

Multivariate Statistical Analysis to Detect Insulin Infusion Set Failure

Ruben Rojas, Winston Garcia-Gabin, and B. Wayne Bequette

Abstract—Multivariate statistical analysis techniques are applied to insulin infusion set failure detection (IISF), a challenging problem faced by individuals with type 1 diabetes that are on continuous insulin infusion pump therapy. Bivariate classification (BC), principal component analysis (PCA), and a combined approach were applied to simulated glucose concentrations for 10 patients, based on a nonlinear physiological model of insulin and glucose dynamics. The PCA algorithm had fewer false alarms than BC, while detecting most drifting (ramp) infusion set failures before complete failure occurred.

I. INTRODUCTION

Diabetes mellitus is a metabolic disease characterized by the inability of the pancreas to regulate blood glucose levels within a normal range (70–150 mg/dL). Diabetic patients have little or no endogenous insulin production (type 1 diabetes), or a combination of insulin resistance and insufficient pancreatic insulin release (type 2 diabetes). Individuals with type 1 diabetes mellitus (T1DM) must receive insulin, either in the form of boluses (shots) or from continuous insulin infusion pumps. The development of a closed-loop artificial pancreas, which would use continuous glucose monitoring (CGM) signals and adjust the infusion rate of continuous subcutaneous insulin infusion pumps, is a major research thrust by a number of groups throughout the world. Challenges in the creation of a closed-loop artificial pancreas are reviewed in [1-4], while [5] reviews algorithms associated with CGM technology.

A common problem encountered by diabetic patients on continuous insulin therapy is insulin infusion set failure, IISF, when Teflon catheters or steel needle infusion sets are worn for long periods of time (more than three days). Common causes of IISF include blocked or dislodged sets, inflammation, or insulin leakage back to the skin surface. IISF can cause the glucose to rise to hyperglycemic levels, even as the control algorithm (or individual) has administered the calculated insulin infusion rate, because the requested insulin was not effectively delivered. Prolonged hyperglycemia in poorly controlled diabetes is associated with long-term complications in eyes, kidneys, nerves, heart, and blood vessels.

The goal of fault detection and isolation is well-known;

detect the malfunction of the safety specific system and take any measures to avoid catastrophes, damages, injuries or simply the system breakdown. There exist several alternative ways to detect and isolate failures. The comprehensive surveys performed by Venkatasubramanian et al. [6-8] indicate that fault detection and diagnosis methods can be categorized as: (i) quantitative model-based, (ii) qualitative model-based, and (iii) process data history-based methods. Much of the focus is on complex, large-scale chemical process systems. Frank [9] and Isermann [10] provide reviews, with selected applications, of quantitative model-based fault detection and diagnosis techniques. Both reviews distinguish between abrupt (or sudden) faults, and slowly developing (drifting) faults. Using fault detection algorithms to detect IISF within as short a time as possible could allow an alarm to be activated to warn the patient to take action.

The major challenge of fault detection in insulin infusion, particularly compared to most manufacturing systems, is the limited number of measurements that are available: (i) continuous glucose signals from the CGM, (ii) insulin infusion rates. Multivariate statistical analysis has been used as a process data history method that could help in such a situation [11].

Principal component analysis (PCA) is a dimension reduction method in multivariate statistical analysis for computing linear latent variables. It has been successfully employed in numerous areas including data compression, feature extraction, image processing, pattern recognition, signal analysis, and process monitoring [12]. Thanks to its simplicity and efficiency in processing different amounts of process data, PCA is recognized as a powerful statistical process monitoring tool and widely used in the process industry for fault detection and diagnosis [8]. Recent developments in the PCA technique include adaptive process monitoring using, for instance, recursive implementation of PCA [13], fast moving window PCA [14] or kernel PCA [15].

This article presents preliminary results using a multivariate statistical analysis approach for dynamic IISF detection in an artificial pancreas framework. The proposed algorithms employ bivariate classification and PCA approaches and have been evaluated *in silico*, using a simulation environment with virtual diabetic subjects. The Food and Drug Administration (FDA) approved University of Virginia/Padova T1DM Simulator [16, 17], which uses subcutaneous insulin delivery, has been used to develop and test the algorithms.

R. Rojas is with Electrical Engineering Department, Universidad de Los Andes, Mérida 5101, Venezuela (phone: 58-274-240-2803; fax: 58-274-240-2906; (e-mail: rdrojas@ula.ve). W. Garcia-Gabin, is with Automatic Control Laboratory, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. (e-mail: wgarcia@kth.se). B.W. Bequette is with the Chemical and Biological Engineering Department, Rensselaer Polytechnic Institute Troy, NY 12180-3590 USA, (e-mail: bequette@rpi.edu).

The paper is structured as follows: section II briefly describes the initial settings and the detection algorithms development. The results obtained in the simulation scenarios are presented and discussed in section III; and finally some conclusions are drawn in section IV.

II. METHODS

A. Initial Settings

The UVa/Padova T1DM simulator [16, 17], was modified to simulate infusion set failures in the second day of a three day scenario, with three regular meals each day. The IISF were simulated as ramp insulin infusion delivery degradations from 100% to 0% in a six hour period, for a patient classified as the average child (see Fig. 1) and 10 other simulated child patients. This ramp infusion degradation was introduced at midnight (12:00 a.m.) when no meals are present, at noon (12:00 p.m.) simultaneous with the lunch meal and at 4:00 p.m., two hours before the dinner meal, causing changes in the glucose time courses (Fig. 2). The CGM signal was sampled every five minutes.

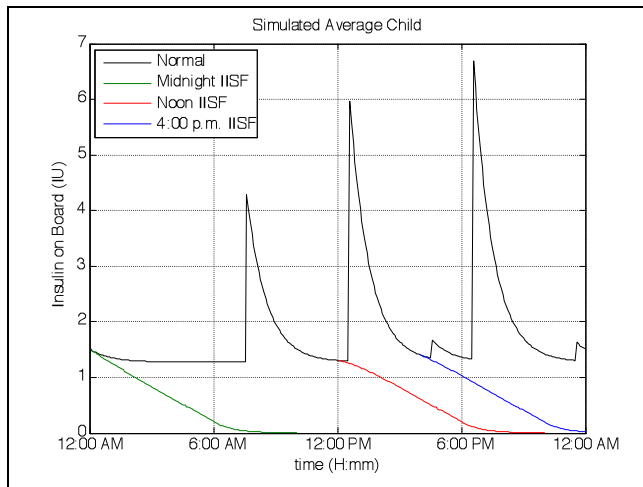


Fig. 1. Simulated insulin set failures as a ramp delivery degradation starting at midnight, at noon and at 4:00 p.m. for the simulated average child.

B. Bivariate Classification

As can be observed in Fig. 2, the IISF causes a change in the plasma glucose time course slope (SLOPE) while it is assumed that the normal insulin infusion is delivered (insulin on board, IOB). In this case, for all the simulated patients the plasma glucose slope was calculated for any point considering the least squares estimation of the slope over a 45 min window:

$$S_i = (x_i - \bar{x}_i)^T (x_i - \bar{x}_i) / [(x_i - \bar{x}_i)^T (T_i - \bar{T}_i)] \quad (1)$$

Where S_i is the estimated slope at the i -point, x_i and T_i are vectors containing the last 45 min data (10 points) for plasma glucose and daily time, respectively. Finally \bar{x}_i and \bar{T}_i stand for the vectors mean values.

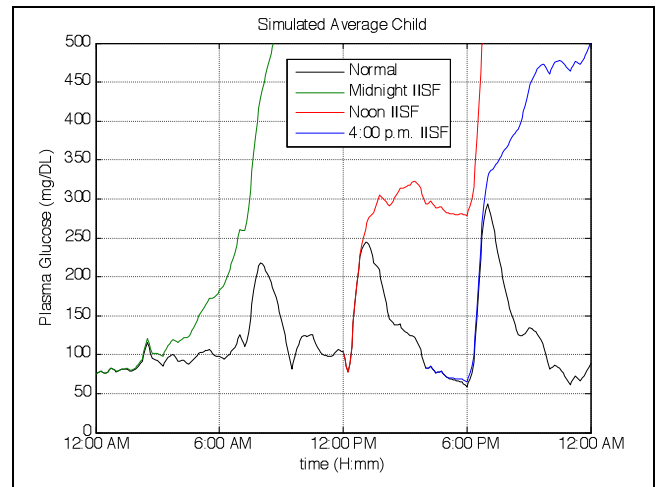


Fig. 2. Plasma Glucose responses for three simulated ramp set failures.

To apply the *BC algorithm*, two classes are assumed: “normal” and “faulty” classes. The idea is to find a way to separate both classes. It is evident that neither variable, SLOPE nor IOB by its own is useful to separate the two classes. Assuming that the simulated average child represents the mean response of the diabetic children population, a scatter plot using these two variables allows a discrimination of both classes (*bivariate classification*).

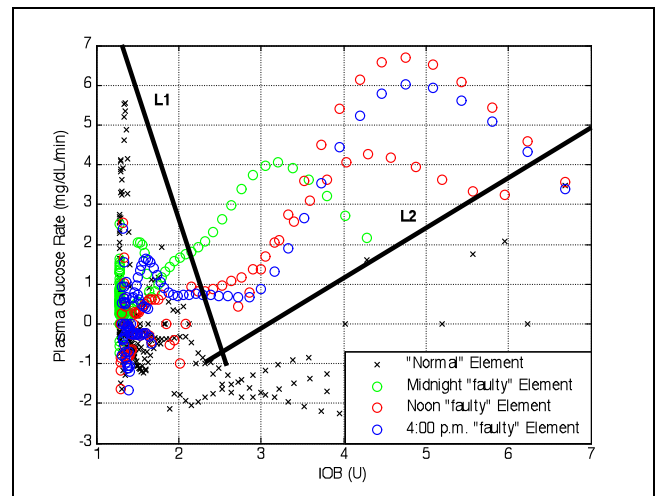


Fig. 3. Bivariate Classification, the “faulty” region is considered for all elements over L1 and L2.

The proposed approach uses the boundary region for the “faulty” class (defined as the region where only faulty elements are present), as a threshold to detect when any patient glucose time course enters the “faulty region” (see, L1 and L2 in Fig. 3). For any child patient simulation, if three consecutive calculated pairs $\langle IOB_i, S_i \rangle$ are over L1 and L2, a IISF is assumed.

C. Principal Components Analysis

In this case, every simulated child data set was arranged in a feature matrix X where each row (object) corresponds to the actual analyzed element x_i , used to calculate the slope (10 plasma glucose measurements over the last 45 min

window). Because x_i only contains plasma glucose measurements, each column (feature) is highly correlated with the others so a PCA reduction to extract the main information of the plasma glucose was carried out.

For the simulated average child, the centered feature matrix Xc was calculated as follows:

$$Xc_j = (X_j - \bar{X}_j) \quad j = 1, 2, \dots, m \quad (2)$$

where j corresponds to the feature matrix columns (in our case $m = 10$). Principal Components of Xc (PCs), were calculated to find that the first component represents 97.7% of the data variation and 99.96% can be represented by the first three principal components.

Using the same approach as the *bivariate classification* a detection algorithm (*PCA algorithm*) was proposed for the elements belonging to the “faulty” class (see, Fig. 4). In this case the boundary region for the “faulty” class (L3) defined a threshold for the first principal component score detecting the IISF when any patient glucose time course enters the “faulty region”.

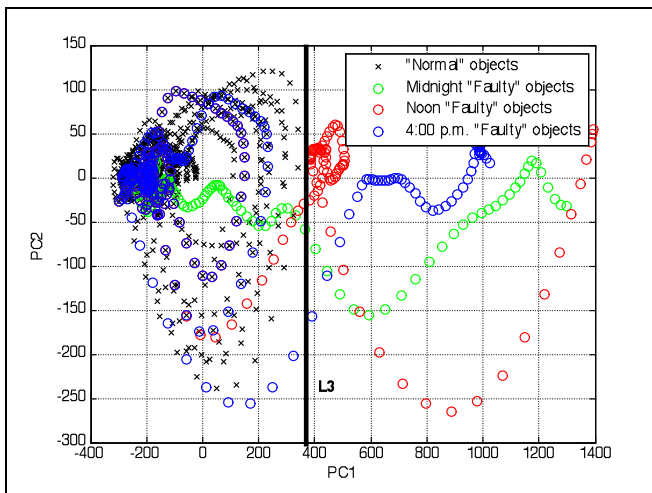


Fig. 4. PCA Classification using PC1, the “faulty region” is considered for all elements to the right of L3.

Finally, another IISF detection algorithm as a combined approach that includes *bivariate classification* and PCA by using the first component, $PC1$, obtained by the PCA analysis and the insulin on board signal, IOB , as variables for the *bivariate classification*. In this case (*PCA IOB algorithm*), the boundary region for the “faulty” class (L4 and L5) defined a threshold for the $\langle IOB_i, PC1_i \rangle$ pairs detecting the IISF if the patient glucose time course enters the “faulty region” (see, Fig. 5).

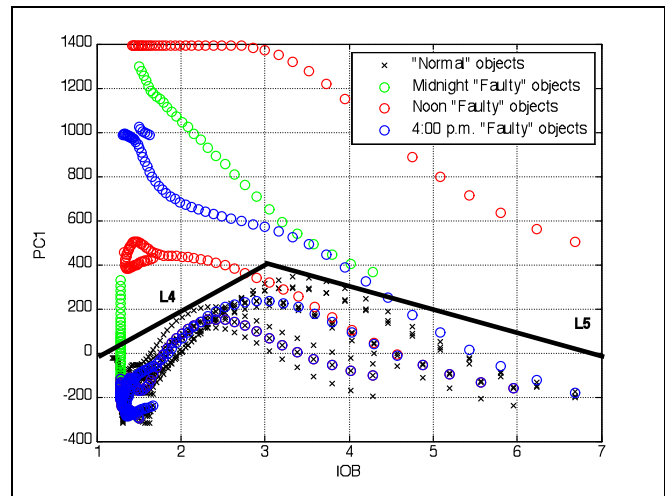


Fig. 5. Bivariate Classification using IOB and PC1, the “faulty region” is considered for all objects above L4 and L5.

III. SIMULATION RESULTS

After applying the initial preprocessing the three proposed algorithms (*BC*, *PCA* and *PCA IOB*) were tested using the 10 simulated child data files with: i) no IISF, ii) ramp IISF starting at Midnight, iii) ramp IISF starting at Noon and iv) ramp IISF starting at 4:00 p.m..

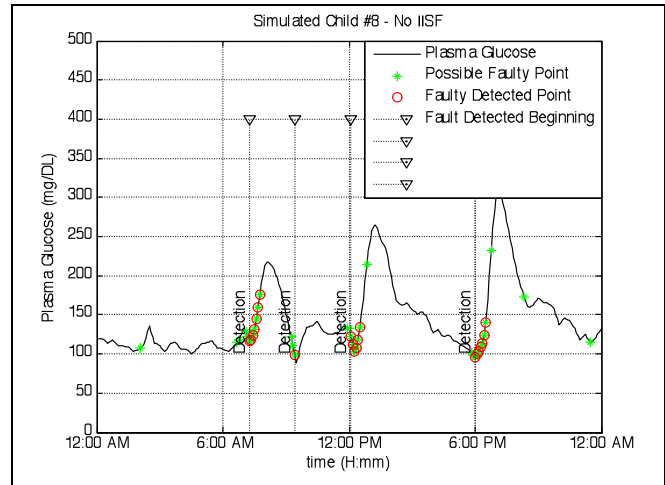


Fig. 6. Plasma glucose time course for the simulated child #8 in normal regulated conditions. Four detection signals were activated (false positives) when the *BC* algorithm was used.

Fig. 6 shows the plasma glucose time course for simulated child #8 for normal condition without IISF when the *BC* algorithm is applied. As can be observed, four detection signals were activated during the simulated 24h period. These signals are considered false positives, FP (the detection signal is activated when no IISF has occur).

Fig. 7 shows the normal and the simulated midnight IISF ramp, for child #2. The IISF detection occurs at 7:00 a.m., seven hours after the simulated ramp IISF starts but just one hour after complete set failure is achieved.

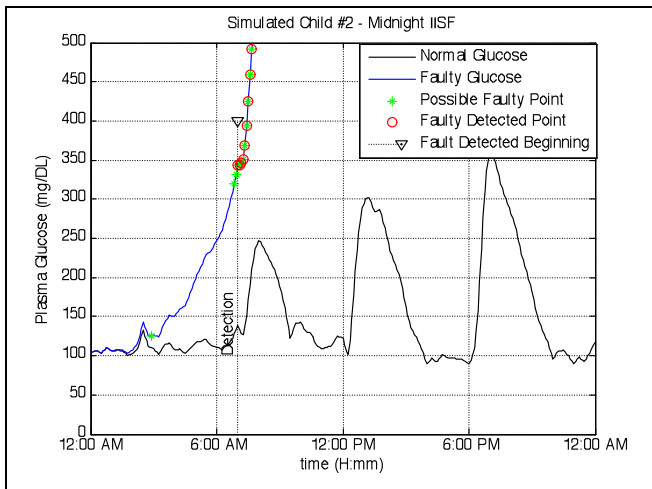


Fig. 7. Plasma glucose time course for the simulated child #2 when a ramp IISF starting at Midnight was simulated. IISF was detected at 07:00 a.m. without FPs when the *BC algorithm* was used.

Fig. 8 shows the *BC algorithm* results when the simulated ramp IISF onset starts at noon. It presents the IISF detection signal at 1:25 p.m. with a previous FP (8:00 a.m.); note that the last detection signal would not occur if the appropriate correction action is performed at 1:25 pm.

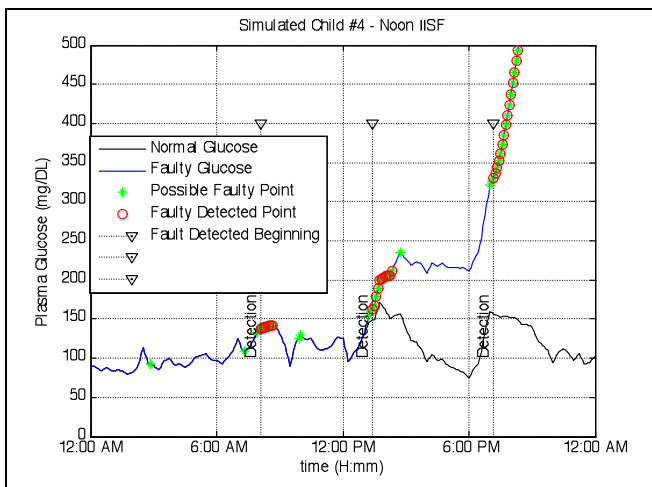


Fig. 8. Plasma glucose time course for child #4 when a ramp IISF starting at Noon was simulated. IISF was detected at 1:25 p.m. with a previous FP when the *BC algorithm* was used.

As can be observed when the *bivariate classification algorithm* was used a high rate of false positives, FP, were detected. From the previous results, the *BC algorithm* is sensitive to changes in plasma glucose slope.

For *PCA algorithm* the following results were obtained: Fig. 9 shows the plasma glucose time course for the simulated child #8 for normal regulated condition without IISF. As can be observed, in this case just one detection signal was activated during the simulated 24h period. This shows a significant decrease in the false positive rate.

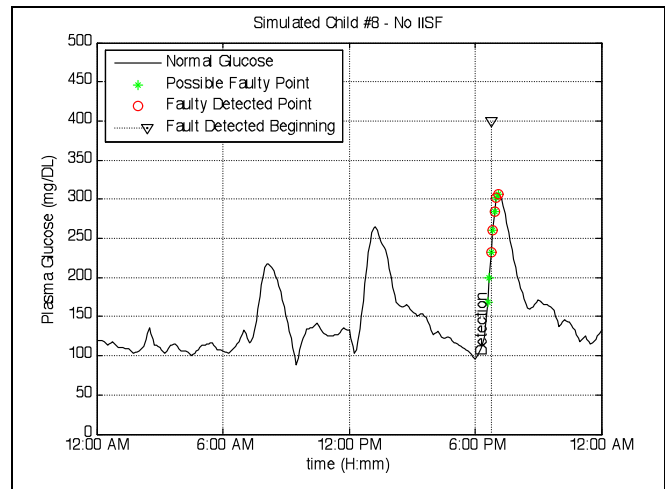


Fig. 9. Plasma glucose time course for the simulated child #8 in normal regulated conditions. Just one detection signals was activated (false positive) when the *PCA algorithm* was used.

Fig. 10 shows the normal and Midnight simulated ramp IISF onset for child #2 when the *PCA algorithm* was used. The IISF detection is at 6:50 a.m., just 10 minutes before that when the *BC algorithm* was used (See Fig. 7).

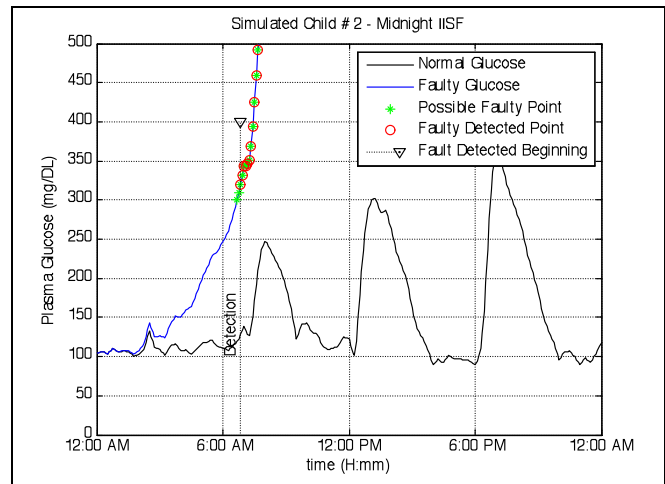


Fig. 10. Plasma glucose time course for the child #2 when a IISF starting at Midnight was simulated. IISF was detected at 6:50 a.m. without FPs when the *PCA algorithm* was used.

Fig. 11 shows the *PCA algorithm* results when the simulated 6-hr ramp IISF starts at noon. It shows the IISF detection signal at 7:10 p.m. that represents a significant delay with respect to that obtained by the *BC algorithm* (Fig. 8), but without FPs. The rate of false positives, FP, decreased when the *PCA algorithm* was used compared to the *BC algorithm*, without significant increment in the actual time of IISF detection.

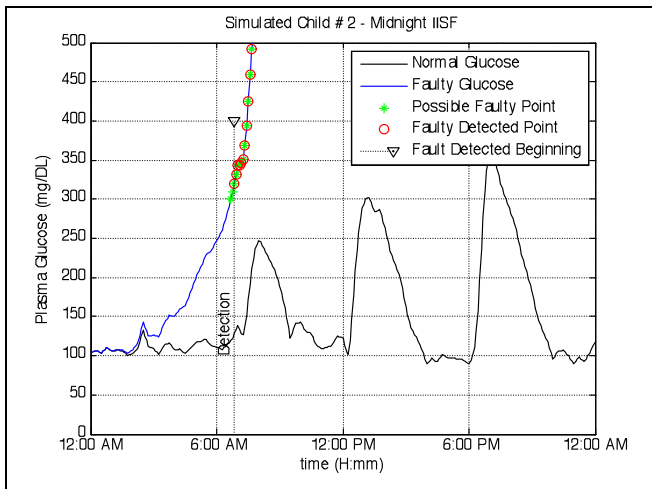


Fig. 11. Plasma glucose time course for child #4 when a IISF starting at Noon was simulated. IISF was detected at 7:10 p.m. but without FP when the *PCA* algorithm was used.

For *PCA_IOB* algorithm the following results were obtained: Fig. 12 shows the plasma glucose time course for the simulated child #8 for normal regulated condition without IISF. As can be observed, in this case two detection signals were activated during the simulated 24h period. This shows an increase in the false positive rate with respect to the *PCA* algorithm.

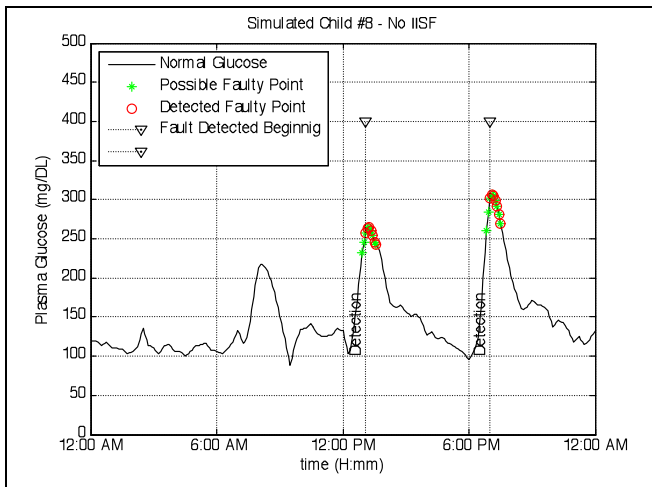


Fig. 12. Plasma glucose time course for the simulated child #8 in normal regulated conditions. Two detection signals were activated (false positives) when the *PCA_IOB* algorithm was used.

Fig. 13 shows the normal and the Midnight simulated IISF onset, for child #2 when the *PCA* algorithm was used. It presents the IISF detection at 4:30 a.m. that represents two hours and twenty minutes before of that when the *PCA* algorithm alone was used (see Fig. 10).

Fig. 14 shows the *PCA_IOB* algorithm results when the simulated IISF onset starts at noon and at 4:00 p.m., respectively. Fig. 14 shows the IISF detection signal at 7:05 p.m., just five minutes before of that obtained by the *PCA* algorithm (see Fig. 11) which no represents a significant difference.

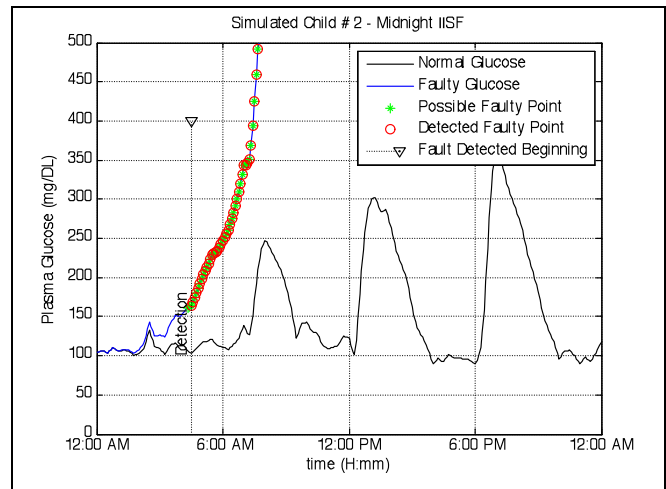


Fig. 13. Plasma glucose time course for the child #2 when a IISF starting at Midnight was simulated. IISF was detected at 4:30 a.m. without FPs when the *PCA_IOB* algorithm was used.

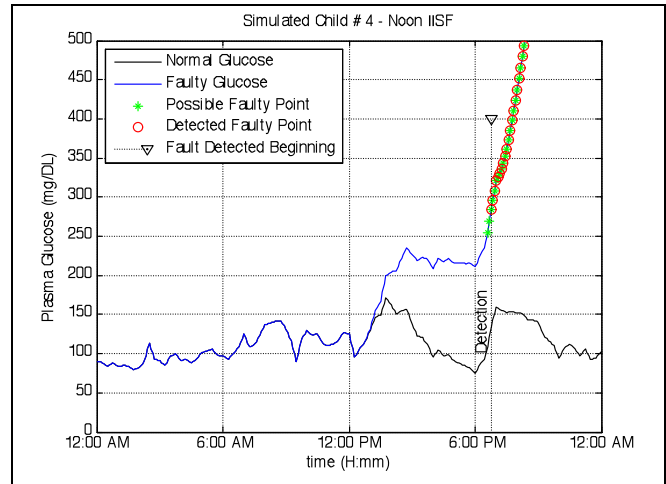


Fig. 14. Plasma glucose time course for child #4 when a ramp IISF starting at Noon was simulated. IISF was detected at 7:05 p.m. but without FP when the *PCA_IOB* algorithm was used.

As can be observed the rate of false positives, FP, increased when the *PCA_IOB* algorithm was used respect to the *PCA* algorithm but the actual time of IISF detection decreased; overall results are summarized in tables 1 and 2.

Table 1 presents the rates of false positives for the three algorithms calculated for each experimental setting where a period without IISF was simulated. Midnight IISF was not included because the experimental simulation window starts with the insulin infusion set failure then there is not time before the simulated IISF. So, for the 24 hour experimental window there are 24, 12 and 16 simulated hours without IISF for the NO IISF, Noon IISF and 4:00 p.m. IISF settings, respectively.

This table shows that the larger mean rate of 0.16 false positives per hour was obtained for the *BC* algorithm while the smaller (0.02 FP/h) was for the *PCA* algorithm but with larger detection times that the *PCA_IOB* algorithm (see Table 2).

Table 1. Rate of False Positives (#FP/hour; Mean [SD])

EXPERIMENTAL CASE	BC ALGORITHM	PCA ALGORITHM	PCA_IOB ALGORITHM
NO IISF (24H)	0.21 [0.13]	0.04 [0.05]	0.08 [0.08]
NOON IISF (12H)	0.13 [0.06]	0.00 [0.00]	0.05 [0.08]
4:00 P.M. IISF (16H)	0.16 [0.04]	0.02 [0.03]	0.07 [0.08]
AVERAGE	0.16 [0.08]	0.02 [0.03]	0.07 [0.08]

Table 2. Detection time (HH:MM; Mean [SD])

EXPERIMENTAL CASE	BC ALGORITHM	PCA ALGORITHM	PCA_IOB ALGORITHM
MIDNIGHT IISF (24H)	7:26 [0:50]	7:48 [1:03]	6:04 [2:03]
NOON IISF (12H)	1:04 [0:11]	3:04 [2:24]	2:05 [2:00]
4:00 P.M. IISF (16H)	2:45 [0:51]	3:49 [0:48]	3:09 [1:35]

IV. CONCLUSIONS

Multivariate Statistical Analysis is a promising tool for insulin infusion set failure detection in an artificial pancreas framework. This preliminary study indicates that the *Bivariate Classification* algorithm is sensitive to changes in plasma glucose slope, firing the detection signals excessively, causing a high rate of false positives. The *PCA algorithm* shows the smallest rate of false positives without significant degradation in the actual time of IISF detection. Better times of IISF detection were obtained when the *PCA_IOB* algorithm was used, but the rate of false positives was higher, giving an overall trade-off between actual detection and false alarms.

V. ACKNOWLEDGMENTS

RR was supported by the Fulbright Visiting Scholars program during his sabbatical stay at Rensselaer Polytechnic Institute, which led to the work presented in this paper. BWB acknowledges support from the Juvenile Diabetes Research Foundation (JDRF) Artificial Pancreas Program. The authors wish to acknowledge many fruitful discussions with Fraser Cameron, Bruce Buckingham and Darrell Wilson.

REFERENCES

- [1] Bequette BW. A Critical Assessment of Algorithms and Challenges in the Development of an Artificial Pancreas. *Diabetes Technology and Therapeutics*; vol. 7, no. 1, pp. 28-47, 2005.
- [2] Doyle FJ III, Jovanovic L, Seborg DE. A tutorial on biomedical process control: glucose control strategies for treating type 1 diabetes mellitus. *J. Proc. Cont.*; vol. 17, no. 7, pp. 572-576, 2007.
- [3] Kumareswaran K, Evans ML, Hovorka R. Artificial pancreas: an emerging approach to treat Type 1 diabetes. *Expert Rev. Med. Dev.* Vol. 6, no. 4, pp. 401-410, 2009.
- [4] Cobelli C, Dalla Man C, Sparacino G, Magni L, De Nicolao G, Kovatchev B. Diabetes: models, signals and control. *IEEE Rev. Biomed. Eng.* vol. 2, pp. 54-96, 2009.
- [5] Bequette, B.W. Continuous Glucose Monitoring: Real-Time Algorithms for Calibration, Filtering and Alarms. *J. Diabetes Sci. Technol.* vol. 4 no. 2, pp. 404-418, 2010.
- [6] Venkatasubramanian, V., R. Rengaswamy, K. Yin and S.N. Kavuri "A review of process fault detection and diagnosis. Part I: Quantitative model-based methods," *Comp. Chem. Engng.*, 27, pp. 293-311, 2003.
- [7] Venkatasubramanian, V., R. Rengaswamy, K. Yin and S.N. Kavuri "A review of process fault detection and diagnosis. Part II: Qualitative models and search strategies," *Comp. Chem. Engng.*, 27, pp. 313-326, 2003.
- [8] Venkatasubramanian, V., R. Rengaswamy, S.N. Kavuri and K. Yin "A review of process fault detection and diagnosis. Part III: Process history based methods," *Comp. Chem. Engng.*, 27, pp. 327-346, 2003.
- [9] Frank PM "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy – a survey and some new results," *Automatica*, vol. 26, no. 3, pp. 459-474, 1990.
- [10] Isermann R. "Model-based fault-detection and diagnosis – status and applications," *Annual Reviews in Control*, 29, pp. 71-85, 2005.
- [11] Varnuza K., Filzmoser P. *Introduction to Multivariate Statistical Analysis in Chemometrics*. CRC Press, Boca de Raton, 2009.
- [12] Jolliffe I. T. *Principal Component Analysis*. New York, Springer-Verlag, 1986.
- [13] Li W, Yue H, Valle-Cervantes S, et al. Recursive PCA for Adaptive Process Monitoring. *Journal of Process Control*, vol. 10, no. 5, pp. 471-486, 2000.
- [14] Wang X, Kruger U, Irwin G W. Process monitoring approach using fast moving window PCA. *Industrial & Engineering Chemistry Research*, 44, pp. 5691-5702, 2005.
- [15] Liu X, Kruger U, Littler T, et al. Moving window kernel PCA for adaptive monitoring of nonlinear processes. *Chemometrics and Intelligent Laboratory Systems*, 96, pp. 132-143, 2009.
- [16] C. Dalla Man, DM. Raimondo, RA. Rizza, and C. Cobelli, "GIM, simulation software of meal glucose-insulin model," *J. Diabetes Sci. Technol.* vol. 1, no. 3, pp. 323-330, 2007.
- [17] B. P. Kovatchev, M. Breton, C. Dalla Man, and C. Cobelli, "In silico preclinical trials: a proof of concept in closed-loop control of type 1 diabetes," *J. Diabetes Sci. Technol.*, vol. 3, no. 1, pp. 44-55, 2009.