Model Based Probabilistic Inference for Intensive Care Medicine

Yusuf B. Erol[1], Romi Phadte [1], Harsimran Sammy Sidhu[1], Claire Asselstine [1],
David Phillips [2], Geoffrey Manley [2], and Stuart Russell [1]
[1]Electrical Engineering and Computer Sciences Department, UC Berkeley
[2]Department of Neurological Surgery, UCSF

Modern intensive care units (ICUs) utilize a multitude of instrumentation devices to provide measurements of various important physiological variables and parameters. While data are valuable, understanding the data and acting upon them is what yields the benefits in terms of improved health outcomes. Due to the uncertainty in our knowledge of patient physiology and the partial, noisy/artifactual nature of the observations we adopt a probabilistic, model-based approach. The core of this approach involves calculating a posterior probability distribution over a set of unobserved state variables given a stream of data and a probabilistic model of patient physiology and sensor dynamics. The probability estimate of the state, includes various physiological and pathophysiological variables which provides a diagnosis on which the nurse or the physician can act. The proposed approach is also capable of detecting artifacts, sensor failures, drug maladministration and other various problems in the ICU setting. The overarching goals of the proposed approach are estimating the current health state of the patient, projecting the future health state, and synthesizing possible intervention plans.

***Index Terms*—Model-based probabilistic inference, intracranial pressure, arterial blood pressure, artifacts, state estimation, pathophysiology.**

I. Introduction

Due to advances in electronics, modern ICUs are capable of collecting and archiving ample amounts of clinical data. An ICU patient is continuously monitored by various sensors as asynchronous interventions, tests and additional measurements are carried out. While data are valuable, *understanding* and *acting* upon them is what provides benefits in terms of improved health outcomes and reduced costs.

Current monitoring technology is largely based on data display. The monitor displays signals from various sensors and provides statistics such as short-term averages. Some monitors also deploy automated alarms based on either simple threshold based rules or single-channel analysis. However, false alarm rates of these monitors often exceed 90%, leading to *alarm fatigue* [1], [2], [3].

An ideal system should not only report all pertinent measurements to clinicians, but should also summarize information by *inferring* the states of various *latent* physiological and pathophysiological variables [4], [5], [6]. We propose a model-based probabilistic inference framework that can handle artifact-ridden data, variability in patient physiology and unknown disease states.

Corresponding author: Yusuf B. Erol (email: yusufbugraerol@berkeley.edu).

We describe the patient's physiology and the sensor dynamics as a probabilistic model using a dynamic Bayesian network. We use existing works on human physiology to describe how the state of the patient evolves over time and employ a nontrivial sensor model that is capable of explaining various artifactual readings in the ICU setting. Then, the main task of the ICU monitoring system is estimating the state of the patient accurately given a sequence of observations. Due to the nonlinear, non-Gaussian behavior of our model; exact inference is intractable. Hence, we resort to approximate inference via sequential Monte Carlo (SMC) methods.

In section 2, we describe the model-based probabilistic inference approach and list motivating reasons for such an approach. In section 3, we apply the proposed approach to the problem of estimating the intracranial hemodynamics of traumatic brain injury (TBI) patients. We then explain in detail the physiology and sensor models and also present some inference results. We conclude by discussing the required extensions to the proposed models and the limitations of the current approach. Our work is the continuation of the framework proposed in [7].

II. Model-Based Probabilistic Inference

Due to the uncertainty in our knowledge of patient physiology, and the partial, noisy/artifactual nature of the observations, we adopt a probabilistic, model-based approach. As mentioned earlier, the core task of the ICU monitoring system then becomes estimating the state of a patient given a sequence of observations by computing a posterior probability distribution. Recent trends in machine learning suggest adopting a purely data-driven, model-free approach. However, this may not be the best fit for the intensive care domain due to the following reasons: 1) The data is complex and artifact-ridden. Data may be missing from some sensors for extended periods and there are many artifactual readings some persisting over extended periods. 2) There is a disparity between the number of available signals and the dimensionality of the state space. We only have access to few measurements whereas the latent state-space is high-dimensional. 3) Models of the underlying physiology are available and sensor dynamics can be explained. By being model-free, the available knowledge and information is not exploited.

A. Probabilistic Modeling

Most physiological models are typically expressed as *determinisitic* differential equations. A classic example is the

Guyton model which aims to describe the whole human circulation system in terms of various interconnected subsystems [8]. Deterministic modeling is not a good fit as patient-specific parametrization is unknown and many pathophysiological states manifest stochastic behavior.

We use two different models: The *transition (physiology) model*, $p(\text{states}_t \mid \text{states}_{t-1}, \theta)$, (where states_t stands for the physiological states at time step t and θ for the patient-specific parameters), describes how the state of the patient evolves with respect to time. The variable state_t may include physiological, pathophysiological states, sensor artifacts and failure states, drug administration and so on. Hence, this approach can be extended to handle various possible scenarios of interest in an ICU setting. The *sensor model*, $p(\text{observations}_t \mid \text{states}_t)$, describes how a patient and sensor state are related to the observations.

We represent the model using the dynamic Bayesian network (DBN) framework. DBNs are concise descriptions of stochastic differential equations and they can handle both discrete and continuous variables [9].

B. State Estimation

As noted previously, the core task of ICU monitoring is calculating a posterior probability distribution over states given an observation history, $p(\text{states}_t \mid \text{observations}_{0:t})$. This task is called *state estimation* and is one of the most widely studied problems in statistics, artificial intelligence, and control. Exact state estimation, however, is intractable for most systems. Our model, being nonlinear, non-Gaussian, and hybrid (containing discrete and continuous state variables), is no exception. To compute the required posterior densities, we resort to an approximation scheme known as sequential Monte Carlo (SMC), or *particle filtering*. Particle filtering has successfully been applied to many different state estimation and tracking tasks [10], [11]. Particle filtering can also be applied to the task of prediction by approximating $p(\text{states}_{t+1} \mid \text{observations}_{0:t})$.

C. Parameter Estimation

Every human has a unique set of parameters. Learning these parameter values, $p(\theta \mid \text{observations}_{0:t})$, is a crucial part of the proposed framework; failing to do so will lead to inaccurate state estimates. Vanilla particle filtering fails under the presence of unknown static parameters due to the *sample impoverishment* phenomenon. To counter this, we resort to computationally intensive static parameter estimation algorithms like Resample-Move [12] and Particle MCMC (PMCMC) [13].

D. Synthesizing Intervention Plans

DBN models may include exogenous intervention variables, allowing them to predict the anticipated effects of hypothetical future actions. Constructing treatment plans under partial observability is a computationally hard problem and falls into the domain of partially observable Markov decision problems (POMDP). The planning problem for the intensive care medicine constitutes of hundreds of state variables, thousands of possible actions and a horizon that may span weeks. This problem is well beyond the capability of state-of-the-art POMDP solvers.

III. Application: Intracranial Hemodynamics

We apply our proposed approach on the neurocritical care of traumatic brain injury (TBI) patients. TBI is the developed world's leading cause of mortality and morbidity [14]. For most purposes, available measurements include intracranial pressure (ICP), arterial blood pressure (ABP), blood oxygen level and in some cases cerebral blood flow velocity (CBFV) collected at a resolution of 1-second (1Hz). The problem is depicted in Figure 1. Our goal is calculating the posterior probability on the latent variables given the observed variables (shaded nodes).

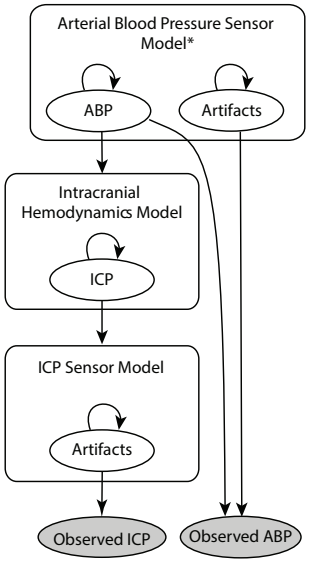


Fig. 1. High-level DBN representation of intracranial hemodynamics

We have three main subsystems: 1) ABP sensor model 2) ICP sensor model 3) Intracranial hemodynamics model. The ABP model is explained in detail in section IV-A. We use a simple random-walk type model to describe the evolution of true diastolic, mean, and systolic ABP values. Our ABP sensor model is also capable of explaining major artifacts like blood draw, zeroing and line clog.

Latent true mean ABP value is an exogenous input to the intracranial hemodynamics model. We are adopting the intracranial dynamics model developed by Ursino and colleagues [15], [16]. This model can be extended to include various pathophysiological states such as autoregulation failure, hematoma, edema, internal bleeding, vasospasm etc. Some results on state and parameter estimation are presented in [7].

Although not shown in Figure 1, transcranial Doppler measurements of CBFV [17] may be available as well. Under these circumstances, noninvasive estimation of ICP via the proposed model-based probabilistic approach using noninvasive measurements like ABP and CBFV is possible. There have been a few recent works in the literature that rely on a model-based approach to achieve the noninvasive reconstruction goal [18], [19].

In section IV-B, we also present a simple ICP sensor model which can describe cerebrospinal fluid drainage artifacts.

IV. Sensor Model

A. ABP Sensor Model

As stated by Aleks et. al., blood pressure informs much of medical thinking and is typically measured continuously in the ICU. The most common way of determining blood pressure in the ICU is to place a transducer on an arterial line, a catheter that is inside one of the patients small arteries; this data is then displayed on a bedside monitor. Due to the high variance of pressure during the cardiac cycle, we use three measurements: systolic (the maximum reached during the cardiac cycle), diastolic (the corresponding minimum), and mean blood pressures[20].

In [20], using the model-based probabilistic inference methodology, over 90% of false ABP alarms (threshold-based) were eliminated while only missing fewer than 1% of the true alarms. The aforementioned work used a minute-resolution DBN. We reproduced their results and then used this model to create a second-resolution blood pressure model. As we predicted, the second-resolution model is able to capture events that the minute-resolution model was unable to capture. For example, most events that lasted less than 10 seconds are not detected by the minute-model, but are picked up by the second-resolution model.

So far, we modeled three major ABP sensor artifacts: zero events, bag events, and line clogging. Zero events occur when an ICU staff calibrates the instrumentation device, effectively zeroing out all the signals. It is seen as a severe and sudden drop in all of the values due to the transducer reading atmospheric pressure. Bag events occur when the transducer reads the intravenous (IV) bag pressure instead of the patient's blood pressure and is seen as a dramatic spike in the observed values. Finally, line clogs occur when there is a kink or an obstruction in the line to the transducer. This zeros the variation in pressure, causing the systolic and diastolic to converge to the mean value.

Figure 2 shows real data from an ICU patient over the span of 100 minutes. From top to bottom, one can see the observed systolic, mean, and diastolic blood pressures. Just by looking at data, it is obvious that handling the artifacts and the missing data is crucial for the success of a real-time decision support system.

1) Model

We developed a one second-resolution model that is capable of generating artifacts and physiological signals. The DBN representation of the developed generative model is depicted in Figure 3.

At the top of the DBN are the true pulse, true mean, and systolic fraction (the ratio of the difference of the diastolic and systolic pressures and the mean). These values give us the true systolic and diastolic values, the ground truth values that our algorithm attempts to infer. We simply used a random-walk type model to describe the evolution of these values. A better cardiovascular dynamics model will significantly improve the performance of the proposed approach.

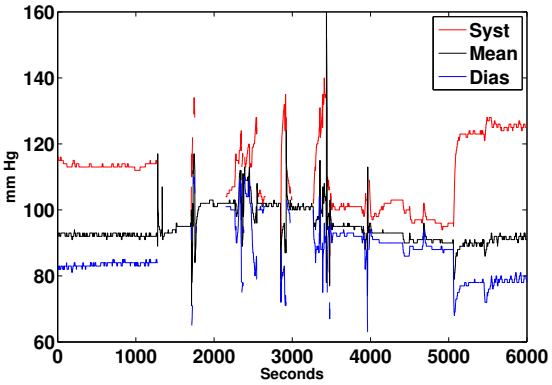


Fig. 2. Sample ABP trace

The values near the bottom are the apparent systolic, diastolic, and mean blood pressure– values that include artifacts and are similar to what the transducer would be measuring. The observed values just below include some additional signal noise and represent what would typically be seen on the monitor.

At any given time, a patient can be in a normal, zero, bag, or clog state. Depending on which state the patient is in, *newPot[S,D,M]* will reflect the new potential that each apparent pressure will converge to. A probabilistic decision is then made on whether patients stays in their current state. Finally, we propagate to the next time step from the current values using our mathematical models. This can be seen in Figure 3 as the arrows point from time step t to $t+1$.

2) Bag Events

One artifact is the bag event, which occurs when the transducer reads the positive pressure from the IV bag elevated at a height. Figure 4(a) shows what the event would look like on an ICU monitor.

The inference results using a particle filter with $N = 20000$ particles is illustrated in Figure 4(b). The three blue lines represent the mean belief of the diastolic, mean, and systolic ABP values, which do not follow the spike of the bag event. The lines at the bottom report the beliefs for each event occurring; in this case, only the bag event registers. The purple shade around the blue lines describe the uncertainty in our estimates.

It is also interesting to see how the posterior density over the bag pressure evolves. The posterior density starts at the prior and gets more and more ambiguous over time. Once the bag event is inferred, the bag value is quickly updated and the particle filter is almost certain of the bag pressure value. This behavior is also theoretically expected.

3) Zero Events

The second artifact incorporated into our DBN is a zeroing event. As mentioned, this occurs when the transducer is exposed to atmospheric pressure which is approximately zero pressure. Figure 5(a) shows an example of such an event.

This event is similar to a bag event in that the sensor no longer reads patient data. A working inference algorithm should detect this event and show that the true diastolic, mean, and systolic values remain in a safe range. Figure 5(b) shows

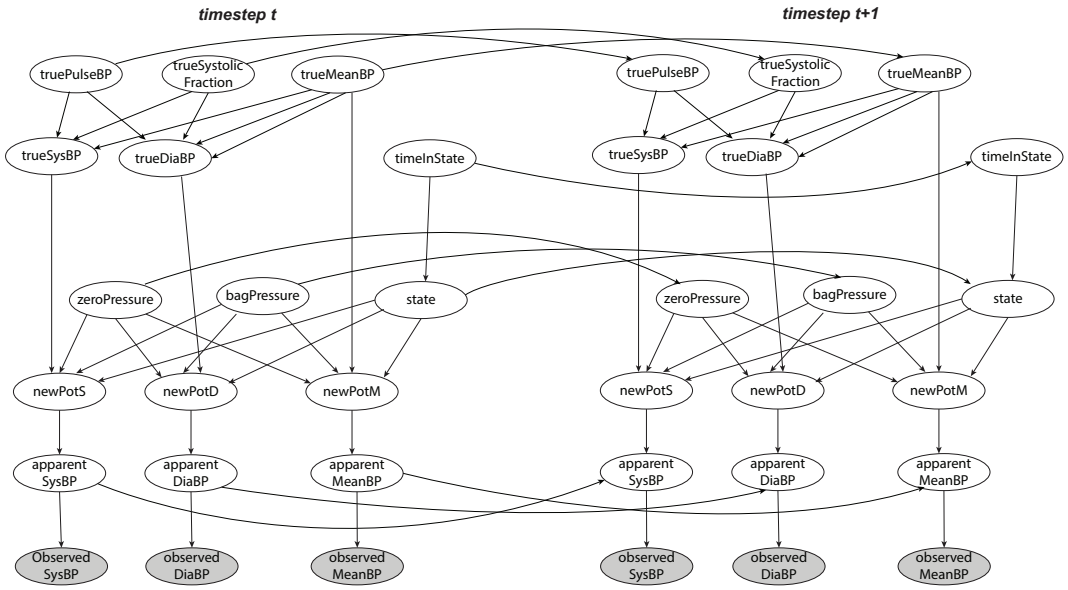
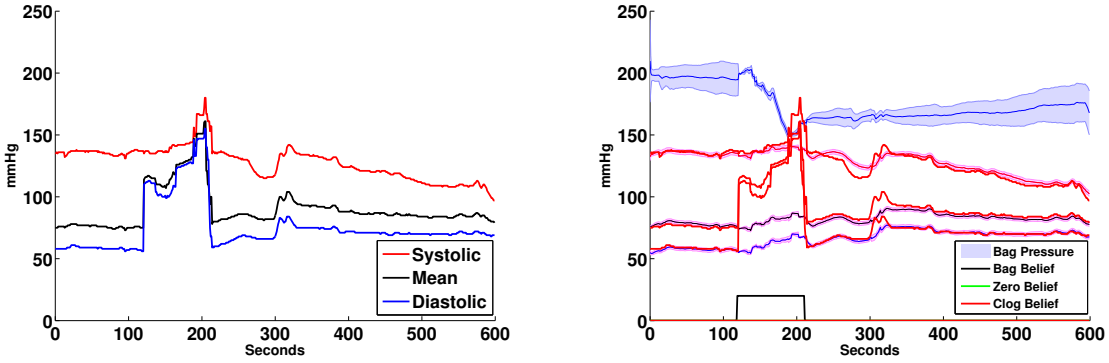
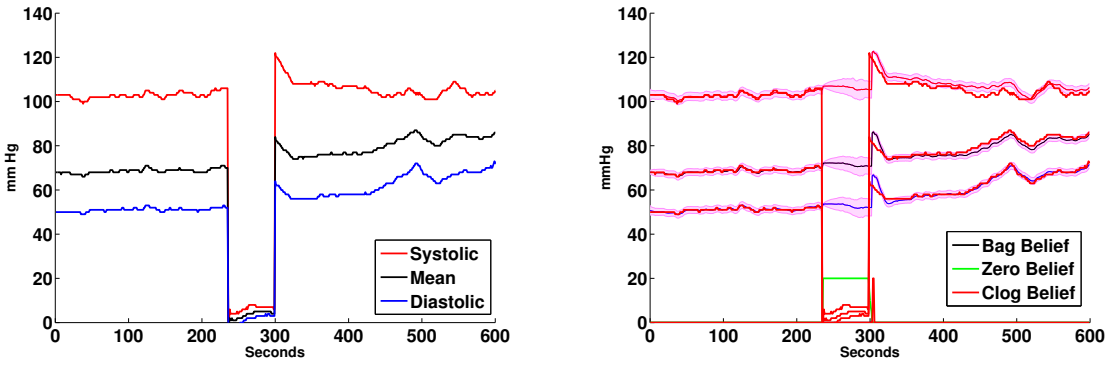


Fig. 3. Dynamic Bayesian network representation of the arterial blood pressure sensor model



(a) Data (b) Inference

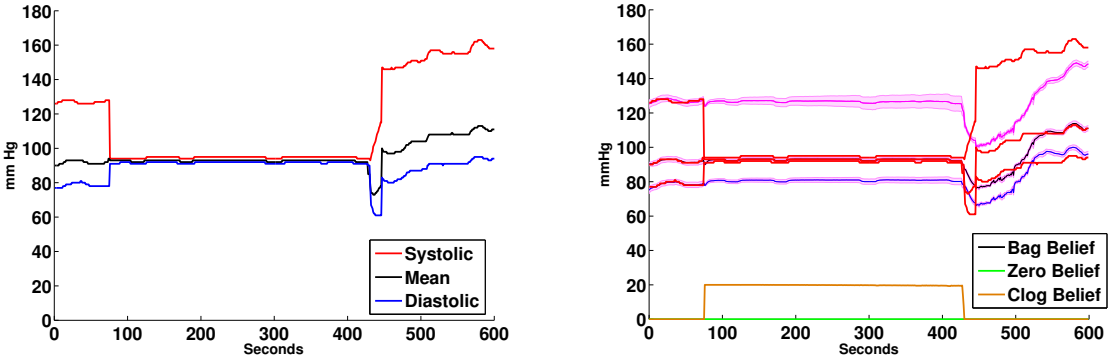
Fig. 4. (a) Sample blood draw trace; (b) Probabilistic belief on systolic, mean, diastolic, and bag latent variables



(a) Data (b) Inference

Fig. 5. (a) Sample zeroing trace; (b) Probabilistic belief on systolic, mean, diastolic, and zeroing latent variables

how the inference performs on this event. The blue lines again represent mean belief values for systolic, diastolic, and mean blood pressures in our inference model. We can see that the zero event is detected and the values reflect that. Note that the purple shade gets wider during the event. This is due to the fact that the inference no longer trusts the current observations and the beliefs on the ground truth variables get more and more uncertain with time.



(a) Data (b) Inference

Fig. 6. (a) Sample line clog trace; (b) Probabilistic belief on systolic, mean, diastolic, and line clog latent variables

4) Clog Events

The last ABP artifact modeled is a line clog, which causes loss of pulse-to-pulse pressure. As seen in Figure 6(a), the observed values for systolic and diastolic blood pressure converge to the mean for the entirety of the event.

The inference results are depicted in Figure 6(b) . The particles are capable of tracking the clogging artifact as well as interpolating the missing systolic and diastolic values.

B. ICP Sensor Model

For the neurocritical care of TBI patients, intracranial pressure (ICP) is the most important measurement for diagnosis and treatment. Current treatment procedure keeps the cerebral perfusion pressure (CPP), which is defined as the gradient between mean arterial pressure (MAP) and ICP, in a safe range in order to keep the patients brain supplied with oxygen. The simplest treatment strategy to keep ICP below a certain level is cerebrospinal fluid (CSF) drainage. CSF is drained when ICP exceeds some set threshold (usually 20 or 25 mmHg). During the drainage, the ICP sensor reads a random pressure value. A sample drainage trace is illustrated in Figure 7.

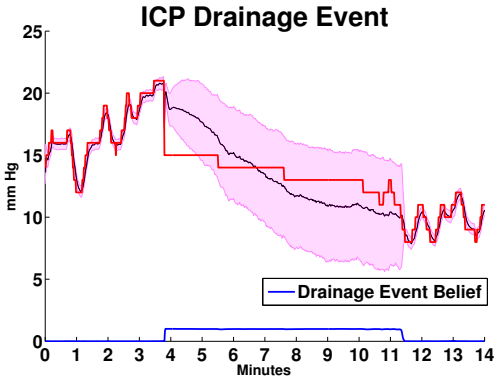


Fig. 7. Probabilistic belief on intracranial pressure during CSF drainage

Unlike blood pressure artifacts, drainage also affects the physiology. During the drainage, the patient's true ICP slowly falls due to the volume loss (Monro-Kellie hypothesis [15]). The particle filter, after noticing the sharp drop from 22 mmHg to 14 mmHg, immediately detects a drainage event since the intracranial physiology model is incapable of explaining such an instantaneous change. It is important to notice that the intracranial pressure belief nicely estimates in a physiologically reasonable way during the drainage event.

V. Conclusion

We described a model-based probabilistic framework capable of representing highly complex physiological phenomena. Using state-of-the-art statistical learning algorithms, we are able to do combined state and parameter estimation. The proposed approach can be used for estimating physiological and pathophysiological states, sensor artifacts and failure states, and drug administration.

The sensor models are still quite inadequate; various other artifacts still need to be added in order to handle real-life clinical data. Artifacts we are considering to add to our DBN in the immediate future are: line flushes, sensor detachment, patient coughing or thrashing, and a nurse rolling the patient. Although our preliminary results seem promising, we still need to validate the artifact cleaning approach on more real data.

The current physiology model we are using is also fairly restrictive as it doesn't describe various pathophysiological phenomena. We need to extend the model provided by Ursino to handle different disease states. Furthermore, we currently do not have a pharmacokinetics model that can explain the dynamics after drug administration. Drugs such as mannitol are frequently used in the clinical care of TBI patients [21]. We would next want to extend our generative model to describe the effects of mannitol administration to infer the onset of drug administration as well as the dosage.

We also still need to validate the performance of the pathophysiological state estimation by comparing inferential results of our algorithms against the physician's diagnosis after a provocative test or intervention. Finally, we should compare our proposed approach for noninvasive estimation against other model-based methods [18] as well as data-driven methods [22], [23].

REFERENCES

[1] B. J. Drew, P. Harris, J. K. Zègre-Hemsey, T. Mammone, D. Schindler, R. Salas-Boni, Y. Bai, A. Tinoco, Q. Ding, and X. Hu, "Insights into the problem of alarm fatigue with physiologic monitor devices: A comprehensive observational study of consecutive intensive care unit patients," *PloS one*, vol. 9, no. 10, p. e110274, 2014.

[2] V. Chopra and L. F. McMahon, "Redesigning hospital alarms for patient safety: Alarmed and potentially dangerous," *JAMA*, vol. 311, no. 12, pp. 1199–1200, 2014.

[3] C. L. Tsien and J. C. Fackler, "Poor prognosis for existing monitors in the intensive care unit," *Critical care medicine*, vol. 25, no. 4, pp. 614–619, 1997.

[4] T. Heldt, B. Long, G. C. Verghese, P. Szolovits, and R. G. Mark, "Integrating data, models, and reasoning in critical care," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*. IEEE, 2006, pp. 350–353.

[5] T. Heldt and G. C. Verghese, "Model-based data integration in clinical environments," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010, pp. 5209–5212.

[6] T. Heldt, G. C. Verghese, and R. G. Mark, "Mathematical modeling of physiological systems," in *Mathematical Modeling and Validation in Physiology*. Springer, 2013, pp. 21–41.

[7] Y. B. Erol, S. J. Russell, A. Sivaganesan, and G. T. Manley, "Combined state and parameter estimation of human intracranial hemodynamics."

[8] A. C. Guyton, T. G. Coleman, and H. J. Granger, "Circulation: overall regulation," *Annual review of physiology*, vol. 34, no. 1, pp. 13–44, 1972.

[9] K. P. Murphy, "Dynamic bayesian networks: representation, inference and learning," Ph.D. dissertation, University of California, Berkeley, 2002.

[10] O. Cappé, S. J. Godsill, and E. Moulines, "An overview of existing methods and recent advances in sequential monte carlo," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899–924, 2007.

[11] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *Signal Processing, IEEE Transactions on*, vol. 50, no. 2, pp. 174–188, 2002.

[12] W. R. Gilks and C. Berzuini, "Following a moving targetmonte carlo inference for dynamic bayesian models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 1, pp. 127–146, 2001.

[13] C. Andrieu, A. Doucet, and R. Holenstein, "Particle markov chain monte carlo methods," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 72, no. 3, pp. 269–342, 2010.

[14] C. Werner and K. Engelhard, "Pathophysiology of traumatic brain injury," *British journal of anaesthesia*, vol. 99, no. 1, pp. 4–9, 2007.

[15] M. Ursino and C. A. Lodi, "A simple mathematical model of the interaction between intracranial pressure and cerebral hemodynamics," *Journal of Applied Physiology*, vol. 82, no. 4, pp. 1256–1269, 1997.

[16] M. Ursino, C. Lodi, S. Rossi, and N. Stocchetti, "Intracranial pressure dynamics in patients with acute brain damage," *Journal of Applied Physiology*, vol. 82, no. 4, pp. 1270–1282, 1997.

[17] L. Steiner and P. Andrews, "Monitoring the injured brain: Icp and cbf," *British journal of anaesthesia*, vol. 97, no. 1, pp. 26–38, 2006.

[18] F. M. Kashif, G. C. Verghese, V. Novak, M. Czosnyka, and T. Heldt, "Model-based noninvasive estimation of intracranial pressure from cerebral blood flow velocity and arterial pressure," *Science translational medicine*, vol. 4, no. 129, pp. 129ra44–129ra44, 2012.

[19] F. M. Kashif, "Modeling and estimation for non-invasive monitoring of intracranial pressure and cerebrovascular autoregulation," Ph.D. dissertation, Massachusetts Institute of Technology, 2011.

[20] N. Aleks, S. J. Russell, M. G. Madden, D. Morabito, K. Staudenmayer, M. Cohen, and G. T. Manley, "Probabilistic detection of short events, with application to critical care monitoring," in *Advances in Neural Information Processing Systems*, 2009, pp. 49–56.

[21] L. Rangel-Castillo, S. Gopinath, and C. S. Robertson, "Management of intracranial hypertension," *Neurologic clinics*, vol. 26, no. 2, pp. 521–541, 2008.

[22] X. Hu, P. Xu, F. Scalzo, P. Vespa, and M. Bergsneider, "Morphological clustering and analysis of continuous intracranial pressure," *Biomedical Engineering, IEEE Transactions on*, vol. 56, no. 3, pp. 696–705, 2009.

[23] S. Kim, R. Hamilton, S. Pineles, M. Bergsneider, and X. Hu, "Noninvasive intracranial hypertension detection utilizing semisupervised learning," *Biomedical Engineering, IEEE Transactions on*, vol. 60, no. 4, pp. 1126–1133, 2013.