

Continuous Venous Oxygen Saturation Estimation via Population-Informed Personalized Gaussian Sum Extended Kalman Filtering

Parham Rezaei¹, Joseph S. Friedberg, Hosam K. Fathy², *Member, IEEE*,
and Jin-Oh Hahn³, *Senior Member, IEEE*

Abstract—Mixed venous oxygen saturation (SvO_2) can play a pivotal role for patient monitoring and treatment in critical care and cardiopulmonary medicine. Unfortunately, its continuous measurement requires the use of invasive pulmonary artery catheters. This letter presents a novel population-informed personalized Gaussian sum extended Kalman filtering (PI-P-GSEKF) approach to continuous SvO_2 estimation from arterial oxygen saturation (SpO_2) measurement. The main challenge in SvO_2 estimation is large inter-individual variability in the cardiopulmonary dynamics, which seriously deteriorates the efficacy of standard EKF. To cope with this challenge, we employ the GSEKF in which individual EKFs are designed using a mathematical model of cardiopulmonary dynamics whose operating points are selected from (i) population-level generative sampling (thus “population-informed”) and (ii) Markov chain Monte Carlo (MCMC) sampling based on a one-time SpO_2 - SvO_2 measurement (thus “personalized”). Using the experimental data collected from 8 hypoxia trials in 4 large animals, we showed the ability of the PI-P-GSEKF to estimate SvO_2 from SpO_2 in comparison with its PI-EKF (EKF with population-level generative sampling as the source of process noise) and PI-GSEKF (GSEKF with population-level generative sampling alone) counterparts (average SvO_2 root-mean-squared error: PI-EKF 4.7%, PI-GSEKF 4.3%, PI-P-GSEKF 3.0%). We also showed that population-level generative sampling and MCMC sampling both had respective roles in improving SvO_2 estimation accuracy. In sum, the PI-P-GSEKF demonstrated its proof-of-principle to enable non-invasive continuous SvO_2 estimation.

Index Terms—Mixed venous oxygen saturation, Gaussian sum filter, Kalman filter, generative sampling, MCMC sampling.

Received 13 September 2024; revised 12 November 2024; accepted 1 December 2024. Date of publication 9 December 2024; date of current version 18 December 2024. This work was supported by the U.S. National Science Foundation (NSF) under Grant OIA2121110 and Grant CNS-2322533. Recommended by Senior Editor L. Menini. (Corresponding author: Parham Rezaei.)

Parham Rezaei, Hosam K. Fathy, and Jin-Oh Hahn are with the Department of Mechanical Engineering, University of Maryland at College Park, College Park, MD 20742 USA (e-mail: prezaei@umd.edu; hfathy@umd.edu; jhahn12@umd.edu).

Joseph S. Friedberg is with the Department of Thoracic Medicine and Surgery, Temple University, Philadelphia, PA 19122 USA (e-mail: joseph.friedberg@tuhs.temple.edu).

Digital Object Identifier 10.1109/LCSYS.2024.3514780

I. INTRODUCTION

IN CRITICAL care and cardiopulmonary medicine, effective patient management frequently requires maintaining the balance between oxygen (O_2) supply and consumption [1], [2]. In this context, mixed venous oxygen saturation (SvO_2) can play a pivotal role for patient monitoring and treatment. A reduction in SvO_2 often signifies the mismatch between O_2 supply and consumption caused by, e.g., a decrease in arterial O_2 content, a decrease in cardiac output (CO), or an increase in O_2 demand [3]. As an example, SvO_2 can offer more immediate and accurate assessment of oxygenation status in circulatory shock than rudimentary vital signs used in clinical practice today [4]. As another example, SvO_2 can reveal post-operative clinical deterioration due to inadequate O_2 supply not apparently indicated by blood pressure (BP) and arterial oxygen saturation (SpO_2) [5]. In this way, SvO_2 can improve the monitoring and treatment of critically ill patients.

The O_2 supply-consumption balance varies dynamically. Hence, continuous SvO_2 measurement has the potential to enable frequent and precise titration of treatment to each patient. However, continuous measurement of SvO_2 requires invasive pulmonary artery catheterization, whose use has declined due to its clinical risks and complications [6], [7], [8], [9]. SvO_2 can also be measured via venous blood sampling [3]. However, it is still invasive. Moreover, it is intermittent, making it inadequate for continuous SvO_2 monitoring.

Given that the difference between inspired oxygen fraction (FIO_2) and SpO_2 may approximately represent O_2 intake, SvO_2 may likely be inferred from these readily available continuous measurements by leveraging the balance between O_2 intake and consumption. An attractive solution may be state estimation based on a mathematical model of cardiopulmonary dynamics. In this context, the main challenge in SvO_2 estimation is large inter-individual variability in the cardiopulmonary dynamics [10]. Such variability acts as uncertainty in state estimation and deteriorates its accuracy. In particular, standard extended Kalman filter built upon nominal cardiopulmonary dynamics may suffer from unacceptably large parametric errors and process noise. One can conceive adaptive state estimation to address this challenge, e.g., by augmenting unknown parameters into the plant dynamics to estimate states and unknown parameters simultaneously [11] or sequentially [12], [13], [14]. However, adaptive state estimation may still result in inadequate accuracy because the augmentation

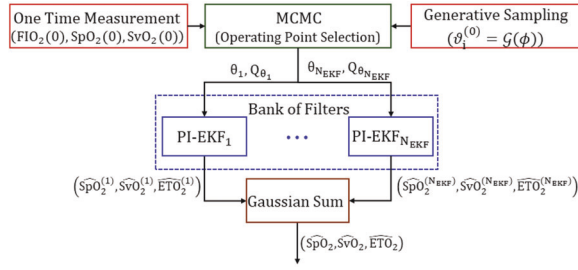


Fig. 1. Population-informed personalized Gaussian sum extended Kalman filter (PI-P-GSEKF). $\vartheta_i^{(0)}$, $i = 1, \dots, N_G$: Generative parameter vectors. θ_i and Q_{θ_i} , $i = 1, \dots, N_{EKF}$: Personalized generative parameter vectors selected as the operating points for the EKFs.

of a large number of parameters as states can degrade the observability of the plant dynamics.

To cope with this challenge, we present a novel population-informed personalized Gaussian sum extended Kalman filtering (PI-P-GSEKF) approach to continuous non-invasive SvO₂ estimation from continuous FIO₂-SpO₂ measurements (Fig. 1). We exploit Gaussian sum filtering (GSF) [15] to achieve good state estimation accuracy while avoiding the worsening of observability caused by state augmentation in adaptive state estimation. We employ a large number of EKFs in the GSF to deal with the nonlinear and uncertain nature of the cardiopulmonary dynamics. A novel aspect of our approach lies in the way the operating points of the individual EKFs are selected. First, we use population-level generative sampling [16] to generate a large number of parameter vector samples, which can replicate diverse plausible cardiopulmonary dynamics when characterizing its mathematical model. Second, we use Markov chain Monte Carlo (MCMC) sampling based on a one-time SpO₂-SvO₂ measurement, which personalizes the generative parameter vector samples toward the cardiopulmonary dynamics pertaining to the SpO₂-SvO₂ measurement. Third, we select a preset number of personalized parameter vector samples as the operating points for designing the EKFs in the GSF. In this way, our approach is population-informed (i.e., via generative sampling) and personalized (i.e., via MCMC sampling) GSEKF (PI-P-GSEKF). We investigated the ability of the PI-P-GSEKF to estimate SvO₂ from FIO₂-SpO₂ using the experimental data collected from 8 hypoxia trials in 4 large animals in our prior work [10].

II. POPULATION-INFORMED PERSONALIZED GAUSSIAN SUM EXTENDED KALMAN FILTER

The PI-P-GSEKF consists of 3 key components: (i) a GSF composed of a bank of EKFs with a broad range of operating points, which are selected by (ii) a population-level generative sampling and (iii) an MCMC sampling driven by a one-time measurement (Fig. 1). Consider a nonlinear system whose dynamics is given by:

$$\begin{aligned} x(k+1) &= f(x(k), u(k), \theta) + w(k) \\ y(k) &= h(x(k)) + v(k) \end{aligned} \quad (1)$$

where $f(\cdot)$ and $h(\cdot)$ are continuously differentiable vector functions, $w(k) \sim N(0, Q(k))$ and $v(k) \sim N(0, R(k))$ are zero-mean Gaussian noises with $Q(k)$ and $R(k)$ as covariance matrices, θ is a vector which contains the parameters characterizing the system dynamics, and $x(k)$ and $u(k)$ are state and input vectors at a time instant k . The PI-P-GSEKF estimates the state $\hat{x}(k)$

using a large number of state estimates $\hat{x}_i(k)$, $i = 1, \dots, N_{EKF}$ furnished by a bank of EKFs:

$$\begin{aligned} \hat{x} &= \sum_{i=1}^{N_{EKF}} \bar{w}_i \hat{x}_i \\ P &= \sum_{i=1}^{N_{EKF}} \bar{w}_i \left[(\hat{x}_i - \hat{x})(\hat{x}_i - \hat{x})^T + P_i \right] \end{aligned} \quad (2)$$

where N_{EKF} is the number of EKFs, $\hat{x}_i(k)$ and $P_i(k)$ are the state estimated by the i -th EKF and its corresponding covariance, $\hat{x}(k)$ and $P(k)$ are the state estimated by the PI-P-GSEKF and its corresponding covariance, $\bar{w}_i(k)$ is the weight pertaining to the i -th EKF. Each EKF estimates its state (i.e., $\hat{x}_i(k)$, $i = 1, \dots, N_{EKF}$) using the standard prediction-correction procedure. Each weight $\bar{w}_i(k)$, $i = 1, \dots, N_{EKF}$ in Eq. (2) is initially set to $\frac{1}{N_{EKF}}$ and subsequently updated at each time instant based on the likelihood of $y(k)$ given $\hat{x}_i^-(k)$:

$$\begin{aligned} w_i(k) &= w_i(k-1)p(y(k)|\hat{x}_i^-(k)) \\ \bar{w}_i(k) &= \frac{w_i(k)}{\sum_{i=1}^{N_{EKF}} w_i(k)} \end{aligned} \quad (3)$$

where $\hat{x}_i^-(k)$ is the state predicted by the i -th EKF at a time instant k . In the PI-P-GSEKF, the operating points of the EKFs (i.e., θ_i , $i = 1, \dots, N_{EKF}$) are selected by a population-level generative sampling and an MCMC sampling. In this context, the generative sampling generates a large number of operating points which can encompass diverse yet plausible cardiopulmonary dynamics, while the MCMC sampling personalizes the generative samples using a one-time measurement. First, the PI-P-GSEKF generates plausible yet wide-ranging random operating point samples from a pre-constructed generative sampler:

$$\vartheta_i^{(0)} \sim \mathcal{G}(\phi) \quad (4)$$

where $\vartheta_i^{(0)}$, $i = 1, \dots, N_G$ are the generative operating point samples while $\mathcal{G}(\phi)$ is the generative sampler in the form of a multi-dimensional probability density function characterized by a set of latent parameters ϕ (which specify the shape of $\mathcal{G}$ [16]). Second, the PI-P-GSEKF generates a sequence of personalized operating point samples from each $\vartheta_i^{(0)}$, $i = 1, \dots, N_G$ using an MCMC sampler (e.g., the Metropolis-Hastings algorithm [17]) so that the operating points included in the MCMC sequence can (i) replicate the available one-time measurement of the system state and (ii) avoid non-sensical state. In this way, random generative operating points are personalized to the system at hand, to the extent where the EKF can be parameterized at the operating points congruent with the one-time measurement. For this purpose, the PI-P-GSEKF uses the following likelihood function in performing the Metropolis-Hastings algorithm:

$$\begin{aligned} L(z(0)|\vartheta_i^{(m)}) &= L_1(z(0)|\vartheta_i^{(m)})L_2(\vartheta_i^{(m)}) \\ L_1(z(0)|\vartheta_i^{(m)}) &= \exp(-\alpha_1 \|z(0) - \hat{z}(0, \vartheta_i^{(m)})\|_2) \\ L_2(\vartheta_i^{(m)}) &= \exp(-\alpha_2 [\|\max(\mathbf{0}, \hat{x}(0, \vartheta_i^{(m)}) - \check{x}(0))\|_1 \\ &\quad + \|\max(\mathbf{0}, \check{x}(0) - \hat{x}(0, \vartheta_i^{(m)}))\|_1]) \end{aligned} \quad (5)$$

where $z(0)$ is the one-time measurement of a subset of the state (i.e., $z(0) \subset x(0)$), $\mathbf{0}$ is a zero vector with the same dimension as x , $\hat{\cdot}$ and $\check{\cdot}$ are element-wise maximum and minimum, and

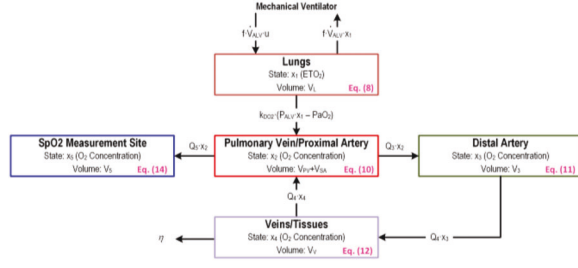


Fig. 2. Mathematical model of oxygenation transport in response to mechanical ventilation.

$m = 1, \dots, N_{MCMC} - 1$. The MCMC sampling yields:

$$\vartheta_i = \left\{ \vartheta_i^{(1)}, \dots, \vartheta_i^{(N_{MCMC})} \right\} \quad (6)$$

where $\vartheta_i^{(j)}$, $j = 1, \dots, N_{MCMC}$ are the personalized operating points in the sequence ϑ_i pertaining to $\vartheta_i^{(0)}$. Third, the PI-P-GSEKF computes the mean and covariance pertaining to all the ϑ_i 's as the candidate operating points and their covariance matrices:

$$\theta_i = \bar{\vartheta}_i, \quad Q_{\theta_i} = \Sigma(\vartheta_i) \quad (7)$$

where $\bar{\cdot}$ denotes mean value and $\Sigma(\cdot)$ denotes covariance. Fourth, the PI-P-GSEKF ranks these N_G population-informed and personalized operating points based on the likelihood in Eq. (5), and adopts the N_{EKF} operating points with the highest likelihood values to construct the PI-P-GSEKF.

In this way, the PI-P-GSEKF can estimate the state even when the ground truth system dynamics is substantially different from its nominal counterpart. This can be achieved by a large number of EKFs whose operating points are diverse (by virtue of generative sampling) yet still close to the ground truth system dynamics (by virtue of MCMC sampling). In this sense, the PI-P-GSEKF may probably be superior to the conventional EKF based solely on nominal system dynamics, especially when the ground truth vs nominal system dynamics are substantially different.

III. CONTINUOUS SVO₂ ESTIMATION VIA PI-P-GSEKF

A. Cardiopulmonary Dynamics: Mathematical Model

We used a mathematical model capable of simulating O₂ transport during mechanical ventilation developed in our prior work [10] as system dynamics. The mathematical model includes 5 states: end-tidal O₂ saturation (ETO₂) (x_1) and O₂ concentrations in arteries (x_2 and x_3), tissues and veins (x_4), and peripheral SpO₂ measurement site (x_5). The input to the system is O₂ provided by mechanical ventilation through the lungs. The output of the system is SpO₂ (Fig. 2).

O₂ in the lungs increases with its supply via mechanical ventilation and decreases with its diffusion to the arterial blood (i.e., pulmonary vein):

$$\dot{x}_1 = \frac{f}{V_L} \dot{V}_{ALV}(u - x_1) - \frac{k_{D,O_2}}{V_L} (P_{ALV}x_1 - P_aO_2) \quad (8)$$

where x_1 is ETO₂ [%], V_L is lung volume [ml], $\dot{V}_{ALV}$ is minute ventilation [ml/s], P_{ALV} is alveolar pressure, P_aO_2 is O₂ partial pressure in the proximal arteries close to the heart [mmHg], k_{D,O_2} is diffusion coefficient between the lungs and the arterial blood [ml/(mmHg·s)], u is FIO₂, and f is the fraction of minute ventilation participating in O₂ exchange in

the lungs [18]. P_aO_2 is related to x_2 via the O₂-hemoglobin dissociation curve [19]:

$$x_2 = K_{O_2} S_aO_2 = K_{O_2} \frac{P_aO_2^\gamma}{P_aO_2^\gamma + P_{50}O_2^\gamma} \quad (9)$$

where S_aO_2 is arterial O₂ saturation, K_{O_2} is a factor to convert O₂ saturation to O₂ concentration (0.204 [ml O₂/ml]), $P_{50}O_2$ is O₂ partial pressure corresponding to 50% O₂ saturation, and γ is cooperativity constant. O₂ in the proximal arteries increases with its supply from the lungs as well as veins and decreases with its transport to the distal arteries:

$$\dot{x}_2 = \frac{Q}{V_{PV} + V_{SA}} (x_4 - x_2) + \frac{k_{D,O_2}}{V_{PV} + V_{SA}} (P_{ALV}x_1 - P_aO_2) \quad (10)$$

where x_2 is O₂ concentration in the proximal arterial blood [ml O₂/ml], V_{PV} and V_{SA} are pulmonary venous and systemic arterial blood volumes [ml], Q is cardiac output (CO) [ml/s], and x_4 is O₂ concentration in the venous (including the tissues) blood [ml O₂/ml]. O₂ in the distal arteries increases with its supply from the proximal arteries and decreases with its transport to the tissues and veins [20]:

$$\dot{x}_3 = -\frac{Q_3}{V_3} x_3 + \frac{Q_3}{V_3} x_2 \quad (11)$$

where x_3 is O₂ concentration in the distal arterial blood [ml O₂/ml], and V_3 and Q_3 are hypothetical volume and blood flow to represent the distal arteries. O₂ in the tissues and veins increases with its supply from the distal arteries and decreases with its consumption in the tissues as well as its transport to the proximal arteries:

$$\dot{x}_4 = -\frac{Q}{V_V} x_4 + \frac{Q}{V_V} x_3 - \frac{\eta}{V_V} \quad (12)$$

where x_4 is O₂ concentration in the veins (including the tissues) [ml O₂/ml], V_V is systemic venous plus pulmonary arterial volume [ml], η is metabolic O₂ consumption rate [ml/s]. P_vO_2 is related to x_4 via the O₂ dissociation curve:

$$x_4 = K_{O_2} S_vO_2 = K_{O_2} \frac{P_vO_2^\gamma}{P_vO_2^\gamma + P_{50}O_2^\gamma} \quad (13)$$

where S_vO_2 is venous O₂ saturation. O₂ at the SpO₂ measurement site is modeled as a delayed version of O₂ in the proximal arteries:

$$\dot{x}_5 = -\frac{Q_5}{V_5} x_5 + \frac{Q_5}{V_5} x_2 \quad (14)$$

where x_5 is O₂ concentration at the SpO₂ measurement site [ml O₂/ml], and V_5 and Q_5 are hypothetical volume and blood flow to represent the SpO₂ measurement site. This mathematical model can be expressed into Eq. (1) upon discretization with $x = [x_1, \dots, x_5]^T$ and $y = x_5$, which is characterized by the following parameter vector θ :

$$\theta = \left\{ V_{PV} + V_{SA}, V_V, V_L, k_{D,O_2}, P_{50}O_2, \gamma, \frac{Q_3}{V_3}, \frac{Q_5}{V_5}, \eta, f \right\} \quad (15)$$

B. Experimental Data

To evaluate the PI-P-GSEKF approach to continuous SvO₂ estimation, we used the experimental data collected from 8 hypoxia trials conducted in 4 male pigs (45-60kg) in our prior work (approved by the Institutional Animal Care and Use Committee at University of Maryland School of Medicine (ID #0121006)). In these trials, we recorded physiological

parameters including SvO₂ and CO at a sampling rate of 0.05Hz, and SpO₂, ETO₂, and minute ventilation at a sampling rate of 1Hz. The animals underwent large dynamic decrease in SvO₂ due to the hypoxia induced by lowering FIO₂ from 20% down to 10%-13%. Hence, the experimental data were ideally suited to the development and validation of the PI-P-GSEKF approach to continuous SvO₂ estimation.

C. Evaluation and Analysis

We evaluated and analyzed the PI-P-GSEKF in the context of continuous estimation of SvO₂ from continuous SpO₂ measurement as follows.

We derived a generative sampler for the mathematical model in Fig. 1 by analyzing all the experimental data using the collective variational inference (C-VI) method developed in our prior work [10]. The C-VI method derives probability density of both hypoxia trial-specific and hypoxia trial-average mathematical model parameter values [16]. We employed the hypoxia trial-average probability density as the generative sampler in this letter.

For each hypoxia trial, the PI-P-GSEKF estimated SvO₂ as follows. First, it used the generative sampler derived above to generate plausible operating point samples (i.e., $\vartheta_i^{(0)}$, $i = 1, \dots, N_G = 100$, each of which represents a vector of mathematical model parameters pertaining to a plausible animal) (Eq. (4)). Second, it adapted the N_G $\vartheta_i^{(0)}$ s using a one-time measurement of pre-hypoxic steady-state SpO₂-SvO₂ pair based on the MCMC sampling (i.e., the Metropolis-Hastings algorithm with the likelihood function in Eq. (5)), so that the resulting MCMC sequences $\vartheta_{i,s}$, $i = 1, \dots, N_G = 100$ in Eq. (7) could replicate the SpO₂-SvO₂ measurement pair when the corresponding FIO₂ and minute ventilation were inputted to the mathematical model parameterized by the samples therein. Although multiple SpO₂-SvO₂ pairs may be advantageous to minimize the adverse influence of measurement noise, only a single SpO₂-SvO₂ pair was considered given that each SvO₂ measurement requires invasive blood sampling procedure. After generating 12,000 samples and excluding the first 6,000 burn-in samples corresponding to transient convergence, the MCMC sequences had the size of $N_{MCMC} = 6000$. Third, it calculated θ_i and Q_{θ_i} in Eq. (7) pertaining to all the $\vartheta_{i,s}$, $i = 1, \dots, N_G = 100$. Fourth, it selected $N_{EKF} = 30$ θ_i - Q_{θ_i} pairs with the highest likelihood values based on Eq. (5). Fifth, it designed a bank of $N_{EKF} = 30$ EKFs, each of which was parameterized with θ_i and used Q_{θ_i} as the source of its process noise covariance matrix similarly to our prior work [21], [22]. In all the EKFs, it used a constant sensor noise covariance: $R(k) = R$, which is the (scalar) variance pertaining to SpO₂ measurement. Sixth, it estimated SvO₂ as the Gaussian sum of the $N_{EKF} = 30$ EKFs as in Eq. (2) using FIO₂, minute ventilation, and SpO₂. Each EKF predicted and corrected the states based on the standard prediction and update procedure. Then, the weights were updated based on the likelihood of the $N_{EKF} = 30$ predicted states to yield the measured SpO₂ as in Eq. (3). We repeated the above PI-P-GSEKF computations across all the 8 hypoxia trials.

We evaluated the efficacy of the PI-P-GSEKF in terms of the root-mean-squared error (RMSE) and its variability caused by the randomness due to generative and MCMC sampling (defined as the standard deviation of RMSE resulting from 10 evaluations), the correlation coefficient between measured vs estimated SvO₂, and the confidence interval (defined as $\pm$ SD envelope pertaining to SvO₂ estimation) pertaining

to SvO₂ estimation averaged across each hypoxia trial. To examine the merit of (i) the GSEKF relative to the EKF in encompassing diverse plausible operating points and (ii) the MCMC sampling in personalizing the operating points pertaining to the EKFs, we designed three competing state estimators. First, we designed a PI-EKF, an EKF designed using the nominal system dynamics and the population-level covariance pertaining to the generative sampler [16] as the source of the process noise covariance matrix similarly to our prior work [21], [22]. Second, we designed a PI-P-EKF, the PI-EKF with its operating point and process noise covariance personalized via MCMC sampling. Third, we designed a PI-GSEKF, a GSEKF designed using a bank of $N_{EKF} = 30$ EKFs. We selected the operating points pertaining to these EKFs by (i) generating $N_G = 1000$ $\vartheta_i^{(0)}$ s using the generative sampler in Eq. (4) and (ii) selecting $N_{EKF} = 30$ samples with the highest likelihood values according to Eq. (5). As in the case of the PI-EKF, we used the population-level covariance pertaining to the generative sampler as the source of the process noise covariance matrix pertaining to all the EKFs in the PI-GSEKF. We evaluated the ability of these state estimators to estimate SvO₂ using all the experimental data (i.e., the 8 hypoxia trials), in terms of the metrics described above. Then, we compared the SvO₂ estimation accuracy pertaining to the PI-EKF, the PI-P-EKF, the PI-GSEKF, and the PI-P-GSEKF. We determined the statistical significance in the difference in the performance metrics using the Wilcoxon's rank sum test with Bonferroni correction for multiple comparisons ($p < 0.0083$).

IV. RESULTS AND DISCUSSION

Continuous measurement of SvO₂ has the potential to advance patient monitoring and treatment in critical care and cardiopulmonary medicine. However, it requires invasive pulmonary artery catheterization. In this letter, we demonstrated the proof-of-principle of non-invasive continuous estimation of SvO₂ from readily available continuous SpO₂ measurement based on a novel PI-P-GSEKF approach. Its key idea is to exploit both population-level generative sampling and MCMC sampling to select the operating points of the bank of EKFs comprising the PI-P-GSEKF in such a way that the EKFs can operate in the vicinity of the ground truth system dynamics against large inter-individual variability in cardiopulmonary system dynamics.

Table I summarizes the performance metrics pertaining to the PI-P-GSEKF, while Fig. 3 shows two representative examples of measured SvO₂ vs SvO₂ estimated by the PI-P-GSEKF. Fig. 4 shows the ground truth operating point, the operating points pertaining to the PI-P-GSEKF, and the nominal operating point pertaining to the generative sampler (and thus the PI-EKF), all associated with the two examples in Fig. 3. The PI-P-GSEKF demonstrated efficacy in continuous estimation of SvO₂ from continuous SpO₂ measurement readily available in clinical settings. In the hypoxia trials, SvO₂ changed 18-47%. Hence, the average RMSE of 3.0% may be viewed as small. In addition, the PI-P-GSEKF significantly outperformed the PI-GSEKF, the PI-P-EKF, and the PI-EKF in terms of RMSE and r value (except the PI-GSEKF in terms of r value). Further, the variability of RMSE pertaining to the PI-P-GSEKF was smaller than the same variability pertaining to the PI-P-EKF and the PI-GSEKF (0.95% vs 1.26% and 1.92%, respectively). All these findings suggest the potential of the PI-P-GSEKF approach to SvO₂ estimation in critically ill patients.

TABLE I

EFFICACY OF PI-P-GSEKF, PI-P-EKF, PI-GSEKF, AND PI-EKF: ROOT-MEAN-SQUARED ERROR (RMSE), R VALUE, AND TIME-AVERAGED CONFIDENCE INTERVAL (CI; $\pm\sigma$ ENVELOPE) PERTAINING TO SvO_2 ESTIMATION (MEDIAN (IQR)). *, †, ‡: $P < 0.0083$ WITH RESPECT TO THE PI-EKF, THE PI-GSEKF, AND THE PI-P-EKF (WILCOXON RANK SUM TEST)

	PI-P-GSEKF	PI-P-EKF	PI-GSEKF	PI-EKF
RMSE [%]	2.96 ^{†‡} (1.90,5.24)	3.84 [*] (3.12,7.39)	4.34 (2.88,5.80)	4.66 (3.43,7.66)
r Value	0.98 ^{†‡} (0.95,0.99)	0.96 ^{†‡} (0.75,0.98)	0.97 [*] (0.94,0.98)	0.96 (0.94,0.98)
CI [%]	1.82 ^{†‡} (1.44,2.12)	1.60 ^{†‡} (1.10,2.50)	0.98 [*] (0.82,1.28)	1.44 (1.42,1.52)

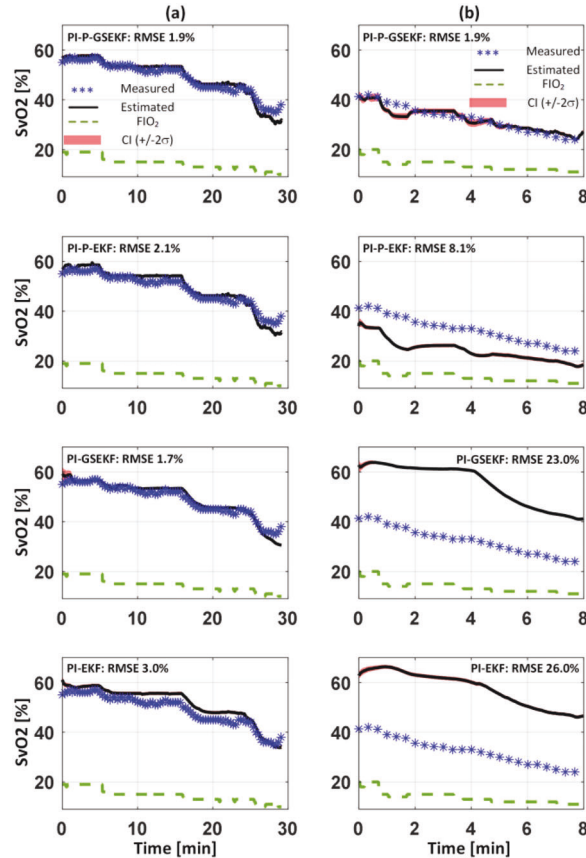


Fig. 3. Two representative examples of measured SvO_2 vs SvO_2 estimated by the PI-P-GSEKF, the PI-P-EKF, the PI-GSEKF, and the PI-EKF. (a) An example where the ground truth operating point was close to the nominal SpO_2 and SvO_2 operating point. (b) An example where the ground truth operating point was far from the nominal SpO_2 and SvO_2 operating point.

Both the population-level generative sampling and the MCMC sampling via one-time SpO_2 - SvO_2 measurement appeared to play meaningful roles in improving the efficacy of continuous SvO_2 estimation via the PI-P-GSEKF. In developing the PI-P-GSEKF, we predicted that the generative sampling would provide diverse operating points to cope with the large inter-individual variability in the system dynamics, and that the MCMC sampling via one-time measurement would adapt the generative operating points toward the ground truth system dynamics. First, the PI-GSEKF exhibited significant improvement in the RMSE of 7% on the average relative to the PI-EKF, which confirmed our prediction on the merit of the generative sampling. Second, the PI-P-GSEKF exhibited

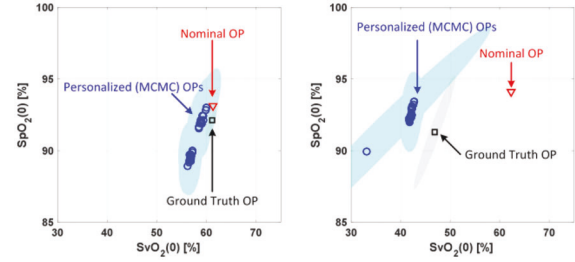


Fig. 4. Two representative examples of ground truth operating point (i.e., experimental values; black square), operating points pertaining to the PI-P-GSEKF (blue circles), and nominal operating point pertaining to the generative sampler (and the PI-EKF) (red triangle) in SpO_2 - SvO_2 space pertaining to Fig. 3. OP: operating point.

an even greater improvement in the RMSE of 31% on the average (which translates into 36% on the average relative to the PI-EKF) relative to the PI-GSEKF, which confirmed our prediction on the merit of the MCMC sampling. In fact, the MCMC sampling personalized the generative operating points to the vicinity of the ground truth (namely, experimental values) in terms of SpO_2 and SvO_2 (Fig. 4). The generative sampling and the MCMC sampling are both crucial to the PI-P-GSEKF: they appear to exert synergistic effects on the efficacy of the PI-P-GSEKF, especially when the ground truth system dynamics cannot be fully captured by the one-time measurement: diverse generative operating points can be adapted to diverse candidate operating points which replicate the one-time measurement, as illustrated in Fig. 4, a subset of which is close to the ground truth operating point. Indeed, the PI-P-GSEKF exhibited significant improvement in the RMSE of 23% on the average relative to the PI-P-EKF with smaller inter-individual variability in all the metrics in Table I, which shows the advantage in leveraging the GSEKF relative to a single EKF. The advantage of the PI-P-GSEKF, which employs a bank of EKFs at diverse operating points commensurate with the one-time measurement, may be further highlighted by the limitations pertaining to the state-of-the-art methods. In particular, our work showed that adaptive techniques such as augmented EKF (where parameters are augmented in the state [11]) and dual-EKF (where the state and the parameters are estimated sequentially [12], [13], [14]) suffered from poor accuracy in continuous SvO_2 estimation, which is likely due to the deteriorated observability as implied by a substantial ill conditioning of the observability test matrix (not shown; not surprisingly, a large number of mathematical model parameters in addition to SvO_2 cannot be inferred from FIO_2 and SpO_2 measurements alone).

All in all, the comparison of Fig. 3 and Fig. 4 illustrates that the PI-P-GSEKF was comparable to the PI-P-EKF, the PI-GSEKF, and the PI-EKF when the ground truth operating point was close to the nominal operating point in terms of SpO_2 and SvO_2 (so that it could be easily generated by the generative sampler) (Fig. 3(a) and Fig. 4(a)), whereas the PI-P-GSEKF appeared to largely outperform the PI-P-EKF, the PI-GSEKF, and the PI-EKF when the ground truth operating point was far from the nominal operating point in terms of SpO_2 and SvO_2 (so that the generative sampler could not generate operating points close to it) (Fig. 3(b) and Fig. 4(b)).

The GSF architecture is known to frequently result in a large confidence interval in state estimation [15]. In contrast, the integration of the generative sampling in the PI-P-GSEKF approach resulted in a small confidence interval (Table I). In the setting of the PI-GSEKF, the operating points are diversely

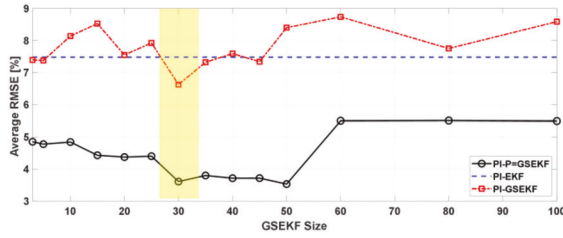


Fig. 5. Dependence of the RMSE pertaining to the PI-P-GSEKF (black) and PI-GSEKF (red) on their size (i.e., the number of EKFs therein). The blue dashed line indicates the RMSE pertaining to the PI-EKF (whose size is fixed to 1).

distributed. Hence, we observed that only a small number (usually 1) of EKFs which are close to the ground truth operating point attained high weights, whereas most of the remaining EKFs attained weights close to zero, which may have contributed to decrease the confidence interval. On the other hand, the integration of the MCMC sampling in the PI-P-GSEKF approach resulted in a larger confidence interval (Table I). In the setting of the PI-P-GSEKF, the operating points are clustered in the vicinity of the ground truth operating point (e.g., Fig. 4). In addition, the size of the process noise covariance pertaining to most (if not all) operating points (i.e., Q_{θ_i} , $i = 1, \dots, N_{EKF} = 30$ in Eq. (8)) is decreased by virtue of personalization. However, there are a good number of equally good MCMC operating points. Hence, we observed that a relatively large number (usually ≥ 3) of EKFs attained high weights, which may have contributed to increase the confidence interval relative to the PI-GSEKF. Nonetheless, the confidence interval was small in the absolute sense ($< 2\%$).

Our analysis also revealed that there may be an optimal size (i.e., the number of EKFs) pertaining to the PI-P-GSEKF. In the context of continuous SvO_2 estimation problem at hand, Fig. 5 shows the dependence of the RMSE pertaining to the PI-P-GSEKF and the PI-GSEKF on their size. In the case of the PI-P-GSEKF, the estimation efficacy initially improved as its size increased (Fig. 5). But, beyond a certain size, the efficacy started to deteriorate (Fig. 5). In the case of the PI-GSEKF, the trend was more sophisticated (Fig. 5). The biphasic trend in Fig. 5 pertaining to the PI-P-GSEKF may be explained as follows. Since the PI-P-GSEKF includes N_{EKF} high-rank operating points based on the likelihood evaluated using Eq. (6), its efficacy improves initially as an increasing number of high-rank operating points are included. However, its efficacy deteriorates as low-rank operating points start to be included. Hence, the size of the PI-P-GSEKF may need to be selected with caution. Remarkably though, the PI-P-GSEKF consistently outperformed the PI-EKF and the PI-GSEKF irrespective of its size, which may be an additional evidence to support the efficacy of the PI-P-GSEKF.

In sum, our work illustrates that (i) SvO_2 estimation via SpO_2 measurement may be a reasonable approach; and that (ii) the PI-P-GSEKF may be a viable solution to continuous SvO_2 estimation problem, with its ability to cope with the large inter-individual variability via the generative sampling and the MCMC sampling to specify diverse yet personalized operating points relevant to the EKFs therein.

V. CONCLUSION

We demonstrated the proof-of-principle of the novel PI-P-GSEKF approach to state estimation in dynamical systems

with large inter-individual variability using continuous SvO_2 estimation as a case study. The PI-P-GSEKF may enable non-invasive continuous SvO_2 estimation, which may enable superior patient management in critical care and cardiopulmonary medicine. Future work must be invested to further investigate the PI-P-GSEKF approach including its computational aspects, compare it with other existing state estimation techniques, and explore its use in other challenging state estimation problems.

REFERENCES

- [1] R. M. Leach and D. F. Treacher, "The pulmonary physician in critical care- 2: Oxygen delivery and consumption in the critically ill," *Thorax*, vol. 57, no. 2, pp. 170–177, 2002.
- [2] D. S. Martin and M. P. W. Grocott, "Oxygen therapy in critical illness: Precise control of arterial oxygenation and permissive hypoxemia," *Crit. Care Med.*, vol. 41, no. 2, pp. 423–432, 2013.
- [3] C. Hartog and F. Bloos, "Venous oxygen saturation," *Best Pract. Res. Clin. Anaesthesiol.*, vol. 28, no. 4, pp. 419–428, 2014.
- [4] F. Bloos and K. Reinhart, "Venous oximetry," *Intensive Care Med.*, vol. 31, pp. 911–913, Jul. 2005.
- [5] S. M. Bradley and A. M. Atz, "Postoperative management: The role of mixed venous oxygen saturation monitoring," *Seminars Thorac. Cardiovasc. Surg., Pediatr. Card. Surg. Annu.*, vol. 8, no. 1, pp. 22–27, 2005.
- [6] M. Hadian and M. R. Pinsky, "Evidence-based review of the use of the pulmonary artery catheter: Impact data and complications," *Crit. Care*, vol. 10, pp. 1–11, Jun. 2006.
- [7] J.-L. Vincent, "The pulmonary artery catheter," *J. Clin. Monit. Comput.*, vol. 26, pp. 341–345, Aug. 2012.
- [8] K. Cruz and C. Franklin, "The pulmonary artery catheter: Uses and controversies," *Crit. Care Clin.*, vol. 17, no. 2, pp. 271–291, 2001.
- [9] S. G. Peters, B. Afessa, P. A. Decker, D. R. Schroeder, K. P. Offord, and J. P. Scott, "Increased risk associated with pulmonary artery catheterization in the medical intensive care unit," *J. Crit. Care*, vol. 18, no. 3, pp. 166–171, 2003.
- [10] P. Rezaei et al., "Hypothermic peritoneal perfusion of cold oxygenated perfluorocarbon may improve the efficacy of extracorporeal oxygenation: A mathematical model-based analysis," *J. Dyn. Sys., Meas., Control.*, vol. 147, no. 2, pp. 1–11, 2025.
- [11] A. Gelb, *Applied Optimal Estimation*. Cambridge, MA, USA: MIT Press, 1974.
- [12] J.-M. Jover and T. Kailath, "A parallel architecture for Kalman filter measurement update and parameter estimation," *Automatica*, vol. 22, no. 1, pp. 43–57, 1986.
- [13] S. Haykin, *Kalman Filtering and Neural Networks*. Hoboken, NJ, USA: Wiley, 2004.
- [14] T. A. Wenzel, K. J. Burnham, M. V. Blundell, and R. A. Williams, "Dual extended Kalman filter for vehicle state and parameter estimation," *Veh. Syst. Dyn.*, vol. 44, no. 2, pp. 153–171, 2006.
- [15] J. L. Crassidis and J. L. Junkins, *Optimal Estimation of Dynamic Systems*. Boca Raton, FL, USA: Chapman Hall/CRC, 2004.
- [16] A. Tivay, G. C. Kramer, and J.-O. Hahn, "Collective variational inference for personalized and generative physiological modeling: A case study on hemorrhage resuscitation," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 2, pp. 666–677, Feb. 2022.
- [17] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, *Handbook of Markov Chain Monte Carlo*. Boca Raton, FL, USA: CRC press, 2011.
- [18] L. M. Brewer, J. A. Orr, and N. L. Pace, "Anatomic dead space cannot be predicted by body weight," *Respir. Care*, vol. 53, no. 7, pp. 885–891, 2008.
- [19] R. N. Pittman, *Regulation of Tissue Oxygenation*. San Rafael, CA, USA: Morgan Claypool, 2016.
- [20] C.-S. Kim, J. M. Ansermino, and J.-O. Hahn, "A comparative data-based modeling study on respiratory CO_2 gas exchange during mechanical ventilation," *Front. Bioeng. Biotechnol.*, vol. 4, p. 8, Feb. 2016.
- [21] W. Yin, A. Tivay, and J.-O. Hahn, "Hemodynamic monitoring via model-based extended Kalman filtering: Hemorrhage resuscitation and sedation case study," *IEEE Control Syst. Lett.*, vol. 6, pp. 2455–2460, 2022.
- [22] Y. R. Chalumuri, X. Jin, A. Tivay, and J.-O. Hahn, "Detection of internal hemorrhage via sequential inference: An in silico feasibility study," *Diagnostics*, vol. 14, no. 17, p. 1970, 2024.