Published in Smart Health, December 2018, vol. 9–10, pp. 101-114, which should be cited to refer to this work. DOI : 10.1016/j.smhl.2018.07.003

Robustness analysis of personalised delivery rate computation for IV administered anesthetic

Alena Simalatsar[a,b,*], Monia Guidi[b,c], Pierre Roduit[a], Thierry Buclin[b]

[a] School of Engineering, University of Applied Sciences and Arts Western Switzerland (HES-SO), Sion, Switzerland
[b] Service of Clinical Pharmacology, University Hospital of Lausanne (CHUV), Lausanne, Switzerland
[c] School of Pharmaceutical Sciences, University of Geneva, University of Lausanne, Geneva, Switzerland

ARTICLE INFO

Keywords:
Individualized anesthesia
Drug delivery
Closed-loop control
Kalman filer
Robustness analysis

ABSTRACT

Controlled delivery of intravenous (IV) anesthetics aims at fast and safe achievement and maintenance of a suitable *depth of hypnosis* (DOH), by ensuring appropriate effect site (i.e. brain) exposure to the drug. Today, such drugs are regularly injected by Target Controlled Infusion (TCI) systems, piloted by an open-loop algorithm based on Pharmacokinetic (PK) models. Yet the inaccuracy of concentration prediction of current TCI can reach up to 100%. The situation could be improved by closing the loop with sensors providing regular real measurements of the anesthetic concentration in body fluids. In this paper we present a closed-loop algorithm based on the classic open-loop algorithm combined with a Kalman filter. The latter estimates plasma drug concentration based on PK model and sensor measurements. The estimates are then used in the open-loop algorithm. To validate our approach measurements are generated by means of modulating the population-based plasma concentration values with the maximum inter- and intra-patient variability of the statistical Eleveld's (Eleveld et al., 2014) PK model. This allows us to stress the system to a maximum level prior to testing it on patients. We also perform robustness analysis of this algorithm by accounting for realistic measurement periods and delays.

1. Introduction

Every year, only in the US, 40 million people undergo anesthesia. One in every 1000 of them remains awake, some still feel pain while not being able to move (Burchum et al.,; Orser et al., 2008) and about one in 200,000 dies only because of anesthesia (Kotsovolis et al., 2009). The toxic actions of anesthetics continue to deserve constant research. Part of this is due to the difficulty of ensuring a leveraged delivery of anesthetics, analgesics and muscle relaxants, drugs administered in a cocktail during the anesthesia process. Anesthetics are hypnotic agents, usually short-time acting, administered intravenously. The controlled delivery of IV anesthetics aims at fast and safe achievement and maintenance of a given level of DOH, correlated with plasma and effect site (i.e. brain) concentration of the administered drugs (e.g. *propofol*, *fentanyl*, etc.). Today, such drugs are regularly injected by TCI systems (*Orchestra Base Primea*), controlled by an open-loop algorithm (Shafer and Gregg, 1992) computing the delivery rate based on plasma drug concentration quantitatively predicted with respect to a set of patient's demographic characteristics by a population PK model. The classical PK model of *propofol* is a three compartments PK model extended with a virtual compartment representing the effect

* Corresponding author at: School of Engineering, University of Applied Sciences and Arts Western Switzerland (HES-SO), Rue du Rawil 47, 1950 Sion, Switzerland.
E-mail addresses: alena.simalatsar@hevs.ch (A. Simalatsar), monia.guidi@chuv.ch (M. Guidi), pierre.roduit@hevs.ch (P. Roduit), thierry.buclin@chuv.ch (T. Buclin).

https://doi.org/10.1016/j.smhl.2018.07.003

site. The population *propofol* PK models supported by TCI systems are those of Schnider et al. (1998, 1999) and Marsh et al. (1991). However, only due to inter-patient variability, their inaccuracy in estimating drug concentrations can reach up to 100% (Eleveld et al., 2014).

Instead of relying on predicted drug concentration in plasma, some attempts to improve the situation were made by introducing a closed-loop control that uses *electroencephalogram* (EEG) monitoring to derive the *bispectral index* (BIS) correlated with the DOH (De Smet et al., 2007; Liu et al., 2011). However, BIS only indirectly depends on drug exposure, is delayed with respect to changes of drug concentration in plasma, can be affected by various sources of noise, and cannot be used during interventions involving the patient's head. Moreover, when BIS is used to compute personalised drug administration as in De Smet et al. (2007), the rate of drug delivery is still computed based on a population PK model. The situation could definitely be improved with sensors providing regular measurements of the concentration of anesthetic agents actually circulating in blood and a closing-loop algorithm able to personalise the delivery rate computation based on these measurements. Some preliminary prototypes of such sensors exist already (Kivlehan et al., 2012; Langmaier et al., 2011; Stradolini et al., 2017), allowing us to estimate expected measurement noise, delay and possible measurement period. However, even when the technology is mature enough, closing the control loop based only on sensor measurements will certainly not be fully reliable, since such measurements will be affected by noise and breakdowns. An algorithm adjusting the delivery rate based on the population PK model and sensor measurements able to deal with huge noise is required to improve the precision of anesthesia delivery.

In this paper, we present a control algorithm for the closed-loop delivery of *propofol*. It is a modification of the classical Shafer's open-loop algorithm (Shafer and Gregg, 1992) combined with a Kalman filter estimating drug concentrations in PK compartments based on both a population PK model and sensor measurements of plasma concentrations. The algorithm calculates the infusion rate by looking in the future from time n after injecting the drug during $10sec$ with a set of reference delivery rates, and choses the delivery rate such that the target effect site concentration is achieved as rapidly as possible without overshoot. In our implementation, the concentration at the effect site is computed based on regularly updated plasma concentration measurements and population PK model using Kalman filter. Rather than using the Schnider et al. (1998, 1999) or Marsh et al. (1991) models supported by standard TCI, we rely on the PK model recently developed by Eleveld et al. (2014). This model is based on measurements performed on a larger population than the Schnider model and includes patients and volunteers (see Section 3.3). Moreover, it provides both inter- and intra- patients variability of PK parameters.

Prior to testing such a system on patients, it is essential to verify the robustness of the closed-loop algorithm against the measurement noise, period, and delay. To do so, we emulate the realistic noisy measurements of plasma drug concentrations obtained in a random patient by using Eleveld's model with both inter- and intra- patients variability of PK parameters. The inter-patient variability is used to simulate a group of 1000 patients having identical demographic characteristics but different PK parameters and, consequently, different concentration-time (CT) profiles. It is important to note that both Schnider and Marsh models, as they are currently implemented in TCI systems, would give a single CT profile (one for each model) for this group of patients under some predefined delivery rate. In turn, the intra-patient variability was used to emulate the random normally distributed measurement noise. Using the maximum possible variability allows us to stress the system to the extreme. Since each measurement may arrive with a certain period (e.g. every $15sec$) and be delayed, such that its time stamp correspond to some time in the past, we also test the algorithm robustness for realistic values of measurement periods and delays. We show that the delivery rate adapts well to administer appropriate amounts of drug for a random individual, avoiding major potential overdosing or unexpected awakening in such patients compared with the standard TCI.

This paper is organised as follows. In Section 2, we present the related work. Section 3 talks about the basics of PK modelling, and explains how we generate the measurements to test our algorithm. In Section 4, we present the control algorithm and evaluate its robustness in Section 5.

2. Related work

Controlled delivery of different drugs, such as inhaled or IV injected anesthetics, analgesics, and muscle relaxants, administered to patients during anesthesia, has received much attention in the past decades from various research groups. The use of PID controller for anesthesia delivery based on *Electromyography* (EMG) measurements was presented in several research articles (Silva et al., 2014; Silva et al., 2015; Zhusubaliyev et al., 2013). Inhaled anesthetic agents are widely used nowadays with feedback control systems based on the control of 'MAC' values, Minimum Alveolar (i.e. fundamental units of the lung) Concentration, associated with the probability of patient movement in response to surgical stimuli (Bailey and Haddad, 2005). Gentilini et al. (2001) present two complicated closed-loop controllers for the delivery of inhaled anesthetic *isoflurane*: one based on the monitoring of arterial pressure and another one of DOH through BIS. For *propofol*, an IV anesthetic agent often preferred to inhaled gases, an adaptive cascaded PID controller for the target driven delivery has been proposed (Simalatsar et al., 2016). The control variable is the plasma concentration of *propofol*. All such approaches require an extensive study of system safety and robustness. In this paper, we also focus on closing the loop for the control of *propofol* delivery.

A recent review (Dumont and Ansermino, 2013) gives a good survey of existing techniques for the controlled delivery of IV anesthetics. It classifies control algorithms into three groups: open-loop, closed-loop feed-forward, and feedback control. Feedback control algorithms are divided into predictive and adaptive control. A classical open-loop algorithm was presented several decades ago by Shafer and Gregg (1992) and its adapted version is implemented in TCI systems used in hospitals for IV anesthesia deliver, e.g. the original Diprifusor from AstraZeneca or the more recent one Orchestra Base Primea TCI system from Fresenius (Orchestra Base Primea). Still, the use of open-loop and feedforward algorithms in controlled anesthesias delivery suffers from large inter- and intra-

patient variability of underlying models (see Section 3.2). According to Eleveld et al. (2014) the predicted CT profile of a patient with the same characteristics of the average one can differ up to 100% from the average CT profile (see Section 3.3) only due to the inter-individual variability.

In this context, there has been multiple attempts to close the loop with measurable parameters, such as EEG (Silva et al., 2010; Rampil et al., 1998) using either PID controllers (Liu et al., 2011) or Bayesian-based closed-loop systems (De Smet et al., 2007). Silva et al. (2015) give a good overview of PID controllers designed for a comprehensive group of drug delivered during anesthesia: anesthetics, e.g. *propofol*, analgesics, e.g. *remifentanil*, and muscle relaxants, e.g. *atracurium* and *rocuronium*. Liu et al. (2011) present a PID controller for controlled coadministration of *propofol* and *remifentanil*, guided by BIS monitor. The controller allows automated delivery of *propofol* and *remifentanil* that have interdependent pharmacodinamics and maintains BIS values in predetermined boundaries during general anesthesia better than manual administration. De Smet et al. (2007) implement a Bayesian method to evaluate the parameters of the Hill sigmoid function describing pharmacodynamics of anesthesia using observed BIS indexes. This allows the computation of the necessary drug concentration at the effect site to achieve a certain value of the BIS index. However, the rate of drug delivery for such systems is based on the population PK model of regular TCI pump and thus suffers from the same large population PK prediction errors. Moreover, all the controllers developed for these drugs are based on controlling either the level of BIS or the EMG signals. This is likely due to the absence of sensors able to continuously measure drug concentrations in plasma, though the technologies are about to come to a mature phase (Stradolini et al., 2017). Inspired by their work, we elaborate similar closed-loop systems for IV delivered drugs. In addition, similar to Abbas et al. (2016), we perform a step of Computer-Aided Clinical Trials (CACT), during which we evaluate the algorithm robustness against the measurement noise, period, and delay by simulating the plasma drug concentration measurements using inter- and intra- patient variability of Eleveld model (Eleveld et al., 2014).

3. Pharmacokinetic modeling

The PK model of *propofol* is usually described by three-compartment model extended with a virtual compartment representing the effect site, i.e. brain (Fig. 1). The central compartment with volume of distribution V_1 represents the plasma and rapidly exchanging interstitial fluids of a patient. The drug is delivered to the central compartment at rate U (changing over time) and is cleared from it with elimination rate k_{10}, defined as CL/V_1 with CL corresponding to drug clearance. The two compartments with volumes V_2 and V_3 and inter-compartmental clearances Q_2 and Q_3 are the rapid and the slow peripheral compartments, respectively. They are needed to model the drug's distribution and its further release into/by various tissues, regulated by k_{12}, k_{21}, k_{13} and k_{31}, the first-order transfer rate constants from compartment i to compartment j.

The following system of differential Eqs. (1), (2), and (3) describes the evolution of drug amounts in the three compartments:

$$\frac{dA_1}{dt} = A_2 k_{21} + A_3 k_{31} - A_1(k_{10} + k_{12} + k_{13}) + U \tag{1}$$

$$\frac{dA_2}{dt} = A_1 k_{12} - A_2 k_{21} \tag{2}$$

$$\frac{dA_3}{dt} = A_1 k_{13} - A_3 k_{31} \tag{3}$$

where A_1, A_2, and A_3 represent the amount of drug in the central, rapid, and slow peripheral compartments, respectively. To get the concentration in a compartment, the amount must be divided by the corresponding volume of distribution ($C_i = A_i/V_i$). The relationships between the microconstants k_{ij} and the classic PK parameters (CL, V_i, and Q_i) are shown in Fig. 1.

The effect site compartment is considered to be too small and thus has no influence on the concentration in the central compartment. It is then added in the PK model as a fourth compartment with volume of distribution $V_4 = V_1/10000$ (Shafer and Gregg, 1992). The microconstants driving central/effect compartments exchange can be defined as $k_{41} = k_{e0}$ and $k_{14} = k_{e0}/10000$, where k_{e0} is the effect site elimination rate constant. The concentration of the drug at the effect site is computed from the concentration in the central compartment as described by Eq. (4):

$$\frac{dC_4}{dt} = (C_1 - C_4)k_{e0} \tag{4}$$

The most widely used PK models for *propofol* are the Schnider et al. (1998, 1999) and Marsh et al. (1991) ones. The principal

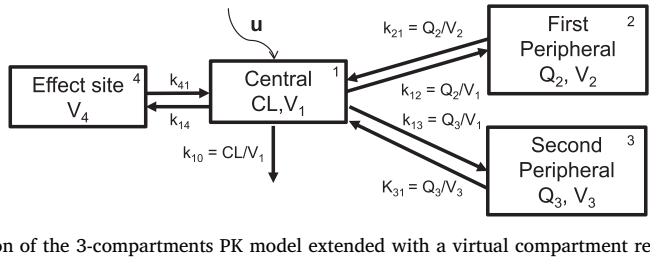

Fig. 1. Schematic representation of the 3-compartments PK model extended with a virtual compartment representing effect site, i.e. brain.

difference between these two models lays in the way the population PK parameters (PAR_k, representing k^{th} parameter out of CL, V_i, and Q_i) are computed based on patients' demographic characteristics represented by weight, height, age and gender of an individual. This difference leads to different sets of k_{ij} constants for these population PK models for patients with identical demographic characteristics. In turn, this results in two different CT profiles for some predefined delivery rate. Recently one additional model was developed by Eleveld et al. It provides not only a new way for population k_{ij} constants computation but also inter- and intra- patients variabilities of these PK parameters. A more detailed comparison of Schnider and Eleveld models is presented in Section 3.3.

3.1. PK model state space representation

The state space representation of such three-compartment PK model with virtual compartment extension is as follows:

$$\dot{x} = Fx + Bu \tag{5}$$

$$y = Cx + Du \tag{6}$$

$$F = \begin{bmatrix} -(k_{10} + k_{12} + k_{13} + k_{14}) & k_{21} & k_{31} & k_{41} \\ k_{12} & -k_{21} & 0 & 0 \\ k_{13} & 0 & -k_{31} & 0 \\ k_{14} & 0 & 0 & -k_{41} \end{bmatrix} \tag{7}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} B = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, C = \begin{bmatrix} 1/V_1 \\ 0 \\ 0 \\ 0 \end{bmatrix}^T, D = 0, \tag{8}$$

where x is a vector of the amounts of drug $x_1 = A_1$, $x_2 = A_2$, $x_3 = A_3$, and $x_4 = A_4$, respectively; F is a state transition matrix, marked as A in classical state space model representation; B and C are the input and output matrices while u is the input vector, i.e. in our case $[U, 0, 0, 0]$ with U being the delivery rate; vector y is the output vector that represents the set of observed value, i.e. the concentration in the central compartment, such that $y = [x_1/V_1, 0, 0, 0]$; D is the feedthrough (or feedforward) matrix, which in cases when the system model does not have a direct feedthrough is a zero matrix. It is easy to show that the system is *controllable* (i.e. it is possible to find inputs that allow to steer the states from any initial value to any final value within some finite time window) and *observable* (i.e. the internal states of a system can be inferred from its external outputs) by showing that $[B\ FB\ F^2B\ F^3B]$ and $[C\ CF\ C^2F\ C^3F]^T$ are full rank matrices.

3.2. Inter- and intra-patient variability

The inter-patient variability is the well-known CT profile difference among individuals having identical demographic characteristics. The main reason of inter-patient variability is that physiological parameters taken into account in PK models only capture a small part of the differences between patients. It is typically described assuming a log-normal distribution for a given PK parameter:

$$PAR_k^p = PAR_k * exp(\eta_k^p) \tag{9}$$

where PAR_k^p is one (k^{th}) PK parameter out of CL, V_i, and Q_i, of the p^{th} individual, PAR_k its average population value and η_k^p is the k^{th} individual component of the inter-patient random effect, an independent, normally distributed variable with 0 mean and ω_k^2 variance. The inter-individual variability is estimated as ω_k. Unfortunately, establishing which parameters are responsible for the variability in a general population strongly depends on the selected group of patients. This implies that published PK models of the same drug can have not only discordant estimates of the inter-individual variability but can also have it associated to completely different parameters.

The intra-patient variability is due to measurement errors in data used during model elaboration, biological fluctuations over the observation period, and inaccuracies inherent to models. For example, a basic assumption says that the compartments are homogeneous and well-mixed, and that the transport of the drug between compartments is linear. However, it is known that underlying processes of drug disposition are not fully linear.

The intra-individual variability is more exhaustively described by a combined additive and proportional error model:

$$Y_n^p = C_n^p + \sqrt{prop^2 * (C_n^p)^2 + add^2} * \varepsilon_n^p \tag{10}$$

where Y_n^p is the plasma concentration measured at time n for the p^{th} individual, C_n^p is the corresponding predicted concentration, $prop$ and add are the proportional and additive error terms and ε_n^p is the np^{th} component of the random effect, an independent, normally distributed variable with 0 mean and 1 variance. A simple additive error model is obtained with $prop = 0$, while simple proportional error model with $add = 0$.

3.3. Schnider vs Eleveld model

The model of Schnider et al. (1998, 1999), implemented in the BasePrimea TCI pump (Orchestra Base Primea), is currently recommended in many hospitals. This model was developed based on *propofol* concentrations collected in 24 healthy volunteers and a

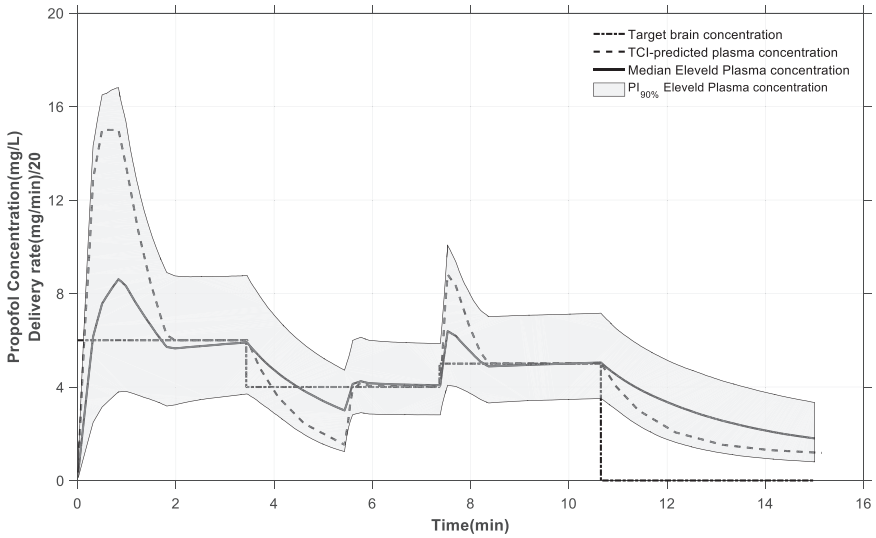

Fig. 2. Comparison of CT profiles based on Schnider's population PK model of and median CT profile with 90% prediction interval for individuals with inter-patient variability following Eleveld's model.

very low inter-individual variability was reported. Body weight, lean body weight, height and age were associated with *propofol* PK. Recently, Eleveld et al. published a model for *propofol* developed over a broader population compared to Schnider's model, of 660 individuals, including children, adults, elderly, and obese patients as well as volunteers, retrieved from 21 studies (Eleveld et al., 2014). Body weight, development, age, and gender had all an impact on *propofol* pharmacokinetics.

In Eleveld's model, distinct PK parameters were associated with patients and healthy volunteers, since clinical conditions might alter *propofol* pharmacokinetics. An important inter-individual variability associated with all the PK parameters (range: 18% to 62%) characterizes *propofol* CT profiles. Intra-individual variability was defined as a mixed proportional and additional error model, with proportional component ranging from 33% to 55% according to each study with an additional inter-study variability of 21% and the additional component set to a fixed value of $1.5x10^{-4}mg/L$. Eleveld's and Schnider's models would give two single CT profiles (one for each model) under some predefined drug dosage regimen for patients with same body weight, height and gender if inter-patient variability is ignored. The profiles obtained with the two models under the same drug dosage regimen differ due to the dissimilarities in their PK parameters.

Fig. 2 compares Schnider's and Eleveld's models, while showing the consequences of the important inter- individual variability characterizing the later model. A virtual 36 year old male patient with a body weight of 70 kg and a height of 170 cm with changing *propofol* target for brain concentration of 6, 4, and 5 mg/L during a 15 min surgical operation was chosen. The *porpofol* infusion rate for the chosen patient to achieve such targets was retrieved from the Orchestra Base Primea TCI pump, which implements the open-loop algorithm described in Section 4 based on Schnider's model without inter-individual variability. The pump was programmed with the indicated patient's parameters and target change profile. According to the TCI pump, the target was well achieved within 1.5 to 2 *minutes* after a change of target value. The dashed line on the plot shows the TCI-predicted plasma concentration response that corresponds to the population plasma concentration computed using Schnider's model. Of note, all patients with the same demographic characteristics as the index subject would have exactly the same PK profiles to achieve the previously defined targets under the same drug delivery scheme using Schnider's model.

The index patient was then simulated 10,000 times under the obtained dosage regimen using Eleveld's model with inter-patient variability. These virtual subjects with identical demographic characteristics would therefore differ in their PK parameters and, consequently, in their CT profiles. Median CT with 90% prediction interval ($PI90\%$), i.e. 5% of the generated virtual patients will have CT profiles below this interval and other 5% above, are shown in the figure. The continuous line on the plot shows the average population (median) plasma concentration response predicted by Eleveld's model to the provided delivery rate. The grey area represents the variability among 90% of 10,000 virtual patients according to Eleveld's model (PI90%). As we can see from the plot, the predicted plasma concentration of some individuals whose CT profile is close to the borders of the grey area can differ from the average one by a factor 2. Due to the huge variability associated with *propofol* pharmacokinetics, open-loop TCI pumps might deliver inadequate *propofol* dosages to patients with possible clinical consequences, which is the main motivation for developing the closed-loop algorithm presented in this paper. Eleveld's model was judged more appropriate than Schnider's one to the purpose of the present work. The population model was used as the initial condition model in the proposed closed-loop control algorithm, while inter- and intra-individual variabilities were used for realistic measurement generation.

4. Controlled delivery

The open-loop algorithm to rapidly achieve and maintain stable drug concentrations at the effect site was presented by Shafer and

Gregg in 1992 (Shafer and Gregg, 1992). Modifications of this algorithm adapted to particular injection platform is widely used in TCI pumps. Here we introduce the essential parts of the algorithm that underwent the modification as presented in Section 4.2.

This algorithm computes the infusion rate based on the extended PK model illustrated in Section 3. The amount of drug in four compartments described with vector x at one unit step n is recursively computed as:

$$x_n = x_{n-1} + Fx_{n-1} + Bu_n \tag{11}$$

The algorithm computes the infusion rate by looking in the future from time n after injecting the drug with reference unit delivery rate during $10sec$ and choses u_n such that the desired target amount at the effect site, A_4, is achieved as rapidly as possible without overshoot. Such computation is repeated every step, e.g. every second, of drug injection.

It is assumed that the pharmacokinetics of anesthetic drug *propofol* is linear. This way, the amount of drug in the effect site over time, following any input represents the superposition of: (1) the amount of drug in the effect site over time, assuming that no input had been administered, and (2) the amount of drug in the effect site over time for the input only, assuming that there had been no drug in the body previously.

Algorithm 1. A_4 with a unit rate during 10 *seconds*.

```
Data: a(0) = [0,0,0,0], j = 1;
Result: t_peak, E[]
while E[j] > E[j-1] do
    if j < 11 then
        u[j] = 1;
    else
        u[j] = 0;
    end
    a[j] = a[j-1] + Fa[j-1] + Bu[j];
    E[j] = a_4;
end
t_peak = j - 1;
```

During the first step (see Algorithm 1), it calculates the amount of drug in the effect site over time during and following a $10sec$ infusion at a rate of $1unit/sec$, administered to the body in the absence of any previously administered drug, up to t_{peak} time (at which the effect site concentration reaches the maximum). This calculation needs to be done only once, and can be done prior to actually starting the infusion. Here, a is a vector of temporary state variables representing the amounts of drug in the four compartments.

During the second step (see Algorithm 2), the amount of drug at the effect site from time n to time $n + t_{peak}$ is computed as if no drug was given. Here, g_i are the temporary state variables representing the amounts of drug in the four compartments created for the purposes of this calculation.

Algorithm 2. $A_4[n: n + t_{peak}]$ with no drug injected.

```
Data: g_1[0] = A_1, g_2[0] = A_2, g_3[0] = A_3, g_4[0] = A_4;
Result: G_4[1 : t_peak]
for j = 1 to t_peak do
    g[j] = g[j-1] + Fg[j-1];
    G_4[j] = g_4;
end
```

According to superposition, the amount of drug in the effect site at time $n + j$ following an infusion of rate u will be $G_4[j] + E[j]u$. Defining j_{peak} as the value of j at peak value of $G_4() + E()u$ for a given u, u can be computed as:

$$u = \frac{targetA_4 - G_4[j_{peak}]}{E[j_{peak}]} \tag{12}$$

Since j_{peak} is not known the equation is solved for j_{peak} and u simultaneously by means of a search algorithm (Shafer and Gregg, 1992).

4.1. Kalman Filter Implementation

Kalman filtering, also known as linear quadratic estimation (LQE), is a recursive filter that estimates the internal state of a linear

dynamic system from a series of measurements containing statistical noise and other inaccuracies, to provide estimation of the system state.

Algorithm 3. Classical Kalman filter implementation.

Predict:
$\widehat{X}_{n|n-1} = F\widehat{X}_{n-1|n-1} + Bu_{n-1}$ // a priori state est.
$P_{n|n-1} = FP_{n-1|n-1}F^T + Q$ // a priori covariance est.
Update:
$\widetilde{d} = y_n - H\widehat{X}_{n|n-1}$
$S_n = HP_{n|n-1}H^T + R_n$
$K_n = P_{n|n-1}H^T S_n^{-1}$
$\widehat{X}_{n|n} = \widehat{X}_{n|n-1} + K_n\widetilde{d}_n$ // a posteriori state est.
$P_{n|n} = (I - K_nH)P_{n|n-1}$ // a posteriori covariance est.

The algorithm (see Algorithm 3) works in a two-step process. In the prediction step, the Kalman filter produces estimates of the current state variables based on the state space model of the linear dynamic system and computes the estimate covariance P based on the covariance of the process noise Q, the state transition model F and the previous estimate of the covariance P.

When applying Kalman filter to PK models with series of plasma concentration measurements, a personalised plasma concentration estimation can be computed based on the population PK model with the covariance of the process noise defined by the inter-individual variability. In the second step, once a new measurement y is available, the measure residual $\widetilde{d}$ is computed. Each measurement is considered to be corrupted with some amount of noise with covariance R. The residual covariance S and optimal Kalman gain K are also updated before computing the a posteriori state estimate $\widehat{X}$ and the a posteriori estimate covariance P. Once the next plasma concentration (corrupted with some amount of error, including random noise) is observed, these estimates are updated, with more weight being given to estimates with higher certainty, such as PK model prediction or the measurement. The two steps algorithm implementing the Kalman filter is presented in Algorithm 3, where $\widehat{X}_{n|n-1}$ and $P_{n|n-1}$ stand for the a priory estimation of the state and estimate covariance, respectively; $\widehat{X}_{n|n}$ and $P_{n|n}$ stand for their a posteriori estimates, and $H = [1/V_1, 0, 0, 0]^T$ is the observation model, which maps the true state space into the observed space.

In the case of plasma concentration estimation, the covariance of the process noise Q is considered to be equal to the maximum element of the variance vector of inter-patient variability (i.e. 0.318 according to the Eleveld model). The measurement error of the sensor cannot be precisely defined yet due to the immaturity of the technologies. However, it can be adjusted since it depends on the quality of the calibration. Currently we can assume the covariance of the process noise R is equal to the variance of the proportional component of intra-patient variability, that is 0.22. Such value was obtained by running Eleveld's model on the original dataset removing the inter-study variability, which represents an average value among all the studies included in Eleveld's analysis.

4.2. Closing the loop with Kalman filter

The modified version of the algorithm still relies on the basic assumption of superposition defined in Section 4. However, the amount of drug at effect site in the classical algorithm is based on the population PK model only. In our modification of the algorithm we use the concentrations of the four compartments estimated with the Kalman filter using the population PK model and measurements of plasma drug concentration. The updated concentration values are then used to recompute the delivery rate using Shafer's algorithm.

5. Robustness analysis

In this section we perform a robustness analysis of the closed-loop algorithm based on Shafer's algorithm and Kalman filter, as presented in Section 4.2. To run experiments, we have chosen a set of 1000 36 year old, 170 cm tall, weighting 70 kg individual female patients, with inter-patient variability defined by Eleveld et al. These virtual individuals with identical demographic characteristics would therefore differ in their PK parameters and, consequently, in their CT profiles. The target effect site concentration is set to 6 mg/L. Let us first assumed that every second we receive a new measurement, estimate the plasma and effect site CT profiles, and then use them to compute the delivery rate for the next step. We have computed the effect site CT profiles for all individuals if they were administered the drug with the delivery rate computed based on population parameters $U_{average}$ and with $U_{individual}$ computed using our closed-loop algorithm.

On Fig. 3 we show an example of the algorithm performance for one individual in the chosen set. We simulate both the effect site CT curves for a female with parameters calculated using Eleveld's model without inter-patient variability (that we further call an average patient), and an identical female with randomly selected parameters distributed with the inter-patient variability defined in Eleveld's model (an individual). Small crosses (Z_1 with $U_{individual}$) represent the set of measurements that were used to estimate the individual plasma CT profile marked with the dashed noisy line that approximates the target effect site concentration (X_{brn} with

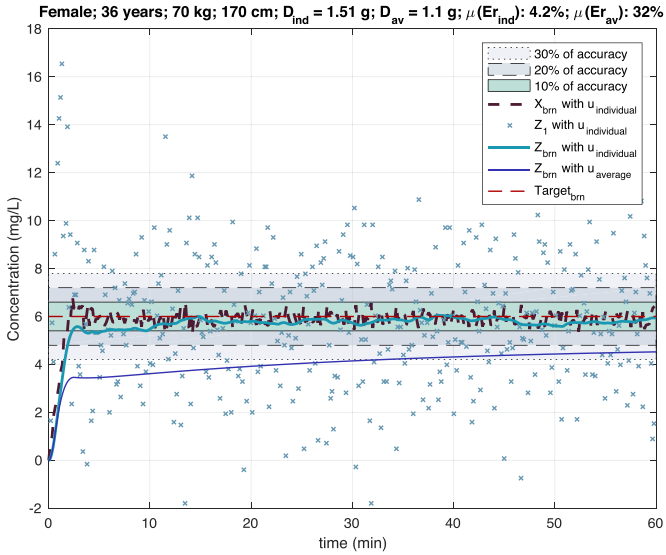

Fig. 3. An individual with faster metabolic rate than average.

$U_{individual}$). The solid line a bit below the target concentration is the CT profile of our simulated individual (Z_{brn} with $U_{individual}$) if she was administered the drug with the delivery rate $U_{individual}$. While the solid line very much below the target is the CT profile of our simulated individual (Z_{brn} with $U_{average}$) if she was administered the drug with the delivery rate $U_{average}$.

As we can see, the drug elimination rate of the selected individual is much higher than the average patient, resulting in an important underdosing with probable awakening during the operation had she been administered the *propofol* with $U_{average}$. The average absolute offset per second (Er) between the concentration Z_{brn} with $U_{average}$ and the target ($\mu(Er_{av})$) is computed in percentage after 2 min of injection, when the estimated effect site and plasma concentrations reach the equilibrium, which is called the steady state. It is equal to 32% for this particular case. A similar offset is computed for Z_{brn} if the chosen individual was administered with individualized delivery rate $U_{individual}$. This offset error is marked as $\mu(Er_{ind})$ and is equal 4.2%. We see that $\mu(Er_{ind})$ is noticeable, however, when compared to $\mu(Er_{av})$, it is more than 7 times smaller. We can also notice a small bias of the actual effect site CT profile (Z_{brn} with $U_{individual}$) towards the average population model (Z_{brn} with $U_{average}$) used for prediction, which results in a slightly larger offset from the target in the beginning that reduces towards to the end, when the population model also approaches the target. This is due to the fact that the estimation with the Kalman filter are based on both the measurements and the population model.

The total amount of drug delivered to a patient, the dose (D), is computed by integrating the deliver rate. We can see that the total dose for an individual (D_{ind}) is 1.51 g, which is 37% greater than the one computed for an average patient ($D_{av} = 1.1$), consequence of the fact that our individual eliminates the drug much faster. For patients with slower elimination rate the situation is the opposite, i.e. Z_{brn} with $U_{average}$ will be placed above the target, average offsets will be positive and D_{ind} will be smaller than the D_{av}.

Both cases described above represent the situations when the CT profile of an individual is either always above or below the average one, and in some sense follows the average shape but has an offset. However, it is possible that the CT profile is quite different and may even cross the CT profile of an average patient as on Fig. 4. In these cases the Kalman filter becomes less stable close to the moment when the CT of the individual crosses the average one and may result in a small overshoot. However, it regains stability after a small period of time.

It should be noted that the measurement noise is very big, having the proportional and additive error terms of Eq. (10), and is equal to 0.22 and 0.00015 respectively. Sometimes the measurements can provide negative values, which is unrealistic. However, we keep these values to preserve the normal distribution of the simulated measurement noise.

Similar to the analysis performed in Section 3.3, we computed the 95% prediction interval ($PI95\%$) for these 1000 virtual individuals being administered with $U_{average}$ (the larger area on Fig. 6) and $U_{individual}$ (a more narrow area on Fig. 6). The $PI95\%$ shows the every second variability among 95% effect site concentration values of these 1000 virtual patients, while 2.5% of concentration values above and 2.5% below this interval are not included. The median CT profiles of these individual administered with $U_{average}$ and $U_{individual}$ are depicted with a solid and a dashed lines, respectively.

The mean absolute difference $\mu(Er_{av})$ among all the simulated individuals administered with *propofol* at $U_{average}$ delivery rate was found to be 17%, while the $\mu(Er_{ind})$ was 3%. The maximal values out of these offsets among all individuals were 92% and 58%, respectively. This means that on average individualization of the delivery rate using our closed-loop algorithm will reduce the target offset error almost by a factor 6, and in the worst-case, by almost a factor 2.

5.1. Period of measurement

The delivery rates presented above where computed assuming that the measurements were arriving every second, i.e. every time

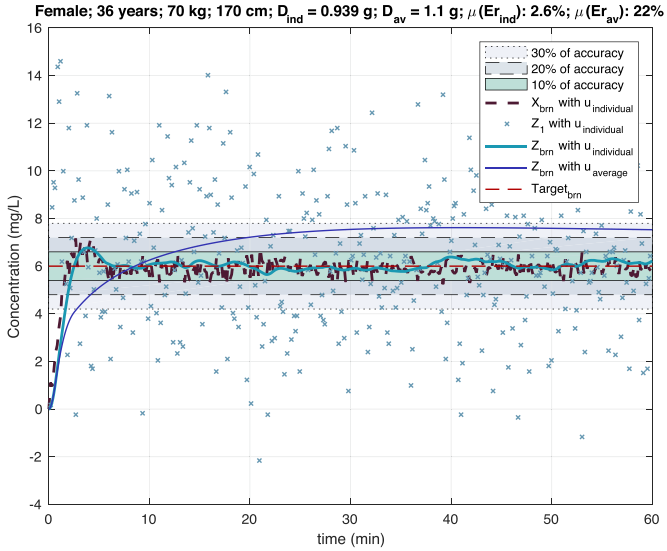

Fig. 4. An individual with an alternative CT profile.

the rate was recalculated. However, new measurements may arrive with a larger period, e.g. every 30 *seconds*. Fig. 5 depicts an example of algorithm performance in this case. As it is well expected, we see that the effect site CT profile Z_{brn} is more noisy than on Figs. 3 and 4.

Figs. 6 to 9 depict the 95% prediction intervals ($PI95\%$) and the median CT profiles for 1000 virtual individuals being administered with $U_{average}$ and $U_{individual}$ for measurement periods of 1, 5, 10 and 30*sec*, respectively.

We can see that the real effect site CT profiles of selected individuals administered with $U_{individual}$ computed using the measurements obtained every second has a tendency to stay within a 10% accuracy area around the target, the lower and upper bounds of which are computed as $Target_{brn} \pm 0.1{*}Target_{brn}$. With the increase of measurement period, the $PI95\%$ interval becomes larger. With 30*sec* measurement period, CT profiles may already exit the 20% accuracy area, however, still remain within the 30% accuracy area. The 30% accuracy area is not indicated on the graphs, however, it can be easily imagined as an area between 4.2 and 7.8 mg/L.

According to the anaesthesiologist's opinion, we can consider the oscillations within the 10% area of accuracy as ideal, while within the 30% area of accuracy as acceptable. Therefore, taking into account the large measurement noise chosen to stress our algorithm, with 30*sec* measurement period the robustness of our algorithm is still acceptable.

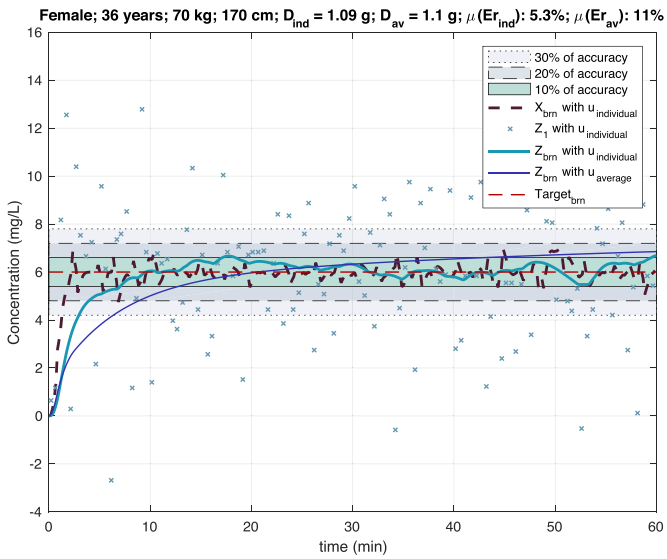

Fig. 5. Computation with measurement period of 30*sec*.

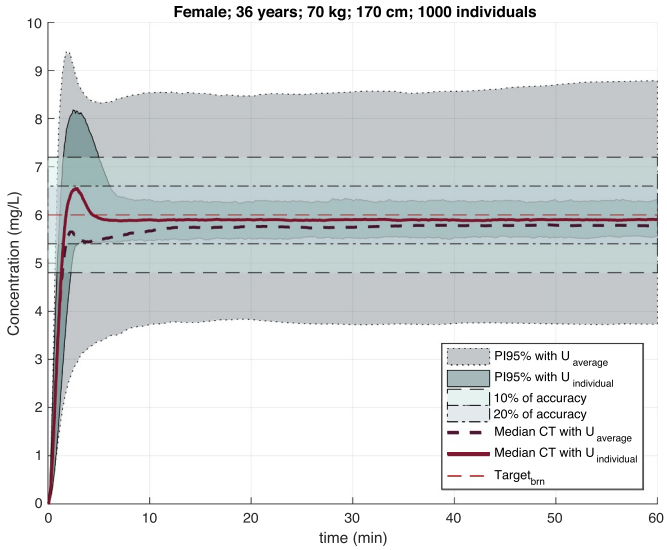

Fig. 6. *PI*95% and median CT profiles, period of 1*sec*.

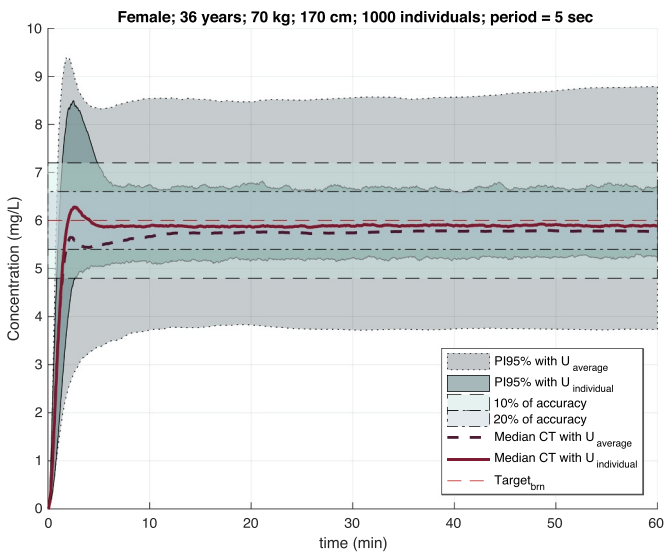

Fig. 7. *PI*95% and median CT profiles, period of 5*sec*.

5.2. Robustness with measurement period and delay

In the previous section, we assumed that there is no delay in the measurements, so that the time stamp of all measurements is equal to the time when it is provided. In reality, the measurement cycle of a sensor quantifying *propofol* concentration would include: (1) the time it takes to withdraw the sample of blood, (2) the measurement delay, e.g. due to electrochemical reaction, (3) the time it take to communicate the measurement to the system running the rate adjustment algorithm, and (4) the time to clean the sensor.

We assume that communication time is in the range of milliseconds, taking into account the performance of modern computing systems, and thus can be neglected, while the blood sample withdrawal can take about 5*sec*. Such sensors must be cleaned from the blood cells that are sticking to the sensor surface causing biofouling, that may affect the accuracy of measurements and must be performed after several measurements. According to recent developments in the domain of electrochemical sensing technologies for *propofol*, the minimum time for sensor cleaning is 10*sec*, while the measurement delay depends on the chosen methodology to perform the measurement. The shortest delay is expected to be 22 s (Stradolini et al., 2017). It means that once we receive a measurement, it would be 22 (for electrochemical reaction) plus 5 (for blood sample withdrawal) seconds old. All this would increase the period of possible measurements if we imagine that the measurements are performed by only one sensor.

However, the measurements period can be reduced by using an array of identical sensors performing *propofol* measurements in parallel. The measurement period then depends on the number of sensors in the array. For example, let us imagine that one sensor has

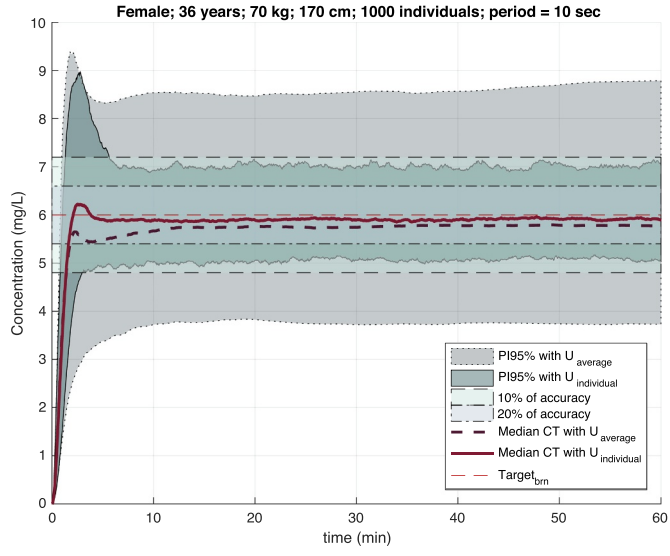

Fig. 8. *PI*95% and median CT profiles, period of 10*sec*.

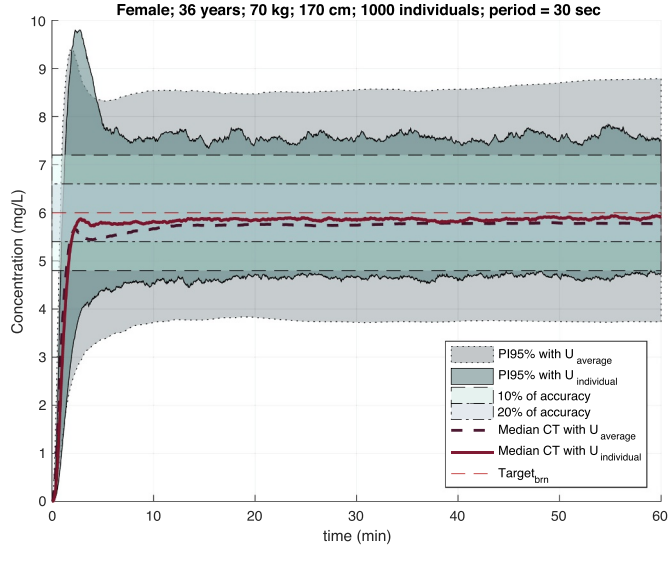

Fig. 9. *PI*95% and median CT profiles, period 30*sec*.

a measurement delay of 25*sec* and every 4 measurements it requires an automatic cleaning procedure that takes another 30*sec*, while blood sample withdrawal takes about 5*sec*. This way, to build an array of sensors able to provide one measurement every 15*sec*, we would need at least 5 sensors. The delay of each measurement will remain unchanged, since the electrochemical reaction that is needed to produce one measurement takes a certain amount of time.

It should be noted that the first informative measurement will arrive at the earliest at a moment greater than the measurement delay. The measurement received at time equal to the measurement delay will correspond to the moment of 0*sec*, and must consequently be equal to 0. However, taking into account that the *propofol* delivery rate during the first 30*sec* is usually set to its maximum value and that the time during which the concentration at the effect site reaches its target is greater than 1 *min*, it is less critical to receive the first measurement with a greater latency (i.e. >30*sec*) than all the consequent ones.

An example of the algorithm performance for a realistic scenario with an array of sensors with a measurement delay of 30*sec* and measurement period of 15*sec* is shown on Fig. 10. With such settings we start to see notable oscillations of the individual real effect site CT profile. This is due to both a large period and the delay, since the dynamics of the system is very fast and by the time the measurement is received, the system state to which we apply the estimated values has already changed. It should be noted, however, that the average per second offset $\mu(Er_{ind})$ is still smaller than the $\mu(Er_{av})$, while the real effect site CT profile has a tendency to stay within the 10% accuracy area.

Fig. 11 depicts the 95% prediction intervals (*PI*95%) and the median CT profiles for the case presented above. It shows that the

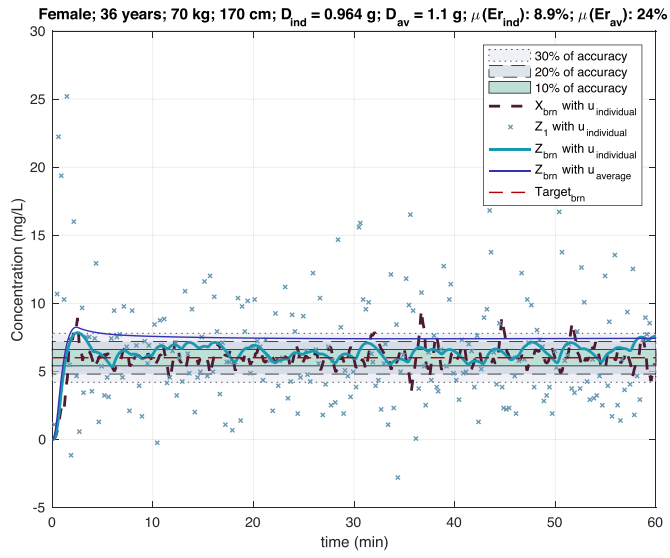

Fig. 10. Measurement period of 15*sec*, delay of 30*sec*, individual example.

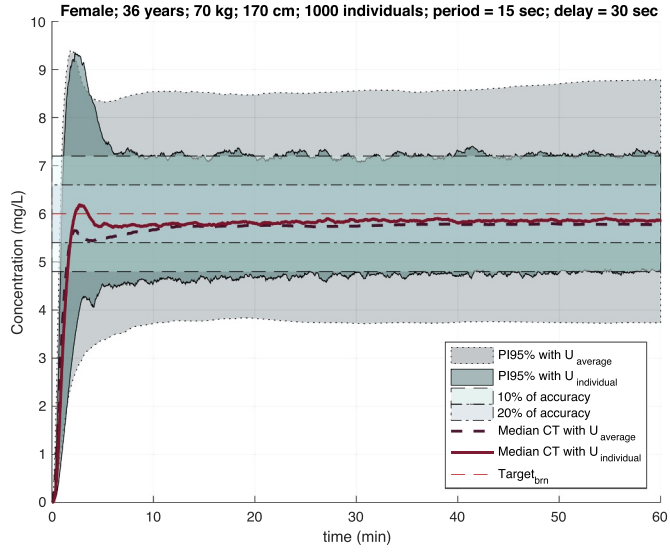

Fig. 11. Measurement period of 15*sec*, delay of 30*sec*, prediction interval PI95%.

prediction interval for individuals being administered with $U_{individual}$ is at the limit of the 20% accuracy area.

It is worth noting that the noise used in our simulations is extremely high. In reality, the measurement variability is expected to be much lower, contributing to a better robustness of the algorithm. Therefore, we have repeated the experiments for a measurement period of 30*sec* with no delay and a measurement period of 15*sec* with delay of 30*sec* with noise level divided by two, setting the variance of the proportional terms (*prop*) of intra-patient variability to 0.055. We compare the algorithm performance for two experiments with values of *prop* = {0.22, 0.055} on Figs. 12 and 13, respectively.

First of all, while looking at these figures, one can notice that the measurement delay has a smaller effect on algorithm robustness than the period, and played a role only before the steady-state was reached. Second, as expected, the variability interval has decreased such that in both cases the effect site CT profiles for 95% of virtual individuals administered with $U_{individual}$ stayed within the 20% accuracy area resulting in effect site CT profile variability being twice smaller than if they were administered with $U_{average}$.

6. Conclusion

Once sensors are available for regular measurement of circulating *propofol* concentrations, they will make it possible to set up closed-loop systems for controlling TCI-based anesthesia. The results presented in Section 5 show the algorithm robustness for different measurement periods and delays with the very high Eleveld's model predicted variability of plasma measurements. In

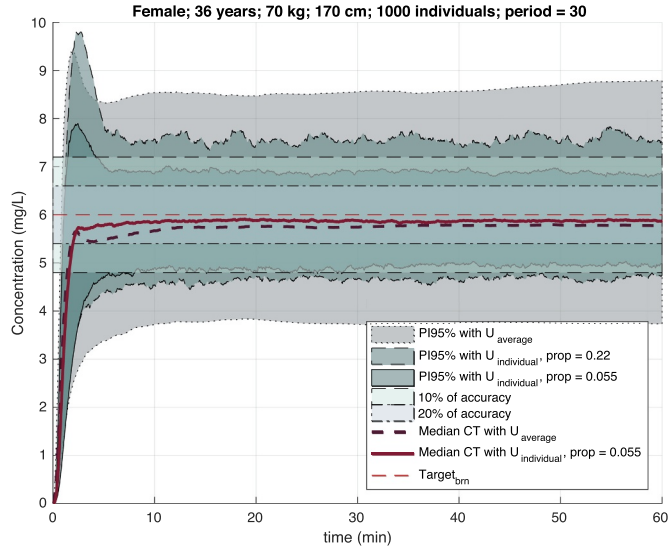

Fig. 12. PI95% for period of 30*sec* and no delay, *prop* = {0.22, 0.055}.

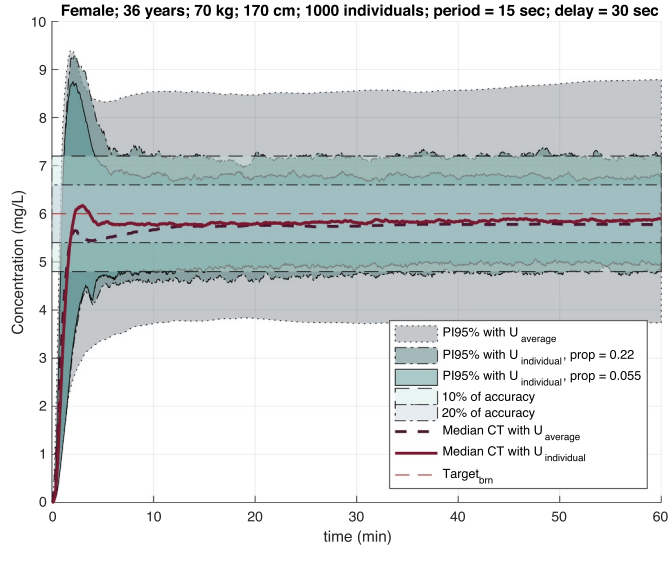

Fig. 13. PI95% for period of 15*sec* and delay of 30*sec*, *prop* = {0.22, 0.055}.

reality, this variability is expected to be much lower, contributing to better robustness of the algorithm. Nevertheless, our work shows that the use of the presented algorithm ensures both more precise and safer plasma and effect size exposures than currently achieved with open-loop TCI pumps. The approach proposed here, based on the Kalman filter, represents a step towards the development of this type of devices, which will undoubtedly become part of daily patient management in operating theaters of the future, leading anesthesia one step closer to the ideal of Precision Medicine.

Conflict of interest statement

None.

References

Abbas, H., Jiang, Z., Jang, K., Liang, J., Dixit, S., & Mangharam, R. (2016). Computer aided clinical trials for implantable cardiac devices, SES 2016: Symposium F-2: Modeling, Design and Safety Analysis in Physiological Closed-Loop Systems.

Bailey, J. M., & Haddad, W. M. (2005). Drug dosing control in clinical pharmacology. *IEEE Control Systems, 25*(2), 35–51. https://doi.org/10.1109/MCS.2005.1411383.

Burchum, J.R., Rosenthal, L.D., Jones, B.O., Neumiller, J.J., & Lehne, R.A. Lehne's pharmacology for nursing care.

De Smet, T., Struys, M. M. R. F., Greenwald, S., Mortier, E. P., & Shafer, S. L. (2007). Estimation of optimal modeling weights for a bayesian-based closed-loop system for propofol administration using the bispectral index as a controlled variable: A simulation study. *Anesthesia and Analgesia, 105*(6), 1629–1638. https://doi.org/10.1213/01.ane.0000287269.06170.0f.

Dumont, G. A., & Ansermino, J. M. (2013). Closed-loop control of anesthesia: A primer for anesthesiologists. *Anesthesia and Analgesia, 117*(5), 1130–1138. https://doi.org/10.1213/ANE.0b013e3182973687.

Eleveld, D., Proost, J., Cortinez, L., Absalom, A., & Struys, M. (2014). A general purpose pharmacokinetic model for propofol. *Anesthesia and Analgesia, 118*(6), 1221–1237.

Gentilini, A., Frei, C. W., Glattfedler, A. H., Morari, M., Sieber, T. J., Wymann, R., Schnider, T. W., & Zbinden, A. M. (2001). Multitasked closed-loop control in anesthesia. *IEEE Engineering in Medicine and Biology Magazine, 20*(1), 39–53. https://doi.org/10.1109/51.897827.

Kivlehan, F., Garay, F., Guo, J., Chaum, E., & Lindner, E. (2012). Toward feedback-controlled anesthesia: Voltammetric measurement of propofol (2,6-diisopropylphenol) in serum-like electrolyte solutions. *Analytica Chimica Acta, 84*(18), 7670–7676.

Kotsovolis, G., & Komninos, G. (2009). Awareness during anesthesia: How sure can we be at the patient is sleeping indeed? *Hippokratia, 2*(13), 83–89.

Langmaier, J., Garay, F., Kivlehan, F., Chaum, E., & Lindner, E. (2011). Electrochemical quantification of 2,6-diisopropylphenol (propofol). *Analytica Chimica Acta, 704*(1–2), 63–67.

Liu, N., Chazot, T., Hamada, S., Landais, A., Boichut, N., Dussaussoy, C., Trillat, B., Beydon, L., Samain, E., Sessler, D. I., & Fischler, M. (2011). Closed-loop coadministration of propofol and remifentanil guided by bispectral index: A randomized multicenter study. *Anesthesia and Analgesia, 12*(3), 546–557.

Marsh, B., White, M., Morton, N., & Kenny, G. N. C. (1991). Pharmacokinetic model driven infusion of propofol in children. *British Journal of Anaesthesia, 67*(1), 41–48. https://doi.org/10.1093/bja/67.1.41〈http://bja.oxfordjournals.org/content/67/1/41.abstract〉.

Orchestra Base Primea TCI system: 〈https://www.fresenius-kabi.de/orchestrabaseprimea.htm〉.

Orser, B. A., Mazer, C. D., & Baker, A. J. (2008). Awareness during anaesthesia. *Canadian Medical Association Journal, 2*(178), 185–188.

Rampil, I. R. (1998). A primer for EEG signal processing in anesthesia. *Anesthesiology, 89*(4), 980–1002.

Schnider, T., Minto, C., Gambus, P. L., Andresen, C., Goodale, D., Shafer, S., & Youngs, E. (1998). The influence of method of administration and covariates on the pharmacokinetics of propofol in adult volunteers. *Anesthesiology, 88*(5), 1170–1182.

Schnider, T. W., Minto, C. F., Shafer, S. L., Gambus, P. L., Andresen, C., Goodale, D. B., & Youngs, E. J. (1999). The influence of age on propofol pharmacodynamics. *Anesthesiology, 90*(6), 1502–1516.

Shafer, S. L., & Gregg, K. M. (1992). Algorithms to rapidly achieve and maintain stable drug concentrations at the site of drug effect with a computer-controlled infusion pump. *Journal of Pharmacokinetics and Biopharmaceutics, 20*(2), 147–169.

Silva, M. M. d., Mendonça, T., & Wigren, T. (2010). Online nonlinear identification of the effect of drugs in anaesthesia using a minimal parameterization and bis measurements. In Proceedings of the 2010 American Control Conference, pp. 4379–4384. 〈http://dx.doi.org/10.1109/ACC.2010.5530791〉.

Silva, M. M., Medvedev, A., Wigren, T., & Mendonça, T. (2015). Modeling the effect of intravenous anesthetics: A path toward individualization. *IEEE Design Test, 32*(5), 17–26. https://doi.org/10.1109/MDAT.2015.2452904.

Silva, M. M. (2014). *Nonlinear modeling and feedback control of drug delivery in anesthesia (Doctoral thesis)*.

Simalatsar, A., Guidi, M., & Buclin, T. (2016). Cascaded PID controller for anaesthesia delivery. In Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 533 – 536. 〈http://dx.doi.org/10.1109/EMBC.2016.7590757〉.

Stradolini, F., Tuoheti, A., Ros, P. M., Demarchi, D., & Carrara, S. (2017). Raspberry pi based system for portable and simultaneous monitoring of anesthetics and therapeutic compounds. In 2017 New Generation of CAS (NGCAS), pp. 101–104. 〈http://dx.doi.org/10.1109/NGCAS.2017.67〉.

Zhusubaliyev, Z. T., Medvedev, A., & Silva, M. M. (2013). Bifurcation analysis for PID-controller tuning based on a minimal neuromuscular blockade model in closed-loop anesthesia, In Proceedings of the 52nd IEEE Conference on Decision and Control, pp. 115–120. 〈http://dx.doi.org/10.1109/CDC.2013.6759868〉.