Zone Model Predictive Control and Moving Horizon Estimation for the Regulation of Blood Glucose in Critical Care Patients

Timothy D. Knab ∗ Gilles Clermont ∗∗,∗∗∗
Robert S. Parker ∗∗,∗∗∗

∗ Department of Chemical and Petroleum Engineering, Swanson School of Engineering, University of Pittsburgh, PA, 15261.
(email:tdk17, cler, rparker@pitt.edu)

∗∗ Department of Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA, 15261

∗∗∗ McGowan Institute for Regenerative Medicine, University of Pittsburgh, Pittsburgh, PA, 15219

Abstract: Critically ill patients commonly suffer from stress hyperglycemia, or elevated glucose levels, following injury or disease. Hypoglycemia, or low glucose level, is a frequent and serious complication of treating hyperglycemia. In order to reduce the incidence of hyper- and hypoglycemia, a linear zone model-predictive controller with moving horizon state estimation and output regulation is developed. Critical care patient data from an observational study was used to construct virtual patients. Closed-loop control in these virtual patients, versus clinical standard of practice, results in a substantial increase in time spent in the target glucose zone and significant reductions in both hyperglycemia and hypoglycemia. Overall, the proposed controller significantly enhances targeted glucose control in critically ill patients in silico, which may translate to improved clinical decision making and patient outcomes in the clinic.

Keywords: Critical Care, Glucose, Insulin, Model Predictive Control, Moving Horizon Estimation, Sensor Error

1. INTRODUCTION

Patients admitted to critical care units commonly present with stress hyperglycemia, or elevated blood glucose levels, resulting from hormonal and inflammatory responses to stress of injury or disease (Van den Berghe et al. (2001); Thorell et al. (2004)). This response upregulates endogenous glucose production (EGP) and a strong pro-inflammatory immune response, both of which contribute to an effective decrease in insulin sensitivity (SI) (Van den Berghe et al. (2001)). The resulting insulin resistance drives hyperglycemia in critically ill patients.

A seminal study by Van den Berghe et al. (2001) indicated that a reduction in hyperglycemia through the use of tight glucose control (80-100 mg/dL) had the potential to drastically reduce morbidity and mortality in the post-surgical ICU, yet these results have not borne out in subsequent studies. Notably, the NICE-SUGAR study (Finfer et al. (2009)), which randomized patients into conventional and targeted glucose control groups, reported no significant difference in measured outcomes between the two groups. Retrospective analysis by The NICE-SUGAR Study Investigators (2012) suggests that an increased incidence of hypoglycemia in the targeted glucose control group contributed to increased mortality in that group. This conclusion is supported by the outcomes of the VISEP (Brunkhorst et al. (2008)) and Glucontrol studies (Preiser et al. (2009)).

An analysis of patients included in the High-Density Intensive Care (HIDENIC) database at the University of Pittsburgh Medical Center (UPMC) identified a blood glucose concentration range between 110 and 130 mg/dL
as being associated with lowest patient mortality (Fig 1). Pre-clinical studies suggesting a causal link between hyperglycemia, hypoglycemia and poorer outcomes, targeting glucose levels within this range appears to be a desirable, data-driven, clinical objective. Yet, this requires close monitoring of blood glucose levels and more frequent intervention than is practical by critical care staff. Furthermore, current care protocols do not explicitly address inter- and intra-patient variability in glucose-insulin dynamics stemming from differing insulin sensitivities. High-frequency sampling of blood glucose levels through continuous glucose monitors, coupled with model-based control, has the capacity to enable zone-targeted glucose control thereby decreasing hyperglycemia while avoiding hypoglycemia ultimately leading to improved patient outcomes (Boyd and Bruns (2014)).

The work described herein details the development and virtual testing of a linear automated model predictive controller with state estimation for the delivery of both insulin and glucose to maintain patient blood glucose levels within a target zone (110-130 mg/dL).

2. METHODS

2.1 Glucose-Insulin Model in Critical Care

The Intensive Control Insulin-Nutrition-Glucose (ICING) model by Lin et al. (2011), originally developed for, and validated in, critical care patients, was used to capture patient dynamics from our study. Many of the patients enrolled in the current study at the University of Pittsburgh received normal and fast-acting insulin via subcutaneous injection and/or continuous infusion; to capture these dynamics, a control-relevant model of subcutaneous insulin dynamics developed by our group (Vilkhovoy et al. (2014)) was incorporated into the ICING model (ICING+SQ).

2.2 Fitting Virtual Patients

The ICING+SQ model was fit to patient data collected through continuous subcutaneous glucose monitoring (CGM) by Dexcom® Platinum™ G4 sensors. Each patient (n=24) enrolled in the study at UPMC had two CGM sensors inserted subcutaneously in the abdominal region. Data records of glucose and insulin infusions and injections (both intravenous and subcutaneous) prior to, and during CGM monitoring were obtained from UPMC medical records for each patient.

Individual virtual patients were developed by estimating endogenous glucose production (EGP) and insulin sensitivity (SI) to CGM data “fused” into a single composite measurement using a Kalman filter. Model inputs included all recorded infusions and injections. An insulin sensitivity (SI) profile was then fit, allowing for SI changes at each 5 minute sampling interval such that model predicted blood glucose closely matched the filtered and fused CGM data. Smooth SI profiles with relatively slow dynamics, representative of the physiological changes during recovery from trauma, were desired. This was ensured by penalizing point to point changes in SI during the estimation process. Patient-tailored profiles of SI were established for each patient. while this is different from a typical identify-and-validate modeling approach, our goal was not to establish the validity of a model but rather to have a population of patients with different dynamics on whom we could test our control system in silico.

Numerical Methods Models were constructed as ordinary differential equations in the Coopr/Pyomo environment (Hart et al. (2012)). Parameter estimation was performed using nonlinear least squares and solved using IPOPT (Wächter and Biegler (2006)) by discretizing the ICING+SQ model into finite elements (5-minute duration elements), with each finite element having three Radau collocation points (Ascher and Petzold (1998)).

2.3 Sensor Noise and Sensor Reconciliation

Sensor error was computed as BG_{Sensor} − BG_{Fingerstick} using all 461 fingerstick blood glucose measurements from patients enrolled in the study. From these computed errors, joint probability distributions for both sensor error vs. fingerstick measurement (true value, Fig. 2(a)) and sensor error vs. sensor reading (measured value, Fig. 2(b)) can be computed. The joint probability distributions in Fig. 2 can be conditioned on virtual patient blood glucose outputs (Fig. 2(a)) or CGM sensor reading (Fig. 2(b)) to generate errors for use in testing the control algorithm, and to assess agreement between the two CGMs for a given patient.

For each given sensor measurement, the joint probability distribution in Fig. 2(b) was conditioned on the measurement to provide a corresponding error distribution around the reported sensor value. Integrating the overlapping area of these distributions for both sensors and normalizing by the total area under both distributions results in a value between 0 and 0.5, which represents the confidence in the joint sensor data streams at a given time. These overlap values are used to weight the fused data points when fitting the virtual patients.

2.4 Model-Predictive Control (MPC)

Given that an optimal range of blood glucose concentrations has been documented in observational studies (Yge-neswaran et al. (2013)), an MPC formulation for control to zone (zMPC), rather than a single set-point is employed.
Let the discrete state-space model be given by:
\[x_{k+1} = Ax_k + Bu_{uk-1} + Bu_{\Delta}u_k + Bdd_k\]
\[y_k = Cx_k\]
(1)
Here, \(x_{k+1}\) is the predicted state vector at the next time step, \(x_k\) is the state vector at the current time, \(u_{k-1}\) is a vector of previous manipulated inputs, \(\Delta u_k\) is the change in manipulated inputs at the current time, \(d_k\) is a vector of disturbance inputs at the current time, and \(y_k\) is the observed output at the current time.

A linear, constrained model predictive controller is formulated as shown in the following quadratic program (adapted from Muske and Rawlings (1993)):

\[
\text{minimize}_{z(k)} \sum_{i=1}^{P} \Gamma_i (\|y_{k+i} - \delta_i\|_2^2) + \sum_{i=1}^{M} \|S \Delta u_{k+i-1}\|_2^2
\]
(2a)
subject to:
\[
x_{k+1} = Ax_k + Bu_{uk-1} + Bu_{\Delta}u_k + Bdd_k
\]
\[y_{k+1} = Cx_{k+1}\]
\[I \Delta u \geq \Delta u_{\text{min}}\]
\[I \Delta u \leq \Delta u_{\text{max}}\]
\[u_{k}, \ldots, u_{k+p-1} \geq 0\]
\[\text{Zone}_{\text{lower}} \leq \delta \leq \text{Zone}_{\text{upper}}\]
where \(z = \Delta u_k, \ldots, \Delta u_{k+N-1}\).

Matrix \(\Gamma\) penalizes model predictions of blood glucose outside the target zone over the control horizon, and \(S\) is a matrix penalizing control moves. \(I\) and \(P\), along with the prediction horizon, \(\Delta u\), and the control horizon, \(M\), are controller tuning parameters. The relative values of \(\Gamma\) and \(S\) are used to normalize the objective function contributions of out-of-zone error for blood glucose and the cost of control actions such that the first and second terms of equation (2a) are appropriately scaled.

This formulation results in targeted zone control by allowing \(\delta\) to move between the its lower and upper bounds (Zone\(_{\text{lower}}\) and Zone\(_{\text{upper}}\), respectively) as given by constraint (2g) to minimize the difference between the measured output and predicted output. While zone control has been previously employed for glucose control (Grosman et al. (2010); Gondhalekar et al. (2013); Harvey et al. (2014)), these studies are primarily in diabetic, not critical care, patients. Furthermore, the control algorithm proposed here uses a more physiologically motivated model and clinically-motivated constraints that differ from those employed in ambulatory diabetic populations.

The \(\Delta u_{\text{min}}\) and \(\Delta u_{\text{max}}\) in constraints (2d) and (2e) are the lower and upper bounds, respectively, on the allowed change in the manipulated input at any given time step. The vector of non-manipulated variables is assumed to remain constant at \(d_k\) over the prediction horizon. Solution of this quadratic program results in a vector of optimal input changes over the control horizon of length \(M\), however only the first suggested input change is implemented by the controller.

2.5 State Estimation

To account for intra- and inter-patient variability, insulin sensitivity is reformulated as a model state and estimated at each five minute interval over the past moving horizon, \(H\). For linear systems, a Kalman filter is an optimal state estimator, however it does not allow for inequality constraints. For this work, a moving horizon estimation (MHE) scheme was chosen as it allows model states to be upper and lower bounded by physiological constraints (Rao et al. (2001)). In the unconstrained case, MHE reduces to a Kalman filter (Rao et al. (2003)).

The MHE for automated glucose and insulin administration system is formulated similar to Robertson et al. (1996). Assuming that model-patient mismatch is due to process noise on both the states (\(\nu_k\)) and inputs (\(\sigma_k\)), as well as measurement noise on the output (\(\upsilon_k\)), the state-space model can be written as:
\[
x_{k+1} = Ax_k + Bu_k + B\sigma_k + \nu_k
\]
\[y_k = Cx_k + \upsilon_k\]
(3)
For the state-space model in equation (3), the MHE is defined by the following quadratic program:

\[
\text{minimize}_{z(k)} \sum_{i=k-H+1}^{k} \nu_i^T R^{-1} \nu_i + \sum_{i=k-H+1}^{k-1} \omega_i^T Q^{-1} \omega_i
\]
(4a)
subject to:
\[
x_{k+1} = Ax_k + Bu_k + B\sigma_k + \omega_k
\]
\[\nu_k = y_k - (C x_k)\]  
(4b)
\[\nu_k = x_{k-H+1} + \frac{A}{k-H+1} x_{k-H+1} - x_{k-H+1} - x_{k-H+1} - \cdots - \times_{k-H+1} - \nu_k - \cdots - \nu_k\]
(4f)
Constrains (4d) and (4e) in the MHE formulation (4) force the process noise on manipulated inputs and states that are not estimated to be zero (through suitable selection of vectors \(\alpha\) and \(\beta\)). \(R\) and \(Q\) are matrices that penalize deviations of the model from measurements and added state noise, respectively, and are used to tune the estimator in conjunction with the estimation horizon, \(H\). \(P\) is weighting matrix that represents the confidence in the state estimation. The calculation and update of \(P\) is carried out as in Robertson et al. (1996).

2.6 Virtual Patient Blood Glucose Control

The controller was tested on a bank of virtual patients, as developed in Section 2.2. The control algorithm administered insulin and glucose to the virtual patient and controller performance was evaluated by comparison to the clinical results.

The ICING+SQ model was linearized around the point corresponding to zero input, the patient’s initial glucose value, as measured by the subcutaneous sensor, and with all other states taken on steady-state values. Blood glucose is the primary output, with the target zone set at 110-130 mg/dL, however glucose and insulin infusion rate were added as secondary outputs with target rates of zero \(mg_{min}\) and zero \(md_{min}\), respectively. The addition of the output regulator serves to mitigate undesirable high frequency control action and ensures that glucose and insulin are not infused unnecessarily. This formulation also: (i) prevents a situation where the controller infuses glucose constantly
while simultaneously infusing insulin at a constant (elevated) rate; and (ii) reduces control effort by minimizing exogenous insulin usage, which is important because continuous insulin infusion suppresses endogenous insulin production, which would thereby increase the control effort required to maintain blood glucose concentrations within the desired zone and, in real world application, increase the probability of hypoglycemia.

The moving horizon estimator computes insulin sensitivity, $S_I$, at every 5 minute interval in the estimation horizon. $S_I$ is assumed to be stepwise constant across the prediction horizon for control purposes. Additional estimated states were insulin concentrations in the blood and interstitium. All estimated quantities were constrained with a lower bound of 0 and an upper bound corresponding to physiologic values.

### 3. RESULTS AND DISCUSSION

#### 3.1 Fitting Virtual Patients

By fitting $S_I$ to CGM data using dynamic optimization tools, with a penalty on point-to-point $S_I$ changes, virtual patients with smooth, slowly varying insulin sensitivity profiles were characterized. We believe the smooth nature and relatively slow dynamics of these $S_I$ profiles are more consistent with expected critical care physiology than previously reported results showing significant changes in $S_I$ over short intervals (e.g., Lin et al. (2011)). The model parameterization captures inter- and within-patient variability, and a representative patient fit to CGM data, with their corresponding $S_I$ profile, is shown Fig. 3.

Overall 24 patients were enrolled in the study, however, only 18 patients had blood glucose data sets over a window of at least 24 hours, which was sufficient to develop virtual patients. Over these 18 patients, the virtual patients exhibited a mean absolute error per point of 1.7893 ± 2.7.  

#### 3.2 Controller Performance

Utilizing the zMPC/MHE control algorithm to automate the delivery of insulin and glucose to the virtual patients resulted in a much higher fraction of time spent within the target glucose zone, a significant decrease in the incidence of hyperglycemia, and dramatic decrease in hypoglycemic events. The results from control on the representative patient are displayed in Fig. 4.

The zMPC/MHE with output regulation algorithm significantly outperforms the clinical protocol in controlling blood glucose levels while completely avoiding any hypoglycemic events. The controller tunings were selected for aggressive (rapidly-adjusted) insulin delivery, thereby allowing the controller to rapidly turn off insulin in the case of predicted hypoglycemia. Mathematically, the lengths of the estimation, prediction and control horizons, as well as the weighting matrices $Q$, $R$, $S$ and $Γ$ were adjusted across the virtual patient population such that time within the zone was maximized without excessive controller effort. The administration of glucose in pulses, lagging insulin pulses, allows the controller to recover from overly-aggressive administration of insulin (e.g. around t=1200-1400 min in Fig. 3) – a key feature that prevents hypoglycemia in our formulation.

#### 3.3 Controller Performance with Sensor Noise

To further test the performance and robustness of the controller, noise sampled from the joint probability distribution in Fig. 2(b) was added in duplicate to the blood glucose concentration of virtual patients under control. This creates two noisy data streams (representative of two CGM sensors) that were fused using a Kalman filter and filtered with a discrete-time exponentially weighted moving average (EWMA) filter before the composite signal is fed to
the controller. If the addition of this randomly-sampled error to the blood glucose concentration of the virtual patient results in a negative blood glucose concentration (a clearly non-physiological condition), the noise sample is discarded and another random error is generated. The effects of the addition of this noise on the control are shown on a representative patient in Fig. 5, and the mean control results over all 18 patients are tabulated in Table 1. An EWMA averaging over the previous three readings with a filter coefficient of 0.75 was found to have the smallest window that did not degrade controller performance.

3.4 Population Performance

Controller performance on the patient population (n=18) reported as the percentage of time spent within the zone, above the zone, and below the zone, for blood glucose signals both with and without additional noise is detailed in Table 1. Across the cohort of critically ill patients in the noise-free case, the control algorithm outperforms clinical practice, resulting in a >50% increase in time spent in the target zone, a >35% decrease in time above the target zone, and perhaps most importantly, an ≈18% reduction in time below the target zone. Furthermore, it is important to note that the clinical protocol induced moderate hypoglycemia (BG ≤ 70 mg/dL) 0.5% of the time, which under control with the zMPC/MHE algorithm was reduced to 0.32%. Neither the clinical protocol nor the zMPC/MHE algorithm resulted in any incidences of severe hypoglycemia (BG ≤ 40 mg/dL (Finfer et al. (2009))).

The presence of sensor noise in our virtual patient cohort degrades controller performance, as expected. The zMPC/MHE algorithm does still outperform the clinical protocol, resulting, on average, in almost double the time spent within the target zone as well as a significant reduction in hypoglycemia with no increase in hypoglycemia incidence. Although the addition of this noise results in degraded performance, it is important to note that this is a worst case noise scenario. The distributions depicted in Fig. 2 are computed from hourly or 6-hourly blood draws; samples compared to blood draws at 5-minute intervals would display a much higher degree of correlation and smaller distribution width in glucose error. Hence, the addition of noise randomly sampled from the Fig. 2 error distributions results in much larger noise magnitudes even after fusion and filtering. This makes it harder for the MHE to generate an accurate estimate of a patient’s state and, correspondingly, for the zMPC to make accurate predictions and corresponding control actions. Higher time-density blood glucose samples with paired CGM measurements in critical care patients are needed to more carefully characterize the distributions of Fig. 2, a topic of future work. Such data would provide a better characterization of critical care sensor noise and improved in silico testing of the control algorithm.

Although the present CGM noise distribution does introduce significant uncertainty into the blood glucose measurements, a recent editorial (Boyd and Bruns (2014)) has suggested that less accurate, but more frequent, measures of glucose levels, such as those provided by CGM systems, result in improved outcomes when targeted glucose control is employed. An analysis of a trial of the CLINICIP system also found that the control system was “advantageously” influenced by a higher sampling frequency versus the control group. The authors go on to state that the 60 minute sampling rate used for control in the standard clinical setting placed an undue strain on nurses, but anything less would result in a marked decrease in control performance.

Table 1. Summary of zMPC/MHE controller performance on our patient population (Pop.) and our patient population with additional noise (Pop.+N) as defined as percentage of time in the target zone (% Z), percentage of time above the target zone (% A), percentage of time below the target zone (% B), and percentage of the time hypoglycemia as defined by BG<70 mg/dL (%H)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Control</th>
<th>% Z</th>
<th>% A</th>
<th>% B</th>
<th>% H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pop.</td>
<td>Clinic</td>
<td>19.72</td>
<td>55.59</td>
<td>24.69</td>
<td>0.50</td>
</tr>
<tr>
<td>Pop.</td>
<td>zMPC/MHE</td>
<td>75.21</td>
<td>19.05</td>
<td>5.75</td>
<td>0.32</td>
</tr>
<tr>
<td>Pop.+N</td>
<td>zMPC/MHE</td>
<td>97.63</td>
<td>26.65</td>
<td>16.21</td>
<td>0.42</td>
</tr>
</tbody>
</table>

3. SUMMARY

A zone model predictive controller with moving horizon estimation and output regulation was developed to control critical care patient blood glucose concentrations within a target range while minimizing the occurrence of hypoglycemia. Virtual patients were constructed by fitting insulin sensitivity in the ICING+SQ model to high-frequency measured blood glucose values from critical care patients. The virtual patients’ glucose dynamics closely
matched measured glucose concentrations over periods of up to 48 hours. The virtual patients were used to test the zMPC/MHE control algorithm, and the controller performed substantially better than the clinical protocol, significantly increasing the time patient glucose concentration spent within the target zone while significantly decreasing hyper- and hypoglycemia. The addition of noise randomly sampled from error probability distributions derived from data collected at UPMC was added to the virtual patient blood glucose signal, and controller performance was still found to out-perform the clinical protocol, though the margin of improvement was less than the noise-free case. The addition of noise in this manner represents a worst case scenario, with higher-frequency paired CGM and blood glucose measurements potentially demonstrating a lower-magnitude error distribution, and correspondingly better closed-loop control. The zMPC/MHE control algorithm therefore provides a means to control patient glucose concentrations to a target range and without significant hypoglycemia – an advance in critical care practice that has the potential to outperform current clinical glucose control protocols.

REFERENCES


