HyMn: Mining Linear Hybrid Automata from Input Output Traces of Cyber-Physical Systems

Imane Lamrani, Ayan Banerjee, and Sandeep K. S. Gupta
iMPACT lab CISDE, Arizona State University, Tempe, AZ
(ilamrani,abanerj3,sandeep.gupta)@asu.edu

***Abstract*—Hybrid systems are versatile in modeling the interaction between the cyber and physical components of cyber-physical control systems (CPS) such as artificial pancreas (AP). They are typically used for analysis of safety of the human centric control systems which have serious consequences of failure. As such hybrid systems are parameterized and the variables often depend on the subject on which the control system is deployed. Traditionally, control systems are initially developed using average statistical estimates of the subject specific parameters. However, such excursions may lead to suboptimal designs. In this paper, we propose HyMn, a hybrid system parameter estimation tool, where the subject specific parameters in a hybrid system are automatically learned from experimental traces of the operation of a human centric CPS control system. We apply HyMn to the AP system and show that the blood glucose control is enhanced using the learned patient specific parameters.**

***Index Terms*—Mining hybrid automata, CPS, Artificial Pancreas, Fisher Information, Cramer-Rao Bound**

I. Introduction

Cyber-physical system (CPS) design and implementation has seen a new revolution of personalization. In the medical domain, one of the primary result of the human genome project is personalized medicine, where diagnosis and treatment for an individual depends on the person's unique clinical, genetic, genomic, and environmental information [14]. The implication being any medical CPS used by an individual should be configured considering the unique parameters of the individual. Given such trends towards personalization or customizations, the CPS verification techniques should also be equipped with capabilities to configure themselves for considering unique parameters of the physical system. One of the versatile tool used for CPS verification is the hybrid system. This paper considers the automated generation of personalized or customized hybrid system models of CPS from observed input output traces of the CPS. In addition, some devices are designed based on data collected from individuals in real world scenarios, for example the AP or closed loop blood glucose controller [8], [10]. However, such excursions may lead to suboptimal design. In this paper, the subject specific parameters in a hybrid system are automatically learned from experimental traces of the operation of a human centric CPS control system, as shown in Figure 1. We show that the artificial pancreas (AP) control system that uses the obtained patient specific parameters learned via HyMn leads to a better control of the system, as depicted in Figure 1. The hybrid system representation of the CPS can be used to perform reachability analysis and argue about safety and stability of the CPS in scenarios where experimental data is not available. Such hybrid system models can potentially be derived from input output traces of a CPS.

Contributions: In this paper, we propose the **HyMn** algorithm to derive hybrid systems of CPS:

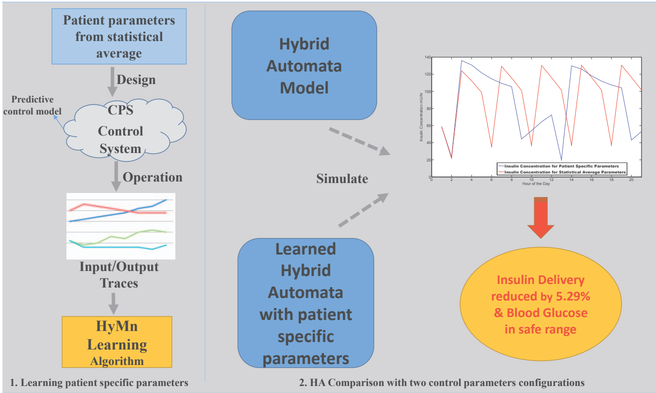

Fig. 1. Overall scheme of the HyMn approach.

- The algorithm takes the following inputs -
  - The time series traces of the controller outputs, and
  - the observed variables that are collected using sensors by the controller and used to compute the next controller output
- It employs a hybrid system mode segmentation methodology to derive the discrete modes of the hybrid system
- It employs Fisher information based analysis and Cramer Rao bound to derive the guards that define transitions between two modes. Such guards are assumed to be urgent [12] and not only limited to rectangular guards but it also includes guards that are linear functions of the continuous system variables.
- For each derived mode, it employs the Cramer Rao bound theorem to derive the flow equations.

We executed the AP control system with two parameter configurations, parameters obtained using average statistical estimates on data collected from individuals in real world scenarios [11] and the patient specific parameters learned using the HyMN approach. We show that the total insulin delivery using patient specific parameters is reduced while the blood glucose level is maintained within a safe range. The paper is organized as follows: Section II discusses competing works towards solving the discussed problem, Section III the system model of CPS and definitions related to hybrid systems while section III-C provides an example of a CPS hybrid system. Section IV shows the HyMn approach, while evaluation and simulation results are given in Section V, and finally Section VI, concludes the paper.

II. Related Work

Several previous work have proposed algorithms and frameworks for mining or learning a hybrid automata. One work proposed an approach for analyzing control code using abstract interpretation and infering a hybrid automaton from an abstract state transition system [2]. Another work creates a tool for translating a simulink model to a hybid automaton [3]. However, our work differs in that we infer a hybrid automata from input-output traces. Similar work has been performed for systems that exhibit input changes in the form of step functions and the derivatives of the continuous state variables are constant, which is often not observed in practice [5]. An algorithm to determine a maximum-likelihood hybrid system model using only continuous output of the system has been proposed [7], but this work assumes that guard conditions are independent of the continuous state, not observed in practice. An extension of this work has been studied by including autonomous mode transitions which are conditioned on the continuous state, but their approach assumes that the guard conditions are given [6]. This paper derives the guard conditions through clustering of the continuous states. Similarly, the problem of generating a finite state machine with a behavior that is close to the one described in the training intput/output traces has been considered [19]. However this works focuses on the inference of Moore automaton since the output actions of the machine are linear combinations of the output actions of the controllers and the guards are assumed to be boolean formulas of boolean predicates that depend on input actions. The closest work to ours is the CHARDA technique [4]. CHARDA and HyMn share same motivation in learning hybrid automata from observed run-time behavior of the system. However, they differ in the fact that CHARDA requires prior knowledge of model templates. Another approach uses clustered symbolic regressions and a machine learning algorithm to infer non-linear symbolic expressions that model the behavior of a dynamical system from unlabeled time-series data. Authors also proposed a transition modeling algorithm that search for non-linear symbolic inequalities to model guard conditions. Unlike HyMn, their work assumes that the guard condition is strictly related to a change in the inputs of the system and that the system can not have two distinct modes with similar behavior. Moreover, the behavior of the system is defined as a strict input-ouput relationship, as opposed to HyMn where behaviors are represented by differential equations. In addition, this works requires a priori knowledge of number of discrete modes, unlike HyMn.

Novelty: The HyMn algorithm differs from previous work in the following key factors:

- it can extract control modes where the controller output is a linear combination of the continuous state variables,
- the continuous state variables follow a set of linear differential equations, and
- the guard conditions can be both rectangular as well as non-rectangular expressed as a linear combination of the continuous state variables.

III. Definitions and Preliminaries

In this section, we discuss the essential components of a hybrid system and other preliminary materials needed to understand the HyMn algorithm.

A. System Model

As show in Figure 2, a CPS control system interacts with the physical system using a set of sensors and actuators. The physical system can be expressed using a set of n continuous variables $\{x_1, x_2, \ldots, x_n\}$. The continuous variables are governed by some differential equations which are also modulated by p controller outputs $\{o_1, o_2, \ldots, o_p\}$. For a linear system, the continuous variables are governed by a set of linear differential equations as shown in Figure 2. The continuous variables are provided as input to the controller based on which the controller decides on the next set of

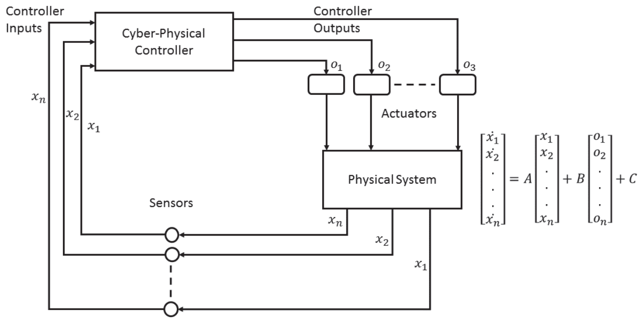

Fig. 2. Cyber-Physical Control Systems.

outputs. The continuous variables or inputs to the controller that are observed using sensors as well as the controller outputs form the input output traces that are taken as input by the proposed HyMn algorithm to extract the linear automaton.

B. Linear Hybrid System

A **linear hybrid automaton** is a formal model of a closed-loop control systems. A controller measures values of the continuous variables representing the plant using a sensor and decides to switch mode if a certain condition is satisfied. This decision is transmitted to the actuator that performs the desired change. We define a linear hybrid automaton (HA) as a tuple of the following components [1].
-$\mathcal{M} = \{m_0 \ldots m_q\}$ is a set of discrete states or *control modes* where m_0 is the initial mode.
-X is the continuous state space in which the continuous variables representing the physical system or the controller inputs $\overrightarrow{x} = \{x_1, x_2, \ldots, x_n\}$ take their values. Hence $X \subset \mathcal{R}^n$, where $\mathcal{R}$ is the set of real numbers.
-a finite set of *Control Switches* in $\mathcal{M} * \mathcal{M}$, where (m_i, m_j) defines the control switch from source mode m_i to target mode m_j.
-a *Flow* function that assigns to each control mode $m \in M$ a set of linear differential algebraic equations that relates the continuous state space variables $\overrightarrow{x}$ to its derivatives and the controller outputs. For every discrete mode m, the equation takes the following form: $\frac{d\overrightarrow{x}}{dt} = A_m \overrightarrow{x} + B_m \overrightarrow{o} + C_m$,where A_m is an $n \times n$ matrix, B_m is an $n \times p$ matrix, whereas C_m is an $n \times 1$ column vector.
-a *Guard condition* is a function that maps every control switch to a guard condition. A mode switch takes place when the corresponding guard condition is satisfied.
-a *Reset* function that maps every control switch to a reset condition. In this paper, $\dot{x}$ and $\frac{dx}{dt}$ both mean differential of x w.r.t time t.

C. Example of a CPS hybrid system

Artificial Pancreas (AP): The AP control system is used for automated control of blood glucose level for Type1 diabetic patients [10]. The controller receives glucose-meter value and outputs the right amount of insulin infusion rate I_t for the infusion pump. The aim is to maintain the prescribed level of blood glucose and avoid occurrence of hypoglycemic/hyperglycemic events. These dangerous events happen as a result of an inaccurate infusion of insulin, e.g. if the glucose concentration B_g goes above $180mg/dl$, it can lead to hyperglycemia while low glucose level i.e. below $60mg/dl$ can cause hypoglycemia. The dynamics of the AP are represented by nonlinear equations 1, 2 and 3, where $\dot{X}$ represents the rate of the variation in the interstitial insulin concentration, $\dot{G}$ is the rate of change of blood glucose concentration (B_g) for the infused insulin concentration X and $\dot{I}$ is the variation in plasma insulin concentration (B_i) [17]. The AP device has three control modes: 1- basal, where $I_t = 5$, 2- braking, where $I_t = 0.5B_g + 44.75$, and 3-correction bolus, where $I_t = 50$. Fig. 3 shows the hybrid system model of the AP. The differential equation expressing the blood glucose and insulin interaction are non-linear.

$$\dot{X} = -k_2X(t) + k_3(I(t) - I_b), \tag{1}$$
$$\dot{G} = -X(t)G(t) + k_1(G_b - G(t)), \tag{2}$$
$$\dot{I} = -k_4I(t) + k_5(G(t) - k_6)^+t. \tag{3}$$

Note that here only the blood glucose and insulin levels are the observed parameters. The parameter X is not observed but plays a significant role in relating blood glucose and insulin. We assume that this model of the physical system is available to us and then we derive the patient specific parameters. We first derive an approximate linear system that matches closely with the real AP system.

Linearization of AP model: The AP system is nonlinear in nature, hence it is necessary to linearize the system. To linearize the AP model we consider the difference in blood glucose, insulin concentration, and the interstitial insulin concentration. We consider a small time interval h and rewrite $G(h) = G(0) + \Delta G$, $X(h) = X(0) + \Delta X$, and $I(h) = I(0) + \Delta I$. We can then ignore the non-linear terms that involve multiplication of ΔX and ΔG. This results in the following linearized equations:

$$\begin{aligned} \Delta\dot{X} &= -k_2(X(0) + \Delta X) + k_3(I(0) + \Delta I - I_b)), \\ \Delta\dot{G} &= -X(0).G(0) - \Delta X G(0) - \Delta G X(0) \\ &\quad +k_1(G_b - G(0) - \Delta G), \\ \Delta\dot{I} &= -k_4I(0) - k_4\Delta I + k_5hG(0) - k_5k_6h. \end{aligned}$$

D. Fisher Information and Cramer Rao Bound

We consider the problem of deriving an unbiased estimator of a continuous variable v from a series of observations. The estimator has design parameters expressed as a vector $\overrightarrow{\theta} = \{\theta_1, \theta_2, \ldots \theta_k\}$. The term unbiased indicates that the expected value of the output of the estimator is the true value of v. Fisher information provides a measure of the information carried by v about an unknown design parameter θ_i. Given a series of observations of the variable v and executions of the estimator, the Fisher information is given by $\frac{\delta ln P(v|\theta_i)}{\delta\theta_i}$, where $P(v|\theta_i)$ is the conditional probability of the observation v given the value of the design parameter θ_i. Larger the value of this Fisher information, larger is the contribution of θ_i in determining the value of v. Hence, an effective method to reduce the number of design parameters that make significant contribution in the estimator for v is to order them in decreasing order of Fisher information and only consider those design parameters that have significantly higher Fisher information. Once, the most significant design parameters are identified, the next logical step is to derive the minimum variance unbiased estimator (MVUE), such that the mean value of the estimator output is the true value of v and the variance of the output of the estimator is minimized. In general, deriving MVUE of a system from a set of observations is an extremely difficult proposition. However, if the underlying design model is linear, then the Cramer Rao Lower Bound (CRLB) [15] theorem can be used to derive the MVUE. The CRLB considers a linear estimator for v such that: $\overrightarrow{v_o} = HD + w$, where $\overrightarrow{v_o}$ is a set of observations for the variable v, H is a set of observations for the design parameters $\overrightarrow{\theta}$, D is the matrix of coefficients for the linear estimator, and w is the observation noise. The CRLB [15] states that the Fisher information matrix is given by:

$$I = \frac{H^TH}{\sigma^2}, \tag{4}$$

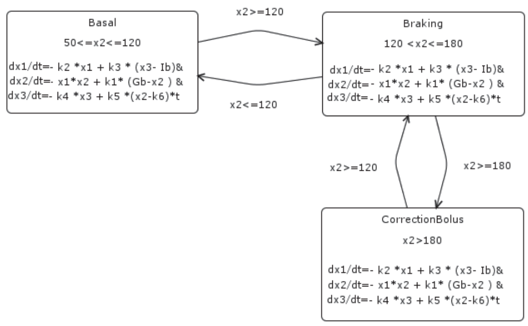

Fig. 3. Hybrid system model of the artificial pancreas.

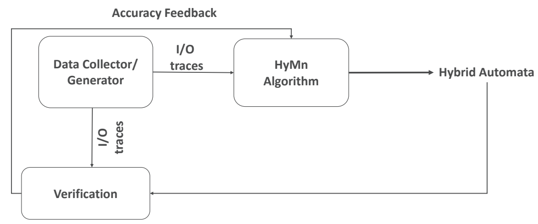

Fig. 4. HyMn inference technique.

where σ is the variance in the observation noise, while the MVUE is given by:

$$D = (H^TH)^{-1}H^Tv. \tag{5}$$

This result will be used in our HyMn algorithm for two purposes: a) to derive flow equations in modes of hybrid system using input output observations, and b) to derive non-rectangular guards which are expressed as linear combinations of continuous state variables of the hybrid system.

IV. HyMn: Hybrid System Mining Methodology

We propose HyMn, a methodology to automatically extract HA from input/output traces collected from the run-time behavior of a CPS. Figure 4 shows the main steps of the proposed technique:

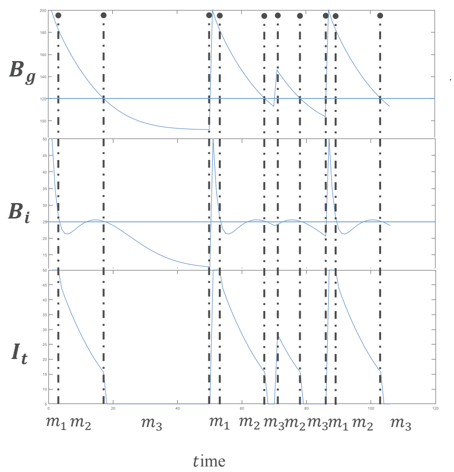

Fig. 5. Input traces (B_g, B_i) and output traces (I_t) for the AP system obtained from the T1D simulator.

Algorithm 1 HA Verification

Input: $\overrightarrow{x} = \{x_1, x_2, \ldots, x_n\}$, $\overrightarrow{o} = \{o_1, o_2, \ldots o_p\}$, and inferred HA $\{M_{inf}, X_{inf}, W_{inf}, E_{inf}, Inv_{inf}, flow_{inf}\}$
Output: A hybrid system model of the form $\{M_{inf}, X_{inf}, W_{inf}, E_{inf}, Inv_{inf}, flow_{inf}\}$

Generate output trace $\overrightarrow{o}_{inf}$ using the inferred HA
if RMSE($\overrightarrow{o}$, $\overrightarrow{o}_{inf}$) $= 0$ **then**
    Terminate
else
    Increment number of collected traces $\overrightarrow{x}$ and $\overrightarrow{o}$
    HyMn($\overrightarrow{x}$, $\overrightarrow{o}$)
end
HA Verification //recursive call

Data Collector/Generator: Input/Ouput traces are collected from the operation of CPS. In this work, we assume that the traces are noiseless. Collected I/O traces are divided into two sets: traces that are used for the inference technique and traces employed to verify the accuracy of the inferred HA.

HyMn: The HyMn algorithm takes the observed continuous states or controller inputs $\overrightarrow{x}$ and the controller outputs $\overrightarrow{o}$ as input and extracts a hybrid system of the form of the tuple $\{M, X, W, E, Inv, flow\}$ according to HA Definition (section III). The methodology has the following steps:

1- I/O Segmentation: The first step is to segment the input output traces considering times at which there is a potential discrete mode change. Whenever there is a discrete transition due to a controller mode change, the controller output changes according to the decisions of the controller. There can be two types of controller outputs for a given mode: a) a step output, where after a transition the controller output changes levels and stays at a given level unless there is another transition, and b) the output is a linear function of the continuous state variables of the physical system. For both types of output, a sudden change in the slope of the controller output indicates change in mode. The timestamps $\overrightarrow{T} = \{t_1, t_2, \ldots t_k\}$ at which such jumps occur are considered to segment the controller inputs $\overrightarrow{x}$ and are marked to be potentially different controller modes. As shown in Figure 6, modes where controller output is constant is characterized by a sharp change in the differential of the outputs. HyMn employes peak detection algorithm on the differential of the outputs and derives the modes that have a constant level as controller output. This gives the time stamps of some of the mode transitions as shown in Figure 6. The time difference between two inflection points comprises of a controller mode.

2- Mode Classification: The second step is to determine the total number of controller modes and cluster the segments into equivalence classes corresponding to each controller mode. The controller strategy or the *jump condition* for each mode can be computed using the following two steps: 1- For each segment where the output differential is zero, the controller strategy is to provide a constant level of actuation obtained from the output trace $\overrightarrow{o}$. 2- For other segments, HyMn utilizes Fisher information theory to derive the linear equation connecting the controller output to the inputs. For each output parameter, HyMn first uses Equation 4 to derive controller inputs whose linear combination gives the considered output. Then it uses Equation 5 to derive the estimator for the controller output. Segments are then grouped into classes based on the derived jump conditions. Each of this equivalent class is a composite mode and represents a unique strategy of the controller.

3- Flow Extraction: For each mode HyMn employs Fisher information and CRLB theorem to derive flow equations. The output of this re-classification are unique modes of the hybrid system, where two modes may have different jump conditions or flow equations.

4- Guard Mining: the guard mining approach takes as input the segmented input output traces where each segment is annotated with a controller mode. HyMn then considers every possible mode transitions $(m \rightarrow m')$ and considers the values of the continuous state variables at the times of transitions, then develops the observation matrix $Go_{m\rightarrow m'}$. $Go_{m\rightarrow m'}$ is an $n \times d$ matrix, where each column corresponds to an observation of the continuous state variables at the time of transition from m to m', and there are d such instances when the same mode transition is observed. In case $Go_{m\rightarrow m'}$ is full rank, HyMn obtains the rows that have constant values over all observation instances and the guard is expressed as a conjunction of equality condition $G_{m,m'} = \bigcap\{x_i = q_i\}$ on all such continuous state variables which have constant values, where q_i is the constant value in the guard observation matrix. For non-rectangular guards, the guard observation matrix will not be full rank. In such a case we consider each continuous variable x_i and express it as a linear combination of the other variables and a constant value, i.e., $x_i = A\begin{Bmatrix} x_1 & x_2 & \ldots & x_n & 1\end{Bmatrix}$, where A is the coefficient matrix. We then use the same Fisher information based analysis to derive the coefficient matrix A. The output of this step expresses guards in the form of equalities. However, we need the half planes which belong to each mode. This means for each transition m, m' we need to find inequalities. For this purpose, for each transition observed, we consider the values of the differentials of the guard expressions. If we have a guard expression as $G_i = \bigcap\{x_i = q_i\}$, then if $\dot{x_i} > 0$, then the condition for x_i is modified from $x_i = 0$ to $x_i \geq 0$. If the guard is expressed as $G_j = \bigcap\{x_i = \sum a_j x_{j_{j\neq i}} + c_i\}$, then we consider the differential of the function $f = x_i - \sum a_j x_{j_{j\neq i}}$. If from the observed I/O trace $\dot{f} > 0$ then the corresponding conjunction is modified as $x_i \geq \sum a_j x_{j_{j\neq i}} + c_i$.

In the final step, for different observations of the same mode transitions m, m' if there is a contradiction in any of the guard conjunction, then such conjunctions are eliminated from all guard expressions. This means that for two $m \rightarrow m'$ mode transition observations let us consider that the corresponding mined guards are $G_1 = \bigcap x_i \approx_1 c_i^1$ and $G_2 = \bigcap x_i \approx_2 c_i^2$, where $\approx_1, \approx_2 \in \{\geq, \leq\}$. Then the following rules must be applied:

-if $\approx_1=\geq$ and $\approx_2=\geq$, then the two terms can be replaced by the term $x_i \geq min(c_i^1, c_i^2)$,

-if $\approx_1=\leq$ and $\approx_2=\leq$, then the two terms can be replaced by the term $x_i \geq max(c_i^1, c_i^2)$,

-if $\approx_1=\geq$ and $\approx_2=\leq$, then the two terms can be eliminated from both the guard expressions if $c_i^2 \leq c_i^1$,

-if $\approx_1=\leq$ and $\approx_2=\geq$, then the two terms can be eliminated from both the guard expressions if $c_i^1 \leq c_i^2$,

Using the above-mentioned rules HyMn mines consistent guards from the observations.

Verification and HA refinement: Once the HA is generated through HyMn, its accuracy verification is crucial to the process. We compare collected I/O traces for verification to those generated using the inferred HA by calculating the root mean square error (RMSE) between the two sets of traces. The matching rate δ defines the accuracy of the inferred automaton that is evaluated according to some predefined rank α and used as a feedback to the HyMn Algorithm. The accuracy of the inferred automaton depends on the number and length of traces. For example, if the length of the trace is too short, then some of the modes can be missed, since these modes are not visible in the trace. HyMn algorithm uses the accuracy feedback to modify its inputs and refine the inferred automaton.

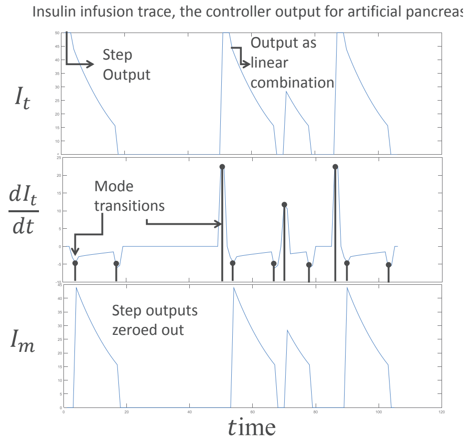

Fig. 6. I/O segmentation and jump condition retrieval example.

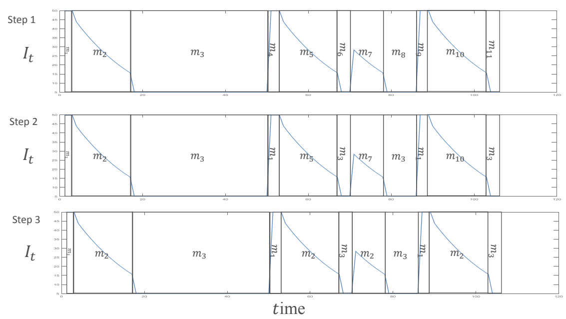

Fig. 7. HyMn mode classification execution example for AP system.

V. Evaluation

In this section, we consider the accuracy of HyMn in extracting a hybrid system. We simulate the hybrid system model of the AP for a given set of initial conditions to obtain input output traces. The simulations were carried out in Simulink and model based T1D simulator [13]. From input output traces, we apply HyMn to obtain the hybrid system model and we compare the actual and inferred tuples for accuracy. In addition, we also evaluate the operation of the two hybrid system in terms of the results of reachability analysis. We use the SpaceEx [11] tool to derive the reach set for both the given and the inferred hybrid system and compare them to find differences. We use the AP example and show results of executing each step of HyMn.

Artificial Pancreas (AP): The first step of the HyMn algorithm is to consider the differential of the controller output I_t as shown in the second part of Figure 7. Employing peak detection the HyMn algorithm initially considers that there are as many modes as the number of peaks. From Figure 7, the HyMn will consider the mode set $M = \{m_1, m_2, \ldots m_{11}\}$, i.e., 11 distinct modes. The HyMn mode classification algorithm then considers the absolute value of I_t to distinguish between modes where I_t is constant or $\frac{dI_t}{dt} = 0$. As a result of this operation HyMn finds that $m_1 = m_4 = m_9$ and $m_3 = m_6 = m_8 = m_{11}$. Hence it reduces the mode set to $M = \{m_1, m_2, m_3, m_5, m_7, m_{10}\}$. It then considers the segments where I_t is not constant as shown in I_m in Figure 7. It employs Equation 4 and 5 to derive the linear relation of I_t with BG and BI. The analysis results in the same equation for modes $\{m_2, m_5, m_7, m_{10}\}$: $I_t = 0.5B_g + 44.75$

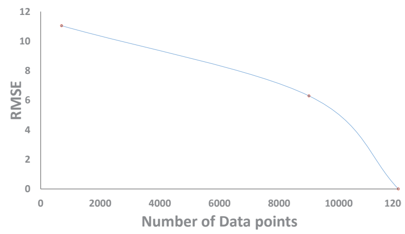

Fig. 8. Variation of the RMSE error with the length of trace for AP.

Since the modes have the same linear equation relating controller output to the inputs, HyMn considers that $m_2 = m_5 = m_7 = m_{10}$. Hence, the total mode set is $M = \{m_1, m_2, m_3\}$. HyMn then considers all the mode transition times and develops the jump conditions. From the input output trace we see that $J_{m_1 \to m_2} = \{\{B_g; B_i; I_t\}, \{B_g; B_i; 0.5B_g + 44.75\}\}$, $J_{m_2 \to m_3} = \{\{B_g;_i I; I_t\}, \{B_g; B_i; 5\}\}$, $J_{m_3 \to m_1} = \{\{B_g; B_i; I_t\}, \{B_g; B_i; 50\}\}$, $J_{m_3 \to m_2} = \{\{B_g; B_i; I_t\}, \{B_g; B_i; 0.5B_g + 44.75\}\}$, and $J_{m_2 \to m_1} = \{\{B_g; B_i; I_t\}, \{B_g; B_i; 50\}\}$. The next step in HyMn is to find the flow equations in every segment using the traces of I_t, B_g, and B_i (Figure 5).

The linearization method used in the paper (III-C) results in a constant bias that depends on the sensed blood glucose, insulin concentration and interstitial concentration values. Hence the bias changes over time. However, the Cramer Rao based estimation only derives coefficients for the difference in the values of the continuous variables. Thus it could not accurately estimate the time varying bias. To circumvent the problem, we add the bias to the estimated constant obtained using Cramer Rao bound each time instant. Based on Equation 4 and 5 HyMn derived the following set of equations:
$\Delta\dot{X} = 45.84\Delta X + 6.89^{-6}\Delta I + 2.47^{-8} - 0.021X(t) + 0.00001(B_g - 10)$
$\Delta\dot{G} = 80.77\Delta X + 45.49\Delta G + 1.21^{-5} - X(t).B_g + 0.031(Gb - B_g)$
$\Delta\dot{I} = 45.59\Delta I + 3.95^{-8} - 0.3B_i + 0.0033hB_g - 0.0033hG0$.

For every segment we obtained the same equation resulting in the conclusion that m_1, m_2, m_3 are unique modes and are not composite. The next step is to determine the guards. Let us consider the transition from m_2 to m_3. The guard observation matrix can be obtained from the traces in Figure 5 as in Equation 6.

$$Go_{m_2 \to m_3} = \begin{Bmatrix} 118.07 & 118.07 & 118.07 & 113.14 \\ 24.8 & 24.8 & 24.8 & 23.8 \end{Bmatrix}. \quad (6)$$

$$Go_{m_1 \to m_2} = \begin{Bmatrix} 177.16 & 177.16 & 171.15.07 \\ 22.76 & 22.76 & 21.4 \end{Bmatrix}. \quad (7)$$

The matrix $Go_{m_2 \to m_3}$ is a full rank matrix. Hence we only consider the row that is constant. However, there is no such row. Hence we have four different expressions for the guard corresponding to each observation in Equation 6. HyMn then considers the derivative of B_g and B_i at the transition point and uses the rules discussed in Section IV. For all the transition points, from Figure 5, we see that $\dot{B_g} < 0$ indicating that the guard for transiting from m_2 to m_3 is $B_g \leq 118.07$. However, $\dot{B_i}$ had both positive and negative values resulting in a contradiction. Thus, the guard expression that uses BI is eliminated from the guard expression. The same operation results in the following guards: $G_{m_1 \to m_2} : B_g < 177.1$, $G_{m_3 \to m_2} : B_g > 118.07$, and $G_{m2 \to m1} : B_g > 177.1$. The invariant set computation was trivial since it only required partitioning of $\mathcal{R}$ using the rectangular guards. The inferred AP hybrid system is almost similar to the given hybrid system discussed in Section III-C. We also used the inferred and the given hybrid system in reachability analysis using the SpaceEx tool. Figure 9 shows reach sets for both hybrid models starting from the same initial set.

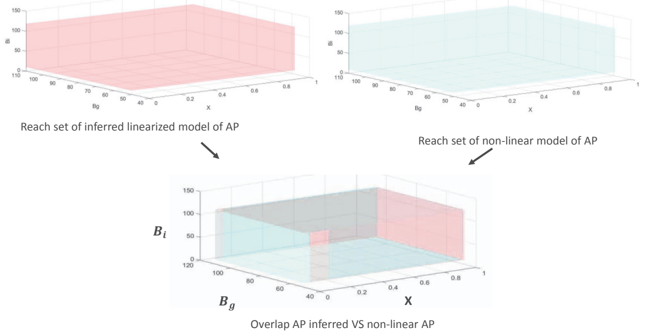

Fig. 9. Reach set of the non-linear model vs inferred linear model of AP.

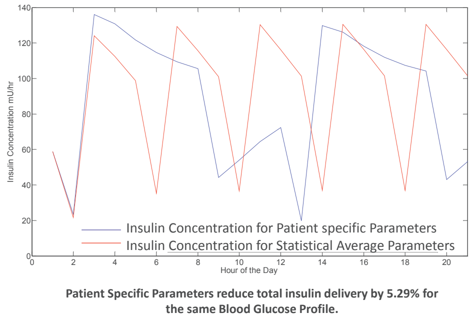

Fig. 10. Comparison of insulin delivery using patient inferred parameters from HyMn vs using statistical average parameters.

Verification and HA refinement Initially, HyMn takes as input a total of 9000 data points of traces of the AP system with a sampling period of 0.005. We generate input/output traces by simulating the inferred HA and compare them to the input/output traces collected for the verification and refinement step. The root mean square error between the two set of traces is approximately equal to 6.3, which is considerably high. Therefore, we increment the number of collected traces by 1000 data points at each iteration and run the HyMn algorithm repeatedly until we achieve a RMSE ≈ 0 as shown in figure 8. The HyMn technique successfully inferred the HA of the AP after 4 iterations (starting from 9000 data points as an initial input set) and a computation time of approximately 0.3 seconds.

Benefits of using HyMn inferred patient specific parameters: We executed the AP system with two parameter configurations: a) taking statistical average parameters obtained from a large pool of T1D subjects as discussed in [13], and b) obtaining the patient specific parameters for a given subject using the HyMN approach. We kept the blood glucose profile the same for both the configurations. Figure 10 shows the plot of the insulin concentration over time for both the configurations. The results show that using patient specific parameters in this scenario reduces total insulin delivery by 5.29%. This is a significant result because the aim of any controller is to achieve normal glucose levels with minimal insulin infusion.

VI. Conclusions and Future Work

The HyMn algorithm presented in this paper can extract linear hybrid systems from I/O traces of a CPS. The key innovation is that the extracted hybrid system can have control modes where the controller output and non-rectangular guards are linear combination of the continuous state variables, and flow equations are not limited to constant functions but are a set of linear differential equations. An important issue is that the accuracy of inferred hybrid system heavily depends on the length of observation of the input output characteristics of the control system. Hence a question for further research is how long do we need to observe a controller so that it is possible to mine an accurate hybrid system representation of the CPS. As a future work, we plan to test HyMn on multiple CPSes to check if the inferred parameters lead to a better control of the system and we plan to use data collected from our collaboration with Mayo clinic for the execution of non-linear AP [18].

Acknowledgments

This work has been partly funded by CNS grant 1218505, IIS grant 1116385, and NIH grant EB019202. Thanks to Yi Zhang and Paul Jones from CDRH, US Food and Drug Administration (FDA) for introducing the authors to the artificial pancreas model and regulatory issues.

References

[1] Alur, Rajeev, et al. "Hybrid automata: An algorithmic approach to the specification and verification of hybrid systems." Hybrid systems. Springer, Berlin, Heidelberg, 1993. 209-229.
[2] Lyde, Steven, and Matthew Might. "Extracting hybrid automata from control code." NASA Formal Methods Symposium, 2013.
[3] Minopoli, Stefano, and Goran Frehse. "SL2SX translator: From simulink to spaceex models." 19th International Conference on Hybrid Systems: Computation and Control. ACM, 2016.
[4] Summerville, Adam, Joseph Osborn, and Michael Mateas. "Charda: Causal hybrid automata recovery via dynamic analysis." (2017).
[5] Medhat, Ramy, et al. "A framework for mining hybrid automata from input/output traces." Proceedings of the 12th International Conference on Embedded Software. IEEE Press, 2015.
[6] Blackmore, Lars, et al. "Model learning for switching linear systems with autonomous mode transitions." Proceedings of the IEEE Conference on Decision and Control (CDC). 2007.
[7] Balakrishnan, Hamsa, et al. "Inference methods for autonomous stochastic linear hybrid systems." International Workshop on Hybrid Systems: Computation and Control, 2004.
[8] Hinshaw, Ling, et al. "Diurnal pattern of insulin action in type 1 diabetes implications for a closed-loop system." Diabetes 62.7 (2013).
[9] Lara, Carlos, Juan J. Flores, and Felix Calderon. "On the hyperboxhyperplane intersection problem." INFOCOMP (2009).
[10] Clarke, William L., et al. "Closed-loop artificial pancreas using subcutaneous glucose sensing and insulin delivery and a model predictive control algorithm: the Virginia experience." (2009): 1031-1038.
[11] Goran Frehse, Olivier Lebeltel, Stefano Minopoli, and Oded Maler, SpaceEx State Space Explorer, http://spaceex.imag.fr/
[12] Minopoli, Stefano, and Goran Frehse. "Non-convex invariants and urgency conditions on linear hybrid automata." International Conference on Formal Modeling and Analysis of Timed Systems 2014.
[13] Man, Chiara Dalla, et al. "The UVA/PADOVA type 1 diabetes simulator: new features." Journal of diabetes science and technology 2014.
[14] Katsios, Christos, and Dimitrios H. Roukos. "Individual genomes and personalized medicine: life diversity and complexity." Personalized Medicine 7, no. 4 (2010): 347-350.
[15] Milanese, Mario, and Gustavo Belforte. "Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors: Linear families of models and estimators." IEEE Transactions on automatic control 27, no. 2 (1982): 408-414.
[16] Nabar, Sidharth, et al. "GeM-REM: Generative model-driven resource efficient ecg monitoring in body sensor networks." Body Sensor Networks (BSN), 2011 International Conference on. IEEE, 2011.
[17] Andersen, Kim E., and Malene Hjbjerre. "A Bayesian approach to Bergmans minimal model." Insulin 50.100 (2002): 200.
[18] Das, Subhasish, et al. "Model Guided Deep Learning Approach Towards Prediction of Physical System Behavior." Machine Learning and Applications (ICMLA), 16th IEEE International Conference 2017.
[19] Buzhinsky, I. P., et al. "Modification of the method of generation of control finite-state machines with continuous actions based on training examples." Journal of Computer and Systems Sciences International(2015).