



## Article

# Mathematical Modeling of RBC Count Dynamics after Blood Loss

Manuel Tetschke <sup>1,\*†</sup>, Patrick Lilienthal <sup>1,†</sup>, Torben Pottgiesser <sup>2</sup>, Thomas Fischer <sup>3</sup>,  
Enrico Schalk <sup>3</sup>  and Sebastian Sager <sup>1</sup> 

<sup>1</sup> Institute for Mathematical Optimization, Otto-von-Guericke-University Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany; patrick-marcel.lilienthal@ovgu.de (P.L.); sager@ovgu.de (S.S.)

<sup>2</sup> Department of Cardiology and Angiology I, Heart Center Freiburg University, Faculty of Medicine, University of Freiburg, 79106 Freiburg, Germany; torben.pottgiesser@uniklinik-freiburg.de

<sup>3</sup> Department of Hematology and Oncology, Medical Center, Otto-von-Guericke-University Magdeburg, 39120 Magdeburg, Germany; thomas.fischer@med.ovgu.de (T.F.); Enrico.Schalk@med.ovgu.de (E.S.)

\* Correspondence: tetschke@ovgu.de; Tel.: +49-391-67-51449

† These authors contributed equally to this work.

Received: 13 June 2018; Accepted: 30 August 2018; Published: 5 September 2018



**Abstract:** The regeneration of red blood cells (RBCs) after blood loss is an individual complex process. We present a novel simple compartment model which is able to capture the most important features and can be personalized using parameter estimation. We compare predictions of the proposed and personalized model to a more sophisticated state-of-the-art model for erythropoiesis, and to clinical data from healthy subjects. We discuss the choice of model parameters with respect to identifiability. We give an outlook on how extensions of this novel mathematical model could have an important impact for personalized clinical decision support in the case of polycythemia vera (PV). PV is a slow-growing type of blood cancer, where especially the production of RBCs is increased. The principal treatment targeting the symptoms of PV is bloodletting (phlebotomy), at regular intervals that are based on personal experiences of the physicians. Model-based decision support might help to identify optimal and individualized phlebotomy schedules.

**Keywords:** erythropoiesis; phlebotomy; modeling; parameter estimation; numerical simulation

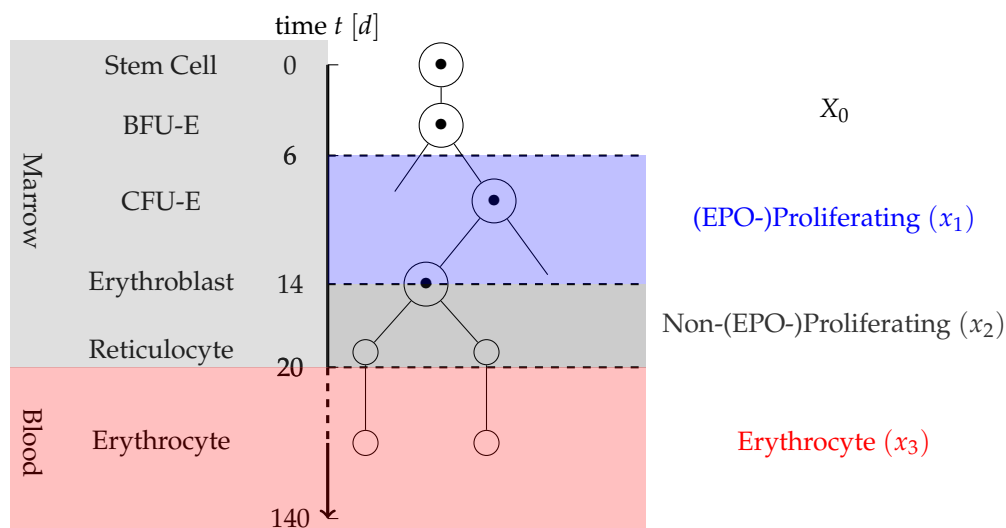
## 1. Introduction

We present a novel mathematical compartment model for erythropoiesis after blood loss in healthy subjects. We see it as a first step towards mathematical models that can be used in clinical practice to predict red blood cell (RBC) counts, for example, for patients suffering from the disease polycythemia vera. After an introduction to erythropoiesis, we provide some motivating details of this disease below, and discuss possible future extensions and usage of the model for optimized phlebotomy schedules in the outlook. However, the main contribution of this paper is a data-challenged mathematical model for erythropoiesis after blood loss in healthy subjects, which is interesting and challenging in its own right.

### 1.1. Overview of Erythropoiesis

RBCs are biconcave discoid cells responsible for the supply of tissue with oxygen from the lungs and the transport of carbon dioxide from tissue back to the lungs. The oxygen transport is realized by the contained protein complex hemoglobin. Healthy adult humans have a total of  $2$  to  $3 \times 10^{13}$  erythrocytes at any given time with men having about  $5$  to  $6$  million erythrocytes per microliter blood and women having about  $4$  to  $5$  million erythrocytes per microliter blood.

The process of production of RBCs beginning at stem cell level is called erythropoiesis. Erythropoiesis takes place in the bone marrow, where multi-potent stem cells are transforming through several stages of erythroid progenitor cells, namely Blast Forming Unit-Hematopoietic (BFU-E) and Colony Forming Unit-Hematopoietic (CFU-E), and several stages of erythroblasts to bone marrow reticulocytes. Afterwards, they are discharged into the blood stream as blood reticulocytes, where they are quickly maturing into fully grown erythrocytes. This process takes about 20 days and includes the removal of nuclei, organelles and mitochondria of the cell to provide more space for hemoglobin. A simplified schematic view of this process is given in Figure 1.



**Figure 1.** Illustration of erythropoiesis with erythropoietin (EPO) dependency of the certain states. The stages (over the age of cell in days) are the stem cells, the precursor cells blast forming unit-hematopoietic (BFU-E), colony forming unit-hematopoietic (CFU-E), several stages of erythroblasts and reticulocytes in the marrow and blood reticulocytes and erythrocytes in the blood stream. Also included is a partition of the erythropoiesis scheme into three compartments, which reflect the dependency of the cells with respect to EPO. This partition will be used for the modeling process.  $X_0$  describes the constant number of cells entering the red blood cell line. The more mature cells in the marrow are divided into proliferating and non-proliferating cells with respect to EPO. In the blood stream, again blood reticulocytes and erythrocytes do reside.

The mean life expectancy of RBCs in the blood stream is about 120 days, after which they are destroyed by phagocytes. This process is necessary as RBCs are not able to further divide or to regenerate themselves due to their missing nuclei [1].

Erythropoiesis is a dynamic process with a tight regulation. Its aim is to control the number of RBCs such that the oxygen supply to the tissue is matched with the current oxygen demand. The rate of erythropoiesis has to be adjusted in reaction to varying environmental conditions like the transition from low to high altitudes or physical exercise. The production also has to compensate for the loss of RBCs due to internal and external bleeding and cellular senescence. In healthy humans, this control is realized by the hormone erythropoietin (EPO). Due to EPO, erythropoiesis has a negative feedback mechanism. The concentration of the hormone is inversely related to the hemoglobin concentration. A high EPO concentration increases the production rate of RBCs in the bone marrow with an especially high amplification of the CFU-E production rate. As additional influence a sufficient iron concentration in the blood stream is necessary for successful erythropoiesis. A more detailed overview of erythropoiesis can be found in [2].

## 1.2. History of Erythropoiesis Modeling and Comparison with Our Approach

Over the last several decades, there have been many different approaches to model the process of erythropoiesis, of which we can only mention a short subjective selection. One early contribution to this field is a model based on delay differential equations (DDE) by Mackey et al. 1977 [3], where a local stability analysis of steady states was performed to explain chaotic behavior using large delays. Another important foundation was laid early by Loeffler et. al in 1989, where an ordinary differential equation (ODE) compartment model of erythropoiesis was developed using experimental results from mice studies [4]. Those were complemented by Belair et al. in 1995 [5] and Mahaffy et al. in 1999 [6], where in both cases an age-structured partial differential equation (PDE) system was used to model human erythropoiesis following a phlebotomy. Later similar models by Ackley et al. (2008) [7] and Fuerstinger et al. (2013) [8] were developed, where more physiological properties were displayed and a more detailed numerical evaluation of these models was done. In addition, discrete time models of RBC production have been developed and early references can be found in Edelstein-Keshet's book [9]. The structure of all these models was often influenced by the underlying and motivating context. Fonseca and Voit 2015 [10] compared several modeling strategies for erythropoiesis and malaria. Other applications include the regeneration of RBCs after a phlebotomy [5–8] or induced anemia in mice [4]. Similar mathematical models arise in the context of white blood cell dynamics [11–13].

As mentioned before, delays play an important role. Erythropoiesis is a physiological process with delayed mechanics like the duration of production in the bone marrow up to a mature erythrocyte of about 20 days or the response by EPO itself. Therefore, many authors considered PDEs with delays [5–8,14] or ODEs with delays [3,10]. Another common strategy has been the usage of ODE compartment models to approximate the characteristic delays associated with maturation [4,11,15,16]. The transition to the next compartment is modeled by a transition probability for each cell which reflects the average maturation time in the compartment.

While more complex and realistic mathematical models have the advantage with respect to biological interpretations and a deep understanding of what is going on, simpler mathematical models that try to capture only the dominant physiological dynamics often have benefits when it comes to identifiability of the model parameters. Currently, the available clinical data is limited to small numbers of RBC count measurements, and no information on the distribution of maturation or EPO concentrations. In addition, reliable values for hematocrit (*Hct*) or total blood volume are difficult to obtain due to imprecise estimation formulas and plasma volume deviations [17–19]. While this may change in the future with better measurement technology, data acquisition for erythropoiesis still implies considerable inconveniences for the patient.

Thus, either large populations or regularizations using values from the literature are necessary for larger numbers of model parameters, initial values, or initial trajectories. Motivated by the current limited amount of available clinical data, and by our long-term goals to use the mathematical models in clinical practice for rough but personalized predictions, we developed a compartment model. Fonseca and Voit 2015 [10] demonstrated that the probabilistic nature of the ODE system may lead to large errors compared to a DDE model. This can be the case if equation properties like cell production rates vary strongly. In our context, we assume production rates to be bounded. The limited amount of oscillations in the clinical data seems to justify this assumption. However, if the available amount of data changes drastically due to progress in clinical practice, the incorporation of delays by DDEs might improve the results presented in this paper. Note also that an analytical or numerical comparison of mathematical models on different levels of complexity may shed further insight, and complement one another in practice.

The novelty of this paper is given by a compartment model that captures the most important medical knowledge, contains few model parameters and has been evaluated using (comparatively) high quality clinical data. We provide population parameters from a nonlinear mixed-effect approach and compare it to point estimates. To our knowledge, there are no similar publications.

The proposed model differs from the other cited compartment models [4,11,15,16] as follows: [15] use only one compartment for cells in the marrow. As the dependency on EPO differs between different cell stages (compare Section 1.2), another compartment is needed to reflect this. Ref. [4] use more compartments and also use a direct dependency on EPO, which cannot be measured in clinical practice due to high fluctuations in a very short amount of time. The model of Friberg et al. [11] differs mainly in the way the dynamics can be influenced by a therapy, as their goal is to use the model for treatment of acute myeloid leukemia (AML). They include a pharmacokinetic/pharmacodynamic (PK/PD) model which influences the first compartment, while we need to influence the RBC dynamics directly. The same holds true for the work of Noble et al. [16], with the difference that they regard the disease acute lymphoblastic leukemia (ALL). Additionally, they use strong nonlinearities in their feedback and a direct dependency on the Mean corpuscular volume of the RBCs.

### 1.3. Motivating Practical Application: Polycythemia Vera

Polycythemia vera (PV), also called primary polycythemia, is a chronic myeloproliferative neoplasm, which means that the production of blood cells is increased to a pathological level. Most prominently, RBCs, also called erythrocytes, are affected which causes the main symptoms of the patients: if the ratio of erythrocytes to the whole blood volume, which in medical terms is called hematocrit (*Hct*), exceeds a certain threshold, the blood cells may clot and thus can cause thromboembolic events. This can lead for example to a stroke, myocardial infarction, vein/arterial thrombosis or pulmonary embolism. If untreated, the mean life expectancy of patients suffering from PV is only about 18 months [20].

Other symptoms of the disease are not fatal, but can strongly reduce the quality of life of the patient. Most prominently, the aquagenic pruritus, a severe itching that patients experience when they have contact to water, can be observed in more than 60% of the cases [21]. As described in [22] in the long run, PV patients have a higher risk to develop other types of blood cancer like AML or myelofibrosis. This risk is associated with the age of the patient and the duration of the disease, respectively.

The basic therapy of PV is blood-letting, also called phlebotomy, of about 500 mL blood on a regular basis. As the body is compensating for the blood loss through blood plasma nearly instantly but needs several weeks to produce new RBCs, the *Hct* can temporarily be reduced using this treatment. In severe cases, this procedure is not sufficient and there is also the need of cytoreductive therapy (or a combination of both). Until now it has not been known how to compute the frequencies and volumes of the phlebotomies to give an optimal outcome for the patient [22]. We will give a speculative outlook to possible extensions of the model and a usage to calculate personalized optimal schedules in Section 5.

## 2. Model

In this chapter, we propose an application oriented model for erythropoiesis of healthy subjects. The ODE system captures only the most essential features concerning erythropoiesis (delays intrinsically via its compartment structure) and contains few identifiable model parameters. Note that such a model may not be able to reproduce or predict exact biological mechanisms. We still prefer such a model of *reduced complexity* given our motivating long-term goal to personalize the mathematical model in clinical contexts to available data. After formulating the most important assumptions on the subjects and the physiological process, we formulate and analyze our novel mathematical model.

### 2.1. Assumptions on the Subject

First assumptions concerning the desired application take into account the subjects, which will lead to a decrease in model complexity:

1. **Healthy adult subject:** The aim of the proposed model is to describe the RBC dynamics of healthy adult subjects. Healthy in this case especially means that the subject has no disturbed erythropoiesis as for example in the case of PV.
2. **Existence of a steady state:** It is assumed that environmental conditions that affect the blood do not dramatically change. If this is the case, a mean value for the erythrocyte count should exist in healthy individuals, which is valid over a longer time period. This will be the *Base* value of the system and the steady state for the erythrocyte count in the model. Using this assumption especially means that the model will not be able to capture large environmental effects to the blood.
3. **Sufficient iron concentration:** The iron concentration in considered individuals is assumed to be sufficient, such that erythropoiesis is not affected by iron shortage. Therefore, iron influence on erythropoiesis is not included in the model.
4. **Reasonable blood loss:** It is assumed that the blood loss in the considered subjects does not exceed a standard blood donation by much. Therefore, emergency reactions of the body in case of severe anemia (for example the release of stress reticulocyte, compare [2]) should not arise and are therefore not included in this model.

## 2.2. Assumptions on the Physiological Process

The process of erythropoiesis consists of a multitude of mechanisms describing its dynamics. As in the data we are using the amount of available information is very limited, some simplifying assumptions are made on the physiological process. This again will contribute to a decrease in model complexity. Some of these simplifications may result in errors, which cannot be avoided if only the very sparse data from clinical standard routines is taken into account.

5. **Average cell:** A compartment modeling approach is used to describe the system dynamics. Therefore, all cells in one compartment share the same properties given by the mean values of the corresponding properties. On the one hand, cell individual information like cell age is lost, which leads to errors especially in early phases where proliferation is very dependent on the state of the cell. On the other hand, often only summarized properties like for example amount of cells or overall amount of hemoglobin is interesting for our application. In addition, influences in early cell stages are often not measured in clinical practice, such that an error in early cell stages is not preventable. The assumption that given mean values reflect summarized properties is justified, as there is a very large number of cells in each compartment and the law of large numbers can be applied.
6. **Partition of progenitor cells:** Measurements only include information on the mature RBC and not on previous cell stages. Therefore, in the application oriented model, a too detailed differentiation of early stem cells should be prevented. As EPO is the main factor for compensation of a blood loss, a partition into two cell types is proposed: cells that are strongly proliferating in dependence on EPO and cells that are not or only slightly affected by EPO. Cells strongly affected by EPO here include CFU-E cells and early erythroblasts. Not affected or less affected by EPO are BFU-E, later erythroblasts and reticulocytes. As BFU-E are not affected by the dynamics, they are modeled by a constant inflow term  $X_0$ , whereas the EPO-dependent and Non-EPO-dependent cells are each described by one compartment. The partition of erythropoiesis can be examined in Figure 1.
7. **Instant blood plasma regeneration:** After a blood loss, both solid and liquid blood components are lost. It is known that blood plasma regenerates very fast after about two to four days [23]. To prevent the usage of delay components in the model, it is here assumed that the regeneration of blood plasma is instant after a reasonable blood loss. This error should be low, as the horizon of about three days is small compared to the average time after a blood donation of about 42 days.
8. **Instant feedback by EPO:** The EPO production in the human body is adapted to a changing demand for erythrocytes. This process does not instantaneously change the production rate of



erythrocytes, as it takes a while for the local EPO production in, for example, the liver to change the overall concentration of EPO. For our modeling approach, it is assumed that this process is instantaneous. This error also should be low, as the observed time scale is very large.

9. **Implicit feedback by EPO:** As stated in Section 1.1, EPO is mainly responsible for the regeneration of RBCs after changing hemoglobin concentrations for example due to blood loss. Unfortunately, EPO concentrations fluctuate strongly on a small time scale and are difficult to measure in clinical practice. Therefore, no direct information about EPO concentration in the subject is available [24]. The influence of EPO on erythropoiesis is therefore indirectly described by a negative feedback term based on the fractional blood loss. We assume the feedback function to be linear, which is certainly valid for small variations in RBC counts. This assumption may however be violated for larger amounts of blood loss and pathological cases, as discussed in Section 5.
10. **Only most necessary processes:** To reduce the model complexity, only the most essential features of erythropoiesis after blood loss are included. Therefore, processes like neocytolysis [2], which has a minor influence on young erythrocytes, is not included.
11. **No self-renewal of progenitor cells:** It is known that stem cells, especially those that are not committed to a hematopoietic lineage, have the capability of self-renewal to maintain the population of cells. The capability of self-renewal decreases with increased maturation progress [2]. There exist mathematical models that consider self-renewal to be even present in progenitor cells (see for example [25]). In contrast to that, here it is assumed that in the considered progenitor cell compartments only differentiation and no self-renewal is present.
12. **Transport rates of cells:** In case of severe anemia, cells might have an increased differentiation rate induced by high EPO concentrations to compensate for the lack of cells [2]. Due to Assumption 4, those reactions do not have to be considered here. Therefore, it is assumed that cells have a constant differentiation rate concerning EPO. Based on [2], it can be assumed that the EPO-Proliferating phase and the Non-EPO-Proliferating phase have an average duration of eight and six days, respectively. The transport rates  $k_1$  and  $k_2$  are set to the respective reciprocal values. As the average lifespan of a mature erythrocyte can be assumed to be 120 days, the death rate for the erythrocyte compartment is set to  $\alpha = \frac{1}{120}$ .
13. **Individual features of erythropoiesis:** Between individuals, there might be a strong variation in the pace of erythropoiesis. For example, a large individual often has more blood than a small individual and therefore should have a faster erythropoiesis to cover for example for fractional cell loss due to cell senescence. This is covered by two amplification variables:  $\beta$  for the overall cell maturation and  $\gamma$  for the reaction to a blood loss.

### 2.3. Model Equations

Using the above assumptions, model equations can be formulated. The compartment model consists of three ODEs: the (EPO-)Proliferating compartment ( $x_1$ ), the Non-(EPO-)Proliferating compartment ( $x_2$ ) and one compartment for the mature erythrocytes ( $x_3$ ).  $X_0$  reflects the absolute amount of cells committing to the erythroid lineage and maturing into the first erythroid progenitor cell compartment. Transition rates and death rates between the compartments are given by  $k_1$ ,  $k_2$  and  $\alpha$  and are independent of EPO. The compensation of a blood loss is amplified by a feedback term of the erythrocytes to the proliferating cells based on the fractional loss of erythrocytes. As mentioned above, additional variables  $\beta$  and  $\gamma$  are introduced, which are used for the description of subject individual features of erythropoiesis. Based on Assumption 2, each subject has a mean erythrocyte count denoted as *Base*. The resulting equations are given as

$$\begin{aligned}\dot{x}_1(t) &= \beta (X_0 - k_1 \times x_1(t)) + Fb(x_3(t)) \times x_1(t), \\ \dot{x}_2(t) &= \beta (k_1 \times x_1(t) - k_2 \times x_2(t)), \\ \dot{x}_3(t) &= \beta (k_2 \times x_2(t) - \alpha \times x_3(t)),\end{aligned}\tag{1}$$

where

$$Fb(x_3(t)) = \gamma \times \frac{Base - x_3(t)}{Base}. \quad (2)$$

Let  $\bar{x} = (\bar{x}_1, \bar{x}_2, \bar{x}_3)$  denote a steady state value of the system. Setting  $\bar{x}_3 = Base$  as the steady state of the erythrocyte count, a unique steady state of the system is given by

$$(\bar{x}_1, \bar{x}_2, \bar{x}_3) = \left( \frac{\alpha}{k_1}, \frac{\alpha}{k_2}, 1 \right) \times Base \quad (3)$$

if  $Fb(Base) = 0$  and  $X_0 = k_1 \times \bar{x}_1$  is fulfilled. Note that, without the physiological assumptions  $x_1, x_2, x_3 > 0$ ,  $\bar{x}_3 = Base > 0$  and the choice of  $Fb$  and  $X_0$ , up to three steady states are possible. The calibration of the model having only one steady state reflects the plausible tendency of the body to reach a healthy *Base* level depending on environmental factors.

Furthermore,  $Fb(x_3(t))$  is a monotone decreasing function in  $x_3$  which is independent of the transport rates  $k_1, k_2$  and  $\alpha$ . This choice of  $Fb$  is a negative feedback function as

$$\frac{d}{dx_3} Fb(x_3) = -\frac{\gamma}{Base} < 0 \quad (4)$$

for  $\gamma > 0$ , which is in accordance with Assumption 9.

The compartment  $x_2$  serves as transition compartment between  $x_1$  and  $x_3$ , where the main dynamics occur. The general model described in [11] for similar applications takes into account a free number of transition compartments. Numerical computations in our case showed that there is nearly no effect on our results in using additional transition compartments. This fact and our goal of finding a model as lean as possible motivate to choose only one transition compartment resulting in three ODEs for the whole system. A full list of all model variables and parameters can be found in Table 1.

**Table 1.** List of model parameters and states with explanation and example steady state values and corresponding units.

State	Explanation	Value	Unit
$x_1$	Erythroid progenitor cells highly proliferating with respect to erythropoietin (EPO)	59.03	[1]
$x_2$	Erythroid progenitor cells not proliferating with respect to EPO	44.27	[1]
$x_3$	Mature erythrocytes and blood reticulocytes	885.42	[g]
Parameter			
$Base$	Mean erythrocyte count in steady state	885.42	[g]
$X_0$	Committed stem cells transitioning into $x_1$	7.3785	$[d^{-1}]$
$k_1^*$	Transition rate from $x_1$ to $x_2$	0.125	$[d^{-1}]$
$k_2^*$	Transition rate from $x_2$ to $x_3$	0.1667	$[d^{-1}]$
$\alpha^*$	Apoptosis rate of $x_3$	−0.00833	$[(gd)^{-1}]$
$\beta^{**}$	Amplifying factor for individual blood regeneration, global factor independent of fraction of blood loss	1.0	[1]
$\gamma^{**}$	Amplifying factor for individual blood regeneration, depending on fractional blood loss	0.3	$[d^{-1}]$
$Fb(x_3)$	Feedback function depending on fractional blood loss	0.0	[1]

\* indicates that this value is not depending on the subject; \*\* are the parameters which will be subject to optimization; dimensionless units are denoted as [1].

The main application of the model should be the prediction of erythrocyte regeneration after a blood loss of the individuals. Therefore, a control is included into the model, which allows the removal of a chosen amount of erythrocytes. For this, the model will be extended as follows:

$$\begin{aligned}
\dot{x}_1(t) &= \beta (X_0 - k_1 \times x_1(t)) + (1-u(t)) \times Fb(x_3(t)) \times x_1(t), \\
\dot{x}_2(t) &= \beta (k_1 \times x_1(t) - k_2 \times x_2(t)), \\
\dot{x}_3(t) &= \beta (k_2 \times x_2(t) - \alpha \times x_3(t)) - u(t) \times P(\beta, \Delta_t, \Delta_{x_3}) \times x_3(t).
\end{aligned} \tag{5}$$

Here,  $u(t) \in \{0, 1\}$  is a control variable with  $u(t) = 1$  for the duration  $\Delta_t$  of the blood donation. It is assumed that the feedback during the blood loss is shut off. This assumption allows for an easy computation of the control variables. This should lead to very small errors as  $\Delta_t$  is usually very small compared to the simulated time horizon. If a steady state is given at the start of the blood donation, only  $x_3$  is affected by the blood loss. Therefore, the ODE for  $x_3$  using the known value of  $\Delta_{x_3} = x_3(0) - x_3(\Delta_t)$  can be solved to obtain a function  $P$  realizing the desired loss of erythrocyte count for each individual. Detailed computation is shown in the Appendix A.1.

Note that the following numerical results could have been achieved without incorporation of the control function  $u$ . Instead, the first measurement after the blood donation could be used for the simulation and parameter estimation to obtain the same results (starting in time 0). The incorporation of the control into the model targets a possible future optimization of phlebotomies (compare Section 5).

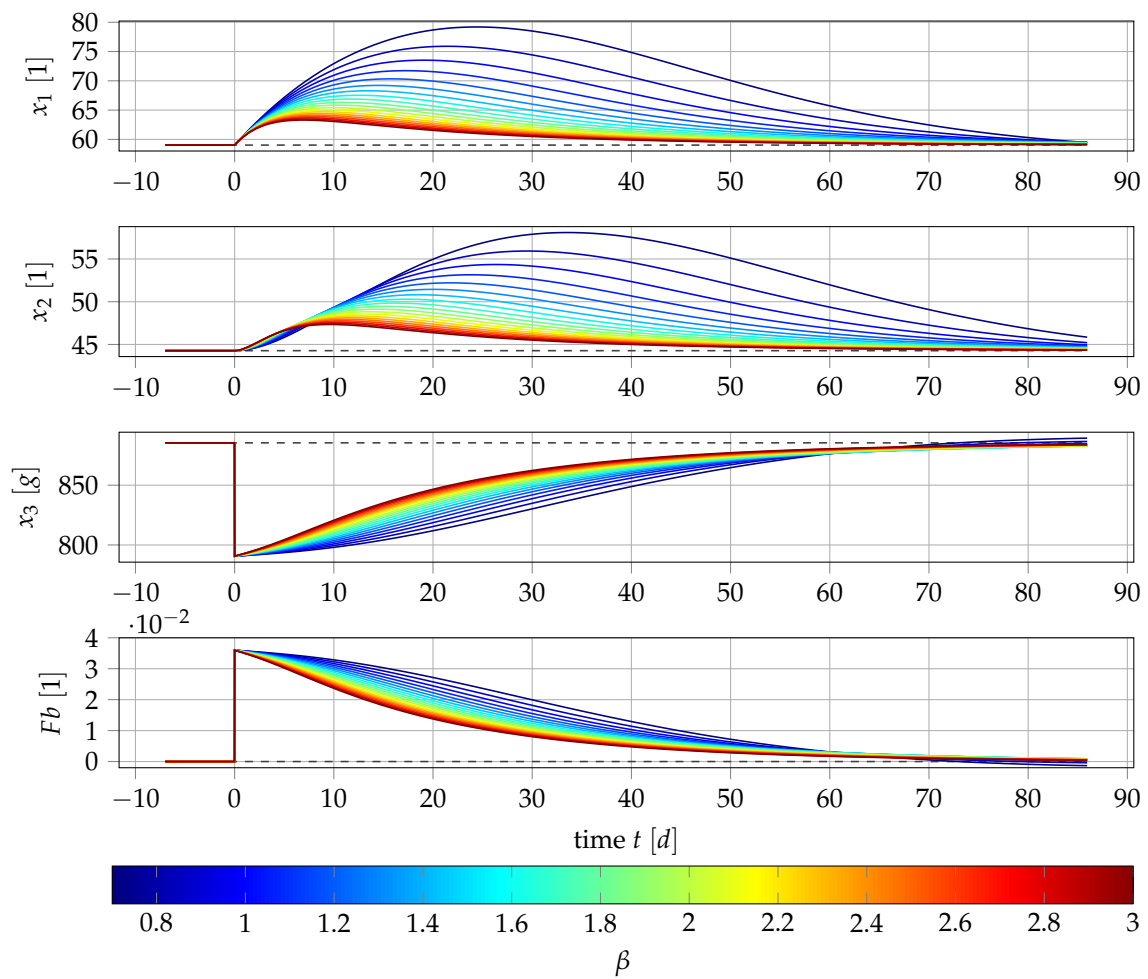
#### 2.4. Sensitivity Analysis

The variables  $\beta$  and  $\gamma$  are designed to reflect subject individual properties of erythropoiesis. Therefore, the influence of those variables on the regeneration behavior after a blood loss will be investigated in the following. As a first step, a blood loss of an exemplary subject is simulated and the model behavior under parameter variation is observed. The results are displayed in two Monte Carlo Plots, one for variation of each parameter, while the other one is fixed. Note that, instead of the number of erythrocytes, we will use the total hemoglobin mass for  $x_3$  throughout all of the results and plots. The connection of the two is detailed in Section 3.2.

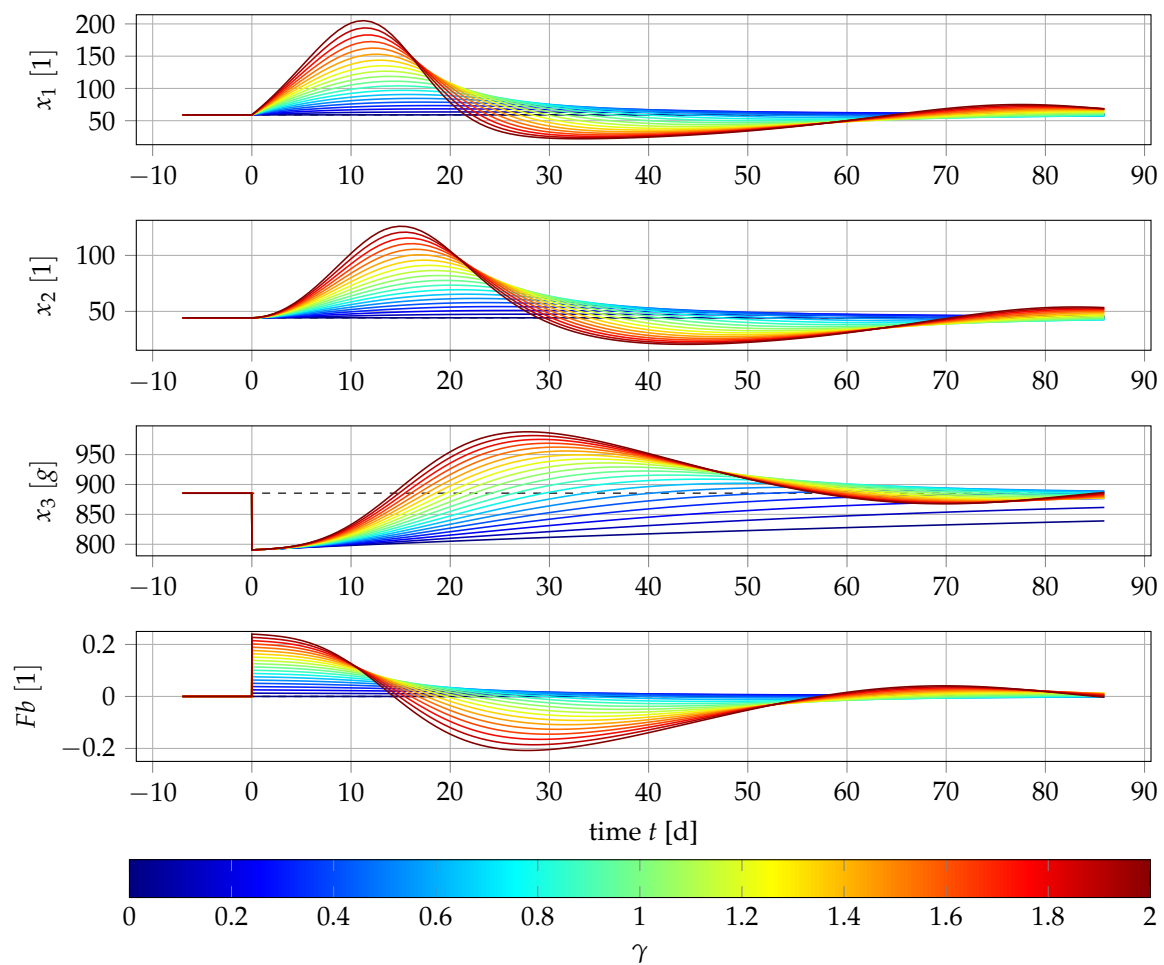
In Figure 2, the variable  $\beta$  describing the global amplification of the cell maturation is varied for a fixed value of  $\gamma$ . It can be seen that, for each value of  $\beta$  after the blood loss, a regeneration of  $x_3$  occurs, which reaches the initial level after about 65 days. Lower values of  $\beta$  tend to have a larger influence on the previous equations than higher values. This might reflect the property of subjects with high  $\beta$  that the default production rate of erythrocytes is almost enough to quickly cover the blood loss without additional production by the feedback function. In subjects with low  $\beta$ , a large fractional blood loss is present, such that the feedback term has a strong influence over a longer period of time. Graphically,  $\beta$  has an influence on the shape of the regeneration curve of the erythrocytes.

In Figure 3, the variable  $\gamma$  responsible for the amplification of the feedback term is varied for a fixed value of  $\beta$ . It can be observed that, for high values of  $\gamma$ , the time of regeneration to the *Base* value is reduced in comparison to lower values of  $\gamma$ . It further can be seen that for high values of  $\gamma$ , an oscillating behavior of each states' curve occurs. Some oscillations might be plausible in a physiological system. Nevertheless, the magnitude of those oscillations should not exceed certain ranges. Higher values of  $\gamma$  tend to have values in magnitude of the blood donation, which usually should not occur and therefore should be prevented. This problem is addressed in Section 2.5.





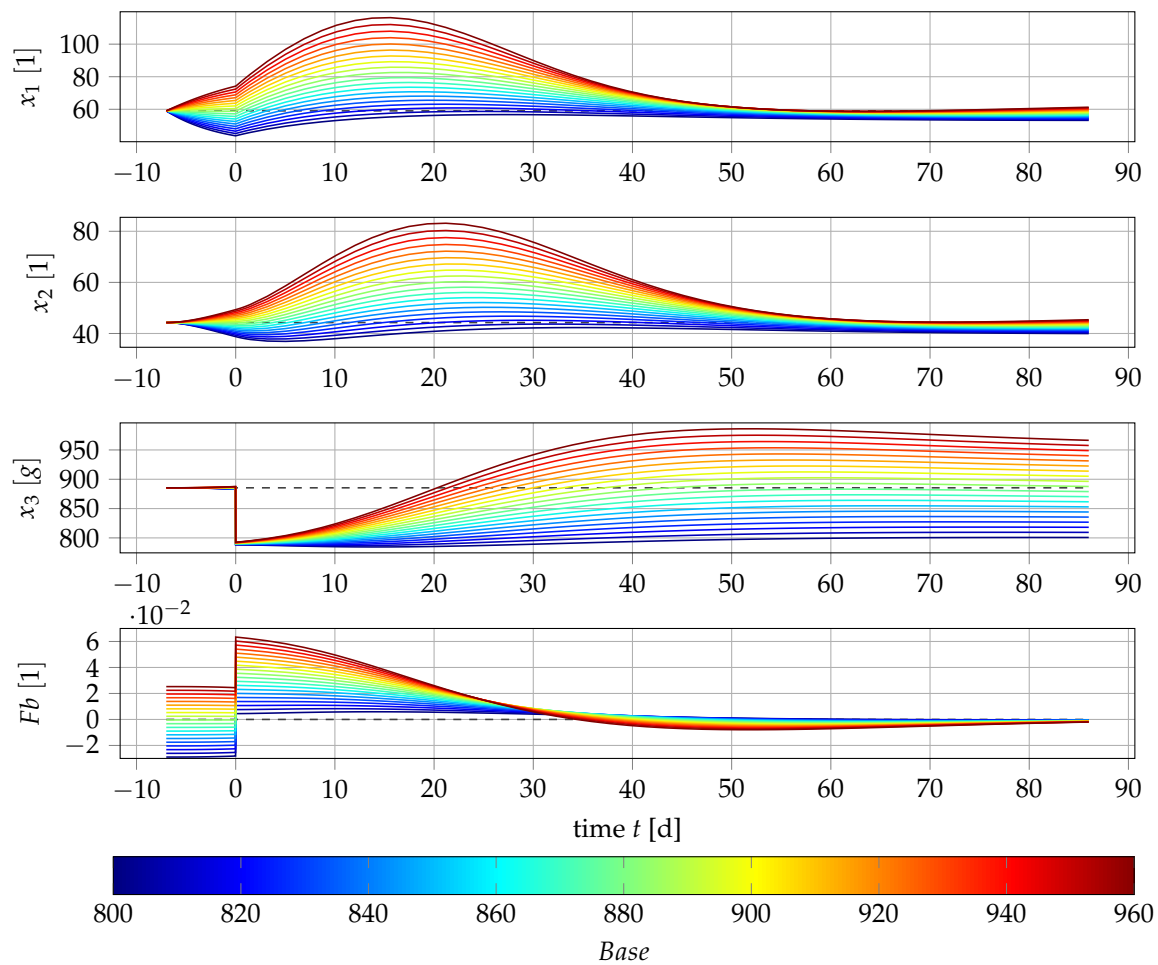
**Figure 2.** Monte Carlo simulation of 20 different values of parameter  $\beta$  for an exemplary subject setting  $\gamma = 0.3$ .  $x_1$  denotes the proliferating cells,  $x_2$  the non-proliferating cells and  $x_3$  the erythrocytes.  $Fb$  is the value of the feedback function at each time point. The simulation starts in the steady state and the simulated blood donation takes place at time 0. The dashed line marks the respective steady state.



**Figure 3.** As Figure 2, but with 20 different values of parameter  $\gamma$  for an exemplary subject setting  $\beta = 1.0$ .

In summary, both variables carry physiologically relevant properties of erythropoiesis and their effects differ from each other. Therefore, it might be a rewarding approach to use those variables for parameter estimation to given data. This is done in Section 4.

In cases in which an individual is not in its steady state, it might also be of interest to estimate the parameter *Base*. Figure 4 shows the system behavior for different values of *Base* for fixed  $\beta = 1$  and  $\gamma = 0.5$  and a given start point at  $t_0 = -7$ . In contrast to the Monte Carlo simulations for varying  $\beta$  or  $\gamma$ , in this case, the feedback  $Fb$  is non-zero already between  $t_0$  and  $t = 0$ . The slightly negative values of  $Fb$  between days 35 and 80 correspond to the findings of a slightly overshooting behavior, which is due to the choice of  $\beta = 0.5 \times \gamma$  as described in Section 2.5.



**Figure 4.** As Figure 2, but with fixed  $\beta = 1.0$  and  $\gamma = 0.5$  and 20 different values for  $Base$ . Note that the initial value  $x_3(0)$  is constant to highlight the sensitivity of the dynamics with respect to  $Base$ .

### 2.5. Local Error Analysis of the System near the Steady State

As can be seen in the previous section, for large values of  $\gamma$ , oscillations occur, which are undesirable if they become too large. The property of a system to have oscillations in case of deviations from the steady state will be locally analyzed to obtain conditions on the model variables and parameters, for which no oscillations occur. More details about this method can be found in [9].

As seen in Section 2.3 under the above posed conditions, a steady state is given in

$$\bar{x} = \left( \frac{\alpha}{k_1}, \frac{\alpha}{k_2}, 1 \right) \times Base. \quad (6)$$

Adding an error vector  $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3)$  to the state vector  $x$  and using Taylor series expansion of first order, one obtains the following differential equation for the error  $\varepsilon$ :

$$\dot{\varepsilon}(t) = J \times \varepsilon, \quad (7)$$

where

$$J = \begin{pmatrix} -k_1 \times \beta & 0 & -\frac{\gamma \times \alpha}{k_1} \\ k_1 \times \beta & -k_2 \times \beta & 0 \\ 0 & k_2 \times \beta & -\alpha \beta \end{pmatrix}. \quad (8)$$

If the error is not too large, it is known that no oscillations occur iff all eigenvalues of  $J$  are real valued. Equivalently, it is sufficient to check the condition if the discriminant  $\Delta_J$  of  $J$  is non-positive. For this, Cardano's Formula [26] can be applied to the characteristic polynomial of  $J$  to determine the necessary coefficients of  $\Delta_J$ . Detailed calculations are given in Appendix A.2.

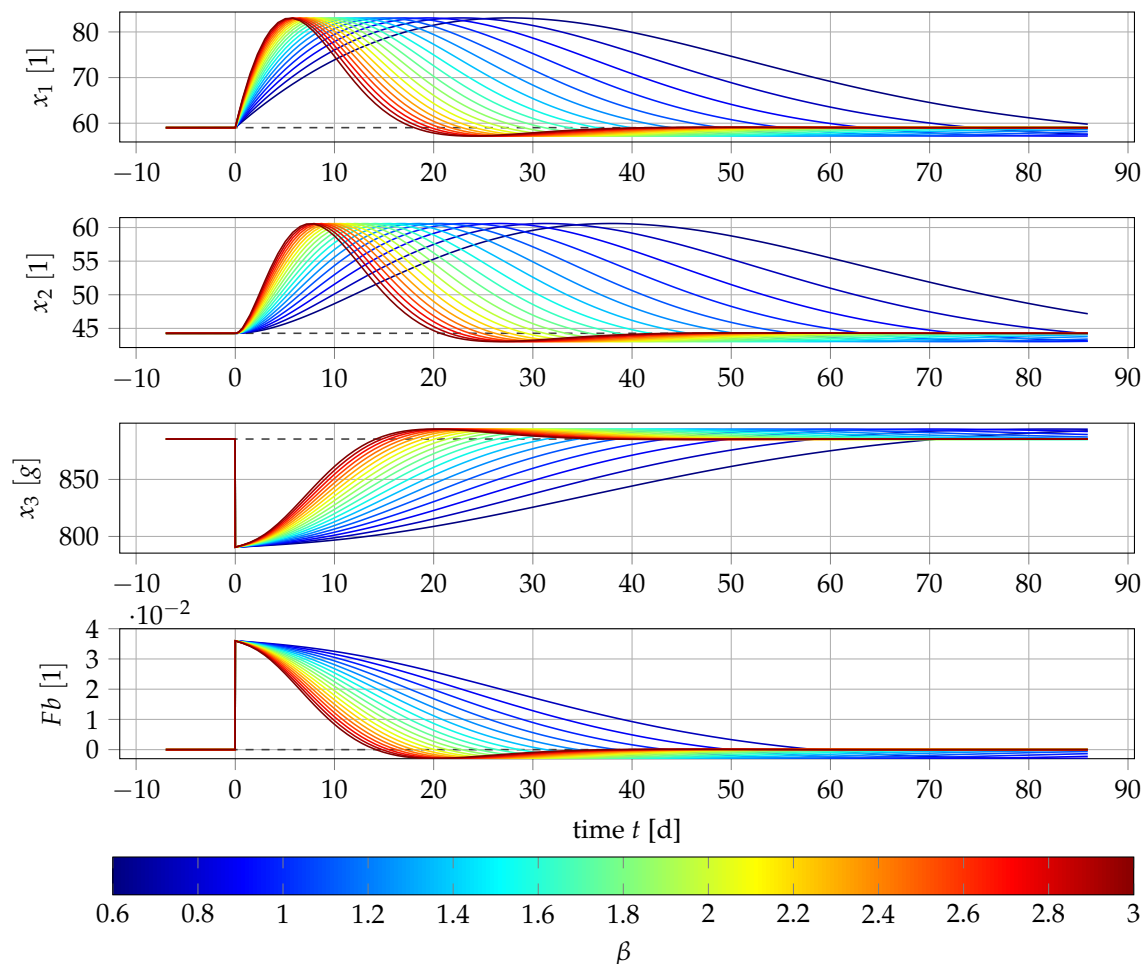
The resulting discriminant is a function depending on model parameters  $\beta$  and  $\gamma$ , which are to be estimated, and on fixed model parameters  $k_1$ ,  $k_2$  and  $\alpha$ . Setting the fixed parameters as described in Section 1  $\Delta_J$  is non-positive, if

$$\gamma \leq 0.263 \times \beta - 0.0003 \quad (9)$$

is fulfilled. As small oscillations might be plausible in a physiological system, we observed that the relaxed inequality

$$\gamma \leq 0.5 \times \beta \quad (10)$$

might be adequate for further numerical evaluation. In Figure 5, both variables are varied using Label (10) with equality. It can be observed that, even for high values of  $\gamma$ , oscillations only occur in low magnitude. Therefore, this relaxed inequality might be used for parameter estimation of the model parameters to subjects data, if large oscillations would occur leaving both parameters free.



**Figure 5.** As Figure 2, but with 20 different values of parameter  $\beta$  for an exemplary subject, where  $\gamma = 0.5 \times \beta$ .

An additional important property is the stability of the steady state in case of small perturbations. To show the stability, it is sufficient to find conditions that make sure the real parts of the eigenvalues

of  $J$  are negative. For this, again the Cardano Formula can be used. Detailed computations can be found in Appendix A.2. The two resulting conditions on  $\beta$  and  $\gamma$  are

$$\gamma \leq 4.9 \times \beta \quad (11)$$

for the first eigenvalue and

$$\gamma \leq 7.8125 \times \beta \quad (12)$$

for the second one. Thus, it is sufficient to take into account Inequality (9).

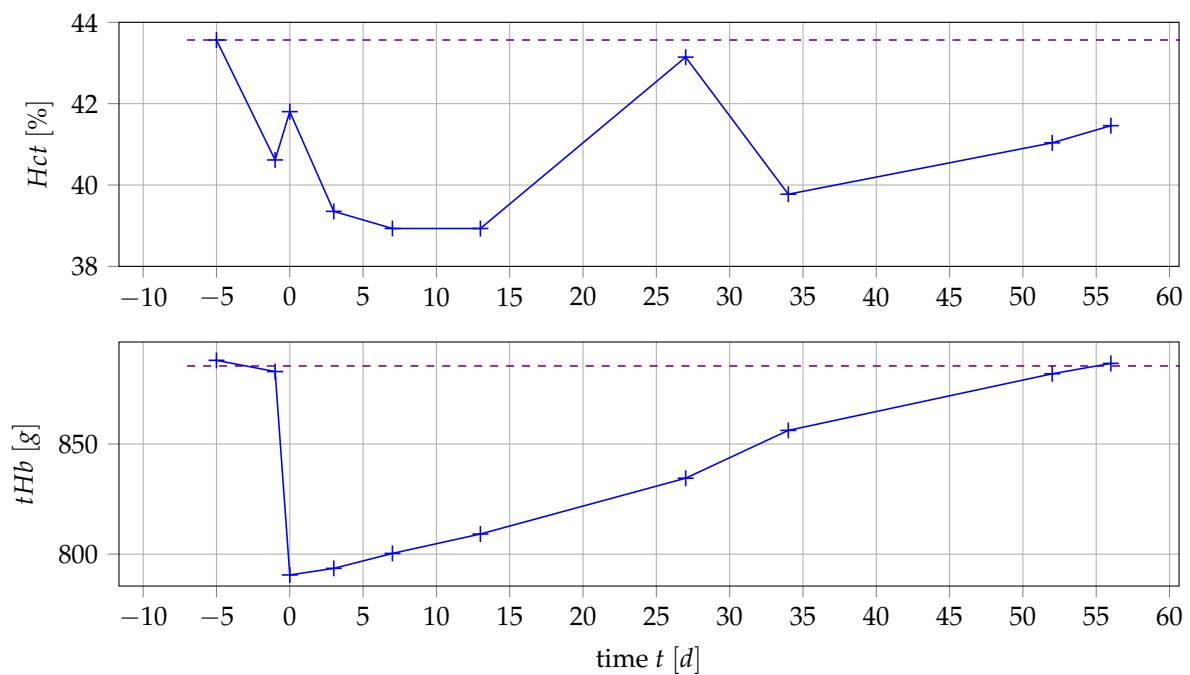
### 3. Methods

In the following, we describe the data used for the numerical results in Section 4, which use total hemoglobin ( $tHb$ ) instead of the direct RBC counts. We explain the connection between these two clinical parameters. Furthermore, we shortly describe the calculation of the *Base* parameter value for the subjects and give a brief description of the mathematical tools used in the calculations.

#### 3.1. Data

The experimental data that are used for the numerical results were obtained in 2008 by Pottgiesser et al. [17]. Here, the recovery time of  $tHb$  after a blood donation in healthy adult humans was investigated. Therefore,  $tHb$  before and after 1-unit (erythrocyte concentrate) standard blood donation was evaluated in 29 male, healthy volunteers ( $30 \pm 10$  years,  $181 \pm 7$  cm,  $76.6 \pm 11.2$  kg). The  $tHb$  measurements were obtained using the optimized CO rebreathing method explained in [27]. The healthy physiological condition of the subjects was ensured by medical examinations and the use of a detailed questionnaire that is standard procedure before blood donations. Amongst others, it can be ruled out that certain physical conditions or environmental changes would have a significant influence on the regeneration rate of the subjects erythrocytes. Therefore, it can be assumed that the erythrocytes reside in the steady state when the procedure takes place. The value can be approximated by the arithmetic mean of available measurements before the blood donation.

The acquired measurements of  $tHb$  in the study were compared to the respective  $Hct$  values. The latter is strongly affected by plasma volume fluctuations, which may occur even during one single day in each human and can be caused by everyday activities. For example, it was shown in [28] that standing still for 30 minutes yields a mean decrease of blood plasma volume of 12%. Thus, when investigating the regeneration of erythrocytes, it is beneficial to use  $tHb$ , as was shown by [17]. An exemplary comparison of  $tHb$  and  $Hct$  is displayed in Figure 6.



**Figure 6.** Comparison of hematocrit ( $Hct$ ) and total hemoglobin ( $tHb$ ) for randomly chosen subject 02.

### 3.2. Relationship of $tHb$ Values and Number of Erythrocytes

In the following, it is advantageous to use  $tHb$  for prediction of the regeneration of RBCs after a blood loss. The clinical parameter  $tHb$  and the number of erythrocytes are related in a highly correlated manner. The connection of the two can be described by the blood parameter mean corpuscular hemoglobin ( $MCH$ ), which is the mean amount of hemoglobin in a single RBC. The RBC count can be computed by

$$\#Erythrocytes(t) = \frac{tHb(t) [g]}{MCH(t) [pg]} = \frac{tHb(t)}{MCH(t)} \times 10^{12}. \quad (13)$$

In the available data from [17], for some of the measurements, venous blood samples were also drawn, which include the respective  $MCH$  values. For each subject's data, the percentage standard deviation (%SD) from the respective mean value was computed. It was observed that, in  $6.62 \pm 1.69$  available  $MCH$  measurements per subject, the %SD from the mean was  $1.85 \pm 0.57$ . As in the  $MCH$  values, for one individual, only very small changes were observed concerning possible measurement errors, and it can be assumed that in Equation (13) the arithmetic mean of all  $MCH$  values can be used. Thus, by Equation (13), the number of erythrocytes and the respective  $tHb$  values are proportional. Therefore, the use of  $tHb$  in our simulations is justified. Similar results for the variation of  $MCH$  values were observed in [23].

### 3.3. Calculation of Base Value and Blood Loss

The subject is assumed to be in a steady state at the start of the experiment (compare Assumption 2 in Section 2.1). Therefore, the *Base* value can be estimated using the arithmetic mean of RBC count measurements before the blood donation. In the underlying data, one to two measurements before the blood donation are available. Due to the low amount of data, comparatively high measurement errors might occur. This problem is further addressed in Section 4.4.

In several (but not all) available data sets, it was observed that the minimal  $tHb$  level after a blood donation was reached after about one week. This collides with the intuitive guess that the minimal value should be attained directly after the blood donation. This observation corresponds to results of other studies—for example, compare [29]. To the best of the author's knowledge, the physiology of



this process until now is not fully understood. In particular, it is unclear how to identify those subjects which do show this behavior and those who do not show it a priori.

The feature of a delayed *tHb* response after a blood donation is not covered by our current model and subject to further inspection in the future. Instead, the control and the variable *P* are chosen such that the *tHb* level after the blood donation is equal to the minimal *tHb* value after the first eight days. Details on how the value of *P* is computed can be found in Appendix A.1.

### 3.4. Parameter Estimation

There are different approaches to estimate model parameters, usually such that they explain measurements for given assumptions with the highest probability. If homogeneous populations of measurement data are available, nonlinear mixed-effect models (nlme) have become a popular choice. For very heterogeneous and sparse data, usually point estimators are applied. Given the homogeneous data described in Section 3.1, we used an nlme approach to calculate population parameters and inter-individual variability. Having future extension of the model towards clinical application and more heterogeneous groups in mind, compare Section 5, we also applied individual point estimation and compared the results.

Parameter estimation was performed in NONMEM (version 7.4, first-order conditional estimation method with interaction) in combination with PsN software (version 4.4.0; Uppsala Pharmacometrics, Uppsala, Sweden). We used an exponential model for inter-individual variability (diagonal OMEGA matrix) and an additive model for residual variability. An exponential as well as combined error model provided the same results, so that we decided to chose the additive error model equivalent to the individual parameter estimation approach. No covariates were included. Data set preparation was performed in Python (version 2.7.6) and analysis of the results was performed in R (version 3.4.4; R Project for Statistical Computing, Vienna, Austria) including the xpose4 package (version 4.6.1) for generating visual predictive checks.

For the point estimation, using the available data and the derived model formulation written compactly as  $F(\cdot)$ , parameter estimation problems of the form

$$\min_p \frac{1}{2} \sum_{i=1}^{n_\eta} \frac{(\eta_i - x_3(t_i, x_3(t_i), p, u(t_i)))^2}{\sigma_i^2} + \phi(p), \quad (14)$$

$$s.t. \quad \dot{x}(t) = F(t, x(t), p, u(t)), \quad (15)$$

$$x(0) = x_0, \quad (16)$$

where

- $n_\eta$  are the number of available measurements of the subjects *tHb*,
- $\eta_i$  is the measurement value of *tHb* at time  $t_i$ ,
- $x_3(t_i)$  is the according model response at time  $t_i$ ,
- $\sigma_i$  is the standard deviation of the measurement at time  $t_i$ ,
- $p$  is the chosen parameter vector,
- $F$  contains the right hand sides of the model Equation (5),
- $x_0$  fixed start values, and
- $\phi(p)$  is a term that can be used to incorporate a priori information, in our setting chosen as 0

were solved with a multiple shooting based Gauß-Newton algorithm coded in the PAREMERA software and an adaptive, error-controlled backward differentiation formulae (BDF) method for integration coded in the software DAESOL, both included in the experimental design package VPLAN [30] developed at the University of Heidelberg. The same integrator was used for all predictions (simulations) in this paper.

The model variable *P*, which is multiplied with the control *u* to realize the blood loss (5), depends on *Base*. In addition, *P* is computed outside of the parameter estimation problem due to limitations of

the used software. Therefore, if *Base* is included into the parameter estimation, the simulation will start at the first measurement point, meaning  $x_0 = (\bar{x}_1, \bar{x}_2, \eta_1)$ . In contrast, if *Base* is not used as a free parameter,  $P$  and  $u$  will be used as described in (5) and  $x_0 = \bar{x}$ . In both cases,  $x_0$  is fixed and thus no state estimation is performed.

Common criteria for evaluating the quality of the results of a parameter estimation are given by the residual value and the covariance matrix with respect to the optimal parameter vector  $p^*$ . The residual value is the value of the least squares functional for  $p^*$ . The covariance matrix for a one-dimensional parameter vector  $p^*$  is the standard deviation of that parameter. The entries of the covariance matrix  $C$  of a two-dimensional parameter vector  $p^*$  are given by  $c_{i,j} = Cov(p_i, p_j)$ . In the following section the robustness of obtained results are displayed taking a large amount of random samples from a 95% quantile around the optimal parameter vector. In the 1-D case the confidence interval is uniquely determined. In the 2-D case the geometry of the chosen confidence ellipsoid is given by the eigenvalues and eigenvectors of  $C$ .

#### 4. Numerical Results

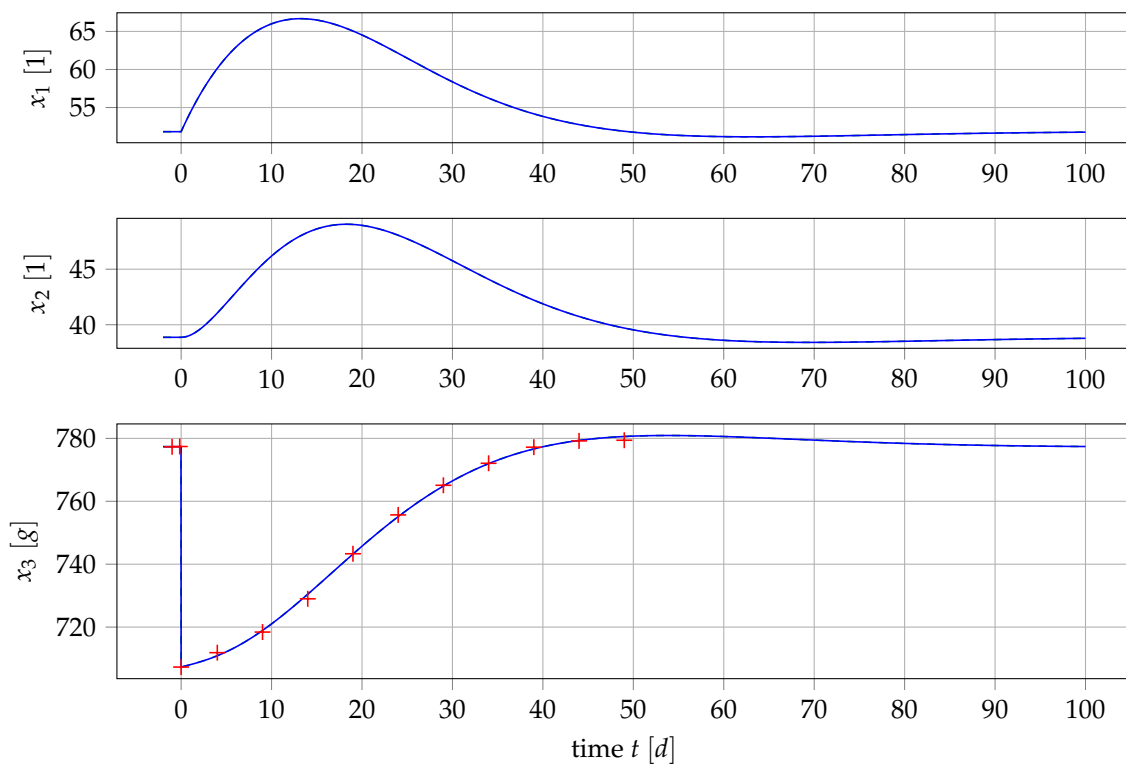
In this section, we present numerical results using the proposed model. First, we compare our model and a state-of-the-art model for erythropoiesis, using the set of free parameters described in Section 2.4. Second, we estimate model parameters for the experimental data from Section 3.1, using both population and point estimation. For point estimation, we examine different combinations of free model parameters.

##### 4.1. Results for Generic Subject Data from Literature

As displayed in Section 2, a simplified approach of modeling erythropoiesis based on three ODEs was used. Here, the proposed model will be compared to a more sophisticated model based on Fuertringer et al. [8]. The model proposed there describes erythropoiesis using an age-structured PDE system describing population densities of erythroid stem cells. The EPO concentration is directly modeled and used as feedback mechanism in an integral constraint depending on the total blood volume in the subject. In addition, the feature of neocytolysis, which is the premature cell death of young erythrocytes in case of oversupply, is included. Model parameters are fit to a generic subject, whose properties are based on average values from literature. The model proposed in Fuertringer et al. covers many physiological properties and is capable of describing the regeneration after a blood loss in a reasonable way. Furthermore, the application regarded there was the restoration of the RBC count after blood donation.

For comparison of the proposed model to the modeling results of Fuertringer, data from the respective plot (Figure 5 in [8]) were extracted. Two points before the blood donation and 10 points after the blood donation every five days are taken into account. The used standard value of  $2.5 \times 10^{13}$  erythrocytes was converted to *tHb* using the average value for  $MCH = 31$  pg. This results in a *tHb* Base value of about 777.4 g. Using the converted data, a parameter estimation was used for finding  $\beta$  and  $\gamma$  to fit the model response of  $x_3$  to the extracted data. The results are displayed in Figure 7.

It can be observed that, after the parameter estimation, the model response is able to fit the extracted data with a high accuracy ( $R^2 = 0.999$ ). The obtained parameters of  $\gamma = 0.56$  and  $\beta = 1.28$  fulfill the inequality (10) for prevention of oscillations and are robust (%SDs of 12.6% and 24.2%). This shows that the proposed model is capable of producing similar results compared to the model of Fuertringer et al. [8] concerning the used case of a blood donation using a more involved model structure. This might justify the use of our less complex model of erythropoiesis, especially for clinical applications.



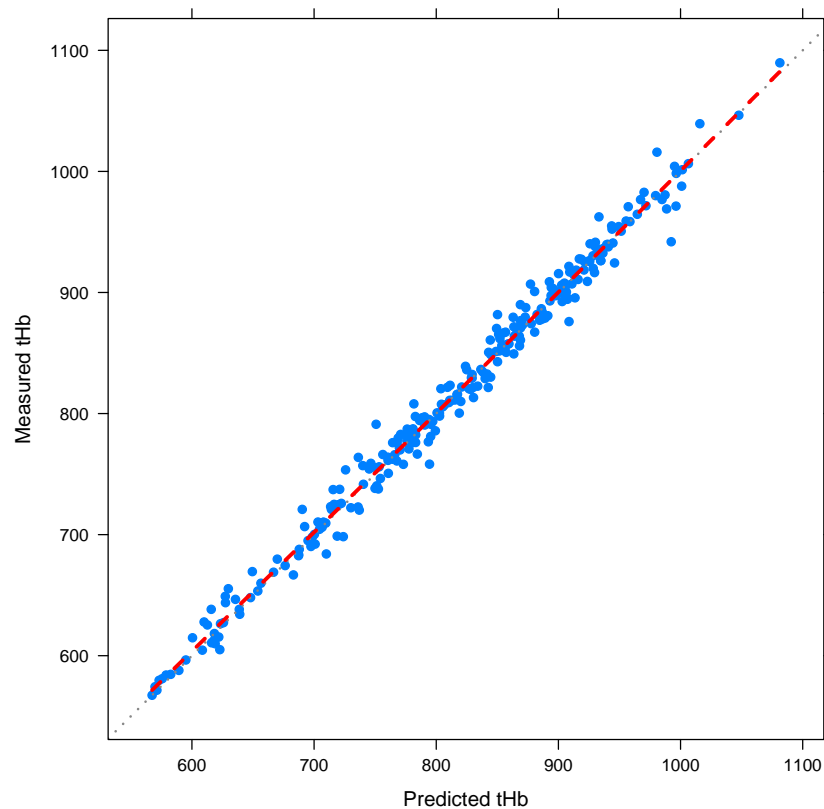
**Figure 7.** Results of parameter estimation using data of the generic subject in Fuertringer et al. [8].

#### 4.2. Nonlinear Mixed-Effects Estimation for $\beta$ and $\gamma$

Based on the discussion from the previous sections, optimal parameters  $\beta$  and  $\gamma$  were estimated using the according experimental data. We did not impose any additional conditions like inequality (10) on the parameters. The steady state values *Base* for each subject were estimated using measurements before the blood donation. The blood donation itself was realized in the model using the mentioned control  $u$  with an accordingly individualized variable  $P$ .

The results from NONMEM for 276 observations from 29 subjects are visualized with a goodness-of-fit plot in Figure 8. The fixed effect of population parameter  $\beta$  was estimated as  $1.02 \pm 0.151$ , the inter-individual variance as  $0.294 \pm 0.125$ . For  $\gamma$ , the fixed effect was estimated as  $0.46 \pm 0.0651$  with an inter-individual variance of  $0.346 \pm 0.148$ .

Given the comparatively high inter-individual variances of more than 40% even for the quite homogeneous population (male, healthy, non-smokers, ...), we concentrate on individual parameter estimations in the following.



**Figure 8.** Goodness-of-fit plot for 276 observations from 29 subjects, using a nonlinear mixed-effects modeling approach. Shown are measured versus calculated total hemoglobin (tHb) values.

#### 4.3. Point Estimation Results for $\beta$ and $\gamma$

We applied a point estimation approach in the same setting as in Section 4.2. Parameters for all but one of the subjects (ID 22) could be estimated. For that subject, no convergence could be achieved with default settings and without a regularization function  $\phi(p)$ , probably due to scattered data.

For four exemplary subjects, the results of the numerical simulation and the respective measurement values, are displayed in Figure 9. In this figure, the steady state values of the respective states have been normalized for a better comparability. The trajectories  $x_3$  are able to capture the measurement values. In particular, the different regeneration times of the chosen subjects were fit nicely by the estimated model responses. The model is capable of including a broad range of regeneration times (in the four depicted subjects 20–60 days). Oscillations only occur at a reasonable pace (compare subject 12) or not at all (compare subject 20). Detailed results are given in the Appendix (Table A1).

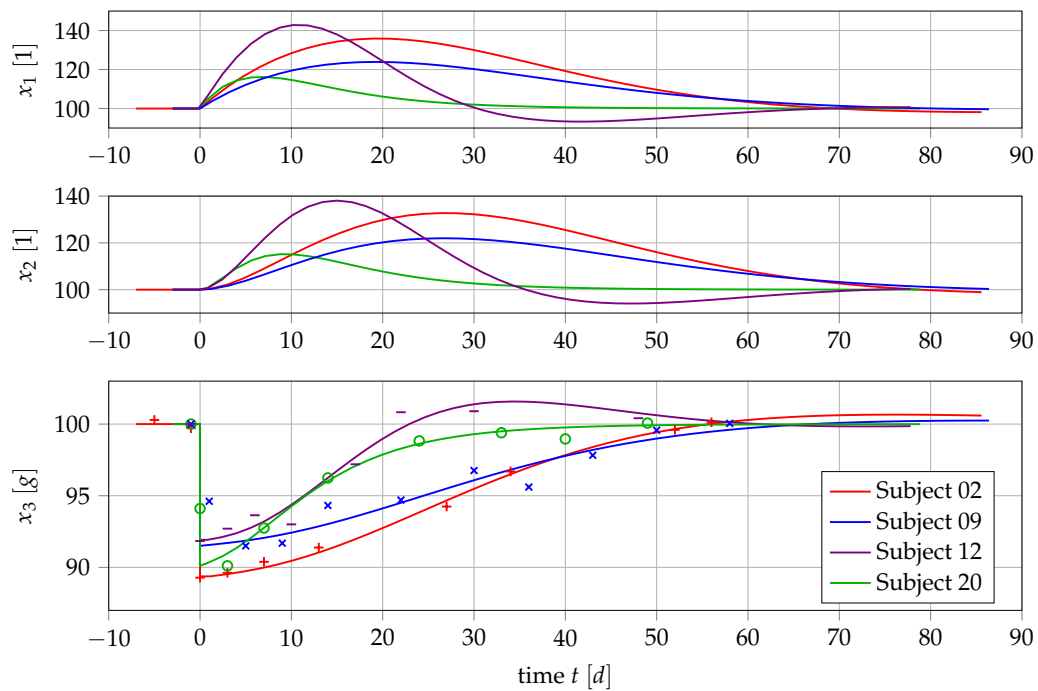
On average, the quality of the fits can be described as high. This can be seen when taking into account the average values for  $R^2$  of  $0.837 \pm 0.151$ . However, the %SDs of the parameters  $\beta$  and  $\gamma$  equal to  $43.5\% \pm 22.9\%$  and  $59\% \pm 85.9\%$ , respectively. In addition, the standard deviation of %SD with respect to  $\gamma$  is very high compared to the mean.

A detailed analysis of the results per subject identifies four subjects with a very high %SD with regard to  $\gamma$  (subjects 05, 10, 13 and 16). We characterized a subject  $S$  as badly estimated if the %SD of at least one parameter fulfilled the following inequality:

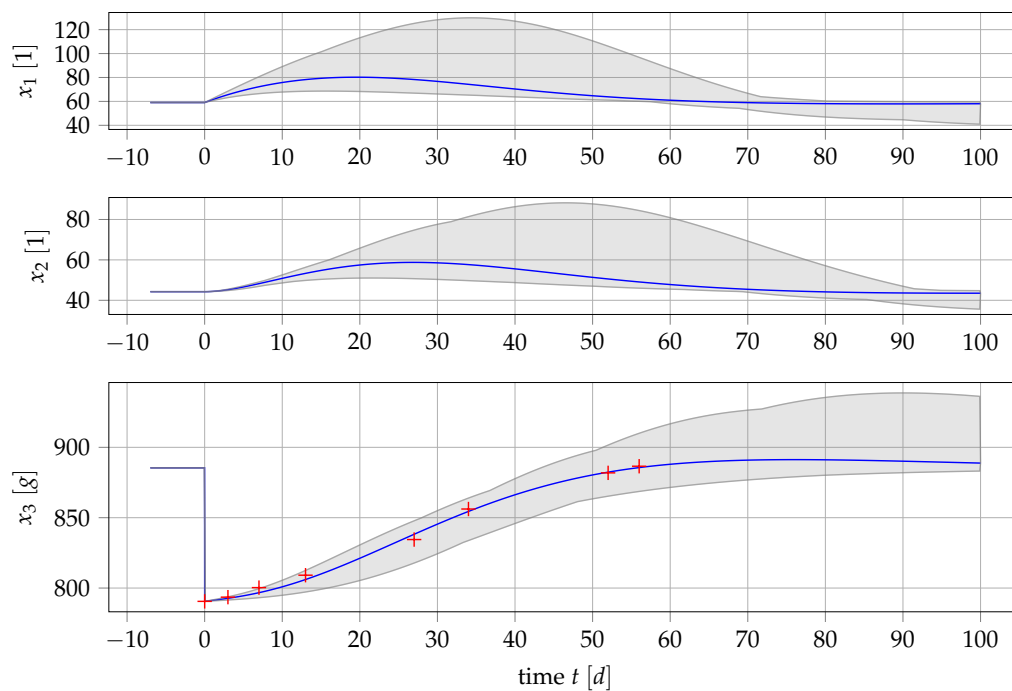
$$\%SD_S \geq \text{Mean}(\%SD) + 1.4 \times SD(\%SD). \quad (17)$$

Considering only the remaining 24 subjects, the quality of the results was very high. Here, the average values for  $R^2$  and the %SD for  $\beta$  and  $\gamma$  were  $0.86 \pm 0.11$ ,  $37\% \pm 15\%$  and  $26.4\% \pm 17.3\%$ , respectively.

As described in Section 3.4, Figure 10 shows an exemplary robustness plot for subject 02. For this, 100 random samples were drawn with respect to the according covariance matrix of the parameter set. All samples that have been included were inside the chosen ellipsoid with a confidence level of 95%. All computed trajectories reside in the confidence band around the optimal solution. The good quality of the estimation is reflected by the tightness of the confidence band, especially for  $x_3$ .



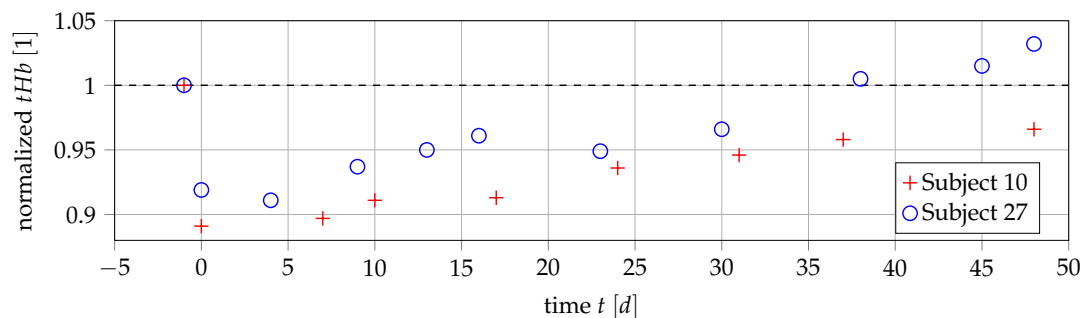
**Figure 9.** Plot of results after parameter estimation for four exemplary subjects.



**Figure 10.** Robustness plot of subject 02.

#### 4.4. Results if Base Is a Free Parameter

One problem that might occur using the data of some subjects is a bad choice of *Base*. This might especially be the case for the four subjects excluded above as it can be seen in the case of subject 10 (compare Figure 11). Here, the measurements seem to converge to a value lower than the assumed steady state value. Even for some subjects that did not have to be excluded (for example subject 27), *Base* might have been chosen unsuitably. In contrast to subject 10, *Base* might have been estimated too low in this case, as can be seen in Figure 11.



**Figure 11.** Normalized *tHb* measurement data of subjects 10 and 27 including assumed *Base* values.

This observation motivates the use of *Base* as a free parameter. Following the approach of the previous section, the intuitive idea would be to use the parameter set *Base*,  $\beta$  and  $\gamma$ . The initial value of *Base* is set to the average of measurements before the blood donation. However, testing this approach leads to identifiability issues. During the corresponding parameter estimations, the covariance matrices came close to singularity, which means that the set of the three parameters was not identifiable. Only in a few cases was the parameter estimation using this parameter set successful. As on average only  $8.65 \pm 1.73$  measurements per subject are available, using three free parameters might be too much. Given the amount of available data, this approach might not be suitable for a successful parameter estimation of the model parameters.

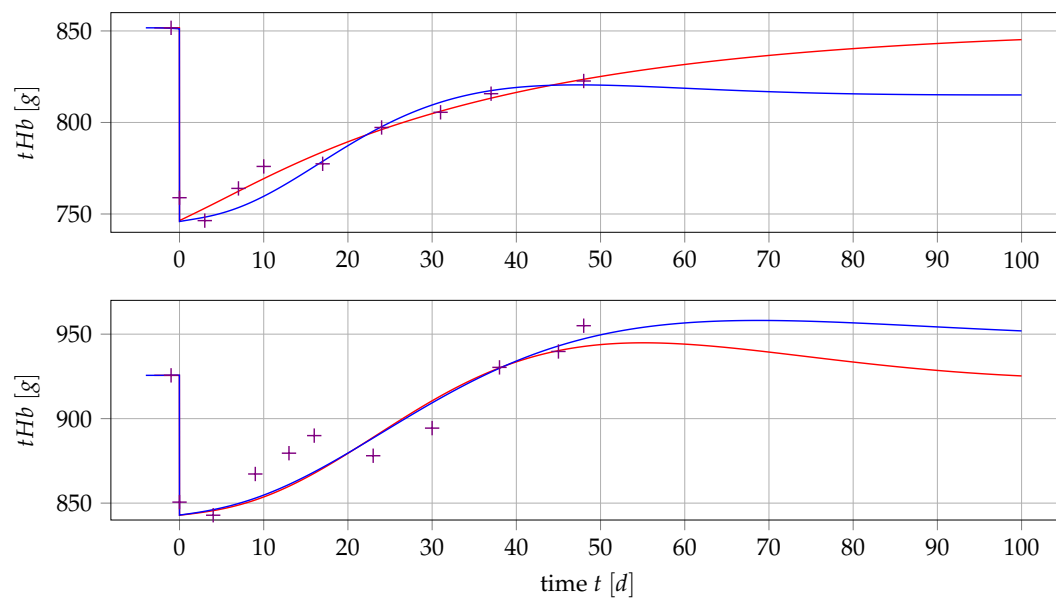
To be able to use only two free parameters and still include *Base* into the parameter set, a linear relation between  $\beta$  and  $\gamma$  can be assumed. However, then the capability of  $\beta$  and  $\gamma$  to reflect two distinct properties of erythropoiesis is strongly restricted. To test this approach, equality is assumed in inequality (10). The behavior of the system dynamics under these conditions is depicted in Figure 5. An exemplary application of this method to the two subjects 10 and 27 leads to the results displayed in Figure 12. Detailed results for all subjects can be found in the Appendix Table A2.

In Figure 12, results for this method and the method used in Section 4.3 are displayed for subjects 10 and 27. The results using also *Base* as a free parameter seem to converge to a more plausible steady state of the system when the measurements are taken into account. Therefore, for these two cases, the results seem to have improved. However, for other cases, this approach leads to worse results. An example is given in Figure 13 for subject 14. The trajectory obtained by the method described in this subsection obviously converges to a physiologically implausible value of *tHb*.

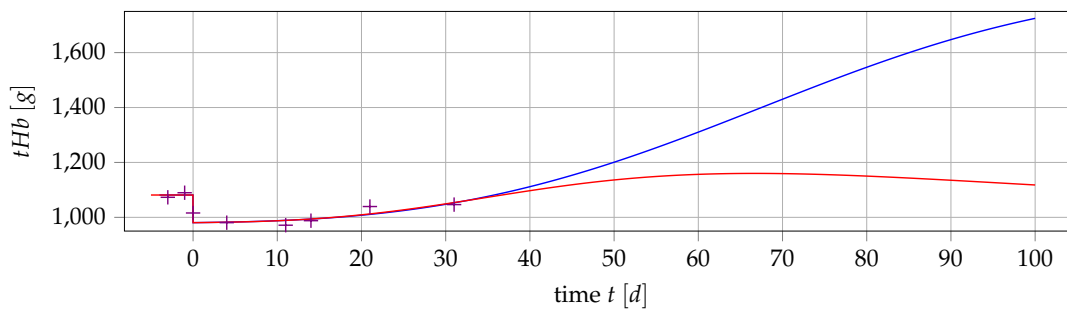
Again, on average, the quality of the parameter estimation can be described as very high with average values for  $R^2$  of  $0.829 \pm 0.141$ . However, the %SD for the parameters  $\beta$  and *Base* of  $29.3\% \pm 20.9\%$  and  $4.8\% \pm 13.7\%$ , respectively, should be discussed critically. This is especially the case for the standard deviation of %SD of *Base*, which is about three times the corresponding mean.

If we again take into account criterion (17) for badly identified parameter sets, two subjects can be excluded (subjects 14 and 17). These two subjects are distinct from the four subjects excluded in the section above. Considering only the remaining 26 subjects, the quality of the results improved. Here, the average values for  $R^2$  and %SD for  $\beta$  and *Base* were  $0.84 \pm 0.13$ ,  $24.8\% \pm 12.7\%$  and  $1.4\% \pm 0.8\%$ , respectively.





**Figure 12.** Results of the two methods described for subjects 10 (above) and 27 (below). The approach using  $\beta$  and  $\gamma$  as free parameters is depicted in red, the one using  $\beta$  and  $Base$  is depicted in blue.



**Figure 13.** Comparison of results of the two methods described for subject 14. The approach using  $\beta$  and  $\gamma$  as free parameters is depicted in red. The approach using  $\beta$  and  $Base$  is depicted in blue.

Overall, the quality of the parameter estimations was similar using the two proposed parameter sets. Both methods had few flaws as some subjects had to be excluded. However, using  $Base$  and  $\beta$  with linear relation between  $\beta$  and  $\gamma$  weakens the individual physiological properties of the latter ones. In addition, the choice of the factor 0.5 in the linear relationship is arbitrary and could be discussed. As the available data quality and quantity does not seem to allow the estimation of three parameters, the first methods using  $\beta$  and  $\gamma$  as free parameters are recommended. As the quality of the results highly depends on a good estimation of  $Base$ , the initial guess of that variable should be improved. This is subject to future research.

A comparison between the population parameters from Section 4.2 and the individual parameters listed in Tables A1 and A2 shows values for  $\beta$  and  $\gamma$  in the same order of magnitude, as expected.

## 5. Outlook: Polycythemia Vera

A motivating future application of our mathematical model is an optimization of phlebotomy schedules for patients suffering from Polycythemia vera (PV). In clinical practice, a phlebotomy is executed in a PV patient if the  $Hct$  is above 45% [2]. As the time when this happens is unknown, the  $Hct$  has to be measured in short time intervals, which implies additional stress and inconveniences for the patient.

The patients would hugely benefit from a therapy strategy that is able to predict the optimal treatment time for the next phlebotomy. Extending the presented results from healthy subjects to patients with PV might help the physician to schedule the therapies individually for each patient based on a set of parameters for each individual. Thus, on the one hand, the probability of severe complications might decrease, if the time until the next measurement was assumed to be too long. On the other hand, in cases where the frequency of two consecutive measurements was assumed to be too low, the patient could benefit from not needing to go to a hematologist and being saved from additional blood withdrawals.

Similar approaches and modeling strategies have been presented in the past for the simulation of white blood cell (WBC) dynamics in the application of AML [11–13]. As some features in WBC dynamics are similar to RBC dynamics, there is hope that decision support tools can also be developed for PV. However, there are at least three different and particular kinds of challenges that will need to be addressed.

### 5.1. Obtaining Data for PV Patients

The clinical parameter *Hct* is highly dependent on the blood plasma concentration, which is subject to huge fluctuations depending on environmental conditions, physical activity and fluid balance (see for example [28,31]). The latter problem could be overcome by instead using the total hemoglobin (*tHb*) mass as described in Pottgiesser et al. [17] and exemplarily visualized in Figure 6. Total hemoglobin is better suited than absolute RBC counts when taking blood samples to scale to the overall number of RBCs in the body, and a valid data source for estimation as shown in this paper. However, measuring *tHb* is complicated, as it involves special measurement devices and procedures, and implies inconveniences for patients. A prospective study will have to find accurate and easy ways to estimate the absolute numbers of erythrocytes in the body.

### 5.2. Extending the Mathematical Model

There are at least two issues that require special attention, when a model extension towards PV is considered. First, PV patients do not have a steady state any more (or it is way above a healthy value), as the hematocrit keeps increasing. Second, the underlying (disturbed) biological mechanism has to be modeled. This is particularly difficult, but also a promising area for mathematical model discrimination techniques, as the mechanisms have not yet been fully understood.

A natural candidate for modeling alternatives is the feedback function  $Fb(x_3)$ . Already in 1978, Eaves and Eaves [32] made the assumption that in PV patients the influence of EPO on the CFU-E progenitor cells is disturbed. It was observed that in vitro the CFU-E population using cells of a PV patient splits into two groups. While the cells in the first group have an EPO response similar to healthy patients, cells in the second group even proliferate in an environment with almost no EPO. This leads to constantly high proliferation and low apoptosis rates of the latter group, despite the typically low EPO concentration in PV patients. This and other assumptions on what goes wrong in PV will have to be modeled and cross-validated with clinical data.

### 5.3. Formulating and Solving Optimization Problems

From a patient's point of view, there may be several risks and inconveniences related to a phlebotomy schedule on a fixed time horizon of, for example, one year. This includes the number of phlebotomies, the number, severity, and length of periods with elevated hematocrit values, the number of measurements (of *Hct* or *tHb*), but also overlap of phlebotomies with professional or personal appointments. In particular, the last aspect has an impact on the patients' life quality that is often underestimated. The days following a phlebotomy are usually dominated by fatigue and dizziness. Therefore, long-term planning allowing the avoidance of, for example, conflicts with an important business meeting or a wedding would be highly desirable.

All of these aspects can be modeled as a part of the objective function or of constraints in an optimization problem. The degrees of freedom would be the blood letting decisions  $u(\cdot)$  introduced above. Because of the combinatorial and dynamic nature of the problem, mixed-integer optimal control problems need to be solved.

## 6. Conclusions

In this work, a three compartment model with a negative feedback mechanism for erythropoiesis was developed. Essential physiological properties were captured in the model, which could be shown with the application of the RBC regeneration after a blood donation.

The proposed model is structurally much easier than other models that can be found in literature. It could be shown that the proposed model has a similar quality regarding the individualization of model parameters as a more sophisticated model with respect to the individualization of model parameters. In addition, the proposed model needs less data for a successful parameter estimation due to the reduced structure. Thus, although biological insight might be lost compared to more complex models, our new model seems to be a valid candidate for decision support in a clinical setting.

A population estimation approach worked fine on the homogeneous data, but might not be suited in an early phase for heterogeneous and sparse data from pathological cases. Point estimation without regularization was successful in most of the cases for two different sets of free parameters. Both sets led to identifiable results for the same 22 of the overall 28 subjects. In the six remaining cases, exactly one of the two methods worked, depending on the initial estimation of the steady state value. A better knowledge of the *Base* value (for example using an increased number of measurements before the blood donation) would simplify the estimation of *Base*.

In case of available data for multiple blood donation cycles, the proposed model should be cross-validated using the first cycle for estimation and predicting the following cycles. This needs to be verified in future clinical trials. This is especially important for a possible future optimization of phlebotomies for PV patients.

The quality of data acquired in the study of Pottgiesser et al. [17] is higher than the data usually generated in clinical practice. This is due to a lower influence of the plasma fluctuations compared to *Hct* measurements. Even using this data set of improved quality, the generation of reliable results is difficult in a few cases. Based on our observations, we are convinced that measurements of *tHb* should be the criterion of choice for investigation of individual properties of erythropoiesis.

**Author Contributions:** M.T. and P.L. developed the model, performed the experiments and wrote the paper; T.F. and E.S. contributed to the design of this research and analyzed data; T.P. provided the clinical data for the experiments and contributed to the design of the research; S.S. contributed to the design of the research and to the discussion of the mathematical model.

**Funding:** This research was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 647573), by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)–314838170, GRK 2297 MathCoRe and was supported by a research grant of the “International Max Planck Research School (IMPRS) for Advanced Methods in Process and System Engineering” in Magdeburg, which is gratefully acknowledged.

**Acknowledgments:** The authors wish to thank Felix Jost, Institute for Mathematical Optimization, Otto-von-Guericke-University Magdeburg, for performing the nonlinear mixed-effect modeling analysis in NONMEM7.4.

**Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

ALL	Acute Lymphoblastic Leukemia
AML	Acute Myeloid Leukemia
BFU-E	Blast Forming Unit-Hematopoietic
CFU-E	Colony Forming Unit-Hematopoietic
DDE	Delay Differential Equation
EPO	Erythropoietin
Hct	Hematocrit
IVP	Initial Value Problem
MCH	Mean Corpuscular Hemoglobin
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
PK/PD	Pharmacokinetic/Pharmacodynamic
PV	Polycythemia vera
RBC	Red Blood Cell
SD	Standard Deviation
tHb	Total hemoglobin
WBC	White Blood Cell
%SD	Percentage Standard Deviation

## Appendix A. Computational Details

### Appendix A.1. How to Calculate $P$

In Equation (5), a control function  $u$  is introduced for incorporating the blood loss into the model. According to Assumption 2, at the beginning of the experiment, the subject's system is in a steady state. The feedback mechanism is shut off for the duration of the blood donation by the model control  $u$ . Therefore, both  $x_1$  and  $x_2$  keep their respective steady state values during that process. For calculation of the necessary constant, an initial value problem (IVP) only using the ODE for  $x_3$  has to be investigated:

$$\dot{x}_3(t) = -(\beta \times \alpha + P) \times x_3(t) + \beta \times \alpha \times Base, \quad (A1)$$

$$x_3(t_{start}) = Base. \quad (A2)$$

Here,  $t_{start}$  is the start time of the blood donation. This IVP has an unique solution given by

$$x_3(t) = \frac{Base}{\beta \times \alpha + P} \times (\beta \times \alpha + P \times \exp((\beta \times \alpha + P) \times (t_{start} - t))). \quad (A3)$$

Depending on  $\beta$ , the variable  $P$  has to be adjusted such that

$$x_3(t_{end}) = \eta_{bl}, \quad (A4)$$

where  $t_{end}$  is the end time of the blood donation and  $\eta_{bl}$  the measured tHb level after blood donation according to Section 3.3. Due to the affine linear dependency between  $P$  and  $\beta$  in the solution (A3), for each subject, a linear equation for  $P$  can be computed:

$$P(\beta) = m \times \beta + n. \quad (A5)$$

### Appendix A.2. Analysis of the Local Error at the Steady State

Calculation of the characteristic polynomial of  $J$  yields

$$\begin{aligned}
0 = \mathcal{X}_J = \det(J - \lambda I) &= \begin{vmatrix} -(\lambda + k_1 \times \beta) & 0 & -\frac{\gamma \times \alpha}{k_1} \\ k_1 \times \beta & -(\lambda + k_2 \times \beta) & 0 \\ 0 & k_2 \times \beta & -(\lambda + \alpha \times \beta) \end{vmatrix} \\
&= -(\lambda + k_1 \times \beta) \times (\lambda + k_2 \times \beta) \times (\lambda + \alpha \times \beta) - \gamma \times \alpha \times \beta^2 \times k_2.
\end{aligned}$$

By rearranging and simplifying, one obtains

$$0 = A \times \lambda^3 + B \times \lambda^2 + C \times \lambda + D, \quad (\text{A6})$$

where

$$\begin{aligned}
A &= 1, \\
B &= \beta \times (k_1 + k_2 + \alpha), \\
C &= \beta^2 \times (k_1 \times k_2 + k_1 \times \alpha + k_2 \times \alpha), \\
D &= \beta^3 \times k_1 \times k_2 \times \alpha + \beta^2 \times \gamma \times \alpha \times k_2.
\end{aligned}$$

The discriminant  $\Delta_J$  of  $J$  is now given by

$$\Delta_J = Q^3 + R^2, \quad (\text{A7})$$

where

$$\begin{aligned}
Q &= \frac{3 \times A \times C - B^2}{9 \times A^2}, \\
R &= \frac{9 \times A \times B \times C - 27 \times A^2 \times D - 2 \times B^3}{54 \times A^3}.
\end{aligned}$$

The discriminant is a function of the model parameters  $\beta$  and  $\gamma$  as well as the model variables  $k_1$ ,  $k_2$  and  $\alpha$ . Setting the model variables to the values mentioned in Chapter 2 and a numerical evaluation on a fine grid shows that  $\Delta_J$  is non-positive if the inequality

$$\gamma \leq 0.26316 \times \beta - 0.0003158 \quad (\text{A8})$$

is fulfilled. As small oscillations are plausible in a physiological system, we observed that the relaxed inequality

$$\gamma \leq 0.5 \times \beta \quad (\text{A9})$$

is adequate for further numerical simulation.

Another important property of a continuous model is the stability of the considered steady state. A steady state is (asymptotically) stable iff the real parts of each eigenvalue are negative (non-positive). In case of small variations from the steady state, positive real parts in some eigenvalues might lead to divergence of the trajectory.

Using the Cardano formula, eigenvalues of  $\mathcal{X}_J$  are given by

$$\begin{aligned}
\lambda_1 &= S + T + \frac{B}{3A}, \\
\lambda_2 &= -\frac{S+T}{2} - \frac{B}{3A} + \frac{i\sqrt{3}(S-T)}{2}, \\
\lambda_3 &= -\frac{S+T}{2} - \frac{B}{3A} - \frac{i\sqrt{3}(S-T)}{2},
\end{aligned}$$

where

$$\begin{aligned}
S &= \sqrt[3]{R + \sqrt{\Delta_J}}, \\
T &= \sqrt[3]{R - \sqrt{\Delta_J}}.
\end{aligned}$$

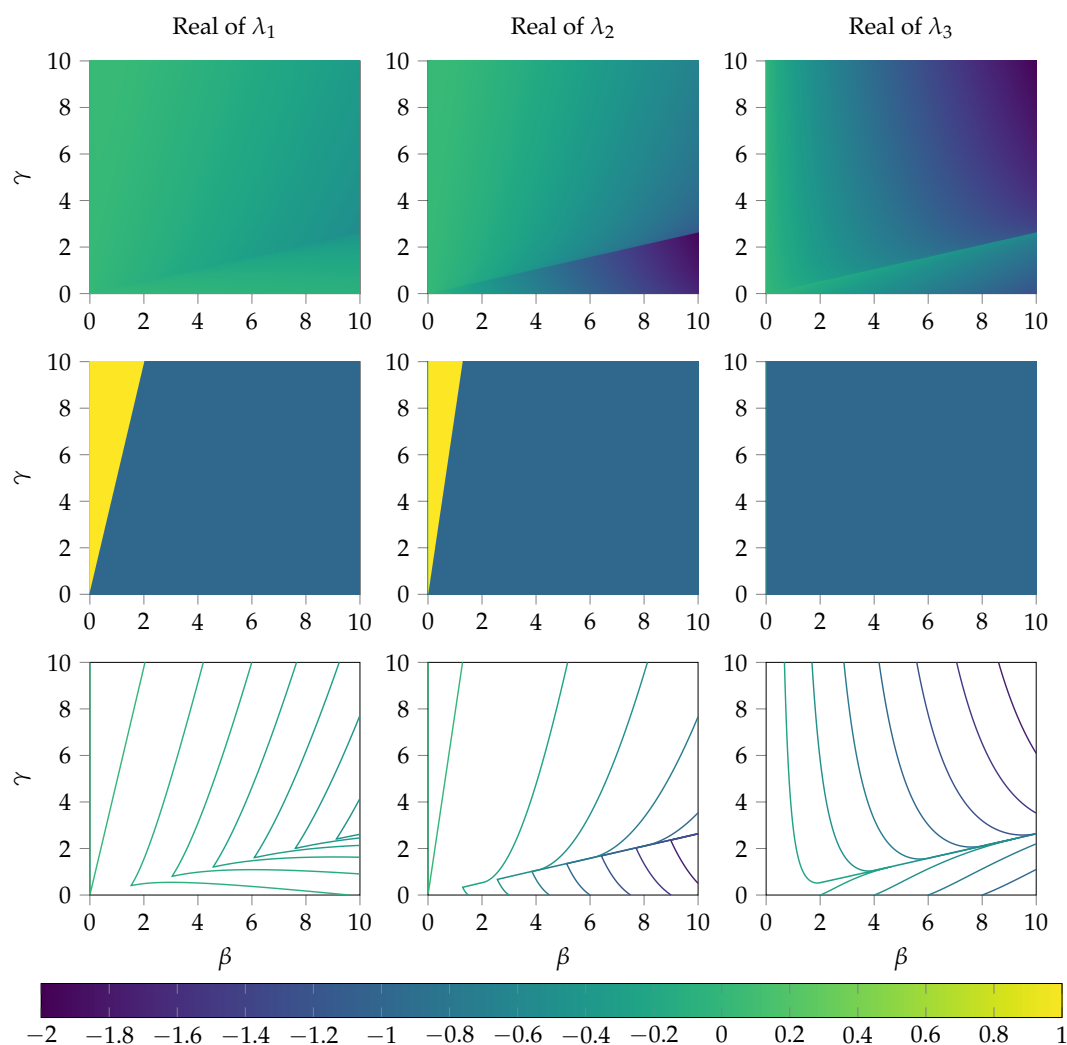
Figure A1 presents the numerical evaluation of the real parts of  $\lambda_1, \lambda_2$  and  $\lambda_3$  on a fine grid. The contour lines of the roots highlight the nonlinear behavior of the roots. Nonetheless, in the (physiologically) relevant interval of  $\beta$  and  $\gamma$ , the negativity of the roots can be described approximately by a linear term, as can also be seen in Figure A2: the real part of the eigenvalue  $\lambda_1$  is negative if inequality

$$\gamma \leq 4.9 \times \beta \quad (\text{A10})$$

and the real part of  $\lambda_2$  is negative if inequality

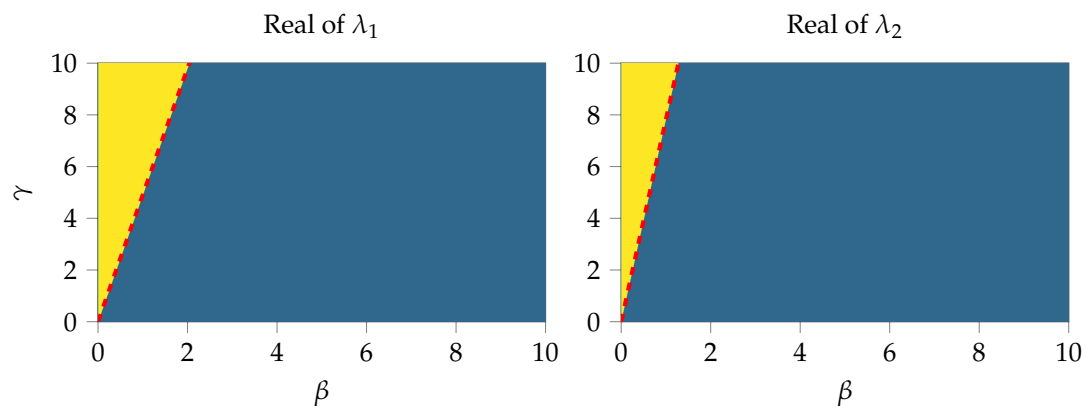
$$\gamma \leq 7.8125 \times \beta \quad (\text{A11})$$

is fulfilled. The real part of  $\lambda_3$  is always negative for non-negative  $\beta$  and  $\gamma$ . If Inequality (9) is fulfilled, both Inequalities (A10) and (A11) are valid. Therefore, no further condition is necessary to guarantee stability of the steady states.



**Figure A1.** Numerical evaluation of the real parts of  $\lambda_1, \lambda_2$  and  $\lambda_3$  on a grid of  $1000 \times 1000$  data points for  $\beta, \gamma \in [0, 10]$ . The top row shows the values of the three roots. In the second row, the sign of the values is presented (negative corresponds to blue, positive to yellow). The third row contains the contour lines of the three roots.





**Figure A2.** As in second row of Figure A1. Additionally, the equality conditions of Inequalities (A10) and (A11) have been included in red.

## Appendix B. Detailed Results

**Table A1.** Results with two free parameters (including standard deviations (SD)).

Subject ID	$\gamma$	SD( $\gamma^*$ ) [%]	$\beta$	SD( $\beta$ ) [%]	$R^2$
01	0.769	13.3	1.65	26.2	0.961
02	0.388	12	0.867	24.4	0.996
03	0.51	15.3	1.617	28.7	0.915
04	0.323	12.4	0.424	24.2	0.801
05	0.061	341.6	1.381	84.9	0.61
06	0.59	20.6	2.615	33.4	0.915
07	0.262	67.9	1.518	69.7	0.887
08	0.324	53.3	2.676	49.1	0.911
09	0.356	16.6	0.891	29.7	0.806
10	0.089	242.2	2.557	61	0.947
11	0.243	20.7	0.925	29	0.897
12	1.003	12.5	1.409	20.1	0.959
13	0.057	178.1	0.879	67.8	0.895
14	0.762	42.7	0.46	53.6	0.507
15	0.344	33.1	2.132	40.9	0.863
16	0.141	257.3	1.661	116.9	0.344
17	0.47	66.2	0.544	69.1	0.842
18	0.525	8.8	0.631	20.1	0.871
19	0.423	44.7	1.525	56.2	0.624
20	0.661	31.1	2.798	47.9	0.812
21	0.686	20.9	1.943	36.7	0.92
23	0.613	36.1	3.142	49	0.897
24	0.421	26.9	1.528	42.1	0.905
25	0.863	15.9	2.078	24.7	0.868
26	0.414	16.3	1.172	29.8	0.703
27	0.635	8.2	0.836	16.8	0.866
28	0.952	15.8	1.596	26.1	0.992
29	0.805	22.2	1.486	39.9	0.925

**Table A2.** Results with Base and  $\beta$  as free parameters ( $\gamma = 0.5 \cdot \beta$ ).

Subject ID	Base	SD(Base) [%]	$\beta$	SD( $\beta$ ) [%]	$R^2$
01	861.014	1.3	1.685	21.2	0.963
02	882.041	1.2	0.838	17.6	0.996
03	856.779	0.7	1.296	14.4	0.907
04	881.885	3.3	0.431	22.9	0.806
05	930.774	1.2	1.306	44.4	0.595
06	986.155	0.5	1.985	12.2	0.899
07	954.28	2.1	0.968	43	0.78
08	692.458	0.8	1.286	14.1	0.861
09	956.621	1.4	0.787	21.2	0.803
10	815.247	0.9	1.345	19.1	0.897
11	980.574	1.1	0.851	17.7	0.902
12	936.771	0.7	1.665	13.9	0.949
13	611.74	2.5	0.64	45.3	0.876
14	1554.83	69	0.324	75.7	0.495
15	922.955	0.5	1.466	13.1	0.833
16	726.472	1.1	2.015	54	0.426
17	973.967	29.8	0.484	100.7	0.842
18	874.356	2.9	0.661	19.9	0.876
19	766.033	2.4	1.569	47	0.622
20	758.275	0.6	2.055	16	0.803
21	896.036	1	2.001	23.9	0.928
23	880.509	0.6	2.245	17.5	0.904
24	680.758	2.3	1.304	33.2	0.895
25	772.443	0.7	1.675	12.6	0.87
26	686.724	1.8	0.925	21.1	0.695
27	947.998	1.5	0.859	16.1	0.884
28	872.556	1.1	1.684	21.1	0.991
29	812.567	3	1.478	41.6	0.925

## References

- Jandl, J.H. *Blood*, 1st ed.; Little Brown: Boston, MA, USA, 1987.
- Lichtman, M.A.; Beutler, E.; Kipps, T.J.; Seligsohn, U.; Kaushansky, K.; Prchal, J.T. *Williams Hematology*; McGraw-Hill: New York, NY, USA, 2006.
- Mackey, M.C.; Glass, L. Oscillation and chaos in physiological control systems. *Science* **1977**, *197*, 287–289. [[CrossRef](#)] [[PubMed](#)]
- Loeffler, M.; Pantel, K.; Wulff, H.; Wichmann, H. A mathematical model of erythropoiesis in mice and rats Part 1: Structure of the model. *Cell Prolif.* **1989**, *22*, 13–30. [[CrossRef](#)]
- Bélair, J.; Mackey, M.C.; Mahaffy, J.M. Age-structured and two-delay models for erythropoiesis. *Math. Biosci.* **1995**, *128*, 317–346. [[CrossRef](#)]
- Mahaffy, J.M.; Polk, S.W.; Roeder, R.K. *An Age-Structured Model for Erythropoiesis Following a Phlebotomy*; Department of Mathematical Sciences, San Diego State University: San Diego, CA, USA, 1999.
- Ackleh, A.; Thibodeaux, J. Parameter estimation in a structured erythropoiesis model. *Math. Biosci. Eng.* **2008**, *5*, 601–616. [[CrossRef](#)] [[PubMed](#)]
- Fuertinger, D.H.; Kappel, F.; Thijssen, S.; Levin, N.W.; Kotanko, P. A model of erythropoiesis in adults with sufficient iron availability. *J. Math. Biol.* **2013**, *66*, 1209–1240. [[CrossRef](#)] [[PubMed](#)]
- Edelstein-Keshet, L. *Mathematical Models in Biology*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2005; Volume 46, pp. 27–33.
- Fonseca, L.L.; Voit, E.O. Comparison of mathematical frameworks for modeling erythropoiesis in the context of malaria infection. *Math. Biosci.* **2015**, *270*, 224–236. [[CrossRef](#)] [[PubMed](#)]
- Friberg, L.E.; Henningsson, A.; Maas, H.; Nguyen, L.; Karlsson, M.O. Model of chemotherapy-induced myelosuppression with parameter consistency across drugs. *J. Clin. Oncol.* **2002**, *20*, 4713–4721. [[CrossRef](#)] [[PubMed](#)]

12. Rinke, K.; Jost, F.; Findeisen, R.; Fischer, T.; Bartsch, R.; Schalk, E.; Sager, S. Parameter estimation for leukocyte dynamics after chemotherapy. *IFAC-PapersOnLine* **2016**, *49*, 44–49. [[CrossRef](#)]
13. Jost, F.; Rinke, K.; Fischer, T.; Schalk, E.; Sager, S. Optimum experimental design for patient specific mathematical leukopenia models. *IFAC-PapersOnLine* **2016**, *49*, 344–349. [[CrossRef](#)]
14. Crauste, F.; Demin, I.; Gandrillon, O.; Volpert, V. Mathematical study of feedback control roles and relevance in stress erythropoiesis. *J. Theor. Biol.* **2010**, *263*, 303–316. [[CrossRef](#)] [[PubMed](#)]
15. Mackey, M.C. A unified hypothesis for the origin of aplastic anemia and periodic haematopoiesis. *Blood* **1978**, *51*, 941–956. [[PubMed](#)]
16. Noble, S.L.; Sherer, E.; Hannemann, R.E.; Ramkrishna, D.; Vik, T.; Rundell, A.E. Using adaptive model predictive control to customize maintenance therapy chemotherapeutic dosing for childhood acute lymphoblastic leukemia. *J. Theor. Biol.* **2010**, *264*, 990–1002. [[CrossRef](#)] [[PubMed](#)]
17. Pottgiesser, T.; Specker, W.; Umhau, M.; Dickhuth, H.H.; Roecker, K.; Schumacher, Y.O. Recovery of hemoglobin mass after blood donation. *Transfusion* **2008**, *48*, 1390–1397. [[CrossRef](#)] [[PubMed](#)]
18. Ertl, A.C.; Diedrich, A.; Raj, S.R. Techniques Used for the Determination of Blood Volume. *Am. J. Med. Sci.* **2007**, *334*, 32–36. [[CrossRef](#)] [[PubMed](#)]
19. Otto, J.M.; Plumb, J.O.; Clissold, E.; Kumar, S.B.; Wakeham, D.J.; Schmidt, W.; Grocott, M.P.; Richards, T.; Montgomery, H.E. Hemoglobin concentration, total hemoglobin mass and plasma volume in patients: implications for anemia. *Haematologica* **2017**, *102*, 1477–1485. [[CrossRef](#)] [[PubMed](#)]
20. Passamonti, F.; Malabarba, L.; Orlandi, E.; Baràtè, C.; Canevari, A.; Brusamolino, E.; Bonfichi, M.; Arcaini, L.; Caberlon, S.; Pascutto, C.; et al. Polycythemia vera in young patients: a study on the long-term risk of thrombosis, myelofibrosis and leukemia. *Haematologica* **2003**, *88*, 13–18. [[PubMed](#)]
21. Siegel, F.P.; Tauscher, J.; Petrides, P.E. Aquagenic pruritus in polycythemia vera. *Am. J. Hematol.* **2013**, *88*, 665–669. [[CrossRef](#)] [[PubMed](#)]
22. Marchioli, R. Vascular and Neoplastic Risk in a Large Cohort of Patients With Polycythemia Vera. *J. Clin. Oncol.* **2005**, *23*, 2224–2232. [[CrossRef](#)] [[PubMed](#)]
23. Ziegler, A.K.; Grand, J.; Stangerup, I.; Nielsen, H.J.; Dela, F.; Magnussen, K.; Helge, J.W. Time course for the recovery of physical performance, blood hemoglobin, and ferritin content after blood donation. *Transfusion* **2015**, *55*, 898–905. [[CrossRef](#)] [[PubMed](#)]
24. Wide, L.; Bengtsson, C.; Birgegård, G. Circadian rhythm of erythropoietin in human serum. *Br. J. Haematol.* **1989**, *72*, 85–90. [[CrossRef](#)] [[PubMed](#)]
25. Marciniak-Czochra, A.; Stiehl, T.; Ho, A.D.; Jäger, W.; Wagner, W. Modeling of asymmetric cell division in hematopoietic stem cells—Regulation of self-renewal is essential for efficient repopulation. *Stem Cells Dev.* **2009**, *18*, 377–386. [[CrossRef](#)] [[PubMed](#)]
26. Spiegel, M.R. *Mathematical Handbook of Formulas and Tables*; McGraw-Hill: New York, NY, USA, 1968.
27. Schmidt, W.; Prommer, N. The optimised CO-rebreathing method: A new tool to determine total haemoglobin mass routinely. *Eur. J. Appl. Physiol.* **2005**, *95*, 486–495. [[CrossRef](#)] [[PubMed](#)]
28. Masoud, M.; Sarig, G.; Brenner, B.; Jacob, G. Orthostatic Hypercoagulability. *Hypertension* **2008**, *51*, 1545–1551. [[CrossRef](#)] [[PubMed](#)]
29. Wadsworth, G.R. Recovery from acute haemorrhage in normal men and women. *J. Physiol.* **1955**, *129*, 583–593. [[CrossRef](#)] [[PubMed](#)]
30. Bock, H.G.; Körkel, S.; Schlöder, J.P. Parameter Estimation and Optimum Experimental Design for Differential Equation Models. In *Model Based Parameter Estimation: Theory and Applications*; Bock, H.G., Carraro, T., Jäger, W., Körkel, S., Rannacher, R., Schlöder, J.P., Eds.; Springer: Heidelberg, Germany, 2013; pp. 1–30.
31. Convertino, V.A. Blood Volume Response to Physical Activity and Inactivity. *Am. J. Med. Sci.* **2007**, *334*, 72–79. [[CrossRef](#)] [[PubMed](#)]
32. Eaves, C.J.; Eaves, A.C. Erythropoietin (Ep) dose-response curves for three classes of erythroid progenitors in normal human marrow and in patients with polycythemia vera. *Blood* **1978**, *52*, 1196–1210. [[PubMed](#)]

