Apprenticeship Learning for a Predictive State Representation of Anesthesia

Pierre Humbert, *Student Member, IEEE,* Clément Dubost,
Julien Audiffren, and Laurent Oudre

***Abstract*—Objective: In this paper, we present an original decision support algorithm to assist the anesthesiologists delivery of drugs to maintain the optimal Depth of Anesthesia (DoA). Methods: Derived from a Transform Predictive State Representation algorithm (TPSR), our model learned by observing anesthesiologists in practice. This framework, known as apprenticeship learning, is particularly useful in the medical field as it is not based on an exploratory process – a prohibitive behavior in healthcare. The model only relied on the four commonly monitored variables: Heart Rate (HR), the Mean Blood Pressure (MBP), the Respiratory Rate (RR) and the concentration of anesthetic drug (AAFi). Results: Thirty-one patients have been included. The performances of the model is analyzed with metrics derived from the Hamming distance and cross entropy. They demonstrated that low rank dynamical system had the best performances on both predictions and simulations. Then, a confrontation of our agent to a panel of six real anesthesiologists demonstrated that 95.7 % of the actions were valid. Conclusion: These results strongly support the hypothesis that TPSR based models convincingly embed the behavior of anesthesiologists including only four variables that are commonly assessed to predict the DoA. Significance: The proposed novel approach could be of great help for clinicians by improving the fine tuning of the DoA. Furthermore, the possibility to predict the evolutions of the variables would help preventing side effects such as low blood pressure. A tool that could autonomously help the anesthesiologist would thus improve safety-level in the surgical room.**

***Index Terms*—predictive state representation, apprenticeship learning, depth of anesthesia, closed-loop control**

I. Introduction

In the early 2010's, the 4th National Audit Project (NAP4) estimated that 2.9 million General Anesthesia (GA) were performed annually in the UK [1]. As this practice carries risks (cardiovascular complication [2], cognitive dysfunction [3] and postoperative delirium [4]), a sustained and intense attention of the anesthesiologists is imperative to evaluate the level of consciousness of the patient, also referred to as the Depth of Anesthesia (DoA). However, its precise estimation remains an open problem and a constant monitoring of many physiological variables such as heart rate or blood pressure is needed to prevent complications. Since this large amount of information is intractable for the human brain, modern monitors provide multiple auditory and visual warnings, to inform and alert anesthesiologists when physiological variables begin to deteriorate. Unfortunately, those additional indications, while originally meant to help, tend to cause information overload [5], and often fail to be fully processed. Moreover, due to the global problematic of cost efficiency and human resource limitations, it has became common for anesthesiologists to manage two surgical rooms at the same time [6]. In this context, the development of autonomous agents[1] which assist the anesthesiologists managing the delivery of drugs during a GA has become crucial to ease the decision making process, reduce the daily workload and personalize the anesthetic administration, all of this allowing a potentially significant improvement in care.

Several methods have been introduced to fully automate a particular task using closed-loop control models. These methods are used in many fields and cover a wide range of applications [7], [8], [9], [10]. The automation of the delivery of drugs in anesthesia is one of them [11], [12], [13], [14]. Conventional control techniques have been proposed, such as proportional integral-derivative control [15]. However, these methods perform poorly when applied to processes with variable time delays, nonlinearities, and non-negligible process noise [16]. More advanced techniques commonly associated with intelligent systems were studied, including bayesian filtering [17], fuzzy control [18], and reinforcement learning algorithms as markov decision processes [19], [20]. The latter are receiving significant interest in the medical community [21], [22] as they provide efficient models and strong training patterns for autonomous agent that are mathematically sound and have already proven their usefulness in other areas (e.g. robotic programming [23], [24]). However, the definition of a proper and accurate reward function – a mandatory part of reinforcement learning methods – is nearly intractable for complex problems [25]. Moreover, while the free exploration of the policies space is a key part of the learning process in reinforcement learning algorithms, this is a prohibitive behavior in healthcare.

The use of apprenticeship learning (also called learning by watching, imitation learning, learning from demonstration) [26] permits to overcome these drawbacks as the learning process in this framework only need observations of experts without the need for exploration. Moreover, models derived from Predictive State Representations (PSRs) [27], such as Transformed PSRs (TPSRs) [28], rely entirely on observable quantities – an especially desirable property when the underlying latent state (in this case, *consciousness*) is complex and poorly understandood. Based on spectral learning algorithms, TPSR increases the compactness of the space of relevant states. From a mathematical perspective, many theoretical results demonstrate the rich expressiveness of these models. For instance, [27] – influenced by [29] – showed that PSRs are as

[1]In the paper we define agents as Apprenticeship Learning based models.

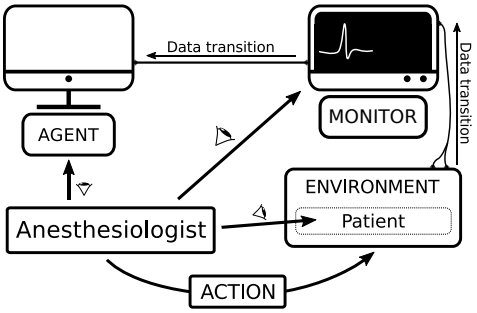

Figure 1: Diagram of the agent and the specific environment. The environment (i.e. the patient) provides observable data (i.e. physiological variables). The monitor records this data and transmits it to the agent. The anesthesiologist chooses an action based on the action suggested by the agent, the values given by the monitor and the behavior of the patient.

flexible and powerful as partially observable markov decision process while providing much more compact representations.

In this study, we introduce a novel decision support tool that predicts in real-time whether anesthesiologists should *reduce the drug dose, do nothing or increase the drug dose* given previous sequences of actions and observations (see Fig. 1 for an illustration). To this end, we combine Apprenticeship Learning principles and TPSR model to solve major problems of control techniques. The resulting approach presents significant advantages, including the fact that the model learns *"how anesthesiologists do"*, instead of trying to learn a complex model of conciousness and deducing *"how anesthesia should work"*. Another major contribution is that our model only relies on a high-resolution recording of the Heart Rate (HR), the Mean Blood Pressure (MBP), the Respiratory Rate (RR) and the concentration of anesthetic drug (AAFi). These four variables are constantly influenced by the drug and are mandatory monitored, making the resulting model suitable for daily use. We also introduce a simple algorithm to homogenize the acquired physiological data and decrease the intra-patient variability. Indeed, the patient's age and gender, as well as disease and surgical intervention are known to affect response to anesthetics [30]. Finally, models were evaluated 1) quantitatively with metrics derived from the Hamming distance and cross entropy 2) with a confrontation to six real anesthesiologists on three cases. This confrontation provides additional metrics to fully evaluate our model and is a mandatory prerequisite for medical application.

This paper is organized as follows. We recall the PSR model and its learning process in Section II. Then, we introduce our main contribution, the construction of a TPSR-based autonomous agent to assist the anesthesiologists managing the delivery of drugs during a GA. (Section III). We also define and discuss our methodology and preprocessing choices. In Section III-C we assess the performance of the model with respect to multiple different metrics (Section III-D) and with three evaluations done by a panel of experts in anesthesiology (Section III-E). Finally, the performances, advantages and drawbacks of our approach are discussed in the last section (Section V).

II. Predictive State Representation

From the angle embraced in this paper we consider a GA as a discrete-time dynamical system where at each time step, the environment (i.e. the patient) generates observable data (i.e. physiological variables) from a set $\mathcal{O}$. Recorded by a medical device, these data are transmitted to the agent which takes an action from a set of possible actions $\mathcal{A} = \{0, 1, 2\} = \{$*reduce the drug dose, do nothing, increase the drug dose*$\}$. Finally, the environment moves to an (unknown) hidden state and produces new observations.

In the present work, we used PSR based models to learn this system. The algorithm of PSR was first introduced by [27]. The authors showed the advantages of this model over Markovian approaches and discussed the improvement brought by possible non-linear models. Following this idea, [31], [32] have focused on improving the learning process of the PSR models. The algorithm used in this paper, called Transformed Predictive State Representations (TPSR), was introduced in [28], where the authors presented the multiple advantages over PSRs, namely removing the problems of local minima in the associated minimization problem and producing a more compact representation. This mathematical model is described below; we refer to [27]–[29], [33] for an in depth presentation.

A. Background on PSR and TPSR

A linear PSR can be seen as a complete description of a dynamical system. Formally, it consists of two infinite countable sets $\overline{\mathcal{H}}$ and $\overline{\mathcal{T}}$ and a system-dynamics matrix $\mathcal{D}$ defined as follows:

- The elements of $\overline{\mathcal{H}}$ (resp. $\overline{\mathcal{T}}$), called *histories* (resp. *tests*) and referring to the past (resp. the future), are defined by
$$\overline{\mathcal{H}} := \left\{h \in (\mathcal{A} \times \mathcal{O})^k \mid k \in \mathbb{N}\right\},$$
$$\overline{\mathcal{T}} := \left\{\tau \in (\mathcal{A} \times \mathcal{O})^\ell \mid \ell \in \mathbb{N}_*\right\}.$$
In other words, they consist in an ordered *sequences* of action-observations pairs $(a, \boldsymbol{o}) \in \mathcal{A} \times \mathcal{O}$, denoted by $h = a_1\boldsymbol{o}_1 a_2\boldsymbol{o}_2 \cdots a_k\boldsymbol{o}_k$ (resp. $\tau = a^1\boldsymbol{o}^1 a^2\boldsymbol{o}^2 \cdots a^\ell\boldsymbol{o}^\ell$).
- The *system-dynamics matrix* $\mathcal{D}$, containing an infinite number of columns and rows, has its elements equal to
$$\mathcal{D}(\tau_i, h_j) = \mathcal{D}_{j,i} := p(\tau_i \mid h_j) = \frac{p(h_j, \tau_i)}{p(h_j)}, \qquad (1)$$
where p denotes the probability associated with the law of the dynamical system for all pairs (τ, h) in $(\overline{\mathcal{T}} \times \overline{\mathcal{H}})$ – in other words, $p(\tau_i \mid h_j)$ denotes the probability of observing τ_i in the future given that h_j was observed in the immediate past. If $p(h_j) = 0$ we set $p(\tau_i \mid h_j) = 0$. The rank of $\mathcal{D}$ characterizes the complexity of the system and is commonly referred to as its *linear dimension*.

- Any family $\mathcal{Q} := \{q_1, \cdots, q_k\}$, $k \in \mathbb{N}$, of linearly independent columns of $\mathcal{D}$ is called a sufficient set of core tests (core set for short) if $|\mathcal{Q}| = \text{rank}(\mathcal{D})$ ($|\cdot|$ denotes the cardinality of a set). The elements of the core set form a base of the vector space spawned by the columns of $\mathcal{D}$. Therefore,

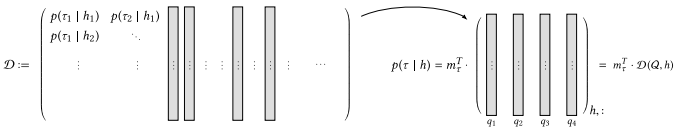

Figure 2: Illustration of the PSR framework. On the left the system-dynamic matrix $\mathcal{D}$. The gray columns involved in the construction of the matrix on the right are core tests.

for any $\tau \in \overline{\mathcal{T}}$, there exists an unique weight vector $\boldsymbol{m_\tau}$ such that for all h

$$\mathcal{D}(\tau, h) = p(\tau \mid h) = \boldsymbol{m}_{\boldsymbol{\tau}}^T p(\mathcal{Q} \mid h). \tag{2}$$

In this equation, $p(\mathcal{Q} \mid h)$ is called the *belief vector* and is defined as

$$\begin{cases} p(\mathcal{Q} \mid h) := \big(p(q_1 \mid h), \dots,\ p(q_{|\mathcal{Q}|} \mid h)\big)^T & \text{if} \quad h \neq \emptyset, \\ p(\mathcal{Q} \mid \emptyset) := \boldsymbol{m}_{\mathbf{0}}^T & \text{otherwise,} \end{cases} \tag{3}$$

with $\boldsymbol{m_0}$ denoting the (unknown) initial condition of the system and $\emptyset$ being the empty history. Similarly, we define $\mathcal{D}(\mathcal{Q})$ as the submatrix of $\mathcal{D}$ that contains the columns relative to the core set i.e. $[\mathcal{D}(\mathcal{Q}, h)^T]_i = [p(\mathcal{Q} \mid h)^T]_i = p(q_i \mid h)$ (see Fig. 2).

***Discovery problem**:* Finding a core set is called the *discovery problem*. This is important as for any such $\mathcal{Q}$, the knowledge of $\mathcal{D}(\mathcal{Q})$ – as well as the initial distribution $\boldsymbol{m_0}$– is enough to fully describe the dynamical system [34].

There are two main approaches to solve this problem and learn PSRs [35]. The first one is a discovery-based technique (e.g. [36], [37], [38]) leading to an explicit knowledge of $\mathcal{Q}$. The second one is a subspace-based technique which is used in this paper and referred to as Transformed PSRs (TPSRs). The latter uses spectral methods to find a subspace isomorph to the vector space generated by $\mathcal{Q}$ instead of determining $\mathcal{Q}$ exactly. To use TPSR model, we applied the spectral algorithm introduced in [33] which learns several matrices (namely B_{ao}, $\boldsymbol{b_\infty}$ and $\boldsymbol{b_*}$, defined below) from sequences of action-observation pairs. This algorithm provides compact and accurate models and permits to predict the most likely future sequences of actions and states efficiently.

We now recall the matrices involved in the algorithm of [33]. For $\mathcal{H} \subset \overline{\mathcal{H}}$ and $\mathcal{T} \subset \overline{\mathcal{T}}$, two finite subsets, let define

- $P_{\mathcal{H}} \in \mathbb{R}^{|\mathcal{H}|}$ that contains the probability of every event in $\mathcal{H}$ i.e. $P_{\mathcal{H}}(h_j) = [P_{\mathcal{H}}]_j := p(h_j)$.
- $P_{\mathcal{T},\mathcal{H}} \in \mathbb{R}^{|\mathcal{T}|\times|\mathcal{H}|}$ where entry (i, j) is the joint probability of (h_j, τ_i) i.e. $P_{\mathcal{T},\mathcal{H}}(\tau_i, h_j) = [P_{\mathcal{T},\mathcal{H}}]_{i,j} := p(h_j, \tau_i)$.
- $P_{\mathcal{T},a\boldsymbol{o},\mathcal{H}} \in \mathbb{R}^{|\mathcal{T}|\times|\mathcal{H}|}$ (one matrix for each unique pair $a\boldsymbol{o}$) where entry (i, j) of $P_{\mathcal{T},a\boldsymbol{o},\mathcal{H}}$ is the probability of the history h_j, the next action-observation pair $a\boldsymbol{o}$, and the subsequent test τ_i i.e. $P_{\mathcal{T},a\boldsymbol{o},\mathcal{H}}(\tau_i, h_j) = [P_{\mathcal{T},a\boldsymbol{o},\mathcal{H}}]_{i,j} := p(h_j, a\boldsymbol{o}, \tau_i)$.

Let $k \in \mathbb{N}$ and $a_1o_1 \dots a_ko_k \in (\mathcal{A} \times \mathcal{O})^k$. For any $t \leq k$, let $h_t = a_1o_1 \dots a_to_t$ and $\boldsymbol{b_t} = p(\mathcal{Q}|h_t)$ the associated belief vector. Thus, the belief vector at time $(t+1)$ can be expressed as $\boldsymbol{b_{t+1}} = p(\mathcal{Q} \mid h_t a\boldsymbol{o}_t)$. The equation binding $\boldsymbol{b_t}$ and $\boldsymbol{b_{t+1}}$ is called the *update rule* and is given by

$$\boldsymbol{b_{t+1}} = \frac{B_{ao_t}\boldsymbol{b_t}}{\boldsymbol{b}_{\boldsymbol{\infty}}^T B_{a\boldsymbol{o}_t}\boldsymbol{b_t}}, \tag{4}$$

where

$$\begin{cases} B_{a\boldsymbol{o}_t} = U^T P_{\mathcal{T},a\boldsymbol{o}_t,\mathcal{H}}(UP_{\mathcal{T},\mathcal{H}})^\dagger & \text{is a transition matrix,} \\ \boldsymbol{b}_{\boldsymbol{\infty}}^T = P_{\mathcal{H}}^T(U^T P_{\mathcal{T},\mathcal{H}})^\dagger & \text{is a normalizer } (\forall h, \boldsymbol{b}_{\boldsymbol{\infty}}^T p(\mathcal{Q} \mid h) = 1), \\ \boldsymbol{b_*} = U^T P_{\mathcal{T},\mathcal{H}} \mathbf{1}_{|\mathcal{H}|} & \text{is the initial state.} \end{cases} \tag{5}$$

Here, $\mathbf{1}_{|\mathcal{H}|}$ is the ones-vector of length $|\mathcal{H}|$, $\dagger$ denotes the Moore–Penrose pseudo inverse and U contains the left singular vectors of $P_{\mathcal{T},\mathcal{H}}$.

***Predictions**:* With the previously defined matrices, for any sequence of u (action, observation) pairs ($u \in \mathbb{N}_*$), we have

$$\begin{cases} p(a_{t+1}\boldsymbol{o}_{t+1} \mid h_t) = \boldsymbol{b}_{\boldsymbol{\infty}}^T B_{a\boldsymbol{o}_{t+1}}\boldsymbol{b_t} \quad (\text{for } u = 1), \\ p(a_{t+1}\boldsymbol{o}_{t+1}, \cdots, a_{t+u}\boldsymbol{o}_{t+u} \mid h_t) = \boldsymbol{b}_{\boldsymbol{\infty}}^T B_{a\boldsymbol{o}_{t+u}} \cdots B_{a\boldsymbol{o}_{t+1}}\boldsymbol{b_t}. \end{cases} \tag{6}$$

This equation is the key to provide an estimator of the probability $p(\cdot)$.

For further discussion on those equations, we refer the reader to the work of [33] where theoretical aspects and relation to the matrices of PSRs were discussed. The methodology to predict actions and/or observations in GA is discussed Section III.

B. Methodological choices

In this subsection, we present our strategy to adapt the TPSR to the problem of closed-loop control of anesthesia. Namely, the introduction of new variables to control the maximum length of each sequence and the use of specific algorithms to compute the different matrices.

***Maximal length of a sequence**:* The computation of the matrices $\overline{\mathcal{T}}$ and $\overline{\mathcal{H}}$ is intractable in practice as they are indexed over an infinite set. To circumvent this problem, we introduced $M_{\mathcal{H}} \in \mathbb{N}_*$ (resp. $M_{\mathcal{T}} \in \mathbb{N}_*$) the maximal length of each history (resp. each test) and restricted ourselves to the learning of $\mathcal{H}_{M_{\mathcal{H}}} := \big\{h \in (\mathcal{A} \times \mathcal{O})^k \mid k \in \mathbb{N}_{\leqslant M_{\mathcal{H}}}\big\}$ and $\mathcal{T}_{M_{\mathcal{T}}} := \big\{\tau \in (\mathcal{A} \times \mathcal{O})^\ell \mid \ell \in \mathbb{N}_{\leqslant M_{\mathcal{T}}} \backslash \{0\}\big\}$. With such a restriction we assumed that $\mathcal{H}_{M_{\mathcal{H}}}$ was sufficient i.e. it allowed to solve the discovery problem – this hypothesis was validated by our experimental results (Section III-C). In the following, we referred those two sets by $\mathcal{H}$ and $\mathcal{T}$ to simplify the notation. It is worth noting that

$$|\mathcal{H}| \approx ((n_{th} + 1)^4 |\mathcal{A}|)^{M_{\mathcal{H}}},$$

and that the same can be stated for $\mathcal{T}$. Consequently, both sets grow exponentially with $M_{\mathcal{H}}$ and $M_{\mathcal{T}}$. The two numbers $M_{\mathcal{H}}$ and $M_{\mathcal{T}}$ were considered as parameters of the problem.

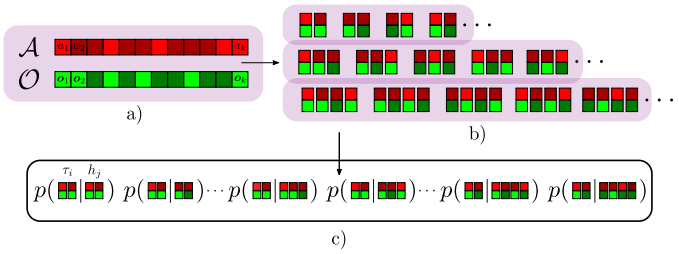

Figure 3: a) Sequence of action in $\mathcal{A}$ and sequence of observations in $\mathcal{O}$. Each shade of red (resp. green) represent a different action (resp. observation). b) Extraction of unique tuples of (actions, observations) of size 2, 3 and 4. c) Example of estimation of the probability of a test of size 2 given all possible histories.

Algorithm 1 Learning problem

Input : M preprocessed trajectories $(\widehat{S_1}, \cdots, \widehat{S_M})$, integers $M_{\mathcal{H}}$, $M_{\mathcal{T}}$, R
Output : $\boldsymbol{b}_{\boldsymbol{\infty}}^T, \boldsymbol{b}_*$ and $\left(B_{a\boldsymbol{o}}\right)_{a\boldsymbol{o}}$
$N \longleftarrow \sum_{m=1}^{M} \sum_{\ell=1}^{M_{\mathcal{H}}} \sum_{k=1}^{|\widehat{S_m}|-\ell} 1$
for $j \in \{1, \cdots, |\mathcal{H}|\}$ **do**
  $\widehat{p}(h_j) \longleftarrow \frac{1}{N} \sum_{m=1}^{M} \sum_{\ell=1}^{M_{\mathcal{H}}} \sum_{k=1}^{|\widehat{S_m}|-\ell} \mathbb{1}_{\{\widehat{S_m}(k:k+\ell)=h_j\}}$
  $[\widehat{P}_{\mathcal{H}}]_j \longleftarrow \widehat{p}(h_j)$
end for
for $(i,j) \in \{1, \cdots, |\mathcal{T}|\} \times \{1, \cdots, |\mathcal{H}|\}$ **do**
  $[\widehat{P}_{\mathcal{T},\mathcal{H}}]_{i,j} \longleftarrow \widehat{p}(h_j, \tau_i)$
  for all $a\boldsymbol{o}$ **do**
    $[\widehat{P}_{\mathcal{T},a\boldsymbol{o},\mathcal{H}}]_{i,j} \longleftarrow \widehat{p}(h_j, a\boldsymbol{o}, \tau_i)$
  end for
end for
$\widehat{U} \longleftarrow$ randomize-SVD$(\widehat{P}_{\mathcal{T},\mathcal{H}}, R)$
$\widehat{\boldsymbol{b}}_{\boldsymbol{\infty}}^T \longleftarrow \widehat{P}_{\mathcal{H}}^T (\widehat{U}^T \widehat{P}_{\mathcal{T},\mathcal{H}})^\dagger$
$\boldsymbol{b}_* \longleftarrow \widehat{U}^T \widehat{P}_{\mathcal{T},\mathcal{H}} \mathbf{1}_{|\mathcal{H}|}$
for all $a\boldsymbol{o}$ **do**
  $\widehat{B}_{a\boldsymbol{o}} \longleftarrow \widehat{U}^T \widehat{P}_{\mathcal{T},a\boldsymbol{o}_t,\mathcal{H}} (\widehat{U} \widehat{P}_{\mathcal{T},\mathcal{H}})^\dagger$
end for

Learning problem: We computed the estimators $\widehat{P}_{\mathcal{H}}, \widehat{P}_{\mathcal{T},\mathcal{H}}$ and $\left(\widehat{P}_{\mathcal{T},a\boldsymbol{o},\mathcal{H}}\right)_{a\boldsymbol{o}}$ of the true TPSR matrices using the entire training set (in other words, all observed combinations were processed). Then, we used a randomized SVD algorithm SVD [39] to compute the Singular Value Decomposition (SVD) of $\widehat{P}_{\mathcal{T},\mathcal{H}}$ and obtain its left singular vectors $\widehat{U}$. Algorithm 1 summarizes *the learning problem* and an illustration is provided in Fig. 3.

The agent predictions were made using a maximum likelihood approach on the distribution given by equation (6).

$$\begin{cases} \arg\max p(a_{t+1}\boldsymbol{o}_{t+1} \mid h_t) \quad (\text{for } u = 1), \\ \arg\max p(a_{t+1}\boldsymbol{o}_{t+1}, \cdots, a_{t+u}\boldsymbol{o}_{t+u} \mid h_t). \end{cases} \tag{7}$$

Ties were broken at random.

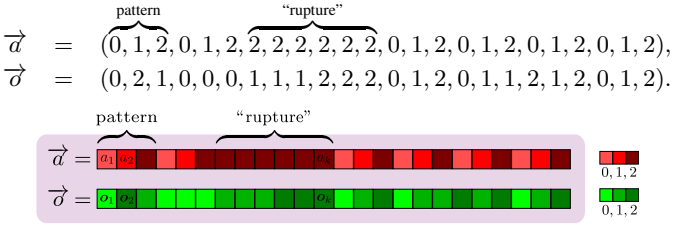

Figure 4: Sequence of actions and of observations constituting the dataset. At each color is associated 0, 1 or 2.

C. Toy example

Here, we give some intuition of the inner working of the TPSR on a simple toy example[2]. The dataset consisted on a sequence of actions $\overrightarrow{a}$ and a sequence of observations $\overrightarrow{o}$ display in Fig. 4.

The sequence of action $\overrightarrow{a}$ presents two interesting features. First, the pattern $(0, 1, 2)$ is repeated almost all the way. Moreover 0 are always followed by 1 i.e $p(1 \mid 0) = 1$. On the contrary, 1 are never followed by 0 i.e $p(0 \mid 1) = 0$. Second, there is a "breakpoint" in the repetition of the pattern with six "2". A visualization of these two sequences is displayed in blue Fig. 5 b).

To emphasize the importance of the observations sequence $\overrightarrow{o}$, we considered two distinct datasets.

- model (A) – Dataset was composed of $\overrightarrow{a}$ and a sequence of observations uniquely composed of 0 which does not bring any information (Fig. 5 (a)),
- model (B) – Dataset was composed of $\overrightarrow{a}$ and $\overrightarrow{o}$ (Fig. 5 (b)).

In both cases, we considered history and test with a maximal size of 2 (i.e. $M_{\mathcal{H}} = M_{\mathcal{T}} = 2$) and computed estimators of the different matrices $\widehat{P}_{\mathcal{H}}, \widehat{P}_{\mathcal{T},\mathcal{H}}$ and $\left(\widehat{P}_{\mathcal{T},a\boldsymbol{o},\mathcal{H}}\right)_{a\boldsymbol{o}}$ (learning part of the algorithm). Then, the core test was found via an SVD (discovery problem). Finally, at any given time t, the agent provided the most probable pair (action, observation) at time $t+1$ using equation (6) and a maximum likelihood approach.

On Fig. 5, we displayed in red the results of the prediction. For the model (A), we observe that the TPSR learned to predict the pattern $(0, 1, 2)$, but cannot anticipate the "breakpoint" sequence of "2" – as no information is brought by the observation in this model. On the other hand, in model (B), we see that the TPSR used the observation information to predict the "breakpoint". Note that since the most present action in the dataset is "2" this is the action predicted at $t = 0$. This underlines the importance of observations for acute prediction of actions.

[2]The source-code is accessible on https://reine.cmla.ens-cachan.fr/p.humbert/TPSR_implementation.

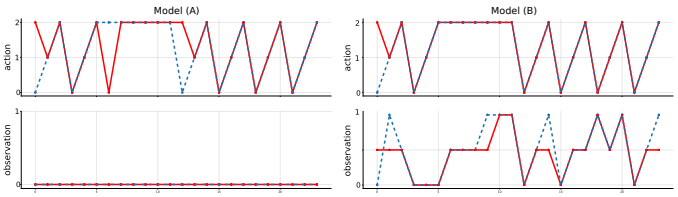

Figure 5: For each figure, on the top are represented the actions and on the bottom the observations. Figures on the left: Resulting curves when considering model (A). Figures on the right: Resulting curves when considering model (B). The blue dot curves are the true sequences. The red curves are predicted by the TPSR model.

Table I

Demographic description of the participants

Sex (F/M)	Age (year)	Weight (kg)	Height (cm)
10/21	60 ± 20	82 ± 14	176 ± 7

The values presented are means and standard deviations.

III. Methods

The goal of our model is to maintain the patient under a deep anesthesia state qualified as "surgical anesthesia". The anesthesia usually requires the use of two types of drugs: morphinomimetic in order to control the pain and hypnotic drugs to ensure that the patient remains asleep. In our model we only focused on the administration of the hypnotic agent (which is made continuously under general anesthesia), in this case the gas sevoflurane. This gas is administered to the patient thanks to the endotracheal tube and rapidly reaches the brain. It is the actions to do on the gas administration that we aimed at modeling, among the three possibilities: decrease, do nothing, or increase the gas concentration.

A. Dataset

Study participants: The study has been approved by the ethics committee of the French society of anesthesiology (SFAR) under the number IRB 00010254-2016-018. Patients were included from March to May 2017 in a single observational center, the Begin military teaching hospital, Saint-Mandé, France. They were included if they were scheduled for an outgoing surgery for inguinal hernia repair under GA, if they gave their consent to the study and if their comorbidity score was low (classified ASA 1 or 2 [40]). They were excluded if they presented complications during the surgery (cardiac arrhythmias, variation of the blood pressure or cardiac frequency more than 20 % compared to the baseline value, or unplanned hospitalization). A summary on the 31 participants is available in Table I.

Anesthesia protocol: The anesthesia protocol was in accordance with the declaration of Helsinki. Four anesthesiologists were included in the study. All the patients were pre-oxygenated via face-mask by 100% oxygen for at least 3 minutes before induction. Sufentanil 0.3 $\mu g/kg$ of ideal-body weight was injected rapidly followed 3 minutes later by $2-4$ mg/kg propofol in combination with ketamine 20 mg. When required for the surgery, patients were paralyzed following induction with a bolus of 0.17 mg/kg of cisatracurium. After tracheal intubation, patients were ventilated with tidal volume of 6 mg/kg ideal-body weight, 5 cmH_2O Positive end-expiratory Pressure (Peep) and a respiratory rate between 10 and 14 to maintain $EtCO_2$ between 30 and 40 $mmHg$. Anesthesia was maintained with sevoflurane MAC age-adjusted (e.g. 1.0), a volatile anesthetic agent [41]. Dose adjustments were made by the anesthesiologist in charge of the patient depending on clinical variables available. Once asleep, patients received a single bolus of local anesthesia when indicated for the surgery.

Table II

Selected variables classified by modules

Variables	Units	abbreviation
Basics Module – 1 Hz		
Heart Rate	/min	HR
Mean arterial blood pressure	mmHg	MBP
Gaz Analysis Module – 1 Hz		
Respiratory Rate	/min	RR
AA Inspiratory Concentration	/100 %	AA FI

For each of the variables, sampling frequency, unit and abbreviation are provided.

Data: During the surgery, patients were continuously monitored with a multiparametric device, the Carescape monitor B850, from General Electrics (GE) HealthcareTM Finland Oy, Helsinki, Finland. Variables were recorded synchronously with a sampling frequency of 1Hz during the anesthesia. We selected 4 standard physiological variables (listed in Table II) providing a dataset of 4 trajectories for each patient. The anesthetics drugs influence all the organs and especially the cardiopulmonary system. Therefore, the four variables that we selected are all constantly influenced by the drug [42]. Moreover, they are mandatory monitored, making the resulting model suitable for daily use since no additional sensors are needed. All these variables are in accordance with the recommendation of the American Society of Anesthesiologists [43], [44]. This choice was also motivated by our aim to provide a decision support tool. Additionally, it should be noticed that the dimension of the system-dynamics matrix $\mathcal{D}$ from the TPSR increases exponentially with the number of variables considered. Therefore, the choice of a restricted number of variables reduce the complexity of the learning problem, acting as an additional regularization.

B. Preprocessing

To homogenize the data, noise and trend of all trajectories were removed via a Simple Moving Average filter (SMA) with a windows size n of $\{5, 15, 30\}$ seconds and no overlap. The random process underlying each physiological variable

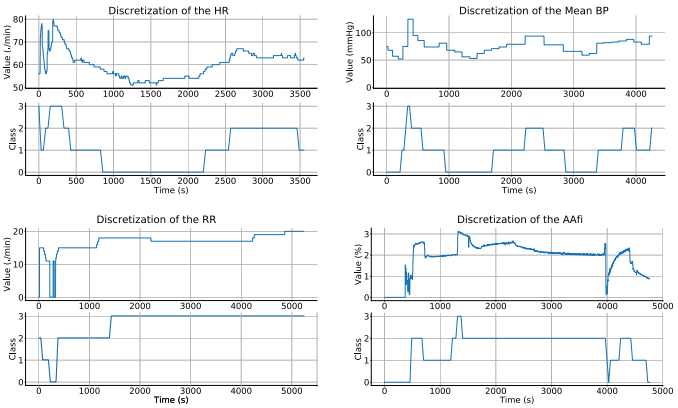

Figure 6: Example of a discretization on the four variables, HR, MBP, RR and AAFi with $n_{th} = 3$. For each variable, on the top the raw signal recorded by the monitor during the GA. On the bottom, its discretization in four classes via CKmean.

was assumed to be locally stationary, as their variations were relatively slow, which justified the use of SMA for small values of n.

Observations: Each observation $\boldsymbol{o} \in \mathcal{O}$ consisted of quadruplets $\big($HR, MBP, RR, AAFi$\big)$ discretized using n_{th} thresholds ($n_{th} \in \mathbb{N}_{>2}$) and taking their values in the set $\{0, 1, \cdots, n_{th}\}$ – where 0 represents low values, and n_{th} high values. The discretization was calculated using Ckmeans, a clustering algorithm based on K-means which has been proven to outperform it in the one-dimensional case [45]. We made an exception for AAFi, which was discretized according to common anesthetic heuristics (i.e. with thresholds between 1% and 3%). The purpose of this calibration procedure was 1) to reduce the inter-patient variability while keeping the intra-patient variability by mapping similar physiological states into the same discretized state– a key part of the problem, as incoherent discretization led to contradictory events, 2) to train a model that automatically adapts to the demographic characteristics of patients (e.g. age, height, weight, BMI). The number of thresholds used in the discretization is a parameter of the model and is evaluated in our experiments. To allow real-time use of the model, preprocessing parameters were estimated during a calibration phase. An example of discretization is displayed Fig. 6.

Actions: The actions were derived from the AAFi variable which represents the amount of drug administrated to a patient. The considered set of possible actions was $\mathcal{A} = \{0, 1, 2\} = \big\{$*Reduce drug dose, Do nothing, Increase drug dose*$\big\}$ formally defined by

- action 0 (Reduce drug dose) – Significant decrease of the AAFi (by at least 10%),
- action 1 (Do nothing) – No significant increase or decrease of the AAFi,
- action 2 (Increase drug dose) – Significant increase of the AAFi (by at least 10%).

More precisely, actions are labeled as follows. Between two regularly spaced sampling points (distant by e.g. 30 s), the action is labeled 2 (resp 0) if the AAFi has increased by at least 10% (resp decreasing by at least 10%). Otherwise, the action is labelled 1.

The data pipeline we used for our model is illustrated in Fig. 7.

Table III

Summary of the adjustable parameters

Names	**Symbols**
Window of the SMA	n
Number of thresholds	n_{th}
Prior on the rank of $\mathcal{D}$	R
Maximal length of a history	$M_{\mathcal{H}}$
Maximal length of a test	$M_{\mathcal{T}}$

Name and symbol of each parameters necessary in the model.

C. Evaluation process

We now present the different experiments made to evaluate the performances of our model. First, we conducted an extensive analysis of the different parameters and their respective influence to identify the best set of parameters, using cross-validation and multiple metrics (see Section III-D). Second, we compared the performance of the resulting model with a Spectral Hidden Markov Model (SHMM) [46], another popular algorithm for learning hidden state based problem. Finally, our model and its associated agent were confronted to a panel of six anesthesiologists assessing three cases.

D. Quantitative Analysis Setup

Prior to any evaluations, the dataset was randomly split into a Training-set (60%), a Validation-set (20%) and a Test-set (20%). We repeated this procedure five times, and average the results over the five random splits.

Classical metrics: In the first experiment, we evaluated the discrepancy between actions predicted by the agent and actions of the experts. The agent predictions were selected using a maximum likelihood approach on the distribution given by equation (6) – ties were broken at random. The metric used in this experiment was the averaged Hamming Distance (HD) between two sequences $(\tau, \widehat{\tau})$ of length μ – a classical metric for PSRs, closely related to the One-Step Prediction Accuracy [47] – (Equation 8).

$$\mathrm{HD}(\tau, \widehat{\tau}) = \frac{1}{\mu} \sum_{i=1}^{\mu} \mathbf{1}_{\tau[i] \neq \widehat{\tau}[i]}. \tag{8}$$

We also computed the distance of actions or observations sequences separately. Let $\tau|_a$ be the sequence of actions provided by the dataset and $\widehat{\tau}|_a$ the one found with the algorithm e.g. $\tau|_a = (a^1 a^2 a^3 a^4 a^5 a^6 a^7 a^8 a^9) = (1, 1, 1, 2, 1, 1, 1, 0, 1)$. We defined the HD of Actions (HD-A) by

$$\text{HD-A}(\tau, \widehat{\tau}) := \mathrm{HD}(\tau|_a, \widehat{\tau}|_a).$$

The HD of Observations (HD-O) is defined similarly. Finally, we used the cross entropy measure in Action 0 –*Reduce drug dose*– or 2 –*Increase drug dose*– and referred it by CE-$\mathrm{A}_{0,2}$.

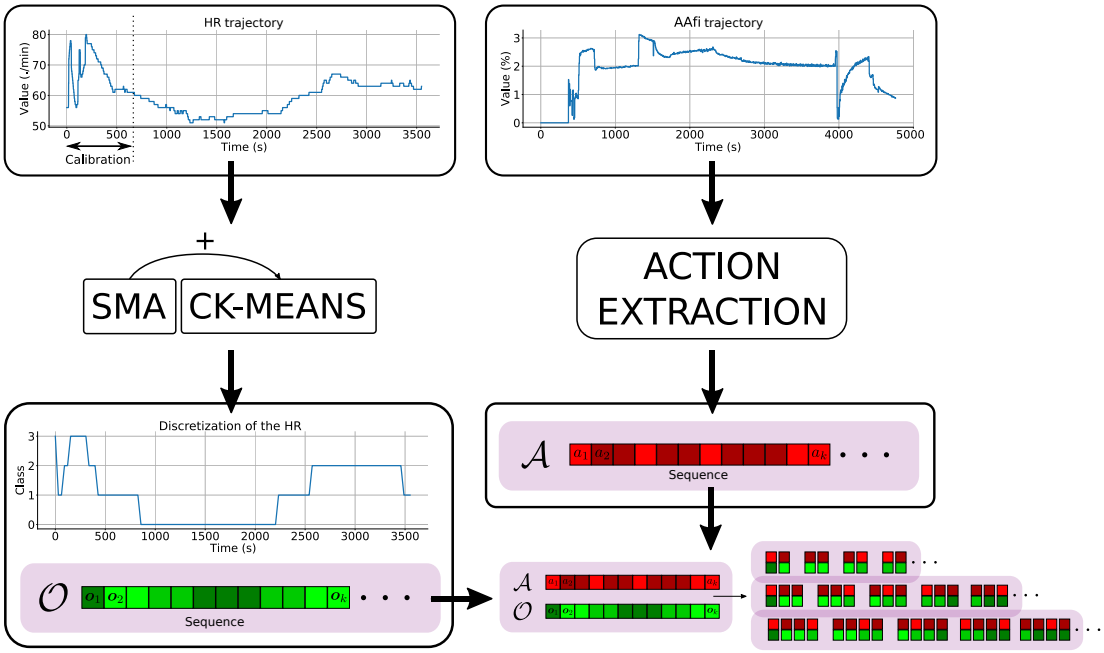

Figure 7: On the left: Preprocessing and discretization procedure for the HR variable. The raw signal (trajectory) is filtered and discretized via the combination of the SMA and CKmean to obtain a sequence of observations. On the right: Extraction of the action from the AAFi variable. The raw signal (trajectory) is filtered and actions are extracted to obtain a sequence of actions. Then, a sequence actions/observations is made in order to fit in the TPSR framework.

This metric is defined as follows. Suppose that at time t the expert takes the action $i \in \{0, 2\}$, then

$$\text{CE-A}_{0,2}(t, i) = -\log\Big(p(\widehat{\tau}|_a(t) = i \in \{0, 2\})\Big). \quad (9)$$

***Metric taking into account a delay**:* Due to anesthetics latency, the action of an anesthesiologist will only be noticed on the recorded variables after a short time delay. Indeed, the time to reach equilibrium point after a modification of the concentration of sevoflorane is approximately 1 minute (considering a supply of fresh gas of 0.4 L/mn) [48]. This phenomenon is not captured by HD-A, HD-0 or CE-$A_{0,2}$. We introduce here a new metric called Sliding Cross Entropy on Action 0 –*Reduce drug dose*– or 2 –*Increase drug dose*– (SCE-$A_{0,2}$) to address this problem. SCE-$A^{(\delta)}_{0,2}$ is defined as follows. Suppose that at time t the expert takes the action $i \in \{0, 2\}$, then let

$$p_{t,i,\delta} = p\Big(\exists s \in [t-\delta, t+\delta] \quad \text{s.t.} \quad \forall s > s' \geq t-\delta, \\ \widehat{\tau}|_a(s') = 1 \text{ and } \widehat{\tau}|_a(s) = i \in \{0, 2\}\Big).$$

In other words, $p_{t,i,\delta}$ represents the probability of the event where the agent takes the correct action, but with a possible time latency of δ – and that the agent only do neutral action (i.e. action 1) before this. Then SCE-$A_{0,2}$ is simply defined as

$$\text{SCE-A}^{(\delta)}_{0,2}(t, i) = -\log(p_{t,i,\delta}). \quad (10)$$

When no delay is considered ($\delta = 0$), SCE-$A_{0,2}$ is CE-$A_{0,2}$. During experiments, the delay was set to 1 minute.

***SHMM comparison**:* In a third step, we compared our TPSR model with the best set of parameters to a tuned SHMM [46] (another family of potent Machine learning algorithms generally used in time series). We used the same metrics as in the previous experiments.

E. Real Expert Evaluation Method

For an exhaustive evaluation of the method, we confronted the best model of the quantitative analysis and its corresponding agent with a panel of six anesthesiologists from the anesthesia-intensive care department of the Begin military teaching hospital. This experiment provides additional metrics to fully evaluate a generative model and is a mandatory prerequisite for medical application. The evaluation was conducted as follows. To begin the confrontation, each anesthesiologist was presented with sequences where only the previous actions and the four discretized selected variables were displayed. Then, the three following experiments were conducted and results collected.

- **Experiment** 1 **–** At each time, and given the real previous sequences, the anesthesiologist chose an action in $\mathcal{A} = \{$*Reduce drug dose, Do nothing, Increase drug dose*$\}$. Those actions were recorded and we measured the disagreement rate between the actions taken by the anesthesiologist and the actions predicted by the agent. This experiment quantifies the capacity of the agent to make the right decisions at the right time.
- **Experiment** 2 **–** At each time, and given the real previous sequences, the agent predicted an action in $\mathcal{A} = \{$*Reduce drug dose, Do nothing, Increase drug dose*$\}$ and the anesthesiologist labeled it as

Table IV

Results of the quantitative analysis

	TPSR			SHMM		
Metrics	$n = 30$	$n = 15$	$n = 5$	$n = 30$	$n = 15$	$n = 5$
HD-A	$\mathbf{0.416}(\pm 0.022)$	$0.438(\pm 0.042)$	$0.458(\pm 0.016)$	$0.592(\pm 0.064)$	$0.542(\pm 0.122)$	$0.599(\pm 0.057)$
CE-A$_{0,2}$	$\mathbf{1.087}(\pm 0.129)$	$1.407(\pm 0.141)$	$1.595(\pm 0.054)$	–	–	–
SCE-A$_{0,2}$	$\mathbf{0.628}(\pm 0.056)$	$0.782(\pm 0.153)$	$1.108(\pm 0.095)$	–	–	–
HD-O						
HR	**0.390**(± 0.032)	$0.445(\pm 0.070)$	$0.479(\pm 0.070)$	$0.759(\pm 0.049)$	$0.741(\pm 0.057)$	$0.820(\pm 0.034)$
Mean BP	**0.342**(± 0.041)	$0.345(\pm 0.026)$	$0.526(\pm 0.082)$	$0.799(\pm 0.052)$	$0.735(\pm 0.106)$	$0.741(\pm 0.081)$
RR	**0.199**(± 0.008)	$0.293(\pm 0.038)$	$0.389(\pm 0.085)$	$0.729(\pm 0.085)$	$0.712(\pm 0.120)$	$0.732(\pm 0.060)$
AAFi	$0.197(\pm 0.051)$	$0.174(\pm 0.034)$	**0.156**(± 0.041)	$0.688(\pm 0.089)$	$0.787(\pm 0.093)$	$0.797(\pm 0.040)$
Mean HD-O	$\mathbf{0.271}(\pm 0.024)$	$0.284(\pm 0.031)$	$0.377(\pm 0.054)$	$0.717(\pm 0.021)$	$0.757(\pm 0.035)$	$0.734(\pm 0.016)$

Results of the quantitative analysis with respect to the n parameter – all the other parameters were optimized with cross validation. For every metrics, the best values were the smallest ones. Metrics reported are the Hamming Distance of Action (HD-A) and Observation (HD-O), the Mean HD-O, the Cross Entropy of Action 0 and 2 (CE-A$_{0,2}$) and the Sliding Cross Entropy of Action 0, 2 (SCE-A$_{0,2}$). On the left, results for our TPSR model, on the right, results for the SHMM. For more details on the metrics, see Subsection III-D

– good: the action is the best choice,
– acceptable: the action is not optimal but still a good choice,
– dangerous: the action may lead to future complications.

We measured the frequency of each label. This experiment provides a qualitative evaluation of the actions of the agent, even if they differ from the real anesthesiologist. Indeed, due to anesthetic latency and the nature of our problem, actions that differ from the anesthesiologist might still be valid choices.

- **Experiment** 3 **–** At each time, and given the previous generated sequences, the anesthesiologist chose an action in $\mathcal{A} = \{$*Reduce drug dose, Do nothing, Increase drug dose*$\}$ and predicted the evolutions of each variables.
  For each variable, we measured the agreement rate between the prediction made by the anesthesiologist to the one made by the agent. This experiment qualitatively evaluate the capacity of our trained model to predict a plausible evolution of the dynamical system given an action.

It should be noted that agreement with human experts in experiment 2 may have been influenced by the lack of a *blind* evaluation. That is why the other two experiments were carefully design to avoid this problem, and their results are in concordance with experiment 2.

IV. Results

A. Quantitative analysis

Results of the quantitative analysis: We evaluated the ability of each set of parameters to predict the right pairs (action, observations) with the metrics defined in Section III-D. For each parameter, the following values were compared: $n \in \{5, 15, 30\}$, $n_{th} \in \{3, 4, 5\}$, $M_{\mathcal{H}} \in \{2, 3, 6\}$, $M_{\mathcal{T}} \in \{2, 3, 6\}$ and R was set to $\{50, 100, 300, 400\}$. It is important to note that n played a very crucial role in our model as it significantly modified the data during the preprocessing. Results of the best set of parameters for each value of n are displayed in Table IV.

The best result was obtained for $(n = 30, n_{th} = 3, M_{\mathcal{H}} = 6, M_{\mathcal{T}} = 3, R = 400)$ (an example of agent sequence is displayed in Fig. 8). This model was used for the confrontation with anesthesiologists. It is interesting to note that the agent tended to predict action and observation with a slight time delay. This aspect was emphasized by the evaluation with the SCE-A$_{0,2}$. Furthermore, the curves of Fig. 8 illustrate that the prediction of physiological variables was accurate and generally differed because of a slight delay.

Contribution of the variables:

- Contribution of AAFi – The AAFi variable is used both as an observation and for the computation of the actions. Hence, the question of whether AAFi influences the model by making the prediction obvious is crucial. To highlight the fact that our model is able to predict the action without simply relying on previous AAFi levels, we conducted additional experiments where AAFi was not included in the model. As a baseline, we also computed the results when no variables were included in the model. These results are presented in Table V. These additional experiments showed that the removal of the AAFi variable in the model only mildly reduced out model performance in term of the **SCE-A**$_{0,2}$ metric: around 0.722 instead of 0.628 for the original model (with AAfi). In comparison, removing all observations (i.e only relying on actions) leads to a **SCE-A**$_{0,2}$ of 0.913. This additional experiment suggests that while AAFi is an important variable for the prediction, it does not trivially contain all the required information. The good results obtained by the agent are therefore not explained by the presence of the AAFi variable in the observations.
- Contribution of RR – The RR is an important variable for monitoring the patient's state. However, in our protocol,

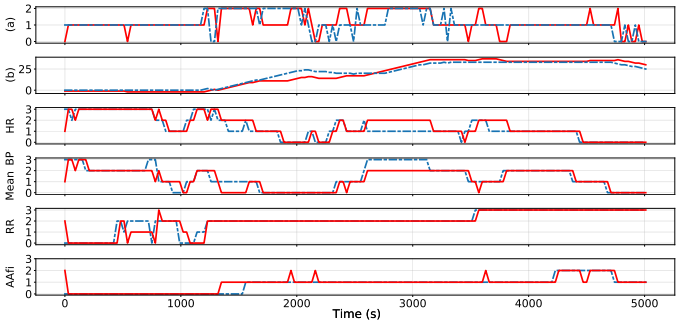

Figure 8: Result of the model with the most promising parameters on one patient. At the top, the two graphs show the results of the prediction of actions. (a) – comparison of the real actions (blue dotted line) with those predicted by our agent (red line); (b) – cumulative sum of the real sequence of actions (blue dotted line) and of the predicted (red line). The next four graphs are the results of the prediction of physiological variables. For each graph, in blue dotted line the real sequences and in red line the predictions.

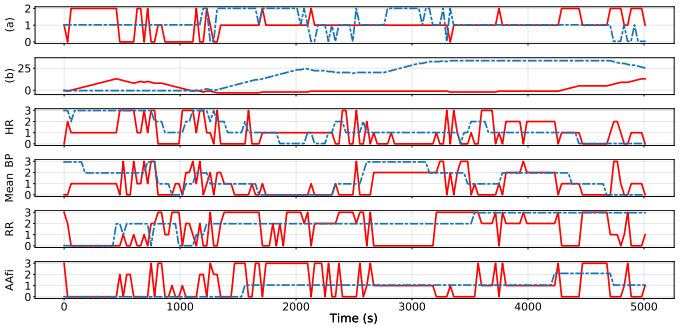

Figure 9: Result of the best SHMM model on the same patient of Fig. 8. At the top, the two graphs show the results of the prediction of actions. (a) – comparison of the real actions (blue dotted line) with those predicted by our agent (red line); (b) – cumulative sum of the real sequence of actions (blue dotted line) and of the predicted (red line). The next four graphs are the results of the prediction of physiological variables. For each graph, in blue dotted line the real sequences and in red line the predictions.

the patient is artificially ventilated, i.e. RR is regulated to maintain $EtCO_2$ at a certain level. To study the importance of this variable, we have computed extra results without the RR variable in the model (see Table IV). It turns out that for $n = 30$, **SCE-A**$_{0,2}$ was equal to 0.635 (against 0.628 when RR is in the model and 0.722 when AAFi is not in the model). This shows that the importance of this variable in our model remains limited. However, we believe that the presence of this variable still makes sense in a clinical setting, especially in critical situations. Indeed, under general anesthesia when the depth of anesthesia is appropriate to perform surgery, patients stop breathing spontaneously. The breathing is thus performed artificially by a ventilator, where the RR is set by the anesthesiologist. In such a condition, the stability of the RR represents the good tolerance of the patient towards the mechanical ventilation and becomes an important indicator of under dosage of anesthesia when the variability increases. In our current experiment, the dataset does not contain any critical situations as every surgery have been unremarkable as regard as the anesthesia. Hence, the RR does not significantly contribute to the model performance at this time. However, we anticipate that this is an important indicator of awakening. Thus, RR can be considered an alert variable, which could be used to introduce hard coded behavior in the model: for instance, when it exceeds a certain threshold, the algorithm could send an alarm and exit the closed-loop system, handing the matter back to the anesthesiologist. This is classical approach to closed-loop system.

SHMM: We evaluated the performances of the SHMM for each of our discretization size and for a maximal rank of 400. The results were reported in Table IV (As the results on classical metrics are unsatisfactory, we do not go deeper with more complex metrics.). We also present in Fig. 9 the prediction of the best SHMM model on the same patient as Fig. 8. Fig. 8 c) is a good representation of the performance of models – the closer the two curves are, the better the model is. It appears that TPSR significantly outperformed SHMM in all the experiments.

B. Real Expert Evaluation

We asked six consultants anesthesiologists to evaluate our best model. Results from the three experiments introduced in Subsection III-E are presented in the Table VI. The results were in accordance with those of the paragraph IV-A.

- Experiment 1 showed an accuracy rate close to the one found in the quantitative evaluation section.
- Experiment 2 showed that 95.7 % of the actions were considered valid by the experts. This high rate of concordance was expected due to the long-latency of the anesthetics drugs.
- Experiment 3 demonstrated that the agent can predict the evolution of the variables in the upcoming minutes secondary to any given action.

V. Discussion and Future Works

Linear dimension: Interestingly, the distribution of the singular values of $\widehat{P}_{\mathcal{T},\mathcal{H}}$ (which is linked to the linear dimension of the TPSR) was found to be similar regardless of the number of included patients. Furthermore, the number of singular values close to zero was significant for several values of horizons, justifying the low rank approximation of the matrix $P_{\mathcal{T},\mathcal{H}}$. Our experiments revealed that models with low rank dynamical system demonstrate strong performances on both predictions and simulations. These results justify the choice of TPSRs over regular PSRs. Moreover, they may have significant consequences in the medical field as the evaluation of DoA through physiological variables could require much less information than presumed i.e. the space of latent states relative to a patient under GA could actually be relatively small.

Table V

Additional results of the quantitative analysis

Metrics	TPSR with AAFi $n = 30$	TPSR without AAFi $n = 30$	TPSR without RR $n = 30$	TPSR with no obs. $n = 30$
HD-A	$\mathbf{0.416}(\pm 0.022)$	$0.439(\pm 0.012)$	$0.419(\pm 0.001)$	$0.456(\pm 0.012)$
CE-A$_{0,2}$	$\mathbf{1.087}(\pm 0.129)$	$1.145(\pm 0.071)$	$1.124(\pm 0.018)$	$1.161(\pm 0.015)$
SCE-A$_{0,2}$	$\mathbf{0.628}(\pm 0.056)$	$0.722(\pm 0.029)$	$0.635(\pm 0.021)$	$0.913(\pm 0.013)$

Additional results of the quantitative analysis. For every metrics, the best values were the smallest ones. Metrics reported are the Hamming Distance of Action (HD-A), the Cross Entropy of Action 0 and 2 (CE-A$_{0,2}$) and the Sliding Cross Entropy of Action 0, 2 (SCE-A$_{0,2}$). For more details on the metrics, see Subsection III-D

Table VI

Evaluation of the model by a panel of anesthesiologists

Exp-1	**Exp-2**		**Exp-3**	
0.371	*Good*	0.632	*HR*	0.914
	Acceptable	0.325	*Mean BP*	0.879
	Dangerous	0.043	*RR*	0.789
			AAFi	0.828

Results of the simulation with the best models ($n = 30, n_{th} = 3, M_{\mathcal{H}} = 6, M_{\mathcal{T}} = 3, R = 400$). Exp-1 : Rate of disagreement between agent and anesthesiologist actions. Exp-2 : rate on actions classify as (good/acceptable/dangerous). Exp-3 : Rate of agreement between agent and anesthesiologists observations. See Section III-E for more details on the three experiments.

Influence of parameters: Throughout our experiments, we observed that different values of $M_{\mathcal{H}}$ and $M_{\mathcal{T}}$ yield similar performances. There might be four possible explanations for this phenomenon.

1) The dynamic system does not have a very long memory. This hypothesis is reasonable, as generally, anesthesiologists do not concentrate on a long period of time, partly because of all the simultaneous tasks required.
2) The population included is homogeneous as we only included patients undergoing inguinal hernia repair under GA. No patient in the population had any significant past medical history nor underwent any side-effect during the GA.
3) Values of the horizon parameters that have a significant impact are large, and thus require significantly larger dataset to observe.
4) The discretization process and the values of the parameter n reduce the long time range dependency of the dynamic system.

Future works might try to evaluate each of these hypotheses.

Additionally, the experiments showed that $n_{th} = 3$ is the best choice for this parameter as it achieve the best trade-off between a) the generalization of the discretization which reduces the inter-patient variability and b) the accuracy of the physiological variable trajectories. Recall that the research of the best set of parameters is more indicative on the behavior of the algorithm than on which parameters need to be actually set for a clinical use. Indeed, the number of patients included is not large enough to properly optimize all the hyperparameters of the models, and current values may change on a larger cohort.

Action and observation prediction: In our experiments, the agent was able to accurately predict the evolution over time of the physiological variables. This performance was expected for discrete AAFi and RR, which exhibit very small variations. However, the small errors on all the observations imply that the agent has learned the complete dynamic system properly. Conversely, predictions were slightly less accurate on actions. This might be explained by the multiplicity of the strategy (policy) exhibited by the experts. Nevertheless, simulations have shown that the actions taken by the agent were validated by the experts. Furthermore, in the first experiments, a significant part of the error was due to small time latency – the agent taking action a few seconds before or after the expert. This behavior was highlighted by the SCE-A$_{0,2}$, a specific metric relevant in our GA scenario. Since those actions would have produced similar results, the good results of the SCE-A$_{0,2}$ demonstrated that the HD metric artificially under estimated the global performance of the agent. The labels of the actions in our model may be seen as relatively inaccurate, since they are restricted to three basics action and that the exact dose of AAFi to be added if necessary is not predicted. Such precision, while theoretically possible by using the continuous extension of the TPSR model [49] would require a significantly larger cohort of patient to be properly calibrated.

SHMM comparison: These experiments highlighted the advantage of our approach over the SHMM. This observation was in line with previous results (see e.g. [34], [50]). One explanation is that, contrary to TPSR, SHMM tends to scale poorly with the complexity of the system to be modeled. However, the implementation of PSRs requires more computational power.

Anesthesiologists feedback: The confrontation with the experts in anesthesiology showed that our agent was coherent and followed an expected policy most of the time. Moreover, all the experts agreed that $n = 30$ appears to be the most realistic value for this parameter. Nevertheless, they also highlighted that there was a latency of the AAFi variable in

some situations, particularly when using low flow of fresh gas.

Clinical relevance: The interest of our agent is double: helping at maintaining a patient at the optimal DoA and predict the occurrence of cardiovascular side-effects (with the idea to avoid them). The workload in the surgical theater imposes that an anesthesiologist is often in charge of two surgery rooms plus the post-anesthesia care unit. A tool that could autonomously help the anesthesiologist would thus improve the safety-level in the surgical room. With such a workload, for a low-risk patient undergoing a low-risk surgery, the anesthesiologist in charge may eventually remember a few characteristics of the patient and usually the pre-induction values of HR and MBP. Once anesthesia level is stabilized and surgery has started, it seems reasonable to consider that the anesthesiologist will leave the patient under the nurse-anesthetist care and will only watch the patient every 10-minute. If we consider that the anesthesiologist will remember the pre-induction, post-induction HR, MBP, RR and AAFi, for one patient we end with: 4 values, every 10-minute meaning 24 values every hour to assess the DoA and status of the patient. As opposed, our agent will take into account all the variables available every second. For a low-risk patient with MBP assesses every 5 minutes this will represent 10.820 values every hour.

Limitations: Despite the strong performances of our model during our experimental evaluations, the PSR approach of the GA setting suffer several drawbacks. First, the model is very dependent on the discretization. Indeed, it is a key component that influences the entire learning process as a too fine or too wide discretization leads to an incorrect estimation of the matrices involved in the model. Second, the lack of a preexisting efficient simulator, as well as a gold-standard for the DoA, greatly limit the possibility to improve the performances above what is observed in the expert trajectories. In its current state, our method is merely a proof of concept for the feasibility of maintaining the anesthesia using carefully trained multimodal algorithm. More experiments and recordings including patients in multiple settings and hospitals will be needed before considering this method as fully valid. It is the authors' belief that the clinical staff will be likely to accept this new approach, as automatic closed-loop anesthesia protocols are already existing, based on the bispectral index [51]. Our method can be seen as an improvement over the exiting protocols, as it takes into account multiple physiological signals as input.

Future works: Beyond the influence of the horizon parameter, we believe that the recording of other relevant physiological variables with additional sensors (e.g. electroencephalogram, muscular sensors, galvanic skin response, ...) could improve the performance of the model. Moreover, a wider range of surgery type in the dataset could bring valuable information on the behavior of the agent. The next step will aim at increasing the population in order to test the generalization of our algorithm in other settings such as in intensive care unit.

VI. Conclusion

In the present paper, we combined apprenticeship learning techniques and model derived from existing PSR, known as TPSR. The resulting agent learned a policy of maintaining the optimal DoA using expert trajectories. The use of machine learning models based on observable variables during GA is pertinent due to the high number of information intractable for the human brain. The performances of the resulting model are promising and convincingly embedded the general behavior of an anesthesiologist. These preliminary results are very encouraging and demonstrate that cardio-pulmonary changes induced by GA can relatively easily be predicted by apprenticeship-learning based algorithm allowing a potentially significant improvement in care.

References

[1] N. Woodall and T. Cook, "National census of airway management techniques used for anaesthesia in the uk: first phase of the fourth national audit project at the royal college of anaesthetists," *British journal of anaesthesia*, vol. 106, no. 2, pp. 266–271, 2010.

[2] M. Golubovic, D. Stanojevic, M. Lazarevic, V. Peric, T. Kostic, M. Djordjevic, S. Zivic, and D. J. Milic, "A risk stratification model for cardiovascular complications during the 3-month period after major elective vascular surgery," *BioMed research international*, vol. 2018, 2018.

[3] Y. Punjasawadwong, W. Chau-in, M. Laopaiboon, S. Punjasawadwong, and P. Pin-on, "Processed electroencephalogram and evoked potential techniques for amelioration of postoperative delirium and cognitive dysfunction following non-cardiac and non-neurosurgical procedures in adults," *Cochrane Database of Systematic Reviews*, no. 5, 2018.

[4] B. A. Fritz, P. L. Kalarickal, H. R. Maybrier, M. R. Muench, D. Dearth, Y. Chen, K. E. Escallier, B. Abdallah, N. Lin, and M. S. Avidan, "Intraoperative electroencephalogram suppression predicts postoperative delirium," *Anesthesia and analgesia*, vol. 122, no. 1, pp. 234–242, 2016.

[5] R. A. Stevenson, J. J. Schlesinger, and M. T. Wallace, "Effects of divided attention and operating room noise on perception of pulse oximeter pitch changesa laboratory study," *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 118, no. 2, pp. 376–381, 2013.

[6] A. F. Merry, J. B. Cooper, O. Soyannwo, I. H. Wilson, and J. H. Eichhorn, "International standards for a safe practice of anesthesia 2010," *Canadian Journal of Anesthesia/Journal canadien d'anesthésie*, vol. 57, no. 11, pp. 1027–1034, 2010.

[7] P. Herrero, T. M. Rawson, A. Philip, L. S. P. Moore, A. H. Holmes, and P. Georgiou, "Closed-loop control for precision antimicrobial delivery: Anin silicoproof-of-concept," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 10, pp. 2231–2236, 2018.

[8] H. M. Romero-Ugalde, V. Le Rolle, J.-L. Bonnet, C. Henry, P. Mabo, G. Carrault, and A. I. Hernández, "Closed-loop vagus nerve stimulation based on state transition models," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 7, pp. 1630–1638, 2018.

[9] Y. Wang, E. Dassau, and I. F. J. Doyle, "Closed-loop control of artificial pancreatic β -cell in type 1 diabetes mellitus using model predictive iterative learning control," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 2, pp. 211–219, Feb 2010.

[10] X. Zhang, J. A. Ashton-Miller, and C. S. Stohler, "A closed-loop system for maintaining constant experimental muscle pain in man," *IEEE Transactions on Biomedical Engineering*, vol. 40, no. 4, pp. 344–352, April 1993.

[11] A. Gentilini, M. Rossoni-Gerosa, C. W. Frei, R. Wymann, M. Morari, A. M. Zbinden, and T. W. Schnider, "Modeling and closed-loop control of hypnosis by means of bispectral index (bis) with isoflurane," *IEEE transactions on biomedical engineering*, vol. 48, no. 8, pp. 874–889, 2001.

[12] C. M. Ionescu, R. De Keyser, B. C. Torrico, T. De Smet, M. M. Struys, and J. E. Normey-Rico, "Robust predictive control strategy applied for propofol dosing using bis as a controlled variable during anesthesia," *IEEE Transactions on biomedical engineering*, vol. 55, no. 9, pp. 2161–2170, 2008.

[13] Y. Sawaguchi, E. Furutani, G. Shirakami, M. Araki, and K. Fukuda, "A model-predictive hypnosis control system under total intravenous anesthesia," *IEEE transactions on biomedical engineering*, vol. 55, no. 3, pp. 874–887, 2008.

[14] G. A. Dumont, "Closed-loop control of anesthesia-a review," *IFAC Proceedings Volumes*, vol. 45, no. 18, pp. 373–378, 2012.

[15] D. A. O'hara, G. J. Derbyshire, F. J. Overdyk, D. K. Bogen, and B. E. Marshall, "Closed-loop infusion of atracurium with four different anesthetic techniques." *Anesthesiology*, vol. 74, no. 2, pp. 258–263, 1991.

[16] K.-S. Tang, K. F. Man, G. Chen, and S. Kwong, "An optimal fuzzy pid controller," *IEEE transactions on industrial electronics*, vol. 48, no. 4, pp. 757–765, 2001.

[17] S. Ching, M. Y. Liberman, J. J. Chemali, M. B. Westover, J. D. Kenny, K. Solt, P. L. Purdon, and E. N. Brown, "Real-time closed-loop control in a rodent model of medically induced coma using burst suppression," *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 119, no. 4, pp. 848–860, 2013.

[18] B. L. Moore, L. D. Pyeatt, and A. G. Doufas, "Fuzzy control for closed-loop, patient-specific hypnosis in intraoperative patients: A simulation study," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 3083–3086.

[19] E. C. Borera, B. L. Moore, and L. D. Pyeatt, "Partially observable markov decision process for closed-loop anesthesia control," in *Proceedings of the 20th European Conference on Artificial Intelligence*. IOS Press, 2012, pp. 949–954.

[20] B. L. Moore, L. D. Pyeatt, V. Kulkarni, P. Panousis, K. Padrez, and A. G. Doufas, "Reinforcement learning for closed-loop propofol anesthesia: a study in human volunteers," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 655–696, 2014.

[21] N. Prasad, L.-F. Cheng, C. Chivers, M. Draugelis, and B. E. Engelhardt, "A reinforcement learning approach to weaning of mechanical ventilation in intensive care units," *arXiv preprint arXiv:1704.06300*, 2017.

[22] B. L. Moore, A. G. Doufas, and L. D. Pyeatt, "Reinforcement learning: a novel method for optimal control of propofol-induced hypnosis," *Anesthesia & Analgesia*, vol. 112, no. 2, pp. 360–367, 2011.

[23] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research*, vol. 4, pp. 237–285, 1996.

[24] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013.

[25] M. Kuderer, S. Gulati, and W. Burgard, "Learning driving styles for autonomous vehicles from demonstration," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*. IEEE, 2015, pp. 2641–2646.

[26] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 1.

[27] M. L. Littman and R. S. Sutton, "Predictive representations of state," in *Advances in neural information processing systems*, 2002, pp. 1555–1561.

[28] M. Rosencrantz, G. Gordon, and S. Thrun, "Learning low dimensional predictive representations," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 88.

[29] R. L. Rivest and R. E. Schapire, "Diversity-based inference of finite automata," *Journal of the ACM (JACM)*, vol. 41, no. 3, pp. 555–589, 1994.

[30] T. W. Schnider, C. F. Minto, P. L. Gambus, C. Andresen, D. B. Goodale, S. L. Shafer, and E. J. Youngs, "The influence of method of administration and covariates on the pharmacokinetics of propofol in adult volunteers," *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 88, no. 5, pp. 1170–1182, 1998.

[31] S. P. Singh, M. L. Littman, N. K. Jong, D. Pardoe, and P. Stone, "Learning predictive state representations," in *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, 2003, pp. 712–719.

[32] M. R. Rudary and S. P. Singh, "A nonlinear predictive state representation," in *Advances in neural information processing systems*, 2004, pp. 855–862.

[33] B. Boots, S. M. Siddiqi, and G. J. Gordon, "Closing the learning-planning loop with predictive state representations," *The International Journal of Robotics Research*, vol. 30, no. 7, pp. 954–966, 2011.

[34] S. Singh, M. R. James, and M. R. Rudary, "Predictive state representations: A new theory for modeling dynamical systems," in *Proceedings of the 20th conference on Uncertainty in artificial intelligence*. AUAI Press, 2004, pp. 512–519.

[35] W. Hamilton, M. M. Fard, and J. Pineau, "Efficient learning and planning with compressed predictive states," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 3395–3439, 2014.

[36] B. Wolfe, M. R. James, and S. Singh, "Learning predictive state representations in dynamical systems without reset," in *Proceedings of the 22nd international conference on Machine learning*. ACM, 2005, pp. 980–987.

[37] M. R. James and S. Singh, "Learning and discovery of predictive state representations in dynamical systems with reset," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 53.

[38] M. R. James, B. Wolfe, and S. P. Singh, "Combining memory and landmarks with predictive state representations." in *IJCAI*, 2005, pp. 734–739.

[39] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM review*, vol. 53, no. 2, pp. 217–288, 2011.

[40] M. Daabiss, "American society of anaesthesiologists physical status classification," *Indian journal of anaesthesia*, vol. 55, no. 2, p. 111, 2011.

[41] S. Patel and K. L. Goa, "Sevoflurane: A review of its pharmacodynamic and pharmacokinetic porperties and its clinical use in general anaesthesia," vol. 51, pp. 658–700, 04 1996.

[42] S. De Hert and A. Moerman, "Sevoflurane," *F1000Research*, vol. 4, no. F1000 Faculty Rev, 2015.

[43] T. A. S. of Anesthesiologists, "Standards for basic anesthetic monitoring."

[44] G. Schneider, D. Jordan, G. Schwarz, P. Bischoff, C. J. Kalkman, H. Kuppe, I. Rundshagen, A. Omerovic, M. Kreuzer, G. Stockmanns *et al.*, "Monitoring depth of anesthesia utilizing a combination of electroencephalographic and standard measures," *The Journal of the American Society of Anesthesiologists*, vol. 120, no. 4, pp. 819–828, 2014.

[45] H. Wang and M. Song, "Ckmeans. 1d. dp: optimal k-means clustering in one dimension by dynamic programming," *The R journal*, vol. 3, no. 2, p. 29, 2011.

[46] D. Hsu, S. M. Kakade, and T. Zhang, "A spectral algorithm for learning hidden markov models," *Journal of Computer and System Sciences*, vol. 78, no. 5, pp. 1460–1480, 2012.

[47] C. Downey, A. Hefny, and G. Gordon, "Practical learning of predictive state representations," *arXiv preprint arXiv:1702.04121*, 2017.

[48] J. H. Philip, B. K. Philip, and S. Leeson, "Paper no: 1306.0 gas man version 4. 1 teaches inhalation kinetics," *British Journal of Anaesthesia*, vol. 108, no. suppl_2, pp. ii215–ii277, 2012.

[49] A. Hefny, W. Sun, S. Srinivasa, and G. J. Gordon, "Predictive state models for prediction and control in partially observable environments."

[50] B. Boots, G. Gordon, and A. Gretton, "Hilbert space embeddings of predictive state representations," *arXiv preprint arXiv:1309.6819*, 2013.

[51] N. Liu and J. Rinehart, "Closed-loop propofol administration: routine care or a research tool? what impact in the future?" 2016.