D(z)$$

with

$$N_0(z) = b_2 z^2 + b_1 z + b_0$$
$$D(z) = z^2 + a_2 z + a_1 + a_0$$
$$N_0(z) = b_2 d z^2 + b_1 d z + b_0$$

Since the two transfer functions have the same denominator, the following input–output representation is obtained

$$\begin{align*}
y(k+1) &= -a_2 y(k) - a_1 y(k - 1) - a_0 y(k - 2) + b_2 u(k) + b_1 u(k - 1) \\
&\quad + b_0 d(k - 2) + b_2 d(k) + b_1 d(k - 1) + b_0 d(k - 2)
\end{align*}$$

Finally, system (5) can be given the following state-space (nonminimal) representation

$$\begin{align*}
\dot{x}_0(k+1) &= A_{00} x_0(k) + B_{00} u(k) + M \xi_0(d(k)) \\
y(k) &= C_{00} x_0(k)
\end{align*}$$

where $x_0(k+1) = [y(k+1)', \ldots, y(k-n+2)', u(k)', \ldots, u(k-n+2)', \ldots, d(k-n+2)']'$ and the matrices $A_{00}, B_{00}, M, C_{00}$ are defined accordingly.

In order to derive the LMPC control law the following quadratic discrete-time cost function is considered

$$J(x_0(k), u(\cdot)) = \sum_{i=0}^{N-1} \left[ \left\| y^*(k+i) - y(k+i) \right\|^2_{Q_0} + \left\| u(k+i) \right\|^2_{R_0} \right]$$

where $N$ is the prediction horizon, $y^*(k)$ the desired output at time $k$ and $Q_0 = Q_0^T > 0$, $R_0 = R_0^T > 0$, $S_0 = S_0^T > 0$ are weighting matrices. The evolution of the system can be rewritten in a compact way as follows:

$$\begin{align*}
Y(k) &= A_{00} x_0(k) + B_{00} U(k) + M_0 D(k) \\
Y(k) &= [y(k+1), y(k+2), \ldots, y(k+N-1), y(k+N)]'. D(k) = [d(k)', d(k+1)', \ldots, d(k+N-1)', d(k+N)']'. U(k) = [u(k)', u(k+1)', \ldots, u(k+N-1)', u(k+N)']'.
\end{align*}$$

4.2. Constrained linear model predictive control

With a relatively small increase of the computational burden it is possible to consider explicitly both input and state constraints by solving a constrained linear quadratic optimization problem. This can be done by solving on-line a quadratic programming problem or by using an explicit solution derived through a multiparametric approach. In this paper the results obtained with constrained LMPC are not reported because they did not show any significant improvement. In fact, the explicit consideration of input constraints alone does not improve the performance of the unconstrained saturated control law, while the fully constrained problem, i.e. also with state constraints, introduces nontrivial problems due to the approximation error caused by linearization and model reduction. Further work is required to explore this approach.

4.3. Nonlinear model predictive control

Unconstrained LMPC with a quadratic cost function provides an explicit control law. The simplicity of this solution is a great advantage in order to implement the controller on real patients.

In the following, a nonlinear model predictive control is proposed. The goal is to evaluate the (upper) performance limit of a predictive control algorithm in the blood glucose control problem. In order to do this, we consider the ideal albeit unrealistic situation:

1. the nonlinear model used in the synthesis of the NMPC exactly describes the system;
2. all the states of the systems are measurable. A state-feedback NMPC is proposed (no observer is used).

The model used in the synthesis of the NMPC is the full nonlinear continuous-time model previously described.

Since nonlinear predictive control allows for the use of nonlinear cost functions, the adopted cost function incorporates...
the kind of nonlinearity entering the LBGI and HBGI indexes previously described. In order to minimize the control energy spent within the optimization horizon, the cost function includes also a term depending on the control action.

The Finite Horizon Optimal Control Problem consists in minimizing, with respect to control sequence $u_1(t_k), \ldots, u_N(t_k)$ ($t_k = kT_S$ with $T_S$ sampling time period), the following nonlinear cost function with horizon $N$

$$J(x_{t_k}, u_1(t_k), \ldots, u_N(t_k)) = \int_{t_k}^{t_k+N} \left\{ 10(q/(\ln(G_p(\tau))))^a - b \right\}^2 + ru(\tau)^2 \, d\tau$$ (6)

Note that a full hybrid solution is developed, i.e. the control inputs $u_i(t_k)$ are constrained to be piecewise constants, $u(\tau) = u_i(t_k), \tau \in [t_{k+1}, t_{k+1}), i \in [1, \ldots, N]$ (see [14]).

The values of $g$, $a$ and $b$ adopted in this case are $g = 0.75$, $a = 1.44$, and $b = 9.02$.

5. Virtual protocol

The performance of closed-loop glucose controllers was tested on a 1-day virtual protocol (see Fig. 3):

1. Patient enters at 17.00 of Day 1. At that time, data gathering for warm-up phase starts. Microinfusor is programmed to inject basal insulin for the specific patient.

2. At 18.00 of Day 1 patient takes a meal lasting about 15 min. The meal contains 85 g of carbohydrates. At the same time, patient receives an insulin bolus $\delta$ according to his personal insulin/carbohydrate ratio (CHO ratio or CR), that is

$$\delta = 85 \cdot CR$$

3. Control loop is closed at 21.30 of Day 1. At this time, the switching phase starts. In this phase the metrics for evaluation are not calculated.

4. At 23.00 of Day 1 regulation phase starts.

5. At 7.30 of Day 2, patient has breakfast containing 50 g of CHO. Meal duration is about 2 min.

6. Experiment finishes at 12.00 of Day 2. Patient is discharged and restarts his normal insulinc therapy.

6. In silico trial

In order to test the proposed control algorithm several in silico trials have been based on the protocol described in Section 5. The main objectives are to verify: (a) the effect of a change in the parameter $q$ that is the only tuning parameter that is individualized; (b) the impact of meal announcement; (c) the robustness with respect to errors in the meal announcement information (e.g. time and amount); (d) the benefit of an explicit consideration of the nonlinear dynamics by means of a NMPC. The CVGA matrix, used to compare population performances, is computed only during the regulation period in order to reduce

![Fig. 3. Nominal scenario for adult patient.](image)

![Fig. 4. CVGA: Experiment L1 (white) versus Experiment L2 (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)](image)

![Fig. 5. CVGA: Experiment L1 (white) versus Experiment L3 (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)](image)
the influence of the open-loop regulation on the closed-loop indexes:

Experiment L1. 100 subjects are simulated using an LMPC control law synthesized with \( T_s = 15 \text{ min} \), \( n = 3 \), \( N = 16 \), \( R_d = 1 \) and \( Q_d = 5q_n = q_n \), where \( q_n \) has been individualized for each subject and the set point is taken equal to 112 mg/dl.

Experiment L2. It is equal to Experiment 1 with \( q = 2q_n \).

Experiment L3. It is equal to Experiment 1, but the LMPC control law is applied without meal announcement.

Experiment L4. The ingested amount of glucose is 125% of the nominal value for all 100 patients. The LMPC control law has the same parameters as those used for Experiment 1 and relies on the nominal glucose dose to decide the feedforward action.

Experiment L5. The ingested amount of glucose is 75% of the nominal value for all 100 patients. The LMPC control law has the same parameters as those used for Experiment 1 and relies on the nominal glucose dose to decide the feedforward action.

Experiment L6. The ingested amount of glucose is delayed of 30 min with respect to the nominal value for all 100 patients. The LMPC control law has the same parameters as those used for Experiment 1 and relies on the nominal glucose profile to decide the feedforward action.

Experiment L7. The ingested amount of glucose is 30 min in advance with respect to the nominal value for all 100 patients. The LMPC control law has the same parameters as those used for Experiment 1 and relies on the nominal glucose profile to decide the feedforward action.

Experiment N8. In order to explore the potentiality of NMPC, some virtual patients undergoing Experiment 1 were also controlled by means of NMPC.

7. Results

The performances obtained in Experiment L1, as shown in Fig. 4, are clearly satisfactory: in fact no hypo phenomenon occurs and also the maximum values are very well regulated. In particular, 59 virtual patients are in the A-zone and 41 are in the B-zone. The comparison between Experiment L1 and L2 (see Fig. 4) shows the effect of a variation of the \( q \) parameter, the only parameter to be tuned. In
Fig. 10. Experiments L1 and L4–L7 for patient 9.

Fig. 11. Experiment L1 (LMPC) versus Experiment N8 (NMPC) for patient 35.
particular, increasing the aggressiveness, i.e. using a larger $q$, all the points in the CVGA move towards the Lower B region reducing the maximum BG value but also the minimum one. 2 patients move even in the lower D-zone and 2 in the lower C-zone. The CVGA reported in Fig. 5 for Experiments L1, L3 shows that in Experiment L3, 9 patients move from A-zone to B-zone. The difference in regulation performance with and without meal announcement information demonstrates the benefit of an anticipative action in the insulin injection exploiting knowledge of the meal profile. The robustness analysis of the proposed control algorithm against wrong information on meal amount or time is performed through Experiments L4–L7 and is shown in Figs. 6–9. In all cases, all the patients are in the A- and B-zones so that. Robustness with respect to this kind of uncertainties is apparent. An example of the time evolution of the glucose and external insulin are reported in Fig. 10 for patient #9. The vertical black lines in Figs. 10 and 11 indicate the beginning of the closed-loop, the beginning of the regulation period and the breakfast nominal time. Finally, in Fig. 11 the glucose and insulin profiles obtained in the Experiment L1 with the LMPC and in Experiment N8 with NMPC are reported for patient #35. As shown in Fig. 11, the adoption of an NMPC scheme, in place of an LMPC one, may bring substantial improvement. In particular the range of glucose concentration variability is reduced and a much smoother insulin profile is achieved. The advantage of the nonlinear predictive control scheme is twofold: constraints on glycemia are explicitly allowed for and knowledge of the nonlinear dynamics is exploited.

8. Conclusions

This paper presents an engineering view of the control of diabetes based on three fundamental clinical perspectives: (1) recent studies have documented the benefits and the technological reliability of continuous glucose monitoring (CGM); (2) intensive insulin treatment attempting to approximate near-normal levels of glycemia markedly reduces chronic complications, but may risk potentially life-threatening severe hypoglycemia; (3) in silico modeling could produce credible pre-clinical information and could substitute the traditional animal trials yielding results in a fraction of the time. Hence, the future development of the artificial pancreas will be greatly accelerated by employing mathematical modeling and computer simulation.

The best way to utilize CGM data would be to devise automated procedures for the interpretation and processing of the continuous data flow, which take into account glucose utilization and the delays related to subcutaneous insulin administration. Automated control algorithms would be needed, which should be based on engineering understanding of the glucose–insulin system in diabetes, e.g. on predictive models. Risk analysis and analysis of temporal glucose variability, including methods such as the BGI and the CVGA presented in this paper, will be fundamental to both the design of open- and closed-loop control algorithms and to the evaluation of the results from their in silico and in vivo testing. A simulator of T1DM, as the one considered in this paper, should be equipped with a cohort of n in silico subjects that span the observed interindividual variability of key metabolic parameters in the T1DM population.

The in silico trial has demonstrated that linear output-feedback MPC achieves satisfactory glycemic regulation in a virtual population of 100 type 1 diabetic patients. The proposed scheme is also robust with respect to errors in the meal announcement signal. Robustness in the face of sensor errors may also be investigated by complementing the simulator with a probabilistic model of sensor noise. Moreover, the benefit brought by individualization of the control parameter $q$ has been demonstrated. For virtual patients, this parameter can be tuned via a trial and error procedure based on in silico experiments. For the real patients it is necessary to find a safe value that can be refined via a day-to-day iterative learning algorithm.

Preliminary tests conducted on virtual patients have shown that state-feedback nonlinear MPC has the potential to introduce further improvements. In fact, although linear MPC yields well-regulated glucose profiles with reasonable insulin administrations, the adoption of the nonlinear controller further reduces glucose variability using a smoother insulin profile. However, for nonlinear MPC to be usable in practice the development of state observers would be required.

References