UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA INDUSTRIALE

CORSO DI LAUREA MAGISTRALE IN INGEGNERIA CHIMICA E DEI PROCESSI INDUSTRIALI

Tesi di Laurea Magistrale in
Ingegneria Chimica e dei Processi Industriali

A Laplace domain approach to the modeling of von Willebrand disease

Relatore: Prof. Massimiliano Barolo
Correlatore: Prof. Fabrizio Bezzo

Laureando: NICOLAS GARCIA SCAGLIONI

ANNO ACCADEMICO 2019 – 2020

Abstract

Von Willebrand disease is an inherited disease that affects the blood clotting process and impacts on 1% of the world's population. Caused by a von Willebrand factor (vWF) deficiency and/or dysfunction, it differs in typologies with intra and inter-individual variability, and is complex to diagnose. Research shows how the use of mathematical models capable of representing the fundamental physiological mechanisms can aid the diagnosis. This study aims to give a new perspective to solve the mechanistic models of vWD. This allows the automatic detection of the disease, a quantitative evaluation of vWF multimers distribution, elucidating the critical pathways involved in the recognition and characterization of the disease.

The Laplace Transform was used to solve and analyze the differential equation model for vWD. This method converts differential equations into algebraic equations and defines new parameters within the Laplace domain. Parameters determined in the Laplace domain were identified using clinical patient data. Principal component analysis (PCA) transformed the parameters identified into a smaller number of principal components to distinguish healthy subjects from vWD patients. The results show that the parameters provide new multimers information that was not visible in the original mathematical model. The PCA model identified 70% of the cases of patients with vWD.

Contents

Introduction

The von Willebrand disease (vWD) is one of the most common hereditary diseases that affect the blood clotting process. It is caused by a deficiency or malfunction of the von Willebrand factor (vWF), a large multimer present in the coagulation process. Mutations allow to classify the disease in different types and this creates complexity in the diagnosis process. The typical symptoms of this disease include bleeding from mucous membranes and minor skin lesions, reaching almost 1% of the world's population.

There are research showing the use of a mathematical model able to represent the multimers physiological mechanisms to aid with the diagnosis. These models are a set of differential equations used to describe multimers profile through time. This study aims to provide a new perspective to solve these mathematical models. The strategy includes the use of Laplace transform technique to solve and analyze the vWD model. This method converts differential equations into algebraic equations and defines new parameters within the Laplace domain. Simulations with the software Simulink (a Matlab® tool) will check the new model and parameters defined within the Laplace domain.

The mathematical model must represent the real system precisely and accurately. The values of the parameters that best fit the real data establish the model quality. Therefore, the model in the Laplace domain must have the new parameters identified in order to reach out the experimental data. The estimated parameters must provide new information regarding the vWF multimers distribution, disease recognition and characterization. To do that, optimization conditions and techniques will be considered to the parameter identification process, providing access to the estimability and identifiability of the model.

The presence of a lot of parameters and variables can create complications when extracting process information, especially in this study aimed to use the model parameters to reach knowledge on the disease diagnosis. The principal components analysis (PCA) is a tool for data compression and information extraction through finding variables combinations that describe major trends in the data. This way, the vWD model parameters can be described in fewer principal components that can be used to help create a diagnosis and achieve this work objectives. The main results obtained, and the criticisms found will be summarized, with considerations about possible future works.

Chapter 1

von Willebrand disease

It is essential to begin this study by explaining the von Willebrand disease (vWD), including the discussion about the von Willebrand factor and its role in the bleed coagulation process. The Chapter starts by analysing the haemostasis and coagulation cascade, followed by the description of the von Willebrand factor role in the mentioned biochemical processes and the types of diseases it could lead to. This Chapter concludes with a discussion on diagnosis for von Willebrand disease and reinforcing this work objectives.

1.1 Introduction

The von Willebrand disease (vWD) is an inherited bleeding disorder discovered for the first time on a 5-year-old girl in 1924 by Dr. Erik von Willebrand in Finland. By examining the girl and members of her family, von Willebrand submitted in 1926 an unrecorded inherited bleeding disorder. At that time he did not have the tools to understand the illness source, but he recognized the symptoms and solved the anemia and bleeding disorder, caused by the disease, with blood transfusions. Only in the 1950's it was discovered that a deficiency or dysfunction of a plasma protein was the cause of the bleeding disorder discovered by Dr. von Willebrand. This protein, named von Willebrand Factor (vWF), mediates the adhesion of platelets at spots of vascular wound, it also binds and stabilizes the blood clotting factor VIII (FVIII) (NHLBI, 2007).

Later, gene sequencing allowed to recognize a greater number of vWD cases and to find variant forms of vWF abnormality. The disease has been reported around many countries, but it is unknown how many people have it, because of diagnose procedure discrepancies and the symptoms might mislead for another bleeding disorder. This highlights the need to improve the vWD clinical data, laboratory diagnostic knowledge and tools (NHLBI, 2007). The prevalence reports published by Rodeghiero et al. (1987) and (NHLBI, 2007) revealed a prevalence rate varying between 0.6 and 1.3 percent of the world population suffering from vWD.

1.2 Hemostasis

After a vascular wound, in order to keep the blood inside the damaged blood vessel, the human body triggers the mechanism called hemostasis (or haemostasis), which is a process that causes the bleeding to stop. The hemostasis happens by forming a clot to curb hemorrhage by a synchronous and sequential process involving blood vessels, platelets, coagulation factors and fibrinolytic factors. The hemostasis process is divided in the following steps: vascular contraction, platelet plug formation, blood clotting followed by an eventual fibrous tissue growth in the clot for closure.

Immediately after injuring the blood vessel, the vascular wall trauma causes the smooth musculature of this wall to contract, in order to reduce temporarily the blood flow through the damaged vessel. Then a mechanical block caused by platelet adhesion begins, when platelets, which normally float in the plasma, bond to the wound area. After that, agglomerated platelets become pointed and sticky and bind to the exposed collagen and endothelial coating, causing the release of granules that contain several active platelet factors. Such effect triggers the von Willebrand factor from plasma into traumatized tissue. All of that stimulate the platelets aggregation, thus forming a platelet cap. The collagen and exposed tissue engages a mixture of proteins and phospholipids that act to assemble a fibrin protein network that stabilizes the platelet cap and forms a clot. The fibrin is the end product of a series of enzyme reactions called coagulation cascade (Hall and Guyton, 2015). Figure 1.1 illustrates the described steps precisely.

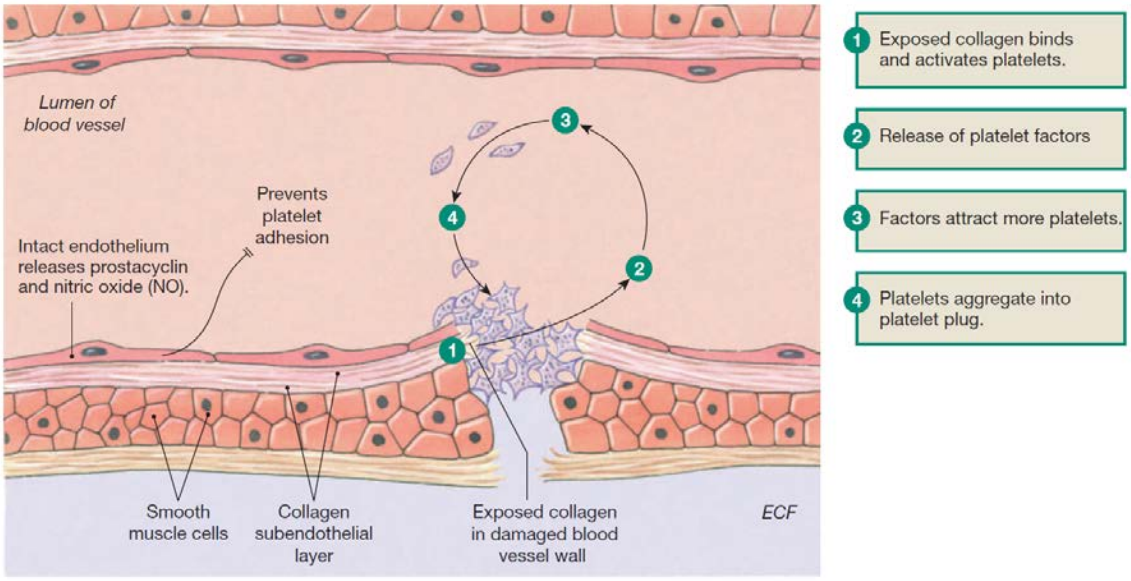


Figure 1.1. *Figure illustrating the platelet plug formation (Silverthorn et al., 2016).*

1.2.1 Blood coagulation cascade

The coagulation is the third major step in hemostasis, where the blood fluid forms a gelatinous clot. The beginning of the coagulation process divides into two pathways that eventually converge into a common one: an intrinsic pathway that starts in the blood when damaged tissue exposes collagen, and an extrinsic pathway that starts with the trauma of the vascular wall and surrounding tissues. At each step, an enzyme converts an inactive precursor into an active enzyme, often with the help of calcium, membrane phospholipids, or other factors. The intrinsic and extrinsic pathway factors interact like a net, instead of a simple cascade (Silverthorn et al., 2016).

The intrinsic pathway, also known as contact activation, uses proteins already present in plasma. Exposure of damaged vascular wall collagen to the blood alters Factor XII and platelets. Factor XII becomes Factor XII activated and, simultaneously, the trauma also damages the platelets, due to its adhesion to collagen causing the release of platelet phospholipids which contain lipoprotein called platelet Factor III that also has participation in later coagulation reactions. The Factor XII with Calcium activates the Factor XI and the Factor XI activates the Factor IX. The Factor IX along with platelet phospholipids, traumatized platelet Factor III and Factor VIII (carried by vWF) activates the Factor X (Hall and Guyton, 2015).

The extrinsic pathway begins when the damaged tissues expose the tissue factor, also called Factor III. The extrinsic pathway is also called the cellular lesion pathway or via the tissue factor. The tissue factor combines with the blood coagulation factor VII and, in the ions calcium presence , acts enzymeatically on the Factor X to form the Factor X activated (Hall and Guyton, 2015).

Activation of the Factor X happens in the intrinsic and extrinsic pathway. The Factor X activated immediately combines with Factor V and with tissue phospholipids to form the prothrombin activator complex. The prothrombin activator, with ion calcium (Ca++) presence, causes prothrombin (present in plasma) convert to thrombin. Thrombin causes polymerisation of fibrinogen molecules in fibrin fibers. Fibrin fibers permeate the platelet buffer and keep erythrocytes within its mesh. Active factor XIII converts fibrin into a cross-linked polymer, which stabilizes the clot. Figure 1.2 extracted from Silverthorn et al. (2016) summarizes the blood coagulation process described by showing both the intrinsic and extrinsic pathways merging into the common pathway to produce fibrin fiber.

1.3 von Willebrand Factor

The endothelial and megakaryocytes cells synthesize the von Willebrand factor (vWF). It circulates as a multimer, a protein complex formed by several subunits, of order size of 500 - 20,000 kDaltons, depending on the number of monomers. In plasma, vWF multi-

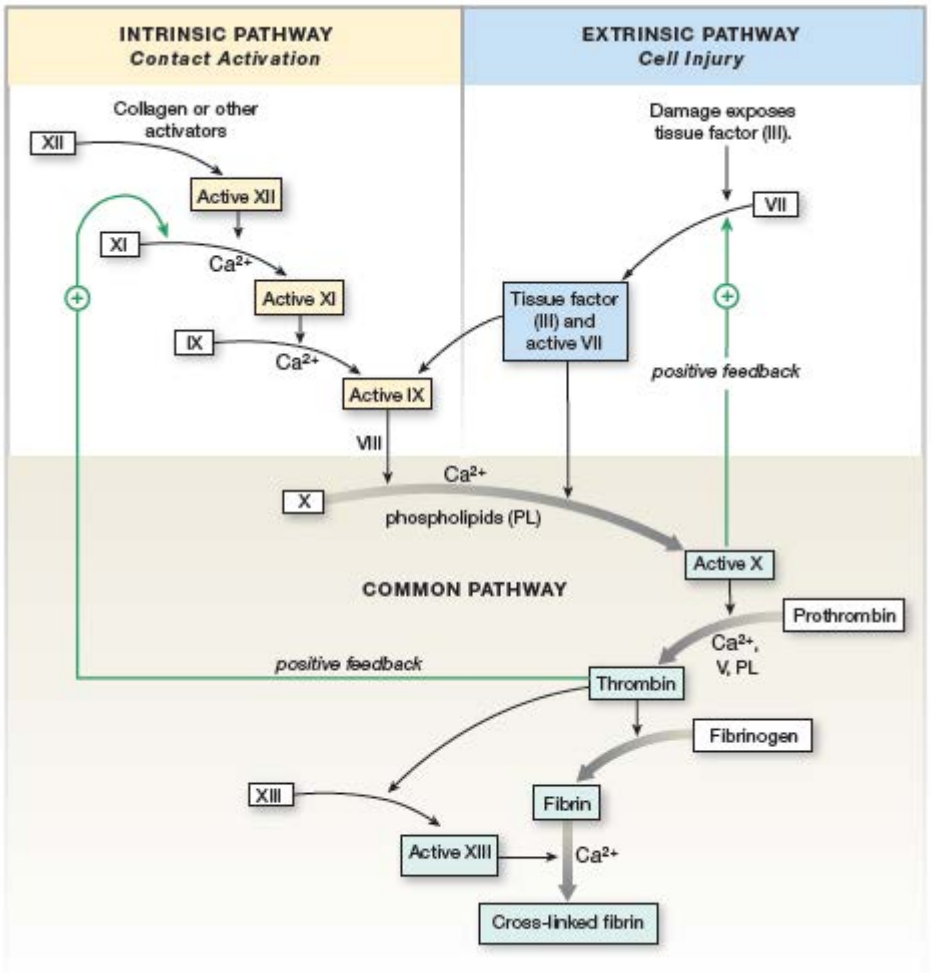


Figure 1.2. *Figure summarizing the Blood Coagulation Cascade (Silverthorn et al., 2016).*

mers are eliminated with a half-life of 12 - 20 hours by an independent size mechanism. The endothelial cells produce vWF to keep constant the basal plasma concentration and to stored in Weibel-Palade bodies in case of released caused by vascular injury or pharmacological stimulation response, increasing the level of vWF in the plasma. The first synthesized vWF units have multimers with ultra large molecular weight. A specific protein called ADAMTS13 (A Disintegrin-like And Metalloprotease domain with ThromboSpondin type I motifs) cleaves the ultra large molecular weight multimers (ULMW) into high molecular weight (HMW) and low molecular weight multimers (LMW) (Sadler, 2008).

The von Willebrand factor acts in two moments of homeostasis; in the platelet adhesion with the damaged vasculature and as a protein carrier for Factor VIII during the coagulation process. The interaction between platelet and damaged vessel are more efficient with ULMW and HMW multimers, because the vWF binds to the collagen exposed in the damaged region creating a binding site for the platelet receptor, serving as aggre-

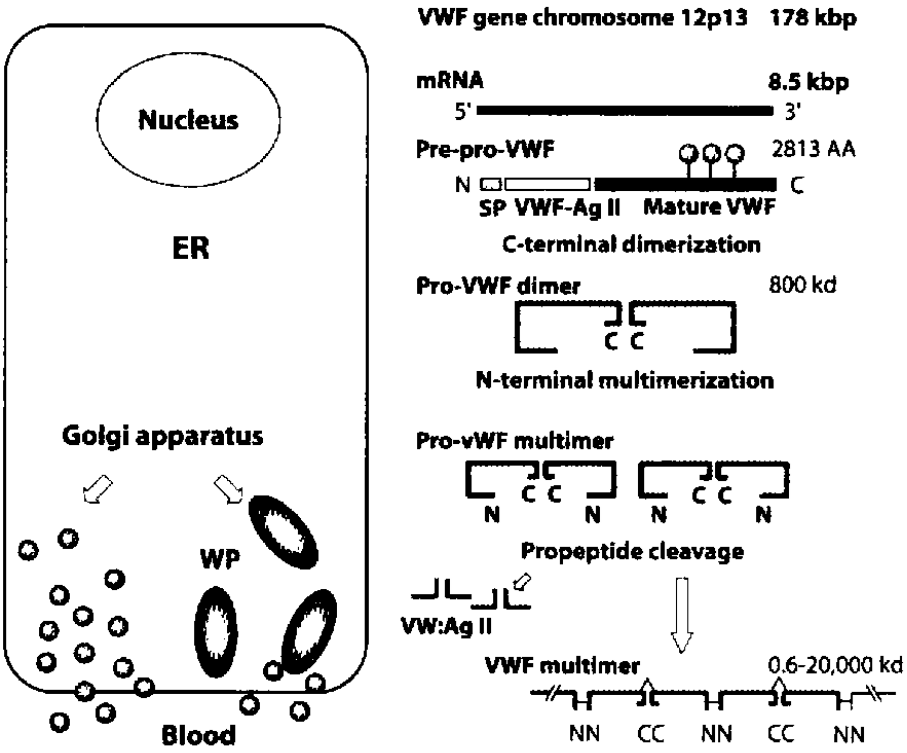


Figure 1.3. *Scheme of vWF biosynthesis in the endothelial cell, where ER stands for endoplasmatic reticulum and WP for Weibel-Palade bodies, which store the synthezied vWF multimers, on the right shows the synthesis and polymerization steps (Schneppenheim and Budde, 2008) .*

gator of platelet and damage vessels. The vWF allows platelets adhesion through their superficial binding glycoprotein Ib (GPIb). Once bonded, the platelets are then activated, passing from the disc shape to the spherical shape with high surface area.

When the vWF acts as Factor VIII carrier, it creates a FVIII-vWF complex, which circulates in plasma as a loosely coiled protein complex that does not interact strongly with platelets or endothelial cells under basal conditions. In regard of this functionality, the multimers size is not relevant. Additionally, the vWF also protects the Factor VIII from protein C and S action, increasing the stability and half-life of the factor (Schneppenheim and Budde, 2008).

Figure 1.4 illustrates the vWF functions. The top image shows the vWF as FVIII carrier where the FVIII-vWF complex are circulating in plasma. The middle image shows the vWF interaction with platelets and damage vessel. The bottom image shows the clotting and platelet aggregation to stop bleeding.

1.4 von Willebrand disease

The disease is the result of a decrease in the vWF levels or a vWF compromised functionality. Some factors influence the vWF levels: age, ethnic group, blood type and

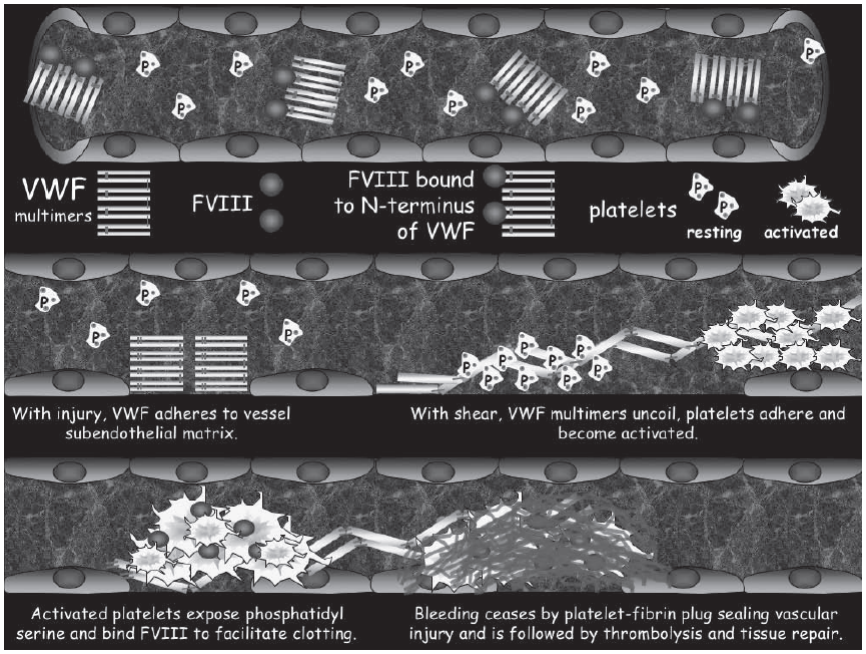


Figure 1.4. *Figure to illustrate the vWF function. The top image shows the vWF as FVIII carrior, the middle image shows the vWF as platelet and damage vessel adhesion, and the bottom image shows the clotting and platelet aggregation to stop bleeding.(NHLBI, 2007)*

hormones. While the vWF synthesis rate remains constant between ABO groups, survival rate reduces if the subject belongs to group O, compared to non-O subjects Gallinaro et al. (2008). The vWD is a very heterogeneous disease characterized by a strong variability classified in types: Type 1, Type 2, Type 3 and Vicenza.

Type 1 is characterized by a partial quantitative deficiency, type 2 by a qualitative deficiency, and type 3 by a total deficiency. vWD type 2 divides into four variants (2A, 2B, 2M, 2N), depending on the qualitative deficiency. Type Vicenza is a specific variation commonly classified under type 1 or type 2M, depending on laboratory test interpretation. According to NHLBI (2007), about 75 percent of symptomatic persons who have vWD are from type 1 and the remaining 25 percent are from type 2, divided between the variants and the prevalence for type 3 is unknown.

1.4.1 vWD type 1

Type 1 vWD includes partial quantitative deficiency of vWF, attributed to a decrease in vWF plasma concentration. They do not show multimers functional abnormalities or any type of selective reduction in large multimers. According to NHLBI (2007), laboratory analysis shows reductions in the antigen vWF protein concentration (VWF:Ag) and collage binding (VWF: RCo). Most people who have type 1 vWD result in normal FVIII and the gels of the vWF multimers do not show a significant decrease in vWF large multimers.

Severe type 1 deficiency relates to the vWF mutations that interfere with intracellular transport of pro-vWF 35-39 dimeric, affecting the release of vWF into the blood. Other deficiency relates to vWF mutations that promote rapid multimers clearance from circulation. People with these mutations usually have low plasma vWF concentration. Increased vWF clearance explains patients who show exaggerated and brief responses to DDAVP administration.

It is difficult to identify vWD type 1 when vWF plasma concentration is not considerably low, and is also close to the lower end of the normal range. Type 1 lacks a qualitative criterion easily to recognise, it is advisable to rely on qualitative decreases in protein concentration, instead. On the basis of vWF levels, the healthy population covers a large range of values and depends on a lot of factors. The symptoms of moderate bleeding are very common in the healthy population, associating bleeding symptoms with a vWF moderately low-level may coincide and be mistakenly correlated with vWD.

1.4.2 vWD type 2

Type 2A refers to qualitative variants caused by a reduced concentration of vWF multimers with high molecular weight, affecting the vWF platelet adhesion. The vWF antigen and FVIII levels might be normal but the functionality can be compromised, which is represented by vWF:RCo. The abnormality might associate with the assembly or secretion of large multimers or by the increase of proteolytic degradation.

The vWD sub type 2B includes vWF variants with increased affinity for platelet glycoprotein Ib. This variant of the disease does not hold the synthesis of large multimers, but after secretion multimers spontaneously bind to platelets and are cleaved by ADAMTS-13. This platelet binding characterize the increased interaction of mutant vWF with platelet glycoprotein Ib. Thus, most patients with vWD sub type 2B have a lower plasma concentration of ultra large vWF multimers and show a high rate of proteolysis.

The vWD type 2M includes impaired interaction of vWF with platelet GPIb or with connective tissue causing decreased platelet and vWF adhesion. This variation does not substantially affect multimer assembly but a mutation in vWF domain A1 shows impair binding of vWF to collagen. Laboratory analysis shows type 2M and type 2A are very similar, being differentiated by multimer gel electrophoresis. Type 2N includes vWF with flawed FVIII binding, which is commonly mistaken for haemophilia, because presents normal values for vWF:Ag and vWF:RCo (Silverthorn et al., 2016).

1.4.3 vWD type 3

vWD type 3 characterizes complete deficiency of vWF. According to NHLBI (2007), usually the vWF:RCo, vWF:CB and vWF:Ag are lower than 5 IU/dL, considered unde-

tectable vWF levels, and FVIII levels are lower than 10 IU/dL, and patients rarely have a measurable response after DDAVP desmopressin administration. The terminology ”Severe vWD” has been used for vWD type 3 and sometimes for symptomatic vWD type 1 characterized by very low vWF levels, but these conditions are almost always clinically distinct.

1.4.4 vWD type Vicenza

This variant is often classified under type 1 or type 2M, but despite this the vWD type Vicenza has an extreme increase of vWF clearance, causing a markedly short half-life of vWF plasma concentration. vWD Vicenza has vWF:RCo and vWF:Ag levels decreased proportionately.

Increased clearance can explain the ultra-large multimer distribution for vWD type Vicenza. A faster clearance rate means shorter time for VWF multimer circulates in plasma, thus less opportunity to be cleaved by ADAMTS-13. Consequently, increased clearance shifts vWF multimer plasma distribution toward ultra-large multimers and few smaller sub units (NHLBI, 2007) (Sadler, 2008).

1.5 Diagnosis

The diagnosis of vWD includes the personal history and the quantitative data, and it segments in three conditions: bleeding history; family history of hemorrhagic disorder; and laboratory tests demonstrating a quantitative and /or qualitative defect of the vWF. Prior to laboratory analysis, which is long and stressful, it is recommended a first evaluation on the patient history for excessive bleeding and any family history.

Preliminary laboratory evaluation of haemostasis includes platelet count, partial thromboplastin time, prothrombin time and evaluation of fibrinogen or thrombin time. If the history of the subject shows a hemorrhagic disorder or abnormal coagulating activity, it is better to proceed with a vWD investigation, which starts by measuring indicators that highlight the disease: the amount of vWF in plasma measured from the vWF antigen (VWF:Ag); the vWF functionality measured by the dosage of the ristocetin co-factor (VWF:RCo) and dosage of the coagulant factor FVIII (FVIII:C). When all dosage levels decrease, vWD type 1 is suspected. When there is a disproportionate decrease in vWF:RCo or FVIII, type 2 is suspected and the virtual absence of vWF:Ag suggest vWD type 3 (NHLBI, 2007). Figure 1.5 shows the electrophoresed image for multimer patterns in vWD.

The vWF:Ag gives a measure of the vWF plasma antigen concentration, this indicator provides the overall plasma vWF concentration, independent of the size of the vWF multimers, and is generally based on the ELISA method (Enzyme-Linked ImmunoSorbent

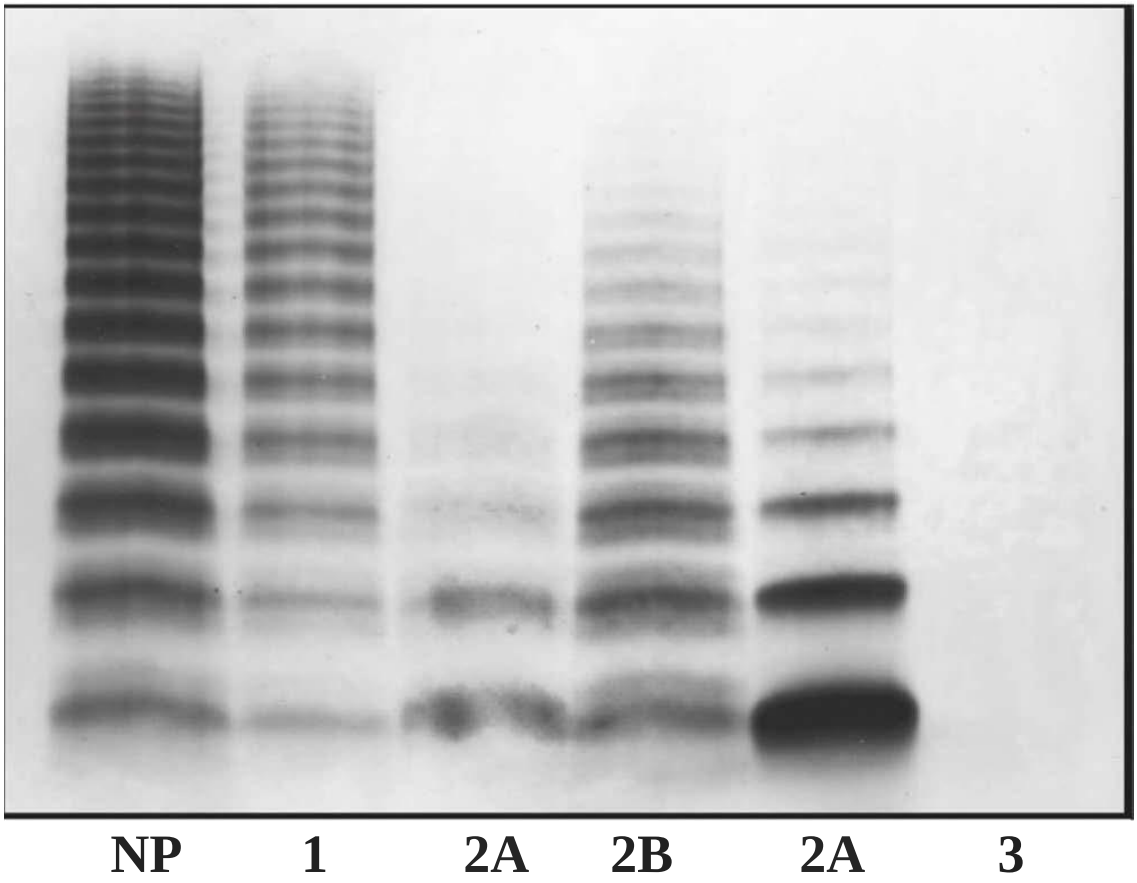


Figure 1.5. *Figure shows the electrophoresed image for multimer patterns in vWD type 1, type 2A (two unrelated patient), type 2B, type 3, and a normal person(NP) (Marder et al., 2013) .*

Assay).

The vWF:RCo test measures ristocetin cofactor activity, testing the vWF functionality to interact with platelets. This test takes a long time to perform and causes thrombocytopenia, because the vWF adheres to platelets generating an aggregate, which removes the platelets from the circulation. For these reasons, researchers replace the vWF:RCo test for the vWF collagen test (vWF:CB) (Sadler, 2008).

The vWF:CB measures the vWF ability to bind to endothelial collagen type I and III. This depends on the multimeric size of the vWF, with larger multimers having a greater affinity for collagen than smaller ones. Multimers of higher molecular weight are the most active and this test provides indications on the large multimers concentration.

The results of these indicators tests are expressed in international units per deciliter (U/dL). The DDAVP (desmopressin, 1-desamino-8-D-arginine vasopressin) is a drug used to treat vWD by promoting the release of vWF. The DDAVP induces the release of vWF from endothelial cells Weibel-Palade bodies. So desmopressin is used in laboratory tests to analyse the vWF indicators. Since patients with vWD type 1 don't have vWF functionality problems, they have a high response to DDAVP, while those with vWD type 2 or vWD

type 3 tend to not respond. In addition, plasma vWF assays during DDAVP administration is useful to resolve preliminary diagnostic ambiguities and in helping to classify the vWD types.

1.6 Study objectives

This Chapter has described the vWF importance for the coagulation system and how mutations and dysfunctions in this protein can cause qualitative and quantitative disorders. These mutations cause dominant effects and produce distinct vWD phenotype. As already seen, changes in vWF release can cause quantitative defects (type 1), as well as qualitative defects that causes binding systems inactive, or regulate platelet affinities, causing other phenotypes correlated with platelet and factor VIII (type 2).

The objective of this study is to help recognize vWD by modeling the vWD plasma concentration. In order to fulfil this task it is important to have a reliable mathematical model based on the patient's physiology that quantitatively decreases the vWF multimer plasma distribution over time. A good mathematical model provides information to better understand complex physiological phenomena and generate new tools for disease diagnosis.

The aim of this thesis is to develop, through a well-known mathematical model, a new approach to assess the von Willebrand disease. A new mathematical approach is evaluated to access whether using the Laplace transform is possible to provide a different way of analyse existing vWF models. In the Laplace domain, transfer functions and parameters constructed from the original model serve as indicators for a different physical interpretation of the kinetic parameters present in the mathematical model. And with that, analyze if new indicators, that help track down the disease and mutations, might emerge. The Laplace transform simplifies the mathematical model resolution, allowing to share information and use a simpler software for the disease diagnosis process.

Chapter 2

Model development in the Laplace domain

This Chapter introduces the Laplace transform and the transfer functions related to it. Then, the Chapter presents the von Willebrand factor pharmacokinetic models and previous studies on the disease modeling. Subsequently, it manipulates the von Willebrand Factor models to the Laplace transform in order to put it in a transfer function form. The Chapter concludes by validating the transfer functions with simulations using Matlab®.

2.1 Introduction

A mathematical model is a useful tool to describe a real process phenomenon. It gives an abstract representation and a physical insight of a process. The development of a mathematical model requires requires some simplifications and assumptions to reduce the complexity of the real process, specially for systems in the physiological field. A good model can represent the dynamic behavior while being no more complex than necessary.

Solving a mathematical model in order to represent the output variable in the time domain can be a considerable effort. Commonly, models are constructed by linear ordinary differential equations (ODEs). The Laplace Transform is a technique to solve and analyze linear differential equation models by converting them into algebraic equations.

With the aim of expressing the physiological behavior of the von Willebrand Factor (vWF) in the plasma, described in the previous Chapter, Galvanin et al. (2014) stated that a good model representation must describe the real pathways of vWF concentration varying in time. Thus, it is possible to recognize the different types of vWD through the multimers distribution in time. A well represented model should also have parameters consistent with the actual process and easy to obtain with the clinical data available.

Galvanin et al. (2014) cites three main mechanism steps during the clinical treatment to diagnose the vWD using the DDAVP vasopressin test. After the drug administration there are:

- release of the SUL (superultra large) multimers of the vWF;
- proteolysis of the SUL into smaller multimers by the enzime ADAMTS13;
- clearance of plasma multimers, which occurs on the liver and is independent of the multimer dimension;

2.2 Laplace transform

The Laplace transform is a mathematical tool used to solve and reduce the effort required to analyze linear differential equations. The technique transforms differential equations of a real variable t (often time) to algebraic equations of a complex variable s. Seborg et al. (2017) define the Laplace transform of a function as:

$$F(s) = \mathcal{L}[f(t)] = \int_0^\infty f(t)e^{-st}dt \tag{2.1}$$

In Equation (2.1), $F(s)$ is the representation of the Laplace transform of the generic function of time $f(t)$, using the so called Laplace Operator $\mathcal{L}$. Once in the Laplace domain, the algebraic equations (that formerly were differential equations) can now be easily manipulated. On the other hand, the inverse Laplace transform $\mathcal{L}^{-1}$, defined by Seborg et al. (2017) in Equation (2.2), operates on the $F(s)$ function and converts it back to $f(t)$ function. Thereby, generically, differential equations can be manipulated and solved.

$$f(t) = \mathcal{L}^{-1}[F(s)] \tag{2.2}$$

Since the Laplace transform involves integration, it is natural for the transform to inherit properties of the integral. One of these properties is linearity, so the Laplace transform is a linear operator and satisfies the superposition principle described in Equations (2.3) and (2.4):

$$\mathcal{L}[k \cdot x(t)] = k \cdot X(s) \tag{2.3}$$

$$\mathcal{L}[ax(t) + by(t)] = \mathcal{L}[ax(t)] + \mathcal{L}[by(t)] = aX(s) + bY(s) \tag{2.4}$$

2.2.1 Transfer functions

Within the Laplace domain, a transfer function ties two process variables, a dependent output variable and an independent input variable. Figure 2.1 shows a scheme representing this connection. Thereby, input and output assume a cause-effect relationship. For a continuous-time linear system, the transfer function is the ratio between the output and input variables in the Laplace domain.

A general linear n th-order differential equation forms a general transfer function

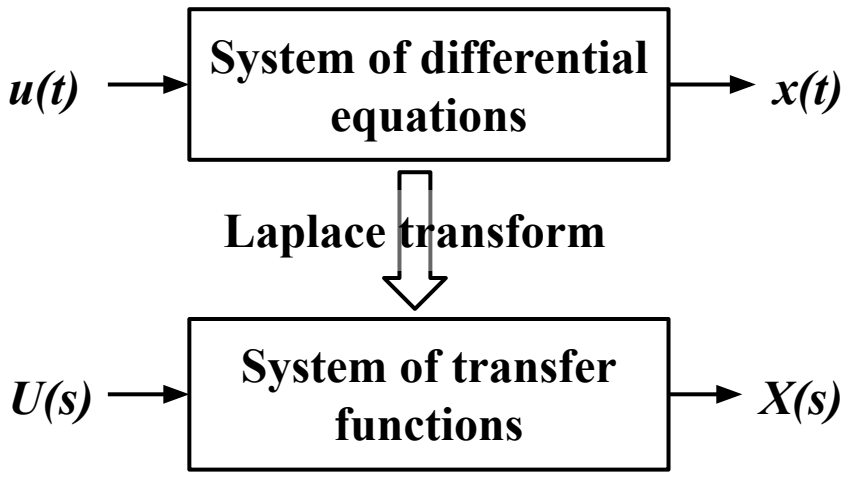


Figure 2.1. *Block diagram representing the Laplace transform of a time-dependent system differential equation into Laplace domain to form a system transfer function*

$G(s)$ represented in Equation (2.5). The letters 'm' and 'n' represent the order of the transfer function, the condition $n > m$ is necessary to ensure the the model is physical realizable.

$$\frac{X(s)}{U(s)} = G(s) = \frac{b_m s^m + b_{m-1}s^{m-1} + b_{m-2}s^{m-2} + \cdots + b_1 s^1 + b_o}{a_n s^n + a_{n-1}s^{n-1} + a_{n-2}a^{n-2} + \cdots + a_1 s^1 + a_o} \tag{2.5}$$

At time $t = 0$ the system is at steady-state and, in this situation, Equation (2.5) becomes $G(s) = b_o/a_o = K$, where K defines the transfer function steady-state gain. Factoring Equation (2.5) by isolating the steady-state gain and the other constants, the parameter τ rises denoting a time constant, well represented by Equation (2.6).

$$\frac{X(s)}{U(s)} = G(s) = \frac{K(\tau_a s + 1)(\tau_b s + 1)\cdots}{(\tau_1 s + 1)(\tau_2 s + 1)\cdots} \tag{2.6}$$

Equation (2.6) provides new process information. The time-constants τ_i and the steady-state gain K provide information about process behaviour, like response speed and shape.

Transfer functions display multiplicative properties for connected processes. For example, suppose two processes with transfer functions $G_1(s)$ and $G_2(s)$ configured in series. The input $U(s)$ connected to $G_1(s)$ produces an output $X_1(s)$, which is the input to $G_2(s)$, and produces the output $X_2(s)$. So the transfer function between the original input $U(s)$ and the output $X_2(s)$ can be obtained by multiplying $G_2(s)$ by $G_1(s)$. This arrangment can be written in equation form, as reported in Equation (2.7) and (2.8).

$$X_1(s) = G_1(s) \cdot U(s) \tag{2.7}$$

$$X_2(s) = G_2(s) \cdot X_1(s) = G_2(s) \cdot G_1(s) \cdot U(s) \tag{2.8}$$

2.3 von Willebrand factor pharmacokinetic model

The models are based on the clinical data of vWF antigen (vWF:Ag) and collagen binding (vWF:CB) measurements followed by the subcutaneous administration of DDAVP vasopressin. The measures of vWF:Ag represent the total concentration of the vWF regardless of the size of the multimeric units. Meanwhile vWF:CB indicates the concentration of high molecular weight multimers in the plasma.

Two models propose the study of the behavior on the average profiles of vWF:Ag and vWF:CB, and the multimers related to these variables. The analysis in this Chapter is based on the model presented by Ferrari et al. (2018), which is a simplification of the model presented by Galvanin et al. (2014). Both of them worked with similar variables and parameters and will be discussed further.

2.3.1 Model proposed by Galvanin et al. (2014)

The first model proposed by Galvanin et al. (2014) is composed of four equations. The model assumes that at the basal state only multimers with high and low molecular weight are present. It also assumes that super ultra large (SUL) multimers cannot be measured directly from experimental measurements, and their release is a consequence of DDAVP injection. The experimental measurement of vWF:Ag is given by the sum of the UL, H and L multimers, and the experimental measurement of vWF:CB is given by the sum of the UL and H multimers. Figure 2.2 shows a scheme of the model variables and parameters represented by Equations (2.9), (2.10), (2.11) and (2.12), where the circular blocks represent the mechanism steps.

Figure 2.2 shows circular blocks representing the multimers SUL, UL, H and L mechanism step, with the arrows representing the kinetics process between them. The parameter described as D represents the effect of DDAVP dose injection phase which causes the release of different multimeric species. The rectangular dotted line represents the vWF antigen concentration (y^{Ag}) as the sum of the UL, H and L multimers and the rectangular dashed line represents the collagen binding concentration (y^{CB}) as the sum of the UL and H multimers.

$$\frac{\mathrm{d}x^{SUL}}{\mathrm{d}t} = k_o D e^{-k_o(t-t_{max})} - k_1(x^{SUL} - x_b^{SUL}) - k_2(x^{SUL} - x_b^{SUL}) - k_3(x^{SUL} - x_b^{SUL}) \tag{2.9}$$

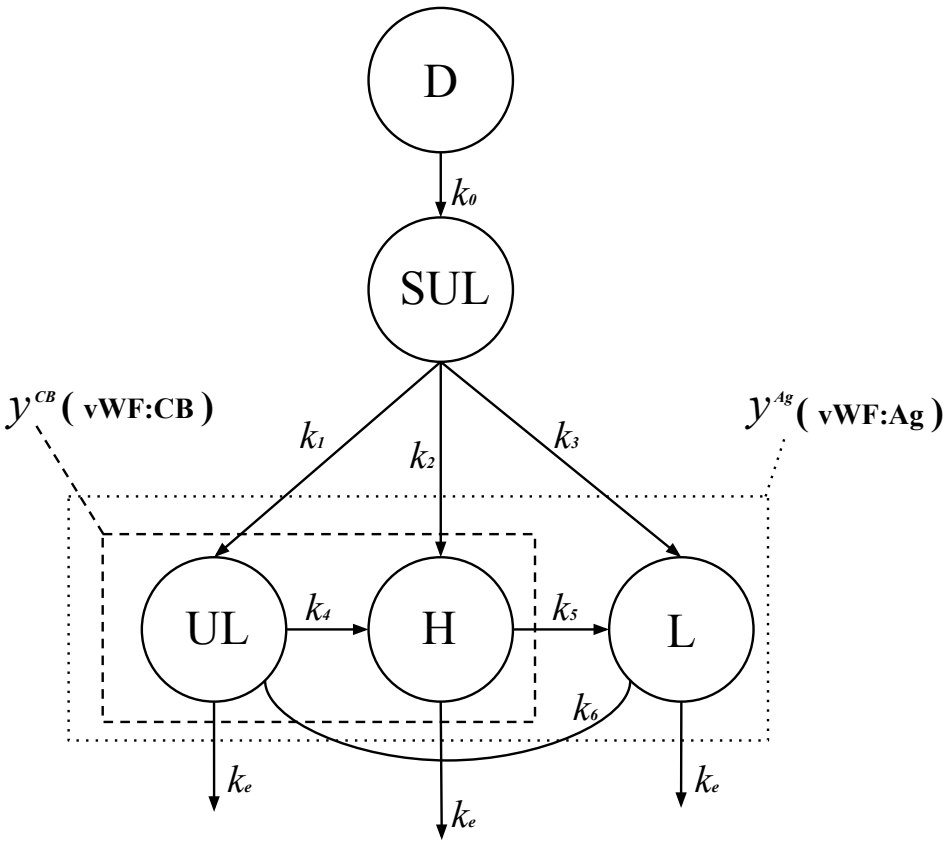


Figure 2.2. *Scheme presented by Galvanin et al. (2014) to represent the pharmacoKinetics model, along with the vWF antigen (vWF:Ag) and collagen binding (vWF:CB) represented in the picture by the dotted and the dashed rectangles respectively. The circular blocks represent the multimers mechanism step. The D block represents the DDAVP dose injection step which causes the release of different multimeric species. The arrows indicate the kinetics involved.*

$$\frac{\mathrm{d}x^{UL}}{\mathrm{d}t} = k_1(x^{SUL} - x_b^{SUL}) - k_4(x^{UL} - x_b^{UL}) - k_e(x^{UL} - x_b^{UL}) - k_6(x^{UL} - x_b^{UL}) \quad (2.10)$$

$$\frac{\mathrm{d}x^{H}}{\mathrm{d}t} = k_2(x^{SUL} - x_b^{SUL}) + k_4(x^{UL} - x_b^{UL}) - k_5(x^{H} - x_b^{H}) - k_e(x^{H} - x_b^{H}) \quad (2.11)$$

$$\frac{\mathrm{d}x^{L}}{\mathrm{d}t} = k_3(x^{SUL} - x_b^{SUL}) + k_5(x^{H} - x_b^{H}) - k_e(x^{L} - x_b^{L}) + k_6(x^{L} - x_b^{L}) \quad (2.12)$$

Equations (2.9), (2.10), (2.11) and (2.12) represent respectively the distribution of SUL, UL, H and L units in the blood plasma. The x^{SUL} [U] represents the amount of super ultra large multimers, x^{UL} [U] represents the amount of ultra large multimers, x^{H} [U] represents the amount of high multimers, x^{L} [U] represents the amount of low mul-

timers. D [U] represents the release of SUL multimers, t_{max} [min] represents the time of maximum response, which is the time where the vWF antigen (vWF:Ag) and collagen binding (vWF:CB) concentration curves achieve the maximum value. k_0 $[min^{-1}]$ is the kinetic parameter at the release step, and k_1 $[min^{-1}]$, k_2 $[min^{-1}]$, k_3 $[min^{-1}]$, k_4 $[min^{-1}]$, k_5 $[min^{-1}]$ and k_6 $[min^{-1}]$ are the proteolysis kinetic parameters. And k_e $[min^{-1}]$ is the clearance kinetic parameter. Subscript "b" describes the initial basal condition.

Equations (2.13) and (2.14) represent the vWF antigen (vWF:Ag) and collagen binding (vWF:CB) concentrations, respectively, by the variables y^{Ag} and y^{CB} [U/dL]. Equation (2.15) represents a correction made by Galvanin et al. (2014) to the y^{CB} in order to adjust the different affinity multimers have to collagen binding for distinct vWD types. Castaldello (2016) also refers to this correction in order to avoid numerical problems considering the experimental measurements variability.

$$y^{Ag} = \frac{(x^{UL} + x^{H} + x^{L})}{V_d} \tag{2.13}$$

$$y^{CB} = \frac{(x^{UL} + x^{H})}{V_d} \tag{2.14}$$

$$y^{CB'} = \alpha y^{CB} \frac{y_b^{Ag}}{y_b^{CB}} \tag{2.15}$$

According to Menache et al. (1996), Equation (2.16) defines the parameter V_d used in Equation (2.13) and (2.14) as the distribution volume expressed in dL. BW corresponds to the patient's body weight in kg.

$$V_d = 0.4 \cdot BW \tag{2.16}$$

Summarizing, the release of SUL step is represented by the parameters D, k_0, and t_{max}; the proteolysis of the multimers SUL, UL and H step is represented by the kinetic parameters k_1, k_2, k_3, k_4, k_5 and k_6; the clearance step is represented by the kinetic parameter k_e. The correction proposed in Equation (2.15) adds parameters α and y_b^{CB}.

2.3.2 Model proposed by Ferrari et al. (2018)

The model proposed by Ferrari et al. (2018), described by Equations (2.17) and (2.18), is a simplified version of the model proposed by Galvanin et al. (2014). In this case the SUL is not taken under consideration given the measurement limitation. Also the Ultra Large and High molecular weight multimers (UL+HMW) were considered together as one unit; this way the proteolysis happens only between the UL+HMW and the

LMW. Figure 2.3 shows the scheme made by Ferrari et al. (2018) to illustrate the model process.

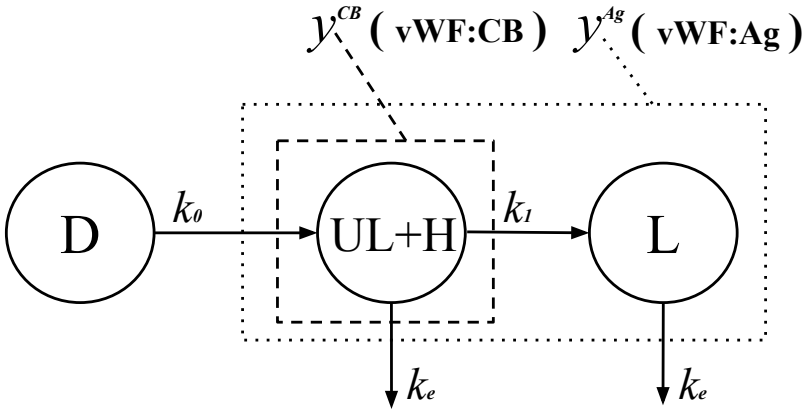


Figure 2.3. *Scheme presented by Ferrari et al. (2018) to represent the pharmacokinetic model and the mutlimers distribution, along with the vWF antigen (VWF:Ag) and collagen binding (VWF:CB) measuremnts represented in the picture by the dashed line*

Figure 2.3 is an adapted scheme of the first one proposed by Galvanin et al. (2014). In the same way, the circular blocks represent the Ultra Large and High molecular weight multimers (UL+H) and Low molecular weight multimers (L), with the arrows representing the kinetics process between them. The circular block with parameter *D* represents the effect of DDAVP dose injection phase, which causes the release of multimeric species. The rectangular dotted line represents the vWF antigen concentration (y^{Ag}) as the sum of UL+H and L multimers, and the rectangular dashed line represents the collagen binding concentration (y^{CB}) as the amount of the UL+H multimers.

$$\frac{\mathrm{d}x^{UL+HMW}}{\mathrm{d}t} = k_o D e^{-k_o(t-t_{max})} - (k_1 + k_e)(x^{UL+HMW} - x_b^{UL+HMW}) \tag{2.17}$$

$$\frac{\mathrm{d}x^{LMW}}{\mathrm{d}t} = k_1(x^{UL+HMW} - x_b^{UL+HMW}) - k_e(x^{LMW} - x_b^{LMW}) \tag{2.18}$$

Equations (2.17) and (2.18) represent respectively the amount of UL+HMW and LMW multimers in the blood plasma. The x^{UL+HMW} [U] represents the amount of ultra large and high multimers, x^{LMW} [U] represents the amount of low multimers. D [U] represents the multimers release, t_{max} [min] represents the time of maximum response, which is the time where the vWF antigen (VWF:Ag) and collagen binding (VWF:CB) concentration curves achieve the maximum value. The k_0 $[min^{-1}]$ is the release kinetic parameter, k_1 $[min^{-1}]$ is the proteolysis kinetic parameter and k_e $[min^{-1}]$ is the clearance kinetic parameter. Subscript "b" describes the initial basal condition.

The vWF antigen (vWF:Ag) and collagen binding (vWF:CB) concentrations are rep-

resented in Equations (2.19) and (2.20) respectively by the variables y^{Ag} and y^{CB}

$$y^{Ag} = \frac{(x^{UL+H} + x^{L})}{V_d} \tag{2.19}$$

$$y^{CB} = \frac{x^{UL+H}}{V_d} \tag{2.20}$$

$$y^{CB'} = \alpha y^{CB} \frac{y_b^{Ag}}{y_b^{CB}} \tag{2.21}$$

Equation (2.21) represents the same correction made by Galvanin et al. (2014) for y^{CB}. According to Menache et al. (1996), Equation (2.22) defines parameter V_d [dL] as the distribution volume, where BW [kg] corresponds to the patient's body weight.

$$V_d = 0.4 \cdot BW \tag{2.22}$$

In this model proposed by Ferrari et al. (2018), the same parameters D, k_0, and t_{max} describe the release stage; but only the kinetic parameter k_1 represents the proteolysis of the multimers UL+HMW into LMW; and the kinetic parameter k_e represents the clearance stage. The correction proposed in Equation (2.21) also adds the parameters α and y_b^{CB}. This way, the model proposed by Ferrari et al. (2018) reduces the number of parameters from 12 to 7.

2.4 Analysis of previous studies on the von Willebrand factor models

Some authors have proposed solutions to the parameter identification for the models presented previously. According to Monte (2013), the initial clinical data were available as von Willebrand factor antigen (vWF:Ag) and collagen binding (vWF:CB) obtained after subcutaneous administration of vasopressin DDAVP. Data are from distinct groups of individuals of different ages and body weight, including healthy subjects of blood type O (HO) and non-O (HnonO) and individuals affected by vWD of type 1, 2A, 2B and Vicenza.

For the Galvanin model, Monte (2013) and Castaldello (2016) used the maximum likelihood technique to estimate the parameter vector set $\theta = [\ k_o\ \ k_1\ \ k_2\ \ k_3\ \ k_4\ \ k_5\ \ k_6\ \ k_e\ \ D\ \ t_{max}\ \alpha\ \ y_b^{CB}\ ]$. The method of maximum likelihood selects the set of values of the model parameters that maximizes the likelihood function. This maximizes the relation between the proposed model and the experimental data. In order to minimize the variability between data, both of them worked with an "average subject", meaning that each

vWD type have a vWF:Ag and vWF:CB concentration profile made from the average of the points collected at the same time from all the individual subjects of vWD type. In other words, each vWD type has an unique vWF:Ag and vWF:CB average subject profile, which was used to estimate the parameter set.

The maximum likelihood technique was also used to estimate the Ferrari model parameters. Castaldello (2016) worked with the average subject and used the parameter set $\theta = [\ k_o \ \ k_1 \ \ k_e \ \ D \ \ \alpha \ \ y_b^{CB} \ \ D/t_{max}]$. In another way, Ferrari et al. (2018) worked with the parameter set $\theta = [\ k_o \ \ k_1 \ \ k_e \ \ D \ \ \alpha \ \ y_b^{CB} \ \ t_{max}]$ and used the maximum likelihood technique to predict the parameters for each single individual. Then the parameters for each vWD type was presented by calculating the mean value of the estimated parameters among each type of vWD individuals.

Table 2.1 shows the values of the parameters found by Ferrari et al. (2018) and Castaldello et al. (2018). Even though both studies estimated the parameters in different ways, the kinetic parameters show minor difference between them, which is a reference for vWD type identification. However the D and t_{max} values between the Ferrari et al. (2018) and Castaldello et al. (2018) differ very much. Castaldello et al. (2018) found that the value of t_{max} is not always the same for the vWF:Ag, and vWF:CB and its value has low influence on the other parameters. This justifies the use of the D/t_{max} in the parameter set and explains the differences. Nevertheless this work, from now on, will use the parameters values estimated by Ferrari et al. (2018).

Table 2.1. *Parameters obtained by Ferrari et al. (2018) and Castaldello et al. (2018), for healthy groups HO and HnonO, and vWD type 2B, 2A and Vicenza. The number in parenthesis below the group's name indicates the quantity of patients available for each group.*

Parameters	**Ferrari et al. (2018)**					**Castaldello et al. (2018)**			
	HO	**HnonO**	**2B**	**2A**	**Vicenza**	**HO**	**HnonO**	**2B**	**Vicenza**
(subjects)	(24)	(18)	(8)	(3)	(9)	(24)	(18)	(8)	(9)
$k_0\ [min^{-1}]$	0.0264	0.0287	0.0177	0.0152	0.0666	0.0264	0.0285	0.0136	0.0423
$k_1\ [min^{-1}]$	0.0006	0.0002	0.0047	0.0042	0.0015	0.0003	0.0001	0.0020	0.0013
$k_e\ [min^{-1}]$	0.0015	0.0007	0.0032	0.0013	0.0082	0.0014	0.0007	0.0037	0.0098
$D[U]$	568	425	597	649	271	141	168	194	199
α	1.01	0.94	0.20	0.06	0.31	0.98	0.97	0.18	0.73
$y_b^{CB}\ [U/dL]$	49.30	77.90	26.10	23.60	2.38	49.97	84.64	24.53	5.10
$t_{max}\ [min]$	60	75	127	135	46.7	110	109	214	65

At this point, it is possible to distinguish the parameters: the ones related with the release of multimers after DDAVP administration, D, t_{max} and k_o; the one related with proteolysis, k_1; and the one related with clearance of multimers, k_e. In Table 2.1, is possible to see the difference between the parameters from the healthy group HO and HnonO, and patients with vWD type 2A, 2B and Vicenza. Each type has some characteristics and

it is related with the parameters.

Even though there are some differences between the healthy group HO and HnonO, Ferrari et al. (2018) reported that only the values of k_e and D are statistically significantly different. For HO patients, the bigger values of k_e is in accordance with a report cited in Ferrari et al. (2018) about a shorter VWF survival in subjects with the O blood group. In groups 2A and 2B an accelerated proteolysis activity is expected, which is evident since k_1 parameter is bigger than the others. The Vicenza group presents a shorter multimers half-life, which is evident by the k_e parameter.

2.5 The von Willebrand model in the Laplace domain

It is possible to apply the Laplace transform (2.1) in the Ferrari model equations (2.17) and (2.18), providing an alternative functional tool to analyze the behavior of the multimers system in the plasma.

2.5.1 Deviation variables

For the sake of simplifying, the following multimeric variables presented in Equations (2.17) and (2.18) will be addressed in the terms of deviation variables. A deviation variable is represented by the variable minus the variable steady-state value. In this context the steady-state is portrayed by the basal state.

- $x'_{ULH} = (x^{UL+HMW} - x_b^{UL+HMW})$ [U]
- $x'_L = (x^{LMW} - x_b^{LMW})$ [U]

being the subscript ULH assigned for UL+HMW variables, the subscript L assigned for LMW variables, the subscript b assigned for basal values and the apostrophe representing the deviation variable. The following Equations (2.23) represent Equation (2.18) in terms of deviation variables.

$$\frac{\mathrm{d}x'_{ULH}}{\mathrm{d}t} = k_o D e^{-k_o(t-t_{max})} + (k_1 + k_e)x'_{ULH} \tag{2.23a}$$

$$\frac{1}{(k_1 + k_e)}\frac{\mathrm{d}x'_{ULH}}{\mathrm{d}t} + x'_{ULH} = \frac{k_o D e^{-k_o(t-t_{max})}}{(k_1 + k_e)} \tag{2.23b}$$

$$\tau_{ULH}\frac{\mathrm{d}x'_{ULH}}{\mathrm{d}t} + x'_{ULH} = u(t) \tag{2.23c}$$

where $\tau_{ULH} = \frac{1}{(k_1+k_e)}$, and $u(t) = \frac{k_o D e^{-k_o(t-t_{max})}}{(k_1+k_e)}$. As described previously, D has unit [U] and k_o, k_1 and k_e has unit [min^{-1}]. A dimensional analysis shows $u(t)$ having unit [U] (like x'_{ULH}) and τ_{ULH} having unit [min]. The following Equations (2.24) represent Equation (2.18) in terms of deviation variables.

$$\frac{\mathrm{d}x'_L}{\mathrm{d}t} = k_1 x'_{UL+H} - k_e x'_L \tag{2.24a}$$

$$\frac{1}{k_e}\frac{\mathrm{d}x'_L}{\mathrm{d}t} + x'_L = \frac{k_1}{k_e} x'_{ULH} \tag{2.24b}$$

$$\tau_L \frac{\mathrm{d}x'_L}{\mathrm{d}t} + x'_L = \frac{k_1}{k_e} x'_{ULH} \tag{2.24c}$$

Equation (2.24c) introduces the new parameter $\tau_L = \frac{1}{k_e}$ [min].

2.5.2 Laplace transform

The next step consists of applying the Laplace transform to Equations (2.23c) and (2.24c) with the intention to put them in the Laplace domain, and find the transfer function that relates $u(t)$ with $x'(t)$. The Laplace transform of the derivative term follows the steps described in Equation (2.25).

$$\begin{aligned}
\mathcal{L}\left[\frac{\mathrm{d}x(t)}{\mathrm{d}t}\right] &= \int_0^\infty \left(\frac{\mathrm{d}x(t)}{\mathrm{d}t}\right) e^{-st}\mathrm{d}t \\
&= \int_0^\infty x(t)e^{-st}s\mathrm{d}t = \left[x(t)e^{-st}\right]_0^\infty \\
&= s\mathcal{L}\left[x(t)\right] - x(0) \\
&= sX(s) - x(0)
\end{aligned} \tag{2.25}$$

It was defined that at the steady state the multimers variables are at basal state. So at $t = 0$, $x'_{ULH}(0) = (x_b^{UL+HMW} - x_b^{UL+HMW}) = 0$. The Laplace transform Equation (2.25) in terms of the deviation variables becomes:

$$\mathcal{L}\left[\tau\frac{\mathrm{d}x'(t)}{\mathrm{d}t}\right] = \tau(sX'(s) - x'(0)) = \tau sX'(s) \tag{2.26}$$

The parameter τ_{ULH} and τ_L of Equation (2.23c) and (2.24c) will follow the definition from Equation (2.3). The defined release term $u(t)$ from Equation (2.23c) became:

$$\begin{aligned}
\mathcal{L}\left[\frac{k_o D e^{-k_o(t-t_{max})}}{(k_1+k_e)}\right] &= \frac{k_o D e^{k_o t_{max}}}{(k_1+k_e)} \cdot \int_0^\infty e^{-k_o t} e^{-st} \mathrm{d}t \\
&= \frac{k_o D e^{k_o t_{max}}}{(k_1+k_e)} \cdot \int_0^\infty e^{-(s+k_o)t} \mathrm{d}t \\
&= \frac{k_o D e^{k_o t_{max}}}{(k_1+k_e)} \cdot \left[-\frac{1}{(s+k_o)} \cdot e^{-(s+k_o)t}\right]_0^\infty \\
&= \frac{k_o D e^{k_o t_{max}}}{(k_1+k_e)} \cdot \left[-0 + \frac{1}{(s+k_o)} \cdot e^0\right] \\
&= \frac{k_o D e^{k_o t_{max}}}{(k_1+k_e)} \cdot \frac{1}{(s+k_o)} = U(s)
\end{aligned} \tag{2.27}$$

So the term $u(t)$ in the Laplace domain has the form of Equation (2.28):

$$U(s) = \frac{\frac{k_o D e^{k_o t_{max}}}{(k_1+k_e)}}{(s+k_o)} \tag{2.28a}$$

$$U(s) = \frac{\frac{D e^{k_o t_{max}}}{(k_1+k_e)}}{(\frac{1}{k_o}s+1)} \tag{2.28b}$$

$$U(s) = \frac{K_U}{(\tau_U s+1)} \tag{2.28c}$$

Equation (2.28c) introduces the new parameters $K_U = \frac{D e^{k_o t_{max}}}{(k_1+k_e)}$ [U.min] (The exponential factor is unitless, the parameter D has unit [U] and (k_1+k_e) has unit [min^{-1}]) and $\tau_U = \frac{1}{k_o}$ [min]. The Laplace transform takes place by replacing Equations (2.26), (2.27) with Equation (2.17) and (2.18). Equation (2.29) shows the result of Equation (2.23c) and the algebraic manipulation at the Laplace domain.

$$\tau_{ULH} s X'_{ULH}(s) + X'_{ULH}(s) = U(s) \tag{2.29a}$$

$$X'_{ULH}(s)(\tau_{ULH} s+1) = U(s) \tag{2.29b}$$

$$\frac{X'_{ULH}(s)}{U(s)} = \frac{1}{(\tau_{ULH} s+1)} \tag{2.29c}$$

Equation (2.29c) introduces, in the form of a transfer function, the parameter $\tau_{ULH} = \frac{1}{(k_1+k_e)}$ [min] representing the time constant. Equation (2.30) shows the Laplace transform and the algebraic manipulation for Equation (2.24c).

$$\tau_L s X'_L(s) + X'_L(s) = \frac{k_1}{k_e} X'_{ULH}(s) \tag{2.30a}$$

$$X'_L(s)(\tau_L s + 1) = \frac{k_1}{k_e} X'_{ULH}(s) \tag{2.30b}$$

$$X'_L(s) = \frac{\frac{k_1}{k_e}}{(\tau_L s + 1)} \cdot X'_{UL+H}(s) \tag{2.30c}$$

$$X'_L(s) = \frac{\frac{k_1}{k_e}}{(\tau_L s + 1)} \cdot \frac{1}{(\tau_{ULH} s + 1)} \cdot U(s) \tag{2.30d}$$

$$\frac{X'_L(s)}{U(s)} = \frac{K_L}{(\tau_L s + 1)} \cdot \frac{1}{(\tau_{ULH} s + 1)} \tag{2.30e}$$

Equation (2.30e) introduces, in the form of a transfer function, the new parameters K_L and τ_L, where $K_L = \frac{k_1}{k_e}$ [unitless] represents a steady-state gain and $\tau_L = \frac{1}{k_e}$ [min] represents a time constant.

2.5.3 von Willebrand model in the Laplace domain

Equations (2.31), (2.32), (2.33), (2.34) and (2.35) summarize the newly constructed parameters.

$$\tau_U = \frac{1}{k_o} \qquad [min] \tag{2.31}$$

$$\tau_{ULH} = \frac{1}{k_1 + k_e} \; [min] \tag{2.32}$$

$$\tau_L = \frac{1}{k_e} \qquad [min] \tag{2.33}$$

$$K_U = \frac{D e^{k_o t_{max}}}{k_1 + k_e} \; [U.min] \tag{2.34}$$

$$K_L = \frac{k_1}{k_e} \qquad [unitless] \tag{2.35}$$

Equations (2.36), (2.37) and (2.38) summarize the input $U(s)$ and the transfer functions $G_{ULH}(s)$ and $G_L(s)$.

$$U(s) = \frac{K_U}{(\tau_U s + 1)} \tag{2.36}$$

$$G_{ULH}(s) = \frac{1}{(\tau_{ULH}s + 1)} \tag{2.37}$$

$$G_L(s) = \frac{K_L}{(\tau_L s + 1)} \tag{2.38}$$

$U(s)$ represents the input function; $G_{ULH}(s)$ and $G_L(s)$ represent the transfer functions connecting $U(s)$ and X'_{ULH} and X'_L respectively. Hence, Equations (2.39) and (2.40) represent , respectively, Equations (2.17) and (2.18) in the Laplace domain.

$$X'_{ULH}(s) = G_{ULH}(s) \cdot U(s) \tag{2.39}$$

$$X'_L(s) = G_{ULH}(s) \cdot G_L(s) \cdot U(s) \tag{2.40}$$

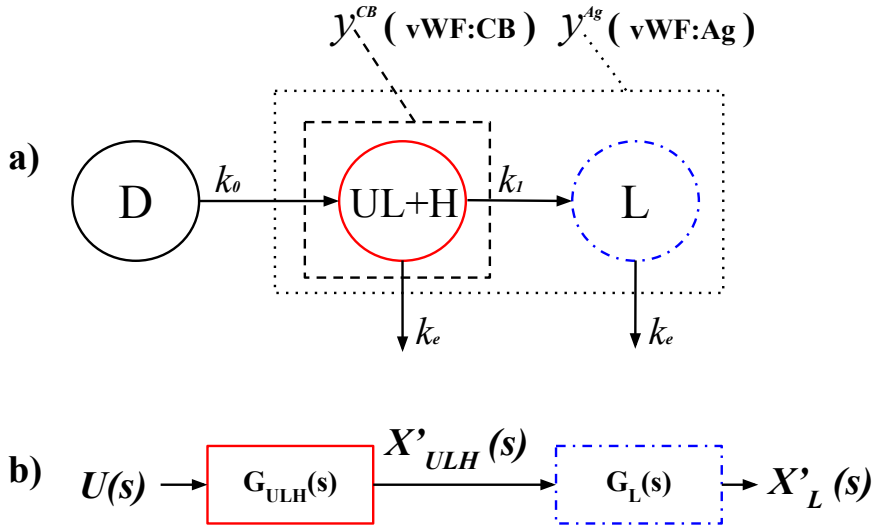


Figure 2.4. *Figure a) shows the model scheme already presented in Figure 2.3 presented by Ferrari et al. (2018) to represent the model and the multimers distribution. Figure b) shows the block diagram representing the Laplace transform from equations 2.39 and 2.40. G_{ULH} and G_L represent the transfer functions, $U(s)$ represents the input function. The X'_{ULH} and X'_L represent the deviation variables for the multimers units in the s domain.*

Figure 2.4a shows the mechanism representation of the model proposed by Ferrari et al. (2018) defined previously by Equations (2.17) and (2.18). This model presents the parameters k_o, k_1 and k_e meaning the rate of multimers reaction in the biological mechanism. That is, these parameters (and the compartmental model representation) indicate the rate at which multimers vary. The parameters arrows in Figure 2.4a indicate how the multimers reaction mechanism works.

Figure 2.4b shows the block diagram for the model in the Laplace domain. In this case, the transfer functions (shown in the Figure as blocks) characterizes how fast the multimer concentration changes, regardless of the mechanism behind the change. The transfer function parameters characterize the multimers dynamics behavior. The time constant pro-

vides information about system response swiftness, in other words, how the multimers will respond to changes in the process input. For instance, the system creates an initial multimers pulse, so τ_{ULH} and τ_L provide information about the system returning to the basal state. The transfer function G_{ULH} models the UL+HMW multimers behavior through time providing the amount UL+HMW multimers by the arrow X'_{ULH}. Since LMW multimers are formed by UL+HMW proteolysis, X'_{ULH} enters the transfer function G_L modeling the LMW multimers behavior through time, providing the amount LMW multimers by the arrow X'_L. While the kinetic model and its constants present information about the multimers rate of reaction mechanism. The model in the Laplace domain gives information about UL+HMW and LMW multimers response facing a stimulus.

Table 2.2 summarizes the parameters estimated by Ferrari et al. (2018) converted in the form of new parameters defined by Equations (2.31), (2.32), (2.33), (2.34) and (2.35).

In the biological process, the input function $U(s)$ shows the parameters K_U and τ_U being different from each patient. So the input $U(s)$ stands for the multimers release induced by DDAVP, represented by the following parameters: K_U embodies the kinetic parameters k_o, D and t_{max} which correlate with the multimers release rate, giving the increase of multimers in the plasma and τ_U embodies the kinetic parameters k_o which correlates the multimers time release.

Table 2.2. *Parameters defined by Equations (2.31), (2.32), (2.33), (2.34) and (2.35), with specific equations, units and values for the healthy groups HO, HnonO, vWD types 1, 2A, 2B and Vicenza. The number in parenthesis below the group's name indicates the quantity of patients available for each group.*

Parameters (subjects)	**Equations**	**Unit**	**HO** (24)	**HnO** (19)	**1** (51)	**2B** (8)	**2A** (3)	**Vicenza** (9)
K_U	$De^{k_o t_{max}}/(k_1+k_e)$	[U.min] $\cdot 10^5$	12.9	39.9	3.01	7.12	3.83	6.28
K_B	k_1/k_e	Unitless	0.41	0.34	0.29	1.46	0.17	0.18
τ_U	$1/k_o$	[min]	38	35	36	56	49	15
τ_{ULH}	$1/(k_1+k_e)$	[min]	466	1063	556	126	483	103
τ_L	$1/k_e$	[min]	658	1420	714	310	565	122

The first transfer function G_{ULH} stands for the UL+HMW multimers combining clearance and proteolysis. This process is represented by the parameter τ_A embodying the kinetic parameters k_1 and k_e, representing the consumption of multimers by the proteolysis and clearance process. Consequently, vWD types 2A, 2B and Vicenza show the lower values of τ_A influencing the total amount of multimers in the plasma.

After the release of ultra large multimers in the plasma, the ADAMTS-13 protein starts breaking the multimers into smaller units. The transfer function G_L represents the LMW multimers and the transfer function steady state gain K_L is the ratio between k_1 and

k_e, the direct proportionality with proteolysis constitutes the consumption of UL+LMW into LMW, while the clearance is the normal multimers decrease. A large value of K_L means that proteolysis prevails on clearance, resulting in a shortage of UL+HMW multimers. This peculiarity is typical of vWD type 2B, which in turn has the largest value of K_L ($K_L = 1.46$). τ_L is the time constant, which is the inverse of clearance kinetics and the vWD type Vinceza has the lowest value $\tau_L = 310$, which represent a very fast clearance for this type of vWD.

2.6 Model validation in the Laplace domain

In order to validate the model in the Laplace domain, Equations (2.39) and (2.40) developed in this chapter were implemented in Simulink® tool from the software Matlab® vR2019b. The simulation was conducted using as reference the parameters shown in Table 2.2. Figure 2.5 shows the image of the model construct in Simulink® to validate the model in the Laplace domain.

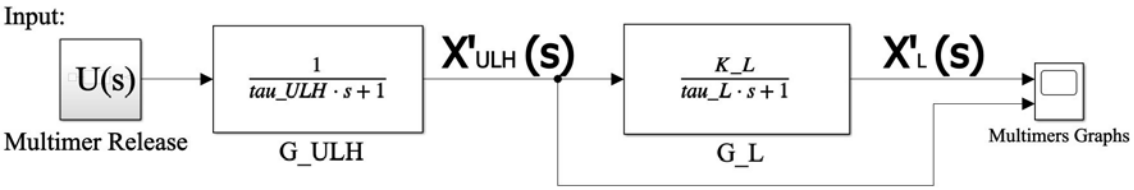


Figure 2.5. *Image of the model construct with Simulink® software to validate the Ferrari et al. (2018) equation in the Laplace domain*

The results obtained from the model of Figure 2.5 are represented in graphs in Figure 2.6, showing the multimeric patterns for healthy patient with blood type O and for patient with vWD type 2B, 2A and Vicenza. The Figure 2.6 is divided in two, where Figure 2.6a shows the graphics published by Ferrari et al. (2018) and Figure 2.6b shows the graphics made by mathematical simulation in Laplace Domain using Simulink® software.

As already mentioned, Ferrari et al. (2018) estimated the parameters for each patient and later calculated the average parameter for each vWD type. Hence, it is assumed the published graphs at Figure 2.6a refers to single subjects parameters. The simulation, instead, uses the published parameters (which is known to be the patients average values for each class) and creates the graphics in Figure 2.6b. This justify the small differences between the graphics from Figure 2.6a and Figure 2.6b, but still validates the methodology in order to use Laplace transform to solve the ordinary differential Equations (2.17) and (2.18). To fix this parameter differences it is important to consider a new parameter identification in the Laplace domain. The parameters defined in the Laplace domain can bring a new point of view in order to discriminate the von Willebrand Disease types. Besides, developing the model in the Laplace domain simplifies the equation resolution and allows news opportunities for the von Willebrand disease diagnosis.

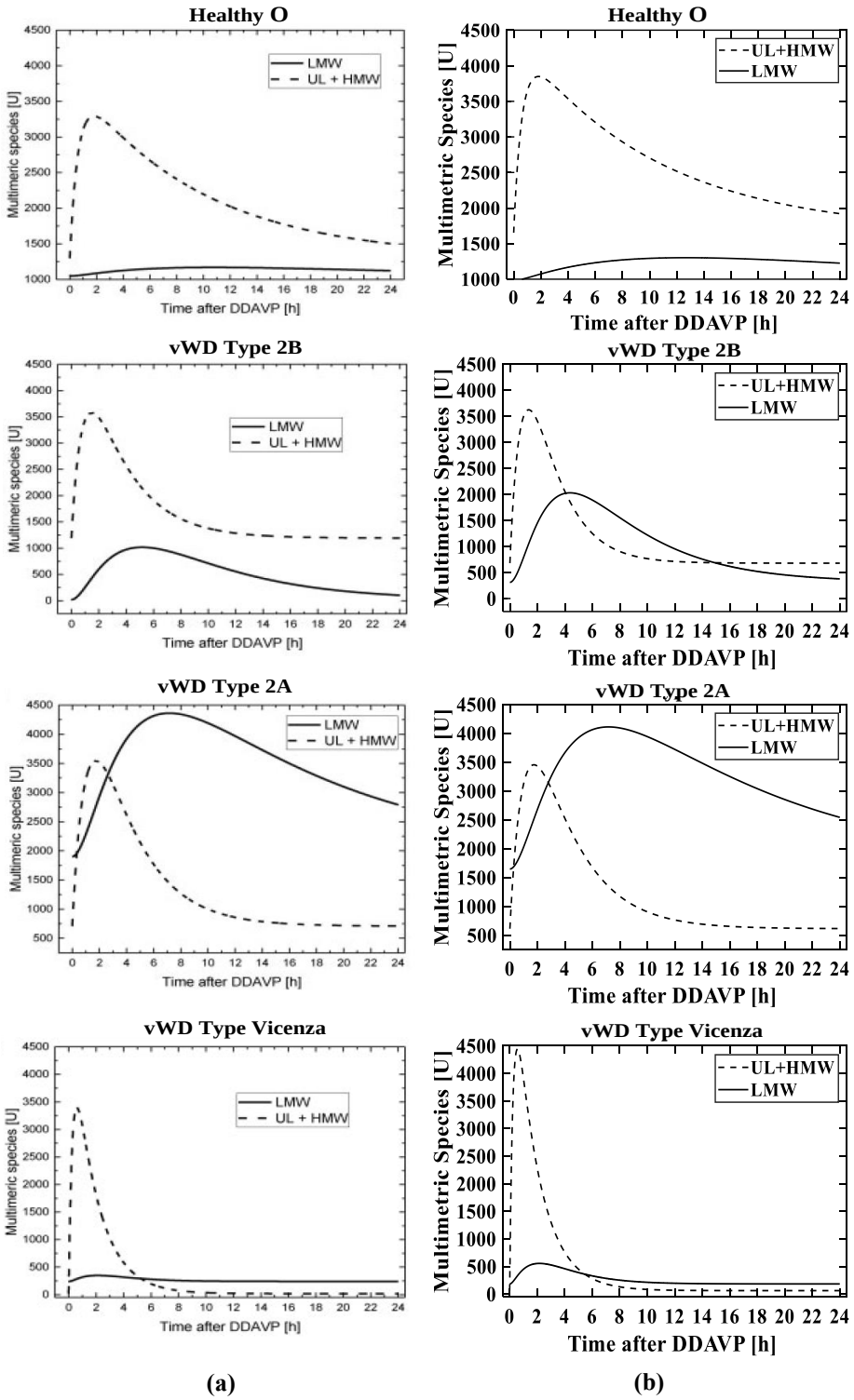


Figure 2.6. *(a) Graphics obtained by Ferrari et al. (2018) with patients data and the model proposed; (b) Graphics obtained at the Laplace domain with Simulink® using the Ferrari et al. (2018) published parameters.*

Chapter 3

Parameter identification methodology

This Chapter describes the parameter identification methodology used in the model presented in the previous Chapter. The parameter identification uses optimization techniques on the model to reach out the experimental data. Therefore, this Chapter begins by presenting the experimental data available and the optimization techniques used by previous research. Then, it presents the conditions and functions used in the model within the Laplace domain to identify its parameters. The Chapter concludes with the presentation of the results obtained with the parameterization.

3.1 Analysis of experimental data available

The Department of Cardiology, Thoracic and Vascular Sciences of the University of Padova provided the vWF data of 114 patients, presented as vWF antigen (vWF:Ag) and vWF collagen binding (vWF:CB) measurements. The data is divided between healthy people with blood type O (24 patients), healthy people with blood type non-O (19 patients), patients with vWD type 1 (51 patients), patients with vWD type 2A (3 patients), patients with vWD type 2B (8 patients) and patients with vWD type Vicenza (9 patients). This information and data were compiled from previous studies made by Monte (2013), Galvanin et al. (2014) and Castaldello (2016).

The vWF:Ag measurement determines the total amount of vWF in the plasma, independently of multimers size, and vWF:CB measurement determines the multimers with high molecular weight. The measurements were made after the subcutaneous administration of 3 $\mu g/kg_{body\ weight}$ of desmopressin (DDAVP). Then, ten samples were collected to measure the vWF concentration within the 24 hour time frame. The collection points were made at pre-established time (0, 0.25, 0.5, 1, 2, 3, 4, 6, 8 and 24h) after the DDAVP administration.

The mechanism proposed by Galvanin et al. (2014) and Ferrari et al. (2018) models after the DDAVP administration starts with the release of ultra large and high molecular weight multimers. Then, the enzime ADAMTS13 causes the proteolysis of high molecular weight multimers into smaller species, and finally occurs the clearance of multimers, which is independent of multimers size.

The measured data for each group (healthy O, healthy non-O, vWD type 1, vWD type 2A, vWD type 2B and vWD type Vicenza) are presented in graphs in Appendix A. Figure 3.1 shows the average value for collected samples in each pre-established time discriminated by group, this average defines the average subject for each group. Figure 3.1a shows the values for vWF antigen (y^{Ag}) and Figure 3.1b shows the values for vWF collage binding (y^{CB}).

Healthy patients with blood type O and non-O present in the plasma similar concentrations for vWF antigen and collage binding, representing a normal multimers distribution. Indeed, the values of vWF:Ag and vWF:CB of these patients are almost equivalent, as can be seen in Figure 3.1. Also, it is noticeable how healthy non-O patients have higher vWF concentrations as compared to healthy O patients.

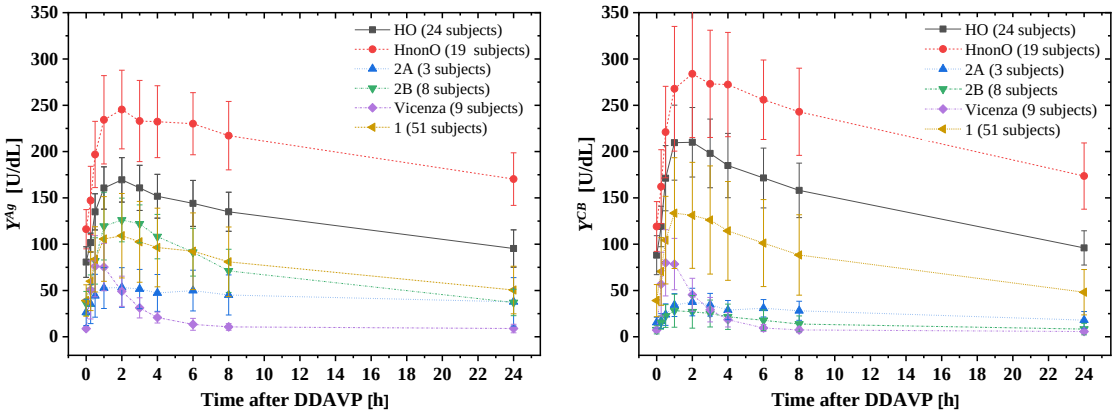


Figure 3.1. *Graphics of mean values for vWF measurements collected at predefined times, separated within each study group. The dots represent the mean values and the error bar shows the variance of those values inside the group. Figure (a) shows the graphic for vWF Antigen (vWF:Ag) concentration expressed as y^{Ag} by the time after the DDAVP administration and Figure (b) shows the graphic for vWF collage binding (vWF:CB) concentration expressed as y^{cB} by the time after the DDAVP administration.*

vWD type 1 is known for partial quantitative deficiency of vWF. The concentration level of vWF in plasma is low, but the remaining vWF works normally. Figure 3.1 shows a smaller concentration for type 1 in both graphics, considering the healthy patients as reference. vWD type 1 has high variability, caused by different types of mutations, resulting in an extra difficulty to distinguish type 1 patients with other types.

Meanwhile vWD type 2 has quality deficiency for vWF multimers. For instance, type 2B has multimers with high affinity for platelet receptors and high protheolysis rate, causing the high molecular weight multimers elimination. Figure 3.1b shows this effect by the lack of collage binding concentration (Y^{CB}) and Figure 3.1a shows a higher vWF antigen concentration (Y^{Ag}). Type 2A is known for multimers assembly and release inter-

ference, causing low level of multimers with high molecular weight. Figure 3.1b shows indeed lower level of vWF:CB, presented as triangle points.

The vWD type Vicenza has mutation which causes a high clearance rate. Indeed Figure 3.1 shows very short levels of both vWF:Ag and vWF:CB (100 U/dL at most) and a fast response after the DDAVP administration (return to basal values after only 4/5 hours). This type of mutation shows a characteristic profile where the released multimers level drops rapidly, as a result of rapid elimination. Since the clearance mechanism is independent of multimers size, both charts in Figure 3.1 show similar profiles for vWF.

3.2 Previous parameter identification

A crucial step in process modeling is investigate the model parameters and its viability. Figure 3.2 shows a scheme for the steps in model formulation process (McLean and McAuley, 2012). After specifying a model within its hypotheses it is required to consider how well the model fits the observed data. The fitting procedure is made by finding the values of the parameters of a model that best fits the data, a procedure called parameter estimation. That is, the use of experimental data to predict parameter values that will make the model represent the real system precisely and accurately. Thus, the reliability of the model is directly related to the estimation of parameters. There are several tools and techniques to estimate model parameters and it is reasonable to review the mechanism used by previous work.

According to Myung (2003) there are two general methods of estimating parameters: one is the least-squares estimation (LSE), the other is the maximum likelihood estimation (MLE). The first is linked to many familiar statistical concepts, such as linear regression, error of the sum of squares and quadratic mean root deviation. This method does not require distributional assumptions. On the other hand, the MLE is a standard approach for parameter estimation and statistical inference. The MLE has optimal estimation properties like efficiency, consistency (true value of the parameter that generate the data recovered) and sufficiency (complete information on the parameter).

Both Ferrari et al. (2018) and Castaldello et al. (2018) used the MLE method for the parameter estimation of the model from Equation (2.17) and (2.18). Ferrari et al. (2018) used the parameter set $\theta = [\ k_o \ \ k_1 \ \ k_e \ \ D \ \ \alpha \ \ y_b^{CB} \ \ t_{max}]$ and performed the estimation using the commercial software gPROMS®, assuming that measurements have equal and normally distributed deviation. The parameters were determined for each individual and the initial guess was determined using a preliminary parameter estimation for the average concentration profile. Given the high number of parameters, an iterative procedure was applied to help convergence. This procedure consisted of two steps, being step zero the estimation of all parameters, step one the estimation of kinetic parameters (k_o, k_1 and k_e) keeping fixed all the other corrective parameters obtained in the previous step (D,

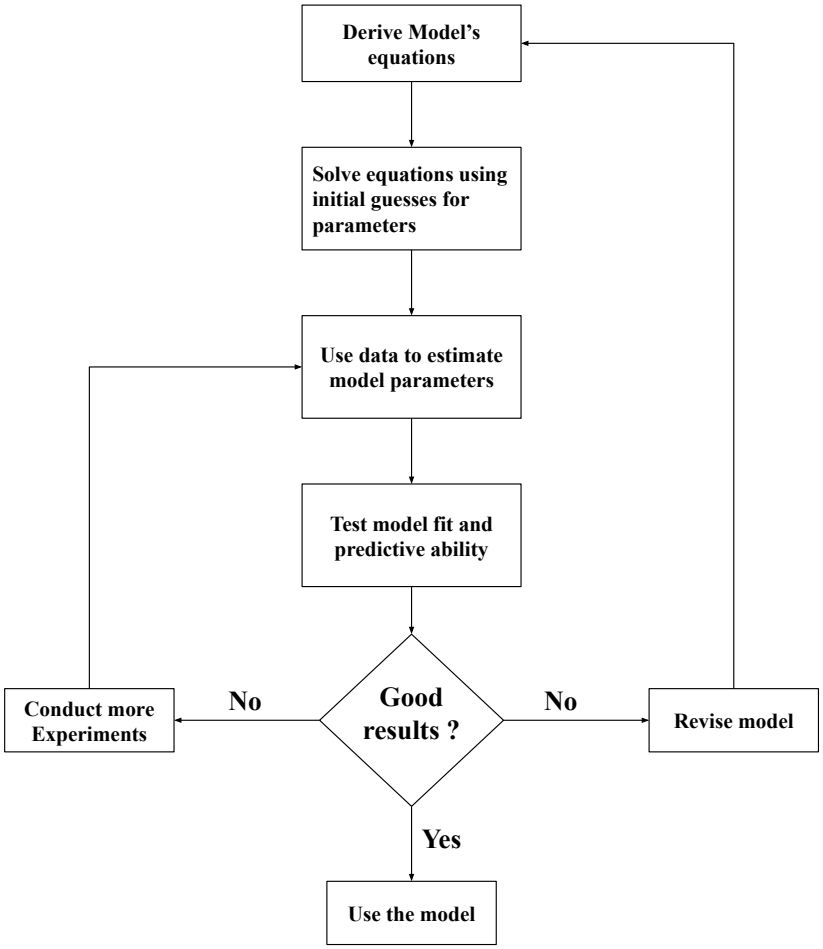


Figure 3.2. *Steps in model formulation (McLean and McAuley, 2012)*

α, y_b^{CB} and t_{max}), and step two the estimation of corrective parameters keeping fixed the kinetic parameters obtained in the previous step. Steps 1 and 2 are repeated iteratively until there is no significant differences between them.

Castaldello et al. (2018) also used the MLE method for the parameter estimation and performed the estimation using the commercial software gPROMS®. The parameter set was defined as $\theta = [\, k_o \;\; k_1 \;\; k_e \;\; D \;\; \alpha \;\; y_b^{CB} \;\; D/t_{max}]$. In a similar way, two step iterative procedure was used for the parameter identification. But differently from Ferrari et al. (2018) research, the parameter t_{max} was obtained in step 0 and not used in the iterative procedures, because it was observed that generally it does not vary significantly between steps. Instead of identifying the parameter of each subject, Castaldello et al. (2018) used the average subject as representative of the corresponding group. The parameter was also normalized allowing to search the values around the unit improving the robustness of the numerical method. The parameter normalization reduces differences between variables groups, since removing the influence of units makes the contribution of each of them equal in the calculation process. Concerning the parameter estimation, this work used the LSE method.

3.3 Considerations prior parameter identification

There are some considerations regarding the parameter identification process. Parameter estimation is based on experimental data of the model variables and sometimes it can be very tricky to find the optimum parameters that allows the model to fit satisfactorily experimental data. Usually, it is possible to discard a prior certain intervals of parameters values, narrowing down the search. For instance, in the model to quantify a person vWF multimers, negative values are unacceptable.

The concepts of estimability and identifiability also needs to be taken into consideration during parameter estimation. Estimability defines whether the model parameters can be determined from the available data and identifiability defines whether exists a unique solution from the optimization process. The process uniqueness has a great importance because independently from the initial parameters guess the identified parameters are unique.

It is necessary that the simulations satisfactorily describe the actual behavior of the biological system, including the multimers distributions and the values of antigen and collagen of each patient. In order to do that, model parameters shall be correctly identified. In the meantime, there are uncertainties regarding the knowledge of the parameters. These uncertainties exist for several reasons, such as: lack of data, variations in measured values over equipment errors and normal variations in vWF antigen and collagen binding numbers. The level of vWF in plasma may vary according to age, blood group, epinephrine, infamous mediators and endocrine hormones (menstrual cycle). For instance, the vWF increases with age, with acute stress or inflammation and increases three to five times the normal level during the third trimester of pregnancy. The vWF is reduced by hypothyroidism and rarely by auto antibodies to vWF.

The parameter identification is made by reducing the residuals. A residual is a measure of how well a line fits an individual data point. They are called ”errors”, not because something is wrong, but because some difference exist between the measured value and the model. The residual is the vertical distance between a data point and the model. Each data point has one residual. They are positive if they are above the regression line and negative if they are below the regression line. If the regression line actually passes through the point, the residual at that point is zero.

The residual defined as e can also be expressed in Equation (3.1). The e is the difference between the predicted value $\hat{y}$ and the observed value y.

$$Residual = Observed\ \ value\ -\ predicted\ \ value \tag{3.1a}$$

$$e = y - \hat{y} \tag{3.1b}$$

One of the techniques to perform parameter identification is the non-linear least-squares. The technique is based on an output error criterion where an existing estimate of the parameters is iteratively improved until the model response is sufficiently close to the measured values, that is when the objective function assumes a value within the imposed tolerance. The criterion used in this approach is generally the residuals minimization. This technique has several optimization algorithms to perform the minimization of errors. In general, an algorithm starts from initial estimates of parameters and improves it iteratively until the system model response is sufficiently close to that of the observations. In Simulink® the function *"lsqnonlin"* uses the the non-linear least-squares technique for optimization.

3.3.1 Nonlinear least-squares optimization function

According to Dennis (2013) the nonlinear least-squares curve fitting solver has the form of Equation 3.2.

$$\min_x f(x) = \sum_{i=1}^{m} f_i(x)^2 \tag{3.2}$$

where $f(x)$ represents the objective function, $f_i(x)$ represent the residuals functions and the index i indicates the particular data points. Residuals are the difference between the measured value and the calculated value using the model in consideration. The function is called nonlinear least-squares because indeed it tracks the minimum of the sum of squares of these residuals functions.

Nonlinear models are more difficult to fit than linear models because they demand an iterative approach that is summarized in the following the steps: determination of an initial guess for the each parameter, non linear models are sensitive to the starting point, so defining the initial guess is a crucial step. Generation of a fitted curve using the current set of parameter coefficients. Adjustment of the coefficients and determination of whether the fit improves. This step depends on the fitting algorithm. *Matlab Tutorial* (2019) indicates two algorithm methodologies. Finally, the process returns to the generation of the fitted curve and does not stop until the specified converge criteria is met.

Matlab Tutorial (2019) indicates the Trust-region and Levenberg-Marquardt as fitting algorithms. Trust-region algorithm must be used if it is necessary to specify coefficient constraints, this algorithm solves difficult nonlinear problems more efficiently than other.

Levenberg-Marquardt algorithm has been used for many years and it work for a wide range of nonlinear models, but it does not allow to create coefficient constraints to improve the fit.

The function *"lsqnonlin"* (nonlinear least-squares) was chosen to identify the model parameters, because considering the available functions within the Optimization tool package present in Simulink, this was the most robust, it works faster and presented lower residual values.

3.4 Optimization procedure at Laplace domain

The model constructed in the last Chapter was based on the multimers deviation variables. But the measured experimental data are in the form of vWF antigen and vWF collage binding. To identify the parameters using the non linear least-square function, the model constructed in Simulink must have the same output variables that were collected experimentally. Therefore, the model needed a modification to convert the deviation variables into vWF:Ag and vWF:CB. In order to do that, Equations (2.19), (2.20) and (2.21) were entered in the Simulink model.

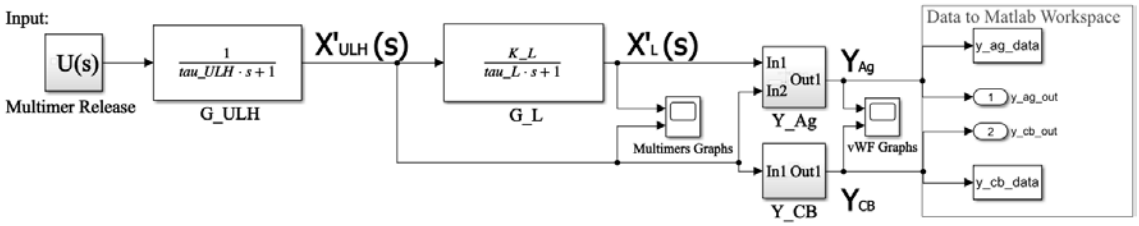


Figure 3.3. *Figure extract from Simulink to show the constructed model used to simulate the vWF model.*

The first step is convert the deviation variables back to the original variable form, to do that the deviation variable are added to their baseline values, defined by Equations (3.3) and (3.4). Equations (3.5) and (3.6) convert the multimers variable into vWF antigen and vWF collage binding variables. Where V_d has the constant value of 0.4, BW is the patient body weight and Y_b^{Ag} is the vWF antigen collect at time zero. Inside Equation (3.6) there are also the correction parameters α and Y_b^{CB} to be estimated. Figure 3.3 shows the Simulink image to simulate the model, where the blocks Y_{Ag} and Y_{CB} represent the functions created to convert the deviation variables into vWF antigen and collage binding respectively.

$$x_b^{UL+HMW} = Y_b^{CB} \cdot V_d \cdot BW \tag{3.3}$$

$$x_b^{LMW} = Y_b^{Ag} \cdot V_d \cdot BW - x_b^{UL+HMW} \tag{3.4}$$

$$Y^{Ag} = \frac{(x'_L + x_b^{LMW}) + (x'_{ULH} + x_b^{UL+HMW})}{V_d \cdot BW} \tag{3.5}$$

$$Y^{CB} = \frac{(x'_{ULH} + x_b^{UL+HMW})}{V_d \cdot BW} \cdot \alpha \cdot \frac{Y_b^{Ag}}{Y_b^{CB}} \tag{3.6}$$

The parameter identification through Simulink software starts by opening the parameter estimation application inside the Simulink Design Optimization toolbox. Once opened some general options must be defined, as explained before, the optimization method chosen was non linear least squares and the algorithm was Trust region reflective. The Parameter tolerance used was 0.001, the function tolerance was 0.001 and the maximum interactions was left 100. Then a new experiment was created by adding the experimental data collected from the patient. To conclude the configuration of the parameter estimation, initial conditions were entered.

3.4.1 Parameter set definition

The parameter set defined for identification was $\theta = [K_u \ \tau_U \ \tau_{ULH} \ K_L \ \tau_L \ Y_b^{CB} \ \alpha]$. The parameters K_u, τ_U, τ_{ULH}, K_L and τ_L are the parameters defined in §2.5 and summarized in Equations (2.31), (2.32), (2.33), (2.34) and (2.35). The Y_b^{CB} and α are considered because represents the parameters used to correct the values of the variable Y^{CB}. This adjustment was made in order to explain the different multimers affinity with collagen observed for different types of vWD. Galvanin et al. (2014) proposed this correction in Equation (2.15) and Ferrari et al. (2018) also used it, described in equation (2.21).

It is necessary to normalize parameters to avoid that their units or dimensions affect the estimation. Furthermore, the parameter normalization allows to search the values in the neighborhood of the unit, improving the robustness of the numerical method. By doing so, it is certain that each parameter contributes equally in the calculation process. In order to do that each parameter value needs to be divided by its initial value. For instance, Equation 3.7 shows the normalization for K_u, where K_u^o represents the initial value and K'_u represents the parameters to be identified.

$$K_u = K'_u \cdot K_u^o \tag{3.7}$$

3.4.2 Initial guess

Parameter identification process requires an initial parameter guess. Literature search or industrial information should be conducted to obtain reasonable initial parameter values, or ranges of values for similar systems. This parameter estimation step often causes the most difficulties for the modeller. Poor initial guesses for parameters can result in convergence to local optima rather than the best set of parameter values. Other estimation problems can result from limited or noisy data, parameters with small influence on model predictions, and correlated effects of parameters.

A non linear function with many coefficients to be optimized together may become very sensitive to the starting point for the minimization process. The greater the number of parameters the model has to improve, the greater the complexity of the optimization process, so the initial value of the parameters is of utmost importance.

Table 3.1. *Parameters obtained by Ferrari et al. (2018) converted in the Laplace Parameters for healthy groups HO, considered as initial guess for the process of parameter identification*

	$K_U^o[U.min]$	K_L^o	$\tau_U^o[min]$	$\tau_{ULH}^o[min]$	$\tau_L^o[min]$	Y_b^{CBo}	α
HO	$12.9 \cdot 10^5$	0.41	38	466	658	49.3	1.0

The previously mentioned researches have already found the parameters values for each vWD group. To identify the parameters within the Laplace domain, it is ideal to use the value already estimated by Ferrari et al. (2018), since it's the same model, just another condition. It is important to start a optimization process from the same point for all kind of subjects, regardless of the vWD condition. Thus, by identifying parameters of a new patient using the proposed model and starting from the same reference point, it is possible to compare this new patient data with those present in the database, and then possibly make a diagnose. So Table 3.1 shows the initial values used to identify the parameter for all patients in the database, this is the value estimated by Ferrari et al. (2018) for healthy patients with blood type O.

3.5 Validation of parameter identification

In order to identify the parameters set that best fits the model into experimental data, the Simulink optimization function conduct several simulations iteratively until achievement of the lowest objective function value. The simulations graphs provide a visual evaluation of the parameters quality. Meanwhile, the objective function and its residual show a quantitative evaluation.

The parameters were identified for each average subject and for all single patients presented in the database. The average subject data were calculated from the measurements

average, for each group studied, for each pre-established time collection point. Table 3.2 shows the parameters values for each studied group, putting together the parameter values estimated in previous studies, converted into parameters used in the Laplace domain, the parameters for the average subject and the parameter average value with their variance for each patient within the group.

Figure 3.4 shows the objective function values of the parameter identification process for all patients. The objective function is the sum of the squares of the residual values. The residual values are normalized in order to have all the values between 0 and 1. Figure 3.4 shows that quantitatively the model fits well with the experimental data, since the objective function values is around 0.1. Patient 51, in particular, has vWD type 1 and presents the highest objective function value, close to 0.5. That happens because this patient in particular has very low values of vWF:Ag and vWF:CB, an average value of 4 U/dL, so small variations produce high residual values compared to the measured values.

vWD type 1 is represented in Table 3.2 and other graphs to show its general behavior. However, it is a vWD type without a multimers qualitative mutation and many mutations within the group. This means that there is no well-defined pattern within the group. Because of that, it is difficult to interpret and compare it with other vWD types. Therefore it must be confronted separately, dividing between blood types and its various types of type 1 inside mutations.

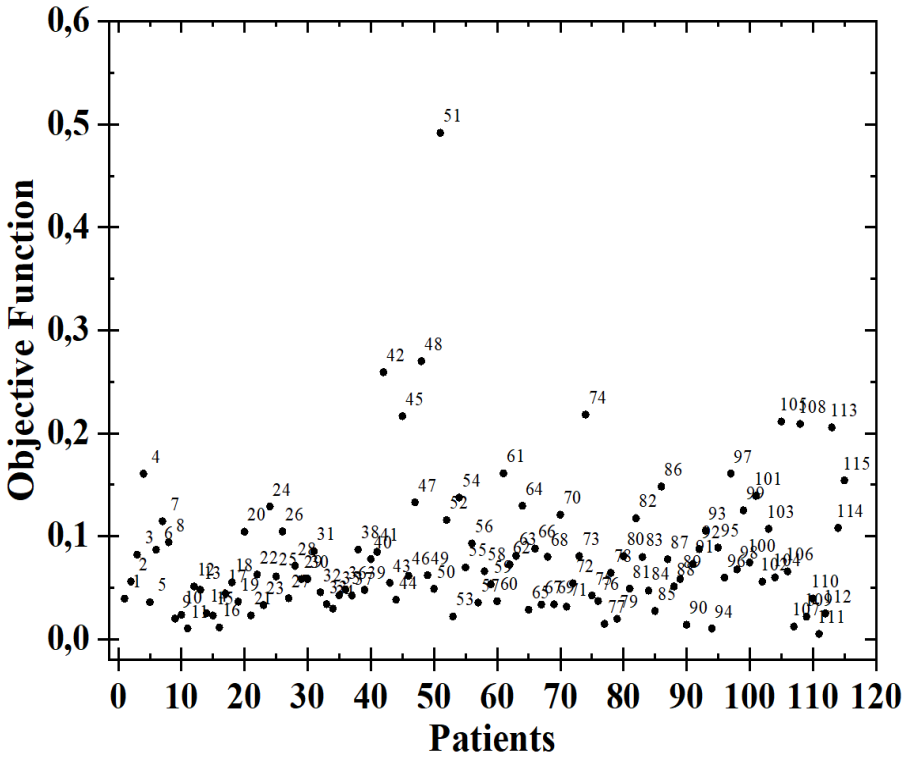


Figure 3.4. *Objective function value for the parameter identification for all patients.*

Table 3.2. *Table showing, for each studied group (healthy blood type O, healthy blood type non O, vWD type 2A, vWD type 2B, vWD type Vicenza and vWD type 1) the parameters identified by: previous work (first line). Average Subject (second line) and average of the parameters identified for each patient, followed by the variance of this average of parameters (third line). The numbers in parentheses below the name of each group indicate the amount of subject present in the group.*

		$K_U \cdot 10^5$ [U]		τ_U [min]		τ_{ULH} [min]		K_L		τ_L [min]		y_b^{CB} [U/dL]		α		Obj.Fun
Initial Guess		12.9		38		466		0.41		658		49.3		1.00		
HO	From Ferrari et al. (2018)	12.9		38		466		0.41		658		49.3		1.01		
(24 subjects)	Average Subject	14.4		34		643		0.23		39		46.2		1.01		0.05
	Single Subject	14.1	± 8.9	41	±15	565	± 226	0.59	±0.91	336	±424	47.9	± 19.8	1.05	±0.21	0.07
HnO	From Ferrari et al. (2018)	38.9		35		1063		0.34		1420		77.9		0.94		
(19 subjects)	Average Subject	44.7		34		1423		0.18		11		72.4		0.97		0.06
	Single Subject	33.9	±13.4	38	±11	1669	±586	0.53	±0.53	1207	±1356	74.5	± 22.3	1.00	±0.12	0.41
2A	From Ferrari et al. (2018)	9.20		49		483		0.17		565		11.0		0.56		
(3 subjects)	Average Subject	7.68		39		907		0.66		3289		14.9		0.48		0.34
	Single Subject	4.15	± 1.90	49	±15	410	±12	0.77	±0.53	2175	±1618	17.7	±8.1	0.56	±0.02	0.44
2B	From Ferrari et al. (2018)	7.12		56		126		1.46		310		26.1		0.20		
(8 subjects)	Average Subject	5.58		36		324		1.21		34		17.3		0.22		0.05
	Single Subject	4.87	±2.23	73	±79	235	±108	2.57	±2.80	70	±48	18.3	±8.76	0.26	±0.13	0.30
V	From Ferrari et al. (2018)	6.28		15		103		0.18		122		2.4		0.31		
(9 subjects)	Average Subject	2.49		22		97		0.36		9		3.9		0.64		0.03
	Single Subject	2.76	±1.77	37	±26	79	±28	0.42	±0.34	36	±50	4.5	± 1.46	0.74	±0.34	0.43
1	From Castaldello et al. (2018)	3.01		36		556		0.29		714		77.9		0.94		
(51 subjects)	Average Subject	11.4		31		625		0.23		72		21.1		0.89		0.07
	Single Subject	10.3	±8.67	53	±50	459	±245	0.75	±1.23	821	±1185	22.1	±10.5	0.94	±0.25	0.35

3.5.1 Average subject

The average subject is formed by the mean measured values and might present a concentration profile that differ from some patients within its group. Figure 3.5 presents the average subject simulated graphs. Figure 3.5a and 3.5b shows, respectively, the graphics for the vWF antigen and collage binding, where the dots are the measured value and the line are simulated data, and Figure 3.5c and 3.5d shows, respectively, the low molecular weight multimers (LMW) and ultra large and high molecular weight (UL+HMW) values calculated using the estimated correction factor parameter for each group, where the dots are the calculated value and the line are simulated data. Figure 3.5 shows that visually all the simulations have a good fit to the measured data.

A high τ_i (i stands for ULH or L) value correspond to a slow system response, without forming any peaks. But as τ_i value decreases, and the system response becomes faster, a peak formation becomes more evident and noticeable, meaning the multimers number come back to the basal state faster. Taking into consideration groups HO, HnO, 2B and Vicenza, the healthy groups HO and HnO presents the highest values for τ_{ULH} and τ_L. That happens because healthy people are expected to keep their vWF level higher for longer. Healthy patients with blood type O and non-O present in the plasma multimers with higher number for ultra large and high molecular weight multimers. For the low molecular weight multimers, although the vWD type 2B group presents the highest peak, the values for Healthy patients stay constant high through time.

Figure 3.5c shows a peak for the LMW multimers simulation for the groups HO, HnO, 2B, Vicenza and type 1, while type 2A has a more flat profile. The vWD 2A has a mutation that interferes with the assembly and release of UL+HMW multimers, so UL+HMW multimers graph has the lowest growth peak during release. This explain LMW multimers not having a value peak after release. The vWD 2B has a low UL+HMW plasma concentration caused by increased affinity for platelet and high LMW plasma concentration caused by high rate of proteolysis, Figures 3.5c and 3.5d show this behaviour by the highest LMW multimer peak and the highest K_L value for this group.

The vWD Vicenza are known for high elimination rate for both UL+HMW and LMW multimers, presenting the lowest τ_L and τ_{ULH} values. Figure 3.5 shows for all graphics that vWD Vicenza presents a peak that returns to the starting point very fast, consistent with the fact that it presents a mutation that creates a high multimers elimination rate. Causing a low multimers number at basal state and a very fast response for induced vWF release.

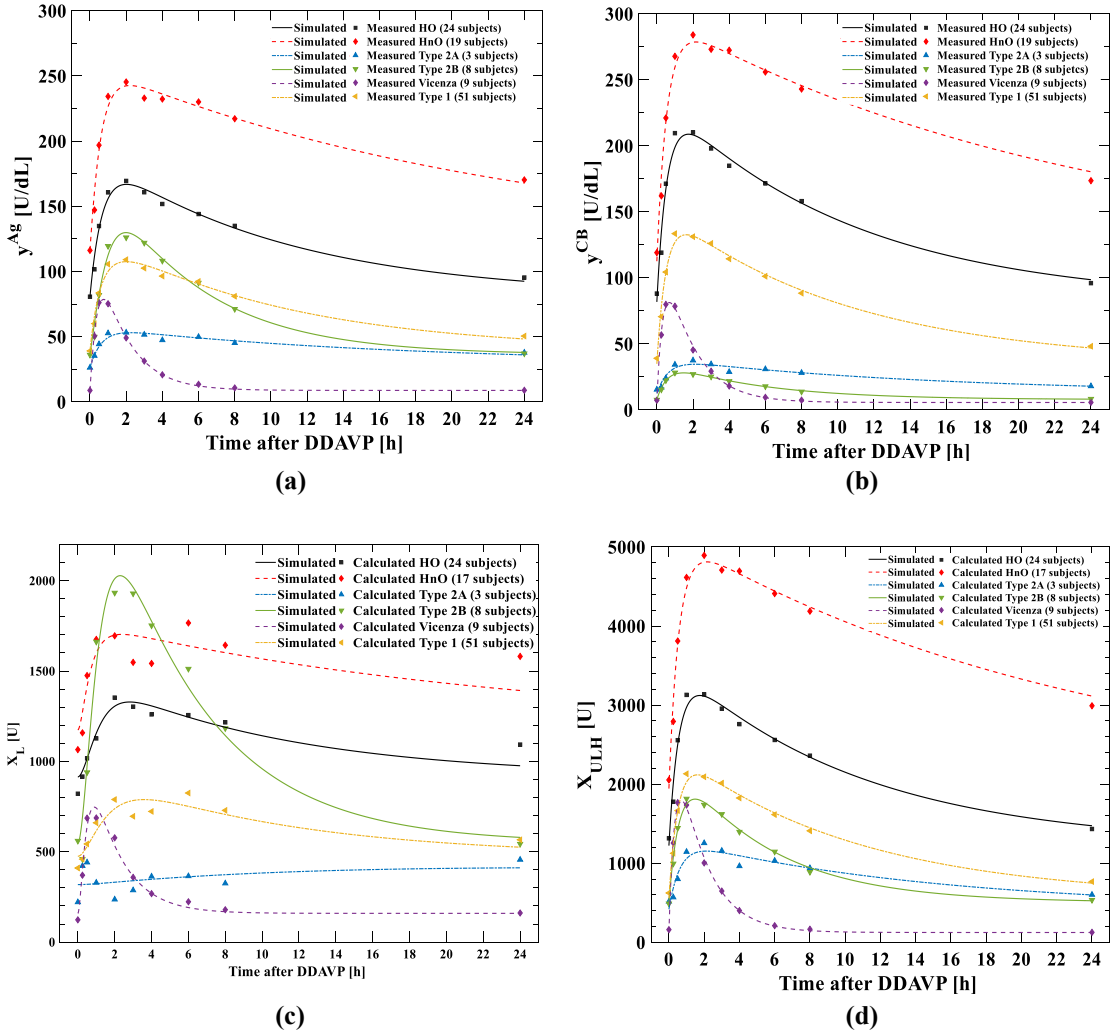


Figure 3.5. *The graphics show the simulations (lines) using the model estimated parameters for each group in order to compare it with the measured values (dots). Figure a) and b) the dots represents mean measured values. Figure c) and d) the dots represents the multimers calculated value using vWF:Ag and vWF:CB mean measurements. Figure a) shows the graphic for vWF Antigen (vWF:Ag) concentration expressed as* y^{Ag} *[U/dL] by the time after the DDAVP administration. Figure b) shows the graphic for vWF collage binding (vWF:CB) concentration expressed as* y^{cB} *[U/dL] by the time after the DDAVP administration. Figure c) shows the graphic for low molecular weight multimers expressed as* X_L *[U] by the time after the DDAVP administration. Figure d) shows the graphic for Ultra Large and High molecular weight multimers expressed as* X_{ULH} *[U] by the time after the DDAVP administration. estimated.*

3.5.2 Identifiability analysis

Table 3.2 shows the parameters of the average subject, the average of each patient and the ones obtained in previous studies for each group. It is noticeable the similarity between them, having the same magnitude order and no substantial difference, with the exception of τ_L.

The previous Chapter defines $\tau_L = 1/k_e$. In Ferrari et al. (2018) model the parameter k_e is the multimers elimination rate. According to Lenting et al. (2004) the multimers biological elimination mechanism is independent of size. Because of that, the parameter k_e is present for both UL+HMW and LMW multimers. Therefore, in the previous studies, the k_e estimated value must be good to fit both multimers groups. At the Laplace domain, the τ_L determines the behavior of LMW multimers and the τ_{ULH} determines the behavior of UL+HMW multimers. Therefore, what was previously predicted to work in both multimers profiles, in the Laplace domain, parameter τ_L should fit for LMW multimers and τ_{ULH} should fit for UL+HMW multimers.

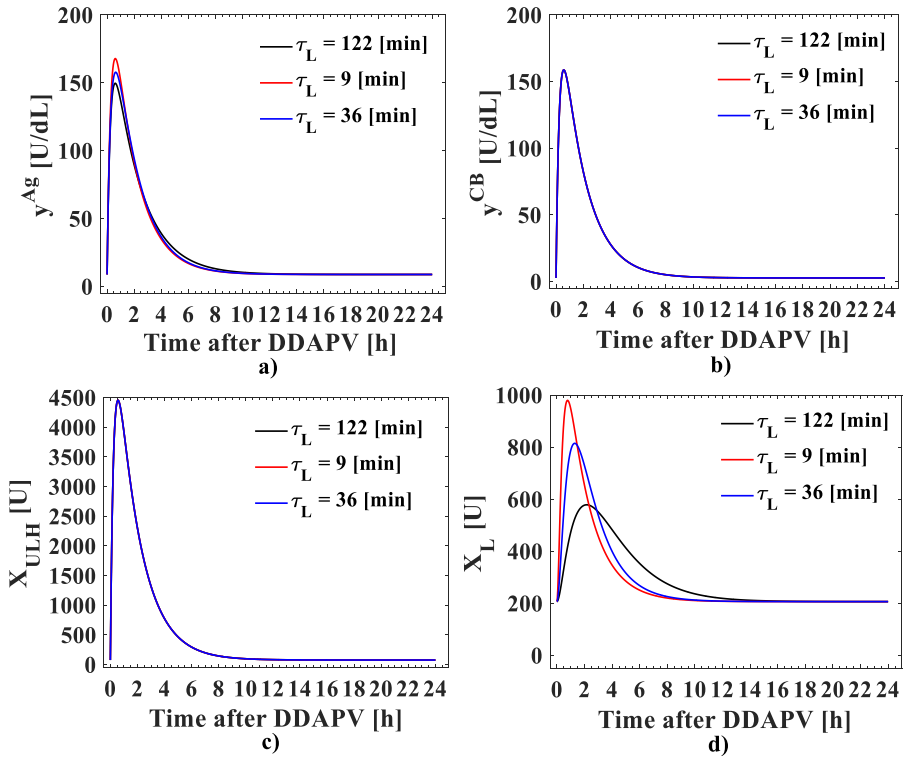


Figure 3.6. *Figure shows three different simulations with three values for tau$_L$. Figure a) presents the vWF:Ag profile, Figure b) presents the vWF:CB profile, Figure c) presents the LMW multimer profile and Figure d) presents the UL+HMW multimer profile.*

Table 3.2 presents different values of τ_L within the same group. This raises questions about the identifiability for this parameter. To assess this, Figure 3.6 shows three simulations with three different values for τ_L, and the other parameters constant, with values

equal to the ones in Table 3.2 subject Vicenza from Ferrari et al. (2018). It shows that τ_L does not influence the values of X_{ULH} and Y^{CB}, so the graphs in Figure 3.6b and 3.6d do not change. However, Figure 3.6c shows that changes in the value of τ_L clearly change the values of X_L, but have no significant changes to the values of Y^{Ag}. That happens because, Y^{Ag} is calculated with the sum of the multimers X_{ULH} and X_L. The X_{ULH} multimer has values much larger than X_L. Thus, the variable Y^{Ag} is much less sensitive to variations in X_L, and as consequence also parameter τ_L has little influence. Because the experimental data is measured in Y^{Ag} and Y^{CB} it became a problem to identify τ_L.

In order to control whether other parameters also have this sensitivity problem, a simulation was conducted with a healthy subject data (as an example, patient 4) considering three scenarios of different initial guess parameters. Case 1 used as initial value the estimated parameters of the average subject of the HO group. Case 2 used as initial value the one previously considered in Table 3.1. Case 3 used as initial value the estimated parameters of the average subject of the Vicenza group. Table 3.3 shows the normalized values used as initial guess for each case, and the values found after the parameters identification. Table 3.3 and Figure 3.7 show that τ_L and K_L parameters have an identifiability problem, meanwhile the other parameters show consistent values. The similarity in the objective function values in Table 3.3 highlights the identifiability problem.

Table 3.3. *Table showing the values for estimated parameters considering different initial guess.*

Initial Guess								
	K_U	τ_U	τ_{ULH}	K_L	τ_L	y_b^{CB}	α	
Case 1	1.114	0.893	1.379	0.560	0.060	0.937	0.924	
Case 2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
Case 3	0.193	0.571	0.208	0.879	0.014	0.078	0.639	
Parameter Identified								
	$K_U \cdot 10^5$	τ_U	τ_{ULH}	K_L	τ_L	y_b^{CB}	α	Objective function
Case 1	28	29	1032	0.07	108.23	25.9	0.62	0.171
Case 2	27	30	962	0.97	1545.1	25.1	0.61	0.161
Case 3	30	30	1070	0.07	113.78	24.5	0.59	0.169

The identifiability problem may results from inappropriate model structure or insufficient experimental data. By looking at Figure 3.7a and 3.7b the model curve fits well the measured values, meaning a well structured model. But converting the measured values of vWF:Ag and vWF:CB into X_L and X_{ULH} multimers, Figure 3.7c shows that the distribution of values does not have a profile as expected, so it is difficult to find K_L and τ_L parameters, revealing that, besides the sensitivity problem, there is also an issue in the experimental data. This is quite evident when finding the parameters of each patient individually, but for the average subject this problem is not so evident. There are several other tests to assess the identifiability of nonlinear models, but global identifiability tests are difficult to implement and are usually restricted to small systems (McLean and McAuley,

2012). Even though these two parameters are not reliable, there are still five parameters to work with.

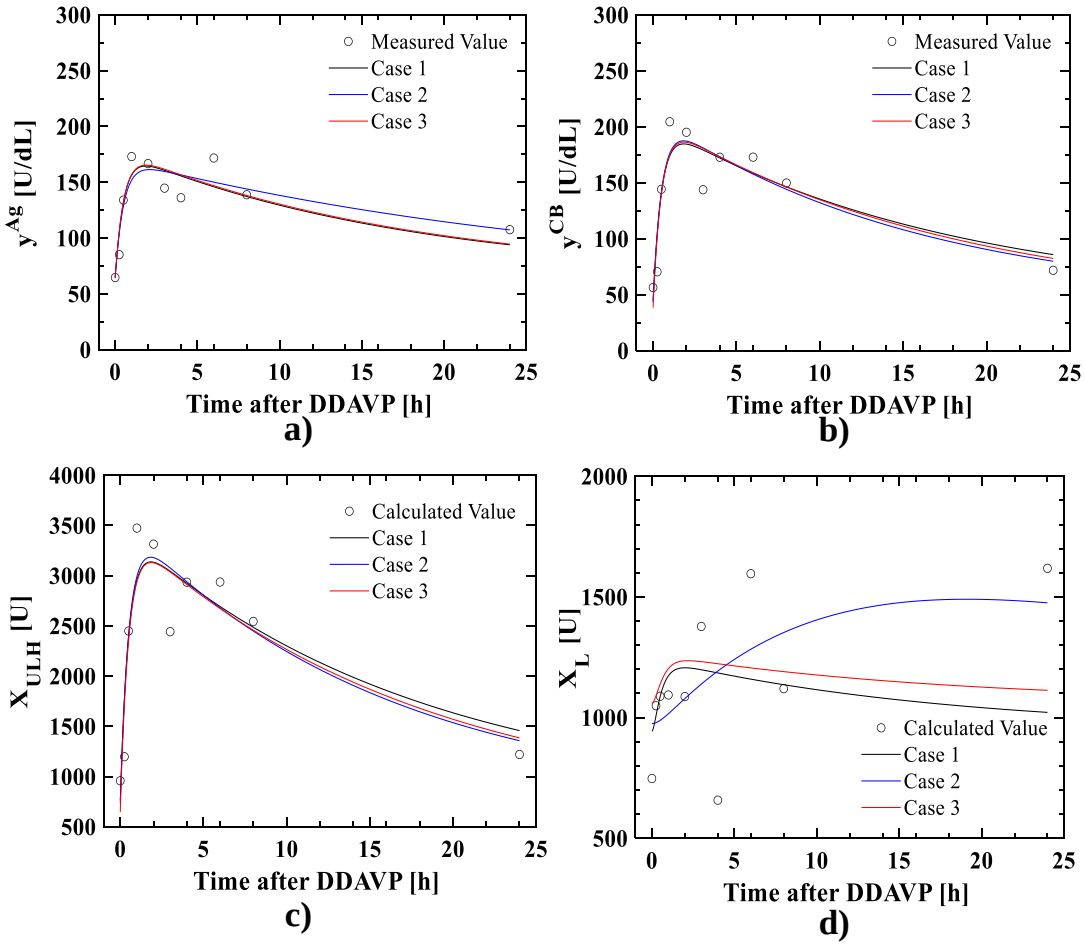


Figure 3.7. *Figure shows three different parameter estimation scenarios for the Healthy Person 4 blood type O. For each estimation case, Figure a) presents the vWF:Ag profile, Figure b) presents the vWF:CB profile, Figure c) presents the LMW multimer profile and Figure d) presents the UL+HMW multimer profile. The dots for Figures a) and b) are the measured values, and the dots for Figures c) and d) are the multimers values calculated with the measured values.*

Chapter 4

von Willebrand disease classification based on principal component analysis

This chapter aims to work with the parameters identified in the previous chapter in order to understand their correlation with the vWD. First, the available parameters and the concept of principal component analysis are described. In order to diagnose the vWD, different strategies of modelling the principal component analysis are presented. The Chapter concludes with the work results and projections for future studies.

4.1 Optimized parameters available

The parameters identified for each patient are in Table B.1, B.2 B.3 and B.4 of Appendix B, separated by groups. The next step is to verify the possibility to describe and diagnose the disease using the identified parameters. In order to do that each group must have a significant number of samples. The healthy blood type O and non-O groups have together data from 43 people. The vWD type 2B has data from eight people and vWD type Vicenza has data from nine people. These groups (HO, HnO, 2B and Vicenza) present sufficient samples to find a pattern within the estimated parameters. The vWD type 2A has data from three people, which does not give much information about the variability of the parameters found for this group. As already discussed, type 1 has very wide variability within the group. Even though there is a large number of people in the database (51), there are many mutations within the group. This means that there is no well-defined pattern within the group and will not be considered.

The previous chapter identified the parameters K_U, τ_U, τ_{ULH}, K_L, τ_L, Y_b^{CB} and α within the model in the Laplace domain. However, the parameters K_L and τ_L present an identifiability problem, but parameters K_U, τ_U, τ_{ULH}, Y_b^{CB} and α have coherent values. Therefore, the total data available to discriminate the vWD lacks two parameters, but there are still five parameters and each one has information and variability for each group.

It is difficult to do a graphical representation of five parameters together. Working with a system of many dimensions makes it difficult to visualize the data or discriminate it between groups. To solve this issue, a principal components analysis (PCA) model

helps to project these parameters on another plane with less dimensions without losing information. In this way it is easier to clarify and distinguish each one of the groups, opening the possibility of vWD diagnosis.

4.2 Principal components analysis

Principal components analysis (PCA) is a tool for data compression and information extraction. PCA finds combinations of variables that describe major trends in the data. The technique that uses linear algebra principles to transform possibly correlated variables into a smaller number of variables is called principal components (PC). It is useful when the data is presented in a matrix with many dimensions and a graphical representation is not possible. In general, PCA seeks to reduce the number of dimensions of the original data set by projecting it in a new plane.

Mathematically, according to Wise et al. (2006), PCA relies upon an eigenvector decomposition of the covariance (or correlation) matrix of the process variables. For a given data matrix $\boldsymbol{X}$ with m rows and n columns, where each variable is a column and each sample is a row, PCA decomposes $\boldsymbol{X}$ as the sum of r components, formed by $\mathbf{t}_i$ and $\mathbf{p}_i$. r is the rank of the matrix $\boldsymbol{X}$ and must be less than or equal to the smaller dimension of $\boldsymbol{X}$ ($r \leq min(m,n)$):

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \cdots + \mathbf{t}_r\mathbf{p}_r^T \tag{4.1}$$

The $\mathbf{t}_i$ vectors are the scores and contain information on how the samples relate to each other. The $\mathbf{p}_i$ vectors are the loadings and contain information on how the variables relate to each other. Generally, the PCA model is truncated after k components and the remaining small variance factors are consolidated into a residual matrix E, as described by Equation (4.2):

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \mathbf{t}_3\mathbf{p}_3^T + \mathbf{E} \tag{4.2}$$

Many times each variable works in different unit, so the columns of X must be autoscaled. That means, adjust the mean to zero and the variance to unit by dividing each column by its standard deviation. In the PCA decomposition, the $\mathbf{p}_i$ vectors are eigenvectors of the covariance matrix. Equation (4.3) shows that each eigenvector $\mathbf{p}_i$ has a eigenvalue λ_i associated with.

$$cov\,(\mathbf{X})\,\mathbf{p}_i = \lambda_i\mathbf{p}_i \tag{4.3}$$

where the covariance matrix of $\boldsymbol{X}$ is defined as

$$cov\,(\mathbf{X}) = \frac{\mathbf{X}^T\mathbf{X}}{m-1} \tag{4.4}$$

Equation (4.5) shows that the score vector $\mathbf{t}_i$ is the linear combination of the original $\boldsymbol{X}$ variables defined by $\mathbf{p}_i$.

$$\mathbf{X}\mathbf{p}_i = \mathbf{t}_i \tag{4.5}$$

where λ_i are a measure of the amount of variance described by the $\mathbf{t}_i$, $\mathbf{p}_i$ pair. Another way to look is variance as information. Because these pairs are in descending order of λ_i, the first pair captures the largest amount of variance. Generally, data can be adequately described using far fewer factors than original variables. Thus, the data overload problem can be solved by observing fewer scores than original variables, with no significant loss of information. In this thesis context, the parameter identified will be used as variables for the PCA model. So PCA turns up the combinations of variables that predicts the distinction of patients in their studied groups.

Figure 4.1 exemplify how the PCA model works by showing the values of three variables. The plotted samples lie on a plane and are enclosed by an ellipse. It is also apparent that the samples vary more along one axis of the ellipse than along the other. The PCA model with two principal components adequately describes all the variation for this case: the first describes the direction of the greatest variation (the main ellipse axis) and the second describes the direction of second greatest variation, (the minor ellipse axis).

The statistical SPE and T^2 values measures the PCA model quality. The squared prediction error (SPE) is the sum of squares of each row (samples) of $\boldsymbol{E}$ (in the Equation (4.2)) using Equation (4.6).

$$\mathbf{SPE}_i = \mathbf{e}_i\mathbf{e}_i^T \tag{4.6}$$

where $\mathbf{e}_i$ is the i$^{\text{th}}$ row of $\boldsymbol{E}$. The SPE statistic indicates how well each sample conforms to the PCA model. It is a measure of the difference, or residual, between a sample and its projection into the k principal components retained in the model. Such contributions can be useful in identifying the variables which contribute most to the sum-squared residual error of a sample. Because the PCA model captures some variables more effectively than others, the residual variance may vary widely between variables.

Figure 4.1 shows SPE as a measure of the distance off the plane containing the ellipse, the Euclidean distance of the data point from the plane formed by the two-component model. The sample on the upper left side of Figure 4.1 presents a large SPE value. The

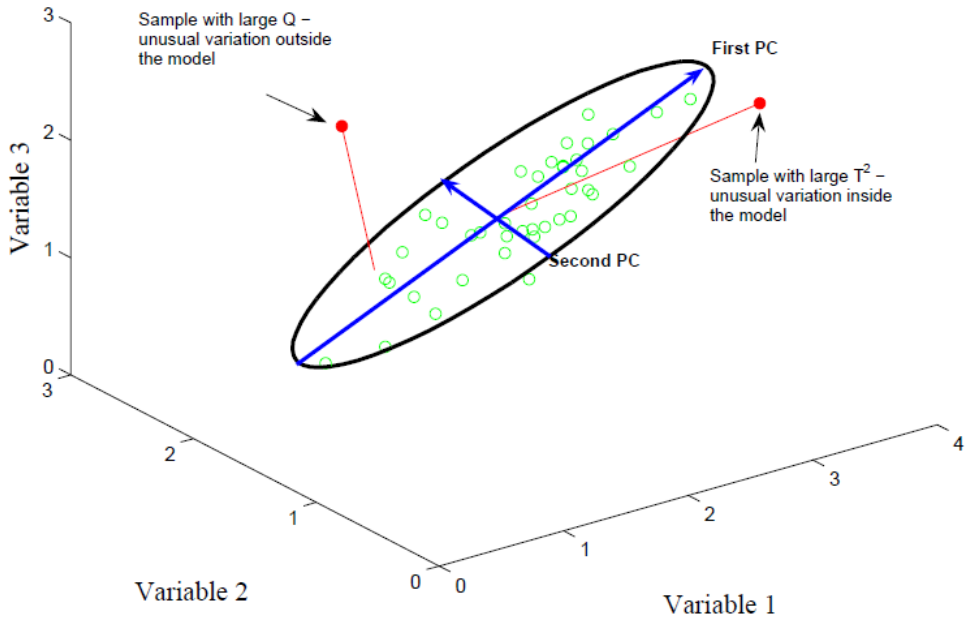


Figure 4.1. *Figure exemplifying the PCA model for a dataset divided by three variables (Wise et al., 2006)*

SPE limit defines a distance of the plane that is considered unusual based on the data used to form the PCA model.

The sum of normalized squared scores, known as T^2 statistic, is a measure of the variation in each sample within the PCA model.

$$T_i^2 = \mathbf{t}_i \lambda^{-1} \mathbf{t}_i^T \tag{4.7}$$

where t_i refers to the ith row of T_k , the m by k matrix of k scores vectors from the PCA model, and λ is a diagonal matrix containing the eigenvalues corresponding to k eigenvectors retained in the model. Figure 4.1 show a large T^2 value on the upper right side. T^2 is a measure of the distance from the multivariate mean (the intersection of the PCs in the figure) to the projection of the sample onto the two principal components. The T^2 limit defines an ellipse on the plane within which the data normally project.

4.3 PCA model from parameters estimated in the Laplace domain

The PCA model was created using a matrix with 60 rows (24 healthy person with blood type O, 19 healthy person with blood type non-O, 8 patients with vWD type 2B and 9 patients with vWD Vicenza) and 5 columns (parameters K_U, tau_U, tau_{ULH}, Y_b^{CB} and α). The PLS_Toolbox Analysis developed by Eigenvector Research® and available to use

with Matlab® was used to create the PCA model.

Before starting the calculation routine, the data has to be loaded in the software and preprocessed to adjust the units of the columns. Since every column has its own units, they have to be scaled to the same variation magnitude. After initiating the PCA model calculation, the number of principal components (PC) to be retained in the model must be analyzed and that step follows a few rules of thumb (Wise et al., 2006). Table 4.1 lists the eigenvalues, the percentage of the variance, the cumulative variance and the root-mean-square error of cross-validation (RMSECV) that each PC captures. Figure 4.2a shows the graphical representation for eigenvalues by the number of principal components and Figure 4.2b shows the graphical representation for RMSECV values by the number of principal components.

Table 4.1. *List of eigenvalues, variance percentage, cumulative variance percentage and RMSECV values that each principal component of the PCA model captures*

PC	Eigenvalue	% Variance	% Variance Cumulative	RMSECV
1	2.59	51.79	51.79	0.926
2	1.07	21.33	73.12	1.096
3	0.73	14.55	87.67	1.759
4	0.37	7.48	95.15	26.07
5	0.24	4.85	100.0	26.07

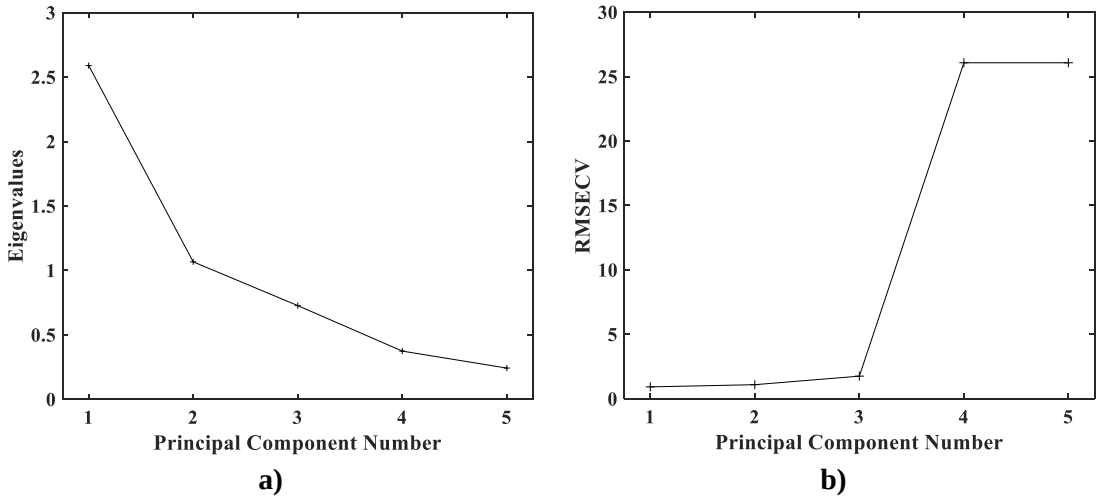


Figure 4.2. *Graphical representation for Eigenvalues that each principal component captured from the PCA model made with the data matrix of 60 patients*

If the plot of eigenvalues shows a "knee" in the line, a sudden jump in the values, the point where that happens indicates to be an appropriate number of PCs for the model. In addition, eigenvalues can be interpreted as the number of original variables each PC is worth. Therefore, a component with an eigenvalue less than one probably does not describes any systematic variance in the system, and should likely not be included in the

model. Eigenvalues greater than 2 usually describe systematic variation, but less than that must be evaluated. It is also possible to choose the number of PCs based on root-mean-square error of cross-validation (RMSECV). When a component that describes the system with a large variance is added to the model, the error decreases. On the other hand, when a component that describes the system with a small variance is added, the error increases. As a rule of thumb, the point where the RMSECV value rises corresponds to the appropriate number of PCs for the model (Wise et al., 2006).

According to the rules presented, Figure 4.2a has inconclusive graphical representation to the eigenvalues, not showing a sudden jump in the values. But Figure 4.2b presents a sudden rise in the RMSECV value after the third component. Confronting Table 4.1, the third component presents eigenvalues less than 1, but it it still has 14% variance and added with the first and second components captures more than 87% of the variance in the data, which suggests that the data are fairly well correlated. This way, the model follows the next steps with a three principal components, capturing over 87% of the variation, leaving 13% non-deterministic variance in the system.

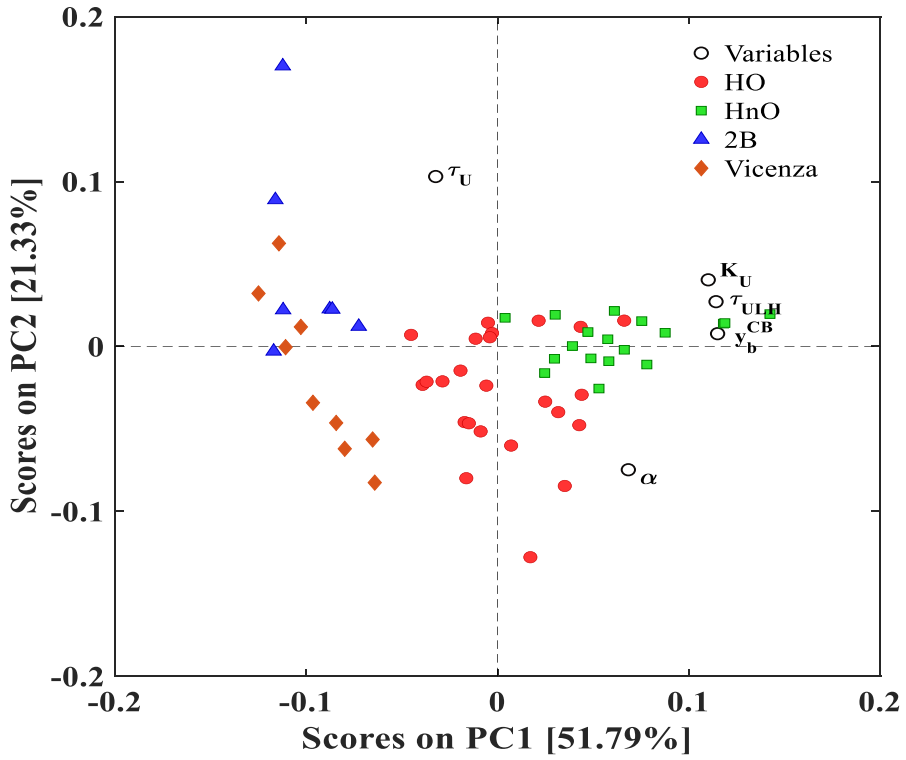


Figure 4.3. *Scores plot from Principal Components Analysis for Component 1 and 2, divided into classes..*

Similar patients must have similar component scores, creating clusters. So it is expected that people with the same characteristics aggregate in the same score cluster. The scores' plot shows the relationship between samples, putting patients with same vWD condition close to each other. Figure 4.3 shows the scores for the first and second component

and Figure 4.4 shows the scores for the first and third component. Figure 4.3 shows that variables which load most significantly into the first PC are K_U, τ_{ULH} and Y_b^{CB}. These variables are the furthest from zero in the right direction on the plot, and since they are on the same side they are positively correlated with each other. Then τ_U and α are closer to zero for the first component and further from zero for the second PC, suggesting that they have opposite correlation with each other for the second component.

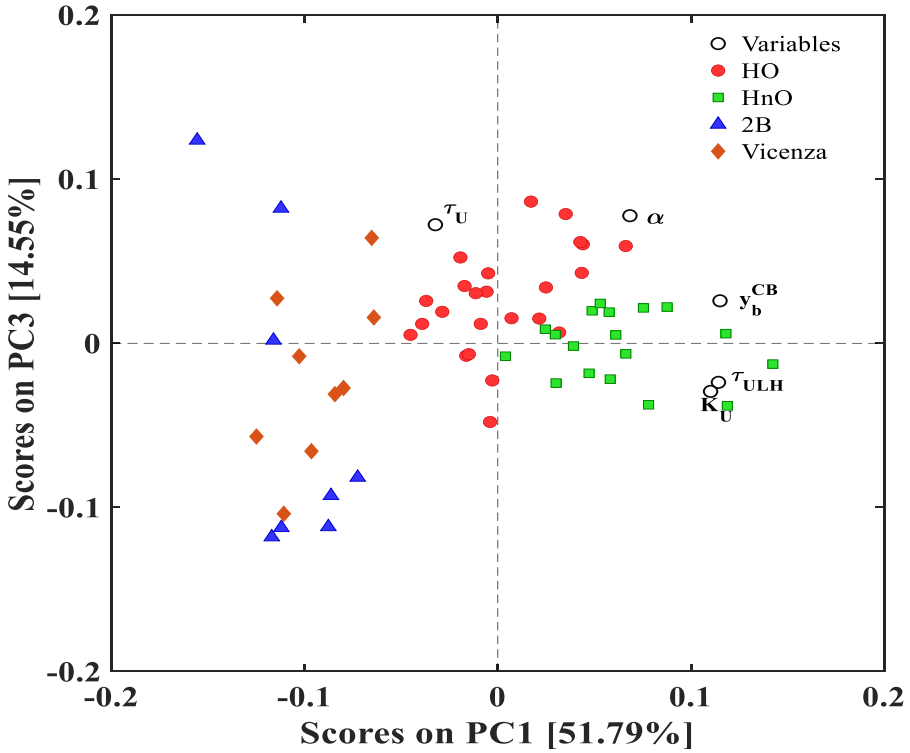


Figure 4.4. *Scores plot from Principal Components Analysis for Component 1 and 3, divided into classes.*

A sample with high values for K_U, τ_{ULH} and Y_b^{CB} tends to move the score towards the right for the first component. Healthy people with blood type O place around zero for both components, meaning that it is a default group. Healthy people with blood type non-O place around zero for the second component, and positively for the first component, which it has already been demonstrated as the group with the highest multimers values, thereby it has high values for K_U, τ_{ULH} and Y_b^{CB}. Alpha is the correction to adjust the values for collage binding vWF. Patients with vWD type 2B have low values of Y^{CB} and the highest contrast between vWF:Ag and vWF:CB, so they are negatively related with alpha. Patients with vWD type Vicenza show negative values for the first component, but for the second component, the scores are around zero divided between negatives and positives. The Vicenza type has low values for both vWF:Ag and vWF:CB profiles, especially for τ_{ULH}. Thereby having a negative value for the first component.

Figure 4.3 and 4.4 show, for the first component, different clusters between healthy

subjects and those suffering from vWD. But the second and third components show no distinction between group values. Especially for 2B and Vicenza samples, the clusters overlap creating a difficulty to divide the vWD type. Therefore, modeling the PCA with the parameters identified in the Laplace domain can simply determine whether a subject is healthy or not, but does not provide a clear separation between vWD groups.

Even though there is a separation for the first component, the scores between the healthy and unhealthy subjects are close, so it is necessary to assess how reliable the model is to diagnose a healthy subject from those who suffer from vWD. To do that, a different strategy is considered. It was used only healthy people data to construct the PCA model. Since it is easier (for the future) to collect data from people not suffering from vWD. Then, by applying the patients' data into the model, it is possible to verify whether the patient conforms to the healthy model.

4.4 PCA model from healthy people parameters estimated in the Laplace domain

With the aim to verify whether a person suffers from vWD or not a PCA model was made with healthy people parameters identified in the Laplace domain. Similarly described in the previous section, this PCA model was created with 43 healthy subjects (24 with blood type O and 19 with blood type non-O). Therefore, the matrix loaded to create the model was composed of 43 rows and 5 columns (parameters K_U, tau_U, tau_{ULH}, Y_b^{CB} and α). It must be considered that in order to have a robust model it is easier to obtain more data to calibrate the model with healthy people. Table 4.2 shows the eigenvalues, variance, cumulative variance, and RMSECV values for the PCA model. Figure 4.5a shows the graphical representation for eigenvalues by the number of principal components and Figure 4.5b shows the graphical representation for RMSECV values by the number of principal components.

Table 4.2. *List of eigenvalues, variance percentage, cumulative variance percentage and RMSECV values that each principal component of the PCA model made with only healthy people.*

PC	Eigenvalue	% Variance	% Variance Cumulative	RMSECV
1	2.31	46.17	46.17	1.404
2	1.27	25.35	71.51	1.466
3	0.87	17.40	88.92	1.888
4	0.34	6.81	95.73	33.89
5	0.22	4.27	100.0	33.89

Using the same rules discussed earlier, Figure 4.5a presents two angles that could determine the number of components: the first component and the fourth component. However, Table 4.2 shows that the first component alone does not capture much of the total

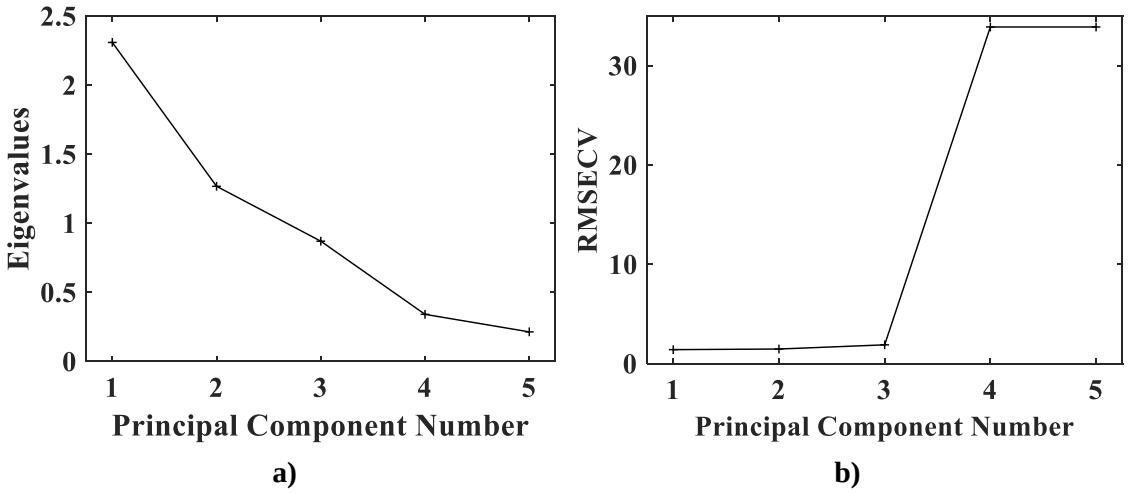


Figure 4.5. *Graphical representation for Eigenvalues that each principal component captured from the PCA model made with the data matrix of 43 healthy people*

data variance and the fourth captures a lot of information that is not important to describe the model. Figure 4.5b shows that the value of RMSECV rises after the third component, so the fourth and fifth components describes such small variance that the error increases. The third component has an eigenvalue of 0.87, but still gets 17% of variance and together with the first and second components it captures 89% of accumulated variation, leaving only 11% variance non-deterministic variance in the system. Therefore, in this case the model is developed with three principal components

4.5 Discriminating the von Willebrand Disease using the PCA model

A PCA model using the parameters identified in the Laplace domain from only healthy people (which will be referenced from now on as healthy PCA model) will provide similar scores and information for each person. As already seen, patients suffering from vWD have characteristics that create differences in the parameters identified in the Laplace domain. Thereby, by entering the parameters of a unhealthy patient into the healthy PCA model, the score should place outside the confidence level for healthy people. This way, by projecting the vWD patients into the healthy PCA model, it is possible to evaluate how many patients place outside the confidence level and verify how reliable this model is to diagnose the vWD. Therefore, eight vWD type 2B patients, nine vWD type Vicenza and three vWD type 2A was projected into the healthy PCA model. For this scenario, each sample is evaluated individually, so the amount of patients does not matter, allowing to consider patients with vWD type 2A.

When project new data to an existing model the SPE plot is usually the most sig-

nificant graphic representation. This plot tells if the new samples are fit by the existing model. Therefore it can tell whether the new data fit as a healthy person or not. The SPE limit defines a distance within the plane that is considered unusual based on the data used to form the PCA model. Figure 4.6 shows SPE plot of the new samples (vWD type 2A, 2B and Vicenza) and the samples used to calibrate the model (healthy people blood type O and nonO). Note that the 95% confidence level line has been added to the plot.

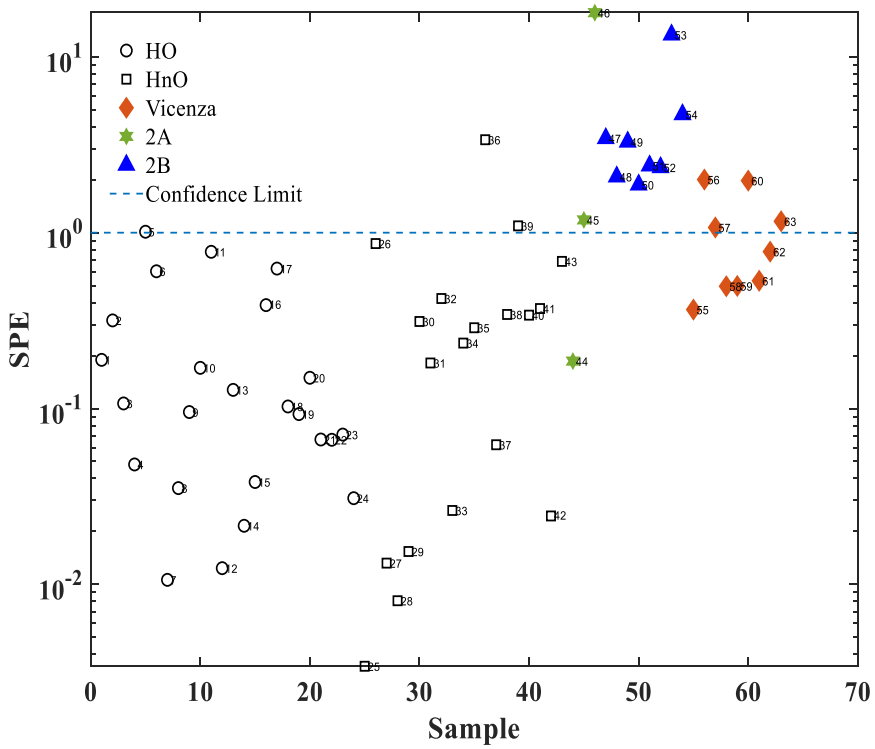


Figure 4.6. *SPE values for new samples (vWD type 2A, 2B and Vicenza) and the samples used to calibrate the model (healthy people blood type O and nonO, described in the legend as HO and HnO respectively) from the data set presented previously. The blue traced line shows the 95% confidence level.*

Figure 4.6 shows that 6 patients (1 with vWD type 2A, and 5 with vWD type Vicenza) out of 20 are below the confidence limit line. That means the model predicts the disease in 70% of cases with 95% confidence. However, being more specific all the patients with vWD 2B are above the confidence line, and only 30% of the Vicenza cases are above the confidence line. The same happens for T^2 values, Figure 4.7 shows the T^2 plot of the new samples (vWD type 2A, 2B and Vicenza) and the samples used to calibrate the model (healthy people blood type O and nonO).

The T^2 is a measure of the variation in each sample within the PCA model. How far the sample distance from the multivariate mean (the intersection of the PCs). Also in this case 6 patients (1 with vWD type 2A, and 5 with vWD type Vicenza) of 20 are below the confidence level line. In both Figure 4.6 and 4.7 there are some calibration data for HO and HnO above the 95% confidence level. This influences the model robustness, considering

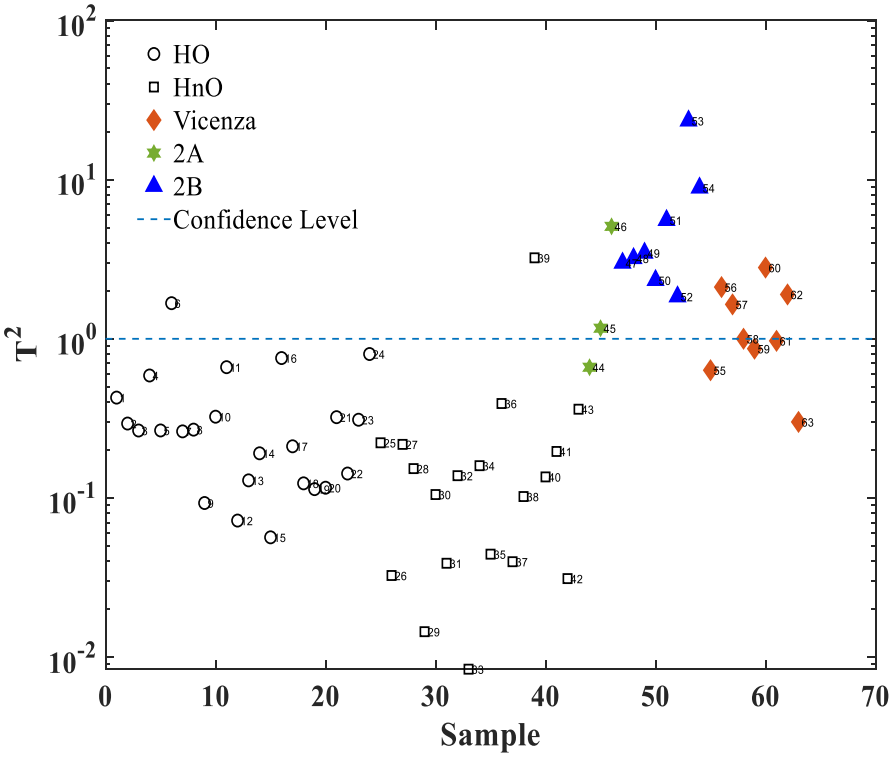


Figure 4.7. *T^2 values for new samples (vWD type 2A, 2B and Vicenza) and the samples used to calibrate the model (healthy people blood type O and nonO, described in the legend as HO and HnO respectively) from the data set presented previously. The blue traced line shows the 95% confidence level.*

this samples as possibles outliers. Another way to evaluate how the PCA model performs with the new data is by plotting the components scores. Figure 4.8 shows the scores for the first and second component and Figure 4.9 shows the scores for the first and third component and Figure 4.10 shows the scores for the second and third component.

Figure 4.8 shows 13 out of 20 samples inside the confidence level for scores on the first and second component. Figure 4.9 shows 7 out of 20 samples inside the confidence level for scores on the first and third component. Especially the first component, the scores for the new data are detached from healthy people scores, but still close enough to be inside the confidence level. The second and third components scores are more distributed, as in Figure 4.10. Figure 4.10 shows 5 out of 20 samples inside the confidence level for scores on the third and second component, 4 samples for vWD type Vicenza and 1 sample for vWD type 2A. Figure 4.11 shows the percentage variance each component captures by variable. Second and third component capture essentially the same variables τ_U and Y_b^{CB}. As already noted, vWD 2B are more sensible to these variables indeed explaining why any sample of this group is below the confidence level line.

The SPE defines a distance on the plane based on the data used to form the PCA model. This means that by adding more data, SPE values may change. There are still common variations within healthy people, for instance the blood type O and non-O show

a significant difference of vWF measured values, which reflects in the PCA model. Other variations like, epinephrine, infamous mediators and endocrine hormone cycles can create outliers for healthy people. Therefore, it is important to maintain a large database, because this provides a robust PCA model to help diagnose people suffering from vWD. This can rise the reliability of the PCA model to diagnose the vWD disease.

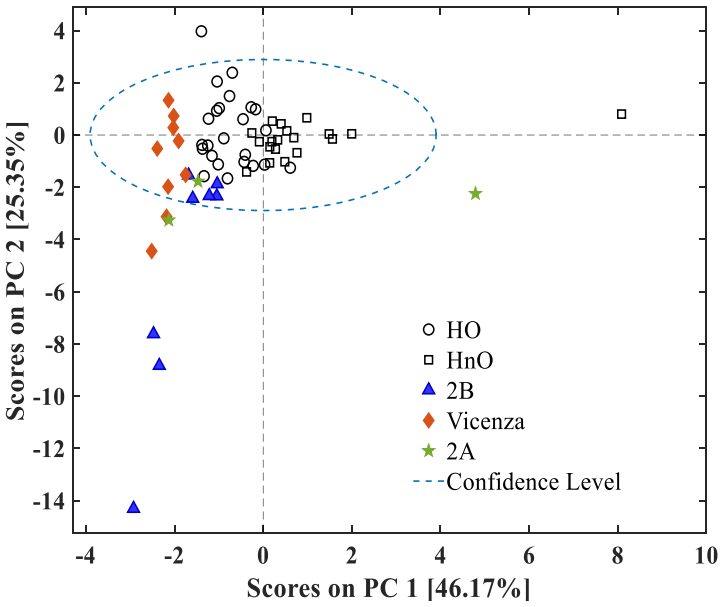


Figure 4.8. *Scores plot for the first and second components from the PCA model made with healthy people samples and applied new data with patients with vWD.*

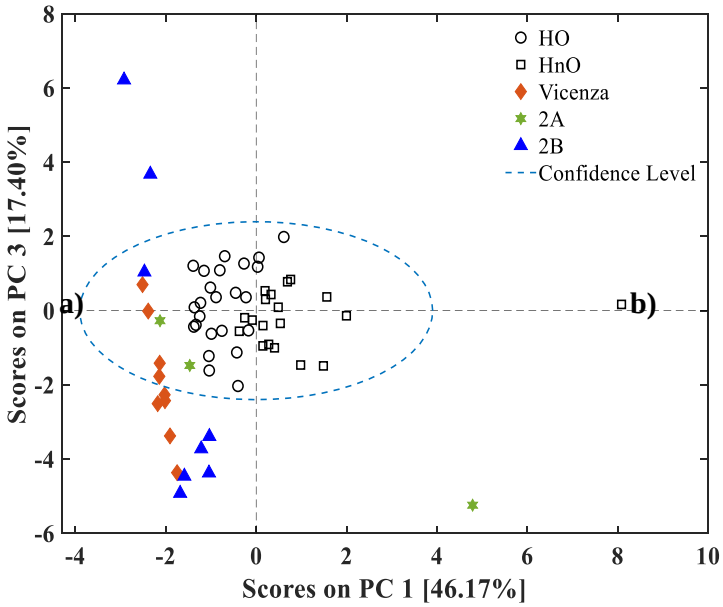


Figure 4.9. *Scores plot for the first and third components from the PCA model made with healthy people samples and applied new data with patients with vWD.*

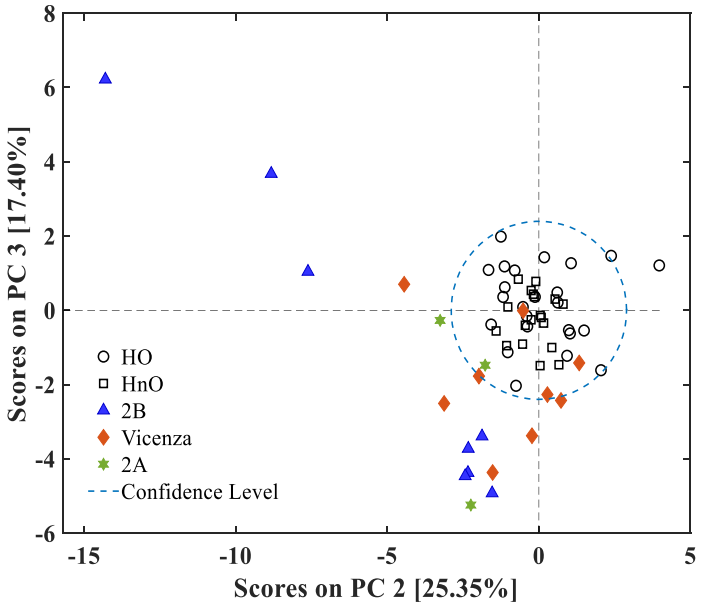


Figure 4.10. *Scores plot for the second and third components from the PCA model made with healthy people samples and applied new data with patients with vWD.*

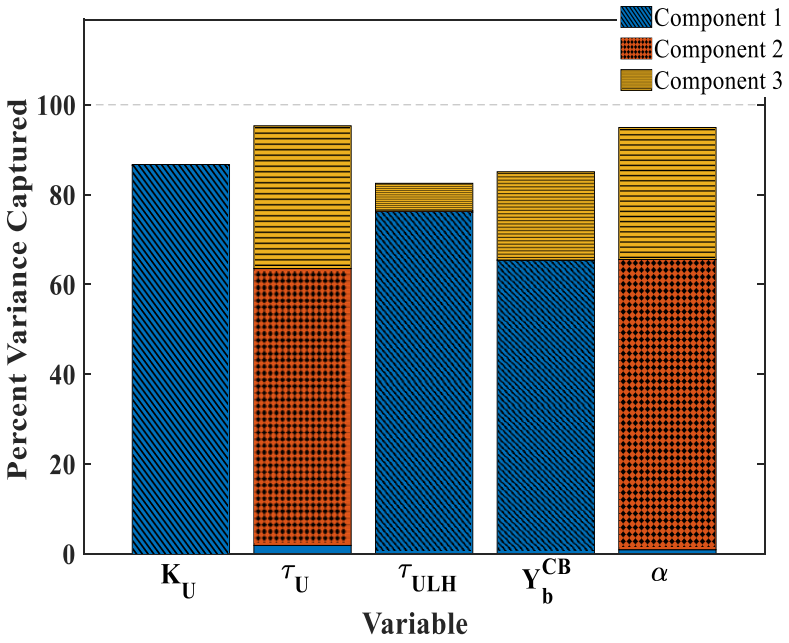


Figure 4.11. *Variance captured for each variable in three component model made with healthy people.*

4.6 Prospective for future work

The Laplace transform was able to solve the mathematical model developed by Ferrari et al. (2018). Simulations using Simulink software provides accurate results. This allows the creation of new study scenarios, such as the simulation of a treatment for a patient with low multimer response, defining the amount of DDAVP needed to achieve normal vWF conditions. For instance, for a vWD type Vicenza that has high elimination

rate, a simulation can define the frequency of application in order to maintain the multimers levels.

The parameters defined in the Laplace domain have information on the behavior of ultra large and small multimers. For instance, the time constants provide insight into multimers response after DDAVPT induction. Even if the model presents a problem for the parameter τ_L, it still shows differences in the speed of response for large and small multimers. It opens a question about the multimers elimination mechanism. An opportunity to study a way to identify the parameter using measured data for the low molecular weight multimers.

The PCA model, the introduction of more healthy people data could build a more robust model by decreasing the presence of outlier. Furthermore, a higher SPE confidence level could increase the reliability of the model as a tool for the diagnosis process of vWD. This database does not provide much information about the blood type from patients suffering from vWD. The healthy PCA model can be separated according to blood type to help diagonsis the vWD.

Conclusions

The von Willebrand disease (vWD) is a hereditary diseases that affect the blood clotting process. This disease affects 1% of the world's population and the diagnosis process is very complex. This work purpose is to use the Laplace transform to solve the vWD mathematical model in order to give a new perspective for the vWD diagnosis protocol.

The use of Laplace transform solved the vWD model developed by Ferrari et al. (2018). The projection of the model into the Laplace domain converted differential equations into algebraic equations, allowing to construct new parameters. The model proposed by Ferrari et al. (2018) shows multimers rate of reaction mechanism, indicating the rate at which multimers vary. The model in the Laplace domain gives information about UL+HMW and LMW multimers response facing a stimulus, defining how the multimers behave over time until return to the basal state. The multimers graphical representation for the model simulated with the software Simulink (a Matlab® tool) provided accurate results. This made it possible to use Simulink to identified the parameters in the Laplace domain, by using the experimental data, in order to have a well defined model.

The model in the Laplace domain had the parameters identified in order to reach out the experimental data. Optimization techniques provided information about the process estimability and identifiability of the model. Parameters K_L and τ_L showed an identifiability problem caused by low model sensitivity and lack of experimental data. Despite this problem, another five parameters were identified with success, still giving good model fit with the experimental data available.

The principal components analysis (PCA) provides instrument to work with many parameters and variables. By creating a PCA model with identified parameters, for the groups 2B, Vicenza, healthy blood type O and healthy blood type non-O, it was not possible to detect a significant difference between the samples scores clusters. An alternative strategy was proposed to get around the problem: a PCA model based on healthy people data was created. When projecting a individual data into the healthy PCA it is possible to say with a certain level of confidence if they are affected or not by the disease. Hence the model was able to distinguish 70% of the 20 vWD patients presented in the database. Even if it is not an expected result, there is still room for improvement of this strategy.

Nomenclature

D	=	multimers release amount [U]
$G_L(s)$	=	low multimer transfer function in the Laplace domain
$G_{ULH}(s)$	=	ultra large and high multimer transfer function in the Laplace domain
k_O	=	release kinetic parameter [min^{-1}]
k_1	=	proteolysis kinetic parameter [min^{-1}]
k_2	=	proteolysis kinetic parameter [min^{-1}]
k_3	=	proteolysis kinetic parameter [min^{-1}]
k_4	=	proteolysis kinetic parameter [min^{-1}]
k_5	=	proteolysis kinetic parameter [min^{-1}]
k_6	=	proteolysis kinetic parameter [min^{-1}]
k_e	=	clearance kinetic parameter [min^{-1}]
K_L	=	parameter constant gain in the Laplace domain
K_U	=	input parameter in the Laplace domain [U.min]
t	=	time
t_{max}	=	time of maximum response [min]
$U(s)$	=	input function in the Laplace domain
V_d	=	volume distribution [dL]
x^{SUL}	=	super ultra large multimers [U]
x_b^{SUL}	=	super ultra large multimers at basal state [U]
x^{UL}	=	ultra large multimers [U]
x_b^{UL}	=	ultra large multimers at basal state [U]
x^{H}	=	high multimers [U]
x_b^{H}	=	high multimers at basal state [U]
x^{L}	=	low multimers [U]
x_b^{L}	=	low multimers at basal state [U]
x^{UL+HMW}	=	ultra large and high molecular weight multimers [U]
x_b^{UL+HMW}	=	ultra large and high molecular weight multimers at basal state [U]
x^{LMW}	=	low molecular weight multimers [U]
x_b^{UL+HMW}	=	low molecular weight multimers at basal state [U]
x'_{ULH}	=	ultra large and high molecular weight multimers deviation variable [U]
x'_{L}	=	low molecular weight multimers deviation variable [U]
y^{Ag}	=	vWF antigen concentration [U/dL]
y_b^{Ag}	=	vWF antigen concentration at basal state [U/dL]
y^{CB}	=	vWF collagen binding concentration [U/dL]
y_b^{CB}	=	vWF collagen binding concentration at basal state [U/dL]

Acronyms

ABO	=	blood group classification system
ADAMTS13	=	enzyme metalloprotease
BW	=	body weight
DAE	=	differential algebraic equation
DDAVP	=	desmovasopressin (1-desamino-8-d-arginine vasopressin)
ELISA	=	enzyme-linked immuno sorbent assay method
FVIII	=	coagulation factor VIII
GpIb	=	platelet binding gycoprotein Ib
HMW	=	high molecular weigh multimers
LMW	=	low molecular weigh multimers
LSE	=	least squares estimation
MLE	=	maximum likelihood estimation
PC	=	principal component
PCA	=	principal component analysis
RMSECV	=	root-mean-square error of cross-validation
ODE	=	ordinary differential equation
SUL	=	super ultra large
vWD	=	von Willebrand disease
vWF	=	von Willebrand factor
vWF:Ag	=	von Willebrand factor antigen
vWF:CB	=	von Willebrand factor collagen binding
vWF:Rc	=	von Willebrand factor activity of the ristocetin cofactor
vWF:FVIII	=	activity of the FVIII with vWF
ULMW	=	ultra large molecular weight multimers
UL+HMW	=	ultra large and high molecular weight multimers

Greek letters

α	=	correction factor
λ	=	eigenvalue
τ	=	time constant (min)
τ	=	time constant (min)
τ	=	time constant (min)
$\mathcal{L}$	=	laplace operator
θ_i	=	i-th parameter

Matrix and Vectors

$\mathbf{E}$	=	residual matrix
$\mathbf{e}_i$	=	i-th vector row from $\mathbf{E}$
$\mathbf{p}$	=	loadings vector or eigenvector
θ	=	model parameter vector
$\mathbf{t}$	=	scores vectors
$\mathbf{X}$	=	Matrix of data

Appendix A

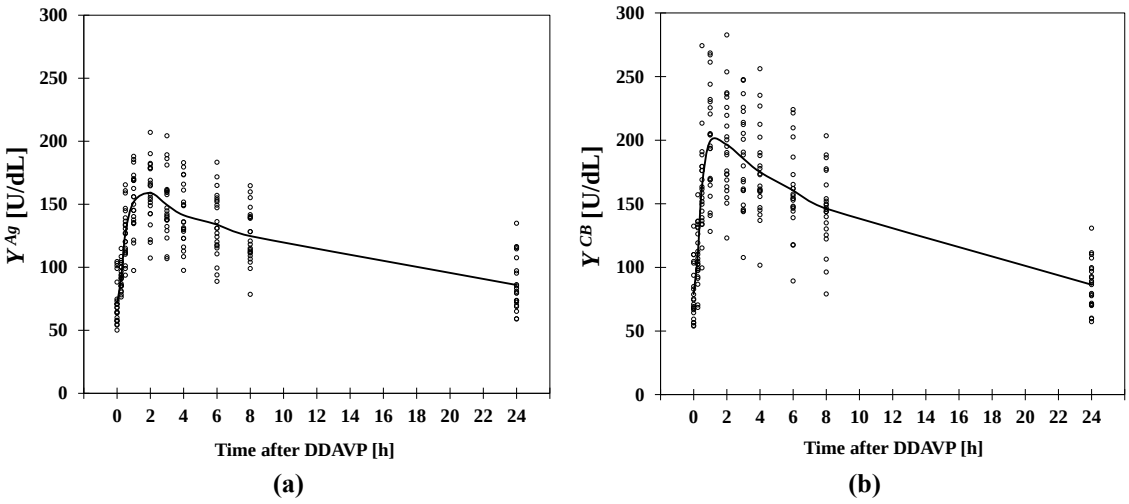


Figure A.1. *Graphic showing as dots the measured data for healthy people with blood type O, and as line the average data value for each establish time, Figure a) shows the antigen vWF concentration (vWF:Ag) as Y^{Ag} and Figure b) shows the collage binding vWF concentration (vWF:CB) as Y^{B}.*

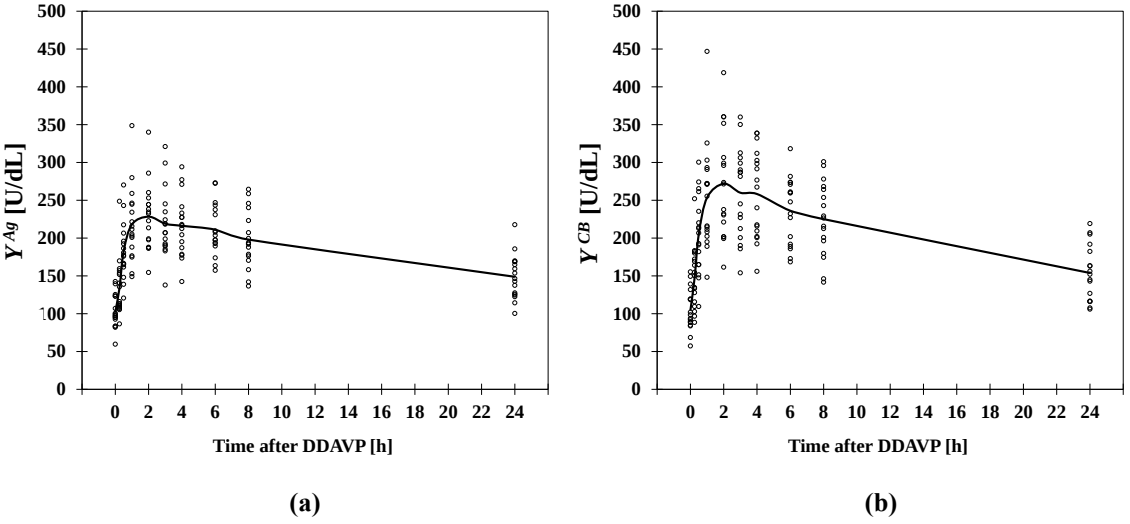
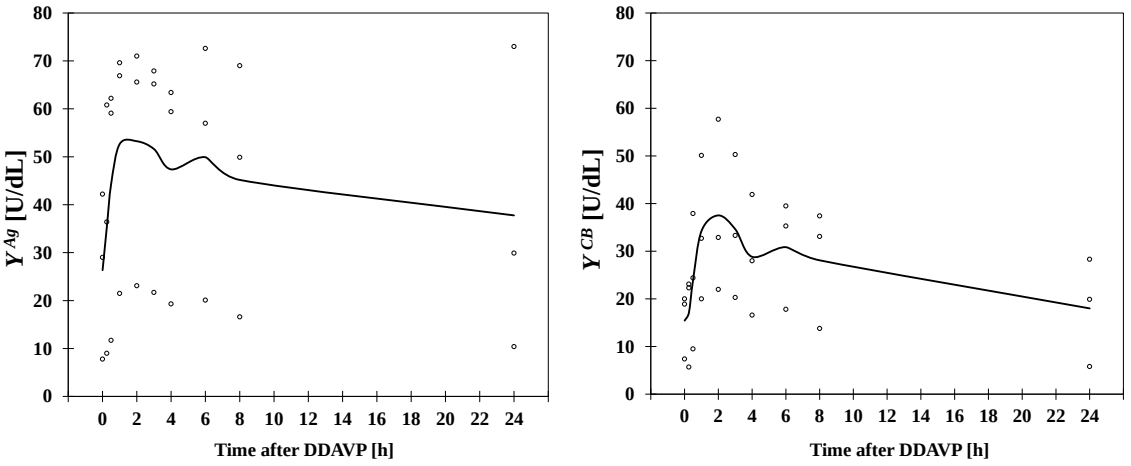
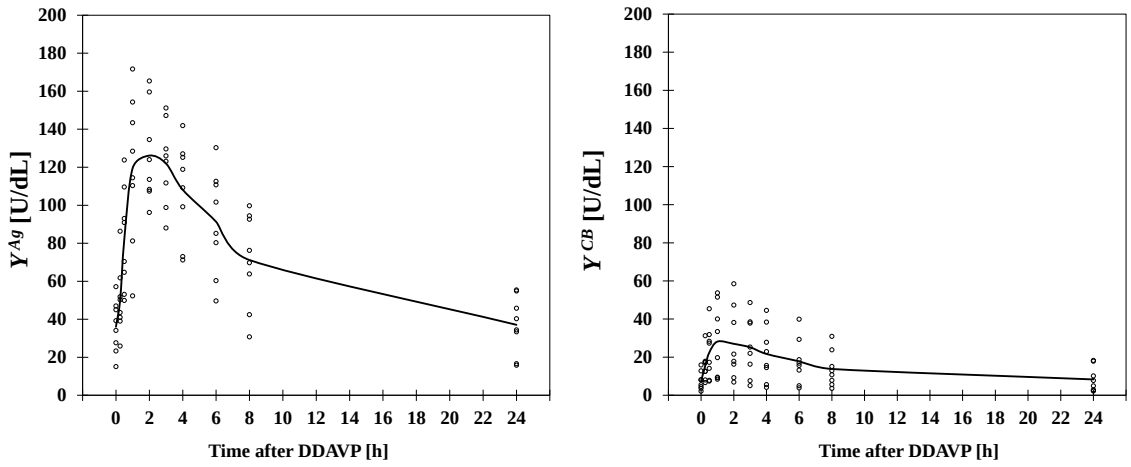


Figure A.2. *Graphic showing as dots the measured data for healthy people with blood type non-O, and as line the average data value for each establish time, Figure a) shows the antigen vWF concentration (vWF:Ag) as Y^{Ag} and Figure b) shows the collage binding vWF concentration (vWF:CB) as Y^{B}.*



(a) **(b)**

Figure A.3. *Graphic showing as dots the measured data for patients with vWD type 2A, and as line the average data value for each establish time, Figure a) shows the antigen vWF concentration (vWF:Ag) as Y^{Ag} and Figure b) shows the collage binding vWF concentration (vWF:CB) as Y^{B}.*



(a) **(b)**

Figure A.4. *Graphic showing as dots the measured data for patients with vWD type 2B, and as line the average data value for each establish time, Figure a) shows the antigen vWF concentration (vWF:Ag) as Y^{Ag} and Figure b) shows the collage binding vWF concentration (vWF:CB) as Y^{B}.*

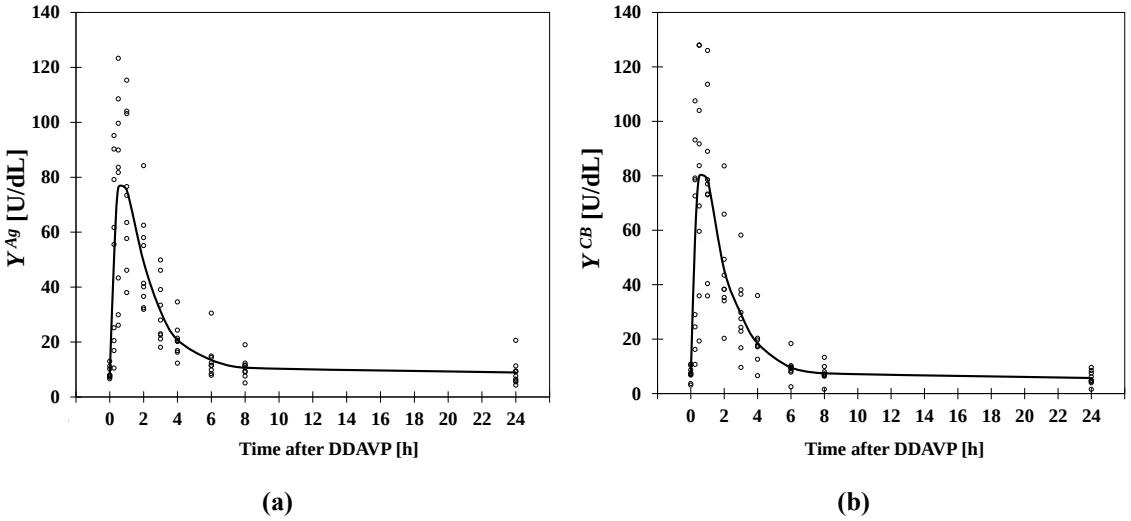


Figure A.5. *Graphic showing as dots the measured data for patients with vWD type Vicenza, and as line the average data value for each establish time, Figure a) shows the antigen vWF concentration (vWF:Ag) as Y^{Ag} and Figure b) shows the collage binding vWF concentration (vWF:CB) as Y^{B}.*

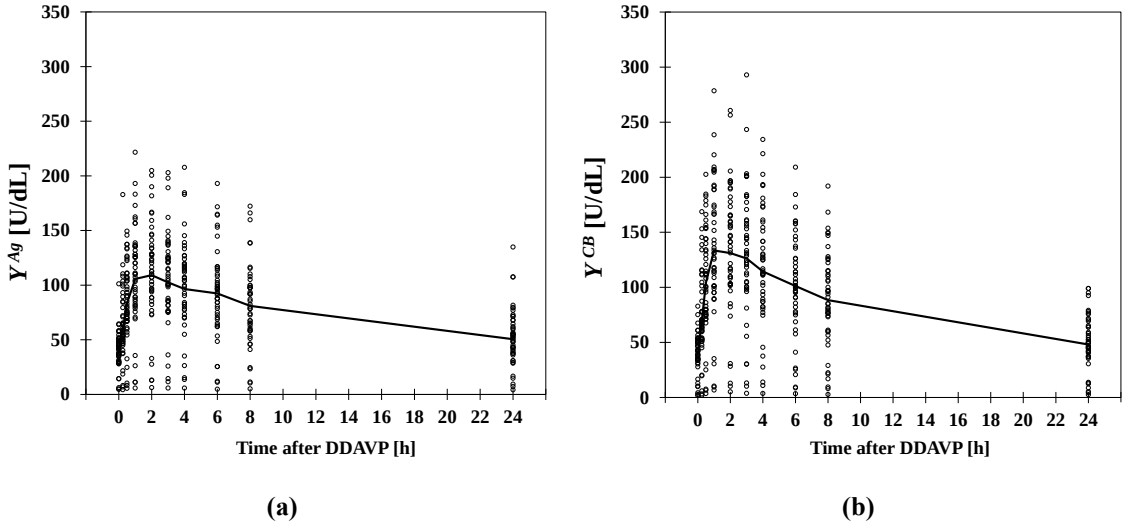


Figure A.6. *Graphic showing as dots the measured data for patients with vWD type 1, and as line the average data value for each establish time, Figure a) shows the antigen vWF concentration (vWF:Ag) as Y^{Ag} and Figure b) shows the collage binding vWF concentration (vWF:CB) as Y^{B}.*

Appendix B

Table B.1. *List of parameters estimated for patients carrying vWD type 2*

Patient	vWD	Mutation	K_u	τ_u	τ_{ULH}	K_L	τ_L	Y_b^{CB}	α	Objec
95	Type 2	A2	0.698	0.947	0.580	6.954	0.682	0.379	7.126	0.166
96	Type 2	A1	0.469	1.171	0.907	0.196	0.000	0.421	0.671	0.437
97	Type 2	A1	0.174	1.944	0.854	2.776	4.920	0.093	0.140	0.448
98	Type 2	A1 (Acquired)	3.971	0.775	15.09	2.630	5.000	0.565	1.405	0.450
99	Type 2	B	0.225	0.440	0.741	9.442	0.060	0.547	6.288	0.278
100	Type 2	B	0.355	0.548	0.566	3.644	0.095	0.133	1.129	0.172
101	Type 2	B	0.060	0.190	0.163	22.813	0.270	0.153	1.020	0.357
102	Type 2	B	0.311	0.655	0.826	6.423	0.034	0.441	2.144	0.156
103	Type 2	B	0.532	2.483	0.202	0.818	0.167	0.234	0.594	0.279
104	Type 2	B	0.371	0.580	0.686	2.655	0.071	0.647	2.295	0.174
105	Type 2	B	0.562	6.235	0.509	3.434	0.053	0.323	2.191	0.421
106	Type 2	B	0.603	4.204	0.336	0.835	0.098	0.495	1.094	0.533

Table B.2. *List of parameters estimated for patients carrying vWD type Vicenza*

Patient	vWD	K_u	τ_u	τ_{ULH}	K_L	τ_L	Y_b^{CB}	α	Objec
107	Vicenza	0.140	0.557	0.127	0.913	0.016	0.125	0.106	0.102
108	Vicenza	0.146	2.467	0.209	0.200	0.000	0.086	0.135	0.940
109	Vicenza	0.235	0.346	0.188	1.282	0.031	0.085	0.150	0.115
110	Vicenza	0.122	0.384	0.161	1.905	0.024	0.111	0.127	0.166
111	Vicenza	0.324	0.541	0.182	0.482	0.056	0.087	0.109	0.193
112	Vicenza	0.554	0.382	0.307	0.259	0.019	0.047	0.184	0.377
113	Vicenza	0.156	1.481	0.121	0.446	0.001	0.124	0.188	0.975
114	Vicenza	0.090	1.268	0.113	2.854	0.257	0.038	0.116	0.308
115	Vicenza	0.158	1.368	0.113	0.785	0.092	0.111	0.084	0.653

Table B.3. *List of parameters estimated for healthy people with blood type O and nonO*

Patient	Blood type	vWD	K_u	τ_u	τ_{ULH}	K_L	τ_L	Y_b^{CB}	α	Objec
1	O	Healthy	0.881	1.722	1.433	0.669	0.121	1.016	1.058	0.040
2	O	Healthy	0.922	1.606	2.529	1.054	0.012	1.377	1.331	0.056
3	O	Healthy	1.389	1.160	1.122	3.048	1.016	0.933	1.163	0.085
4	O	Healthy	2.030	0.817	2.042	1.468	2.000	0.561	0.843	0.161
5	O	Healthy	0.704	1.108	0.854	1.562	0.196	1.917	1.738	0.108
6	O	Healthy	0.277	0.524	0.739	11.38	1.613	0.684	0.412	0.088
7	O	Healthy	0.943	0.582	1.362	0.612	0.060	0.852	0.740	0.124
8	O	Healthy	0.278	1.485	0.673	2.164	0.432	1.021	0.973	0.113
9	O	Healthy	0.498	1.168	1.084	1.358	0.335	0.664	0.678	0.022
10	O	Healthy	0.747	1.394	0.991	0.780	0.047	0.579	0.725	0.023
11	O	Healthy	2.297	1.794	1.422	0.155	0.000	2.110	2.210	0.037
12	O	Healthy	1.793	1.008	1.374	0.147	1.196	1.094	0.948	0.056
13	O	Healthy	0.660	1.087	0.738	0.451	0.892	0.604	0.644	0.050
14	O	Healthy	0.866	0.694	1.183	0.412	0.040	0.724	0.664	0.027
15	O	Healthy	1.076	1.186	0.897	0.109	0.000	0.975	0.917	0.027
16	O	Healthy	0.378	0.166	0.875	1.350	0.034	0.723	0.679	0.015
17	O	Healthy	1.740	1.561	1.369	1.082	0.271	0.690	0.686	0.057
18	O	Healthy	1.455	0.658	1.672	0.151	0.000	1.133	1.052	0.057
19	O	Healthy	0.804	0.990	0.698	0.979	1.113	0.768	0.672	0.037
20	O	Healthy	0.702	1.208	0.394	2.511	0.849	0.762	0.791	0.105
21	O	Healthy	1.072	0.629	1.524	0.661	0.037	0.492	0.477	0.024
22	O	Healthy	3.217	1.358	1.977	0.394	2.000	0.936	1.187	0.064
23	O	Healthy	0.760	1.173	1.188	0.311	0.000	1.424	1.109	0.034
24	O	Healthy	0.682	0.804	0.947	1.736	0.001	1.256	0.837	0.137
25	nonO	Healthy	1.531	1.393	1.623	0.980	0.327	1.169	1.417	0.178
26	nonO	Healthy	2.732	1.098	0.878	0.852	0.696	1.986	1.928	0.301
27	nonO	Healthy	2.295	1.115	2.182	0.308	0.003	1.172	1.357	0.431
28	nonO	Healthy	3.851	0.649	2.711	0.279	2.909	0.813	1.015	0.355
29	nonO	Healthy	1.668	1.058	1.920	1.787	5.000	1.149	1.181	0.217
30	nonO	Healthy	1.819	1.445	3.006	0.350	0.566	1.306	1.336	0.404
31	nonO	Healthy	1.875	0.901	1.861	0.853	5.000	1.536	1.374	0.647
32	nonO	Healthy	2.838	1.117	2.165	2.724	2.775	2.002	1.980	0.364
33	nonO	Healthy	2.163	0.981	1.554	0.112	0.429	1.027	1.042	0.858
34	nonO	Healthy	3.701	1.267	1.748	1.133	1.049	1.812	1.798	0.329
35	nonO	Healthy	3.568	1.279	2.422	0.000	5.000	1.100	0.990	0.519
36	nonO	Healthy	1.218	0.830	6.790	4.610	0.026	1.006	1.050	0.109
37	nonO	Healthy	1.465	1.046	1.973	3.756	4.998	1.216	1.314	0.199
38	nonO	Healthy	3.606	1.284	2.838	0.480	0.559	1.365	1.122	0.350
39	nonO	Healthy	56.71	0.699	22.76	0.857	0.031	3.204	2.817	0.365
40	nonO	Healthy	1.211	1.315	2.465	1.065	0.000	0.986	0.993	0.301
41	nonO	Healthy	4.513	0.886	4.277	0.361	0.489	1.808	1.869	0.924
42	nonO	Healthy	4.084	0.362	2.829	1.217	0.003	2.637	2.210	0.343
43	nonO	Healthy	3.085	0.463	2.034	2.780	4.994	1.411	1.600	0.530

Table B.4. *List of parameters estimated for patients carring vWD type 1*

Patient	Blood	Mutation	K_u	τ_u	τ_{ULH}	K_L	τ_L	Y_b^{CB}	α	Objec
44	O	no mutation	0.526	0.475	0.852	0.615	0.080	0.303	0.292	0.232
45	O	p.P2063S	0.801	0.573	1.323	0.402	0.917	0.432	0.584	0.606
46	A	p.R1145C	0.596	0.588	1.084	2.067	0.917	0.409	0.523	0.326
47	O	p.P2063S	0.563	5.153	0.428	2.846	0.965	0.547	0.693	0.397
48	A	c.7056C>T	0.073	2.387	0.199	20.67	0.168	0.821	0.488	0.656
49	A	p.R1819	0.215	0.757	0.694	0.823	0.191	0.169	0.242	0.173
50	A	no mutation	0.717	0.594	0.936	0.000	4.994	0.525	0.559	0.385
51	B	p.C2362F	2.346	0.567	359.7	0.000	4.668	0.046	0.082	1.036
52	A	p.R854Q	0.654	2.962	0.365	0.865	1.141	0.388	0.347	0.522
53	O	no mutation	0.264	0.809	0.502	3.383	4.997	0.466	0.400	0.157
54	O	no mutation	1.933	1.046	1.213	0.554	0.046	0.579	0.595	0.628
55		RNE	0.674	1.100	1.285	1.075	0.196	0.299	0.289	0.234
56		RNE	0.290	1.951	0.159	1.162	0.578	0.172	0.193	0.606
57	A	no mutation	0.601	1.449	0.664	0.186	0.613	0.417	0.459	0.278
58	O	no mutation	1.017	1.660	0.740	0.607	0.010	0.554	0.486	0.287
59	O	c.1534-3C>A	1.132	1.998	1.554	0.633	0.031	0.786	0.758	0.303
60	O	c.1534-3C>A	0.102	1.246	1.455	0.111	0.000	0.080	0.122	0.373
61	O	c.1534-3C>A	2.007	0.827	1.996	2.954	5.000	0.541	0.839	0.469
62	O	c.1534-3C>A	0.135	1.890	1.522	2.878	5.000	0.037	0.083	0.239
63	O	p.C2362F	0.421	0.874	1.064	0.946	0.126	0.444	0.427	0.317
64	O	p.C2362F	0.552	1.700	0.514	1.399	0.562	0.511	0.691	0.502
65	O	p.C2362F	0.535	1.375	1.215	1.315	0.156	0.378	0.331	0.144
66	O	p.C2362F	0.426	1.439	0.993	0.470	0.214	0.413	0.409	0.255
67	O	RNE	0.795	0.936	0.717	0.313	0.017	0.748	0.695	0.397
68	O	no mutation	0.803	8.863	0.104	0.749	0.023	0.356	0.408	0.198
69	O	p.P2063S	0.284	0.874	0.583	0.663	0.000	0.560	0.502	0.211
70	O	no mutation	0.672	0.330	1.103	1.455	0.537	0.403	0.567	0.521
71	O	no mutation	0.526	0.627	0.657	0.000	5.000	0.534	0.568	0.633
72	B	p.C1130F	0.374	0.839	0.391	3.116	0.772	0.184	0.364	0.149
73	O	p.G2705R	0.391	1.395	0.751	0.920	0.012	0.663	0.499	0.462
74	O	p.G2705R	0.464	0.538	1.651	2.748	0.027	0.571	0.354	0.573
75	O	no mutation	1.774	0.964	0.998	3.792	5.000	0.434	0.564	0.131
76	O	p.C524Y	0.077	0.846	0.570	0.358	0.000	0.222	0.210	0.315
77	A	RNE	3.644	1.101	1.921	0.993	5.000	0.843	0.917	0.622
78	O	p.P2063S omo	0.795	1.026	1.209	0.652	0.004	0.586	0.597	0.284
79	O	p.P2063S	1.821	0.579	1.765	0.497	0.729	0.498	0.590	0.301
80		p.R1564W	0.619	1.687	0.801	0.890	0.006	0.687	0.670	0.322
81	B	p.C2362F	0.486	1.609	0.907	2.169	0.697	0.343	0.381	0.196
82	O	RNE	0.859	1.070	1.133	0.608	0.000	0.639	0.616	0.342
83		RNE	0.364	2.105	0.297	6.779	2.756	0.263	0.209	0.220
84		RNE	0.178	0.929	0.400	3.830	0.208	0.244	0.184	0.131
85	O	p.R670C	0.843	0.548	1.194	1.287	0.055	0.427	0.485	0.111
86	O	no mutation	1.780	0.226	2.019	1.447	0.008	0.912	0.846	0.263
87	O	no mutation	0.720	1.509	0.644	1.550	0.355	0.265	0.369	0.202
88	A	p.C2362F	0.598	2.438	0.654	4.613	2.043	0.469	0.533	0.158
89	O	c.1534-3C>A	1.016	0.792	1.323	0.000	5.000	0.300	0.419	0.664
90	O	p.C1130F	0.194	0.837	0.315	1.951	0.701	0.188	0.386	0.065
91	O	c.1534-3C>A omo	0.211	1.198	1.107	1.142	1.586	0.174	0.200	0.297
92	O	c.1534-3C>A	1.400	1.165	1.132	3.005	0.998	0.933	1.163	0.329
93	A	p.E2539X	1.119	1.037	1.884	0.918	0.011	0.587	0.557	0.320
94	A	p.E2539X	1.398	1.286	2.268	0.624	0.504	0.547	0.695	0.066

References

Castaldello, C. (2016), Progettazione di test clinic per la diagnosi della malattia di von willebrand, Master's thesis, Universtità degli studi di Padova.

Castaldello, C., Galvanin, F., Casonato, A., Padrini, R., Barolo, M. and Bezzo, F. (2018), A model-based protocol for the diagnosis of von willebrand disease, *Can. J. Chem. Eng.* **96**, 628–638.

Dennis, J. (2013), *"Nonlinear Least-Squares." State of the Art in Numerical Analysis*, d. jacobs, academic press edn.

Echahdi, H., Hasbaoui, B. E., Khorassani, M. E., Agadr, A. and Khattab, M. (2017), Von willebrand's disease: case report and review of literature, *PanAfrican Medical Journal* **27**(147).

Favaloro, E. J., Thom, J., Patterson, D., Just, S., Dixon, T., Koutts, J., Baccala, M., Rowell, J. and Baker, R. (2009), Desmopressin therapy to assist the functional identification and characterisation of von willebrand disease: Differential utility from combining two (vwfcb and vwfrco) vonwillebrand factor activity assays?, *Thrombosis Research* **123**, 862 – 868.

Ferrari, M., Galvani, F., Barolo, M., Daidone, V., Padrini, R., Bezzo, F. and Casonato, A. (2018), A mechanistic model to quantify vonwillebrand factor release, survival and proteolysis in patients with von willebrand disease, *Thromb Haemost* **118**(2), 309–319.

Gallinaro, L., Cattini, M. G., Sztukowska, M., Padrini, R., Sartorello, F., Pontara, E., Bertomoro, A., Daidone, V., Pagnan, A. and Casonato, A. (2008), A shorter vonwillebrand factor survival in o blood group subjects explains how abo determinants influence plasma vonwillebrand factor, *Blood* **111**(7), 3540 – 3545.

Galvanin, F., Barolo, M., Padrini, R., Casonato, A. and Bezzo, F. (2014), A model-based approach to the automatic diagnosis of vonwillebrand disease, *AIChE Journal* **60**(5), 1718 – 1727.

Hall, J. E. and Guyton, A. C. (2015), *Guyton and Hall Textbook of Medical Physiology E-Book*, Guyton Physiology, Elsevier Health Sciences.

Lenting, P. J., Westein, E., Terraube, V., Ribba, A.-S., Huizinga, E. G., Meyer, D., de Groot, P. G. and Denis, C. V. (2004), An experimental model to study the in vivo survival of von willebrand factor, *Journal of biological chemistry* **279**(13), 12102–12109.

Marder, V. J., Aird, W. C., Bennett, J. S., Schulman, S. and White, G. C. (2013), *Hemostasis and Thrombosis*, sixth edition edn.

McLean, K. A. P. and McAuley, K. B. (2012), Mathematical modelling of chemical processes — obtaining the best model predictions and parameter estimates using identifiability and estimability procedures, *Can. J. Chem. Eng.* **90**, 351 – 366.

Menache, D., Aronson, D., Darr, F., Montgomery, R., Gill, J., Kessler, C., Lusher, J., Phatak, P., Shapiro, A., Thompson, A. and White, G. (1996), Pharmacokinetics of von willebrand factor and factor viiic in patients with severe von willebrand disease (type 3 vwd): estimation of the rate of factor viiic synthesis. cooperative study groups, *Br J Haematol* **94(4)**, 740–745.

Monte, A. (2013), Un approcio ingegneristico all'identificazione di modelli fisiologici per la descrizione della malattia di von willebrand, Master's thesis, Universtità degli studi di Padova.

Myung, J. (2003), Tutorial on maximum likelihood estimation, *Journal of Mathematical Psychology* **47**, 90–100.

NHLBI (2007), *The diagnosis, evaluation, and management of von Willebrand disease*, U.S. Dept. of Health and Human Services, National Institutes of Health, National Heart, Lung, and Blood Institute, USA.

Palta, S., Saroa, R. and Palta, A. (2014), Overview of the coagulation system, *Indian Journal of Anaesthesia* **58**(5), :515–23.

Rodeghiero, F., Castaman, G. and Dini, E. (1987), Epidemiological investigation of the prevalence of von willebrand's disease., *Blood* **69**(2), 454–459.

Sadler, J. E. (2008), von willebrand factor: two sides of a coin, *Journal of Thrombosis and Haemostasis* **3**, 1702 – 1709.

Schneppenheim, R. and Budde, U. (2008), *Von Willebrand Disease and von Willebrand Factor. Current Aspects of Diagnosis and Treatment*, first ed. edn, Uni-Med Science.

Seborg, D. E., Edgar, T. F., Mellichamp, D. A. and Doyle, F. J. (2017), *Process dynamics and control*, John Wiley Sons, Inc.

Silverthorn, D. U., Johnson, B. R., Ober, W. C., Silverthorn, A. C. and Ober, C. E. (2016), *Human physiology : an integrated approach*, seventh edition. edn.

Wise, B. M., Gallagher, N. B., Bro, R., Shaver, J. M., Windig, W. and Koch, R. S. (2006), *PLS_Toolbox Version 4.0 for use with MATLAB*, Eigenvector Research, Inc.

Yeh, W. W. (1986), Review of parameter identification procedures in groundwater hydrology' the inverse problem, *Water Resources Research* **22**(2), 95–108.

Website:

Matlab Tutorial (2019).
`https://it.mathworks.com/help/optim/ug/least-squares-model-fitting-algorithms. html#f204` (last access: 11/10/2019)

Greetings

First, I would like to thank Prof. Massimiliano Barolo and Prof. Fabrizio Bezzo for giving me this incredible thesis opportunity, making me grow very much by their ever-present advice. And to all CAPE-Lab team for the patient and support.

My heartfelt thanks to my parents Mara Lucia and Giovani, and especially my sister Luiza, for having supported me without reserve in this path. An immense thank you goes finally to my friend Ana which has always been a rock to all my emotional crisis and a true role model inspiration. And finally to my friends: Silvia, Veronica, Paulo, Andre, Marco, Federica, Ilaria and Kattyla.

”Education does not transform the World.
Education changes People.
People change the World.”
Paulo Freire