UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM MESTRADO EM ENGENHARIA ELÉTRICA

LARISSA APARECIDA BARATIERI

ASSESSMENT OF THE USE OF THE KALMAN FILTER AS SUPPORT FOR AN ELECTROCARDIOGRAM CLASSIFIER BASED ON SYNTHETIC SIGNALS

DISSERTAÇÃO

CORNÉLIO PROCÓPIO
2021

LARISSA APARECIDA BARATIERI

ASSESSMENT OF THE USE OF THE KALMAN FILTER AS SUPPORT FOR AN ELECTROCARDIOGRAM CLASSIFIER BASED ON SYNTHETIC SIGNALS

Avaliação Do Uso Do Filtro De Kalman Como Suporte À Classificação De Eletrocardiograma Com Base Em Sinais Sintéticos

Dissertação apresentado(a) como requisito para obtenção do título(grau) de Mestre em Mestrado em Engenharia Elétrica, do Programa de Pós-Graduação em Mestrado em Engenharia Elétrica, da Universidade Tecnológica Federal do Paraná (UTFPR).

Orientador: Doutor Paulo Rogério Scalassara
Coorientador: Doutor Cristiano Marcos Agulhari

CORNÉLIO PROCÓPIO

2021

Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Cornélio Procópio

LARISSA APARECIDA BARATIERI

AVALIAÇÃO DO USO DO FILTRO DE KALMAN COMO SUPORTE À CLASSIFICAÇÃO DE ELETROCARDIOGRAMA COM BASE EM SINAIS SINTÉTICOS

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestra Em Engenharia Elétrica da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Sistemas Eletrônicos Industriais.

Data de aprovação: 22 de Junho de 2021

Prof Paulo Rogerio Scalassara, Doutorado - Universidade Tecnológica Federal do Paraná

Prof Bruno Augusto Angelico, Doutorado - Usp-Universidade de São Paulo

Prof Cristiano Marcos Agulhari, Doutorado - Universidade Tecnológica Federal do Paraná

Prof Victor Baptista Frencl, - Universidade Tecnológica Federal do Paraná

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 22/06/2021.

ACKNOWLEDGEMENTS

First of all, I would like to thank God, because without Him nothing would be possible.

I would like to thank my supervising teachers, Dr. Paulo Rogério Scalassara and Dr. Cristiano Marcos Agulhari for having given me a lot of support during the research, they were much more than just supervising teachers.

I would like to thank my husband, Vinicius Baratieri Suterio for following this journey with me.

Finally, I would like to thank CAPES for funding the research. And to thank all those who helped me in an indirect and direct way to develop this work.

“All that we have to decide is what to do with
the time that is given us”.
(Gandalf, The Grey)

RESUMO

BARATIERI, Larissa. **Avaliação Do Uso Do Filtro De Kalman Como Suporte À Classificação De Eletrocardiograma Com Base Em Sinais Sintéticos**. 2021. 61 f. Dissertação (Mestrado em Mestrado em Engenharia Elétrica) – Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2021.

As doenças cardiovasculares estão entre as doenças que mais matam no mundo. Para evitar esse cenário, é necessário a realização de exames cardíacos periodicamente com a finalidade de realizar de tratamento preventivo caso haja alguma doença. Seguindo essa motivação, o objetivo deste trabalho é avaliar se o KF pode ser usado como um classificador de cardiopatias, ou seja, estudar uma nova possível função para esse filtro, uma vez que ele é um estimador de estados de equações diferenciais lineares. Então, a metodologia aplicada nesse trabalho consiste em: primeiro, desenvolver o filtro capaz de estimar ECGs saudáveis, entretanto um sinal de ECG possui um comportamento não-linear, sendo assim, será utilizado uma atualização do KF, o EKF. Para o desenvolvimento desse filtro é necessário que o ECG seja descrito por equações diferenciais, com isso o algoritmo do EKF será baseado em equações diferenciais ordinárias que formulam um sinal de ECG sintético de ritmo sinusal normal. A ideia de construir o filtro baseado apenas em um sinal sintético de ECG saudável é de que o filtro "aprenda" somente a estimar um sinal saudável de ECG, assim quando o ECG mudasse para uma cardiopatia o EFK detectaria algum erro ou anomalia, e isso indicaria uma doença. Para a validação do objetivo, o EKF deverá estimar um sinal construído com amostras saudáveis e amostras doentes de um ECG real. Dentre as doenças existentes foram escolhidas o fluter ventricular e a taquicardia supraventricular, por causa dos seus comportamentos. E por fim, serão utilizados dois critérios de avaliação para que o objetivo se conclua: o primeiro critério é uma análise qualitativa do sinal do erro do estado a ser estimado (sinal criado a partir das amostras de ECG reais) com o sinal estimado (saída do EKF); e o segundo critério é baseado numa análise quantitativa, em que as amostras desse sinal serão divididas em janelas (doentes e saudáveis) e será calculado a energia do erro de cada janela. Uma vez definido se o KF pode atuar como um classificador de cardiopatias, serão sugeridas melhorias para essa nova funcionalidade da ferramenta.

Palavras-chave: filtro de Kalman. eletrocardiograma sintéticos. classificador. cardiopaticas.

ABSTRACT

BARATIERI, Larissa. **Assessment of the use of the Kalman Filter as Support for an Electrocardiogram Classifier Based on Synthetic Signals**. 2021. 61 p. Dissertation (Master's Degree in Master of Electrical Engeneering) – Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2021.

Cardiovascular diseases are among the diseases that kill the most in the world. To avoid this scenario, it is necessary to carry out periodic cardiac exams in order to carry out preventive treatment if there is any disease. Following this motivation, the objective of this work is to evaluate whether KF can be used as a classifier of heart diseases, that is, to study a new possible function for this filter, since it is an estimator of states of linear differential equations. So, the methodology applied in this work consists of: first, developing a filter capable of estimating healthy ECGs, however an ECG signal has a non-linear behavior, thus, an update of the KF, the EKF, will be used. For the development of this filter it is necessary that the ECG is described by differential equations, thus the EKF algorithm will be based on ordinary differential equations that formulate a synthetic ECG signal of normal sinus rhythm. The idea of building the filter based only on a synthetic healthy ECG signal is that the filter "learn" only to estimate a healthy ECG signal, so when the ECG changed to a heart disease the EFK would detect some error or anomaly, and that would indicate an illness. For goal validation, the EKF should estimate a signal constructed from healthy and diseased samples from a real ECG. Among the existing diseases, ventricular flutter and supraventricular tachycardia were chosen because of their behavior. And finally, two evaluation criteria will be used to complete the objective: the first criterion is a qualitative analysis of the error signal of the state to be estimated (signal created from real ECG samples) with the estimated signal (output of the EKF); and the second criterion is based on a quantitative analysis, in which the samples of that signal will be divided into windows (sick and healthy) and the error energy of each window will be calculated. Once defined whether KF can act as a heart disease classifier, improvements will be suggested for this new functionality of the tool.

Keywords: Kalman filter. synthetic electrocardiogram. classifier. cardiopathies.

LIST OF ACRONYMS

ECG	electrocardiogram
CVDs	cardiovascular diseases
FIR	Finite Impulse Response
KF	Kalman Filter
EKF	Extended Kalman Filter
MIT	Massachusetts Institute of Technology
MSE	mean square error
MIMO	multiple input multiple output

LIST OF FIGURES

LIST OF TABLES

SUMMARY

1 INTRODUCTION

The electrocardiogram (ECG) began as an object of study in the mid-'60s, mainly in the sense of detecting and preventing heart diseases (Blackburn *et al.*, 1960). Along with, in 2016 a survey by the World Health Organization was developed indicating that heart diseases are the number one cause of death globally: more people die annually from cardiovascular diseases (CVDs) than from any other causes, an estimate of 31% of the world's death (WHO, 2017). Since that, the researches in this field have only increased, not only in filtering noise for better comprehension and manipulation, but mainly in the use of artificial intelligence for the classification of cardiopathies in order to assist the health professional to have a correct diagnosis.

As previously stated, the CVDs are among the diseases that kill the most in the world, and to prevent this it is necessary to perform cardiac exams periodically to find out if there is any disease and the need for treatment for preventive purposes. Following this idea, there are works that aim to use computer tools to classify heart diseases in order to help the health professional not to make mistakes. For example, the work done by Cardenas *et al.* (2019), wherein was proposed a method for classifying cardiac arrythimias by using a Fuzzy Cognitive Maps. Another work that can be mentioned, by Gutierrez-Gnecchi *et al.* (2017), they developed an algorithm based on quadratic wavelets to identify individual ECG waves and obtain a fiducial marker array in order to classify some arrythmias using a Probabilistic Neural Network.

In the case of filtering ECG noise, there are works that explore some tools in order to remove noises from the ECG signal, when it collected, to do a better comprehension from the signal. It is possible to cite some recent works like Almalchy *et al.* (2019), wherein the authors tested several models of Finite Impulse Response (FIR) filters of low-pass and high-pass to determine the best band-pass filtering model to reproduce a ECG signal that closely resembles the actual heart function of a patient. In Tayel *et al.* (2018), a new multi-stage combined adaptive filtering design based on Kernel Recursive Least Squares Tracker and Kernel Recursive Least Squares with Approximate Linear Dependency algorithms was proposed for removing artifacts and noise sources, while preserving the low frequency components and the tiny features of the ECG signal.

So, motivated by the researches involving ECGs, databases were created to provide data for study. A well known database is the PhysionBank Massachusetts Institute of Technology

(MIT) database (Moody; Mark, 2001), composed of ECG signals collected in several situations and grouped into categories. However, some obstacles arise when considering data acquired from patients, such as the difficulty to infer how the performance may vary in different clinical contexts with a range of noise levels and sampling frequencies. Also, if the research requires different classes of ECG signals not available in the database, it would be necessary to collect the ECG from human patients, which can pose several problems. So, in this context, there are works to create synthetic ECGs.

The first work that started with this possibility was McSharry *et al.* (2003). This work presents a dynamical model based on three coupled ordinary differential equations which are capable of generating realistic synthetic ECG signals with realistic morphology and prescribed heart rate dynamics. This model aims to provide a standard realistic ECG signal with known characteristics, which can be generated with a specific desired statistic such as the mean and standard deviation of the heart rate and frequency-domain characteristics of the heart rate variability. This synthetic ECG can be generated with different sampling frequencies and different noise levels to analyze the performance of a given technique. Besides, this work received some updates such as Clifford e McSharry (2004), which demonstrates how the same model can be used to generate blood pressure and respiration signals with realistic inter-signal coupling between the respiration, blood pressure and ECG time series. In Clifford *et al.* (2005), a technique is described to simultaneously filter, compact and classify the real ECG, based on the synthetic ECG and quadratic minimum adjustments. And finally, in Sameni *et al.* (2007), the model is adapted to maternal and fetal ECGs.

After that, in the last decade, other studies that explore different methods for generating synthetic ECGs have emerged. The method developed by Kovács (2012) uses different kinds of base functions such as polynomials or rational functions to model ECG curves. More precisely, the author uses piecewise polynomial approximation based on 15 control points. The piecewise approximation approach has the following benefits: the geometrical parameters of the curve can be piecewise adjusted; the ability to set diagnostical parameters of the ECG irrespectively of its geometrical parameters; and the ability to use error measures based on different ECG features like diagnostical intervals and amplitudes. Even if it proves to be an efficient method, this work is restricted to studying only the main points of a normal ECG, these being the P, Q, R, S, and T segments (the heart beat morphology and its segmentes will be better explained in Chapter 2). Furthermore, the author concludes that perhaps this model does not serve to generate pathologies

and is not able to assist in the filtering of ECG noise.

In Kubicek *et al.* (2014), the authors propose the creation of a synthetic ECG by dividing the signal into various parts, which are substituted by elementary functions. The individual ECG segments are approximated by continuous curves, and the subsequent development of these periodic functions was solved by Fourier series.

One of the most recent article of generation of synthetic ECGs is Dolinský *et al.* (2018), where a mathematical model is proposed to generate an artificial synthetic ECG signal based on the geometric characteristics of a real ECG signal. Each specific wave, corresponding to P, Q, R, S, and, T, is modeled using an elementary trigonometric function or a Gaussian monopulse, this method allows the generation of some cardiopathies. Even showing to be an efficient algorithm, the method studied presents small flaws in the generation of Q and S waves that do not present the desired performance.

Among the models aforementioned about synthetic ECG generation, the model developed by McSharry *et al.* (2003) proves to be more advantageous in terms of ease of replication, the quantity and duration of the signal, and the possible application on other projects, mainly in the study of filtering noise from ECG signals.

Joining the concept of synthetic ECGs and filtering, one method that has stood out in recent years is the use of the Kalman Filter (KF) to eliminate noise from ECG signals (Mneimneh *et al.*, 2006; Ting; Salleh, 2010; Vullings *et al.*, 2011). Actually, the KF is an estimator whose purpose is to use measurements obtained over time (contaminated with noise and other uncertainties) and generate results that tend to approximate the real values of the measured quantities and associated values (Kay, 1993). Other works involving both Kalman filter and synthetic ECG are Sameni *et al.* (2005) and Sameni *et al.* (2007), which use the theoretical basis of the generation of synthetic ECG from McSharry *et al.* (2003) to create a modified KF suitable for the estimation of noisy ECG signals.

So, based on some of the references presented, the objective of this work is to evaluate the possibility of using KF as a classifier of heart disease, thus, this research is a study of a possible new functionality of KF, since it is used as estimator of optimal states.

To achieve this goal, a KF algorithm capable of estimating real ECG signals will be developed. As will be seen throughout this dissertation, the KF algorithm is mainly based on linear differential functions, however an ECG signal does not have linear characteristics, it is necessary to use an update of the FK, the Extended Kalman Filter (EKF) which has the same

characteristics as the KF, however is specific to nonlinear equations.

However, to develop the EKF algorithm, it is necessary to transform the human ECG into differential equations, so in this case the work developed by McSharry *et al.* (2003), in which the authors managed to create a mathematical model based on three differential equations capable of replicating an ECG, will be used as a basis. of normal sinus rhythm, that is, a healthy ECG, in a very realistic way. From the differential equations that create a synthetic ECG and from the EKF theory, it is possible to develop an algorithm that estimates a real ECG.

For the validation of the EKF as a possible classifier of heart diseases, the following technique will be applied: the EKF must estimate a signal built with healthy samples and sick samples from a real ECG.

As there are several heart diseases, the criteria created to choose the diseases was as follows: when observing heart diseases, it is clear that they can be divided into two large classes, the first class being composed of those heart diseases that modify the structure pattern of points P, Q, R, S, and, T that a healthy ECG has (which will be further explained in Chapter 2), with ventricular flutter being chosen to represent this class; and the second class of heart diseases is composed of those diseases that have P, Q, R, S, and T points well defined as well as a healthy ECG, but the heart rate is changed, speeding up or slowing down the heartbeat, being chosen supraventricular tachycardia to represent this class.

In order for the EKF to estimate only healthy samples, its algorithm must be based entirely on a healthy synthetic ECG, as the expectation is that the filter will only "learn" how to estimate a healthy ECG signal, just as when the ECG changes to heart disease the EFK would detect some error or anomaly, and that would indicate an illness.

With the form of filter validation defined, two evaluation criteria will be used if the objective can be concluded. The first criterion is a qualitative analysis of the error signal from the state to be estimated (signal created from the actual ECG samples) with the estimated signal (EKF output). And the second criterion is based on a quantitative analysis, in which the samples of this signal will be divided into windows (sick and healthy) and the error energy of each window will be calculated.

It should be emphasized that the objective of this work is an evaluation of a new function for the KF, that is, as a cardiopathy classifier. Once it has been proven whether or not it can be used like this, improvements to this technique will be suggested in order to fix some flaws and complications that appeared during the research.

This dissertation is due as follows: Chapter 2 explains the basics of how a heartbeat works and how the differential equations developed by McSharry *et al.* (2003) generate a synthetic electrocardiogram signal; Chapter 3 explains the KF and the EKF methodology and algorithm. Chapter 4 explains the methodology applied in the work, mainly how the EKF was developed from ordinary synthetic ECG equations, this section is based on the works developed by Sameni *et al.* (2005) and Sameni *et al.* (2007). Chapter 5 presents the results found. And finally, Chapter 6 presents the conclusion and proposals for future work.

2 SYNTHETIC ELECTROCARDIOGRAM SIGNALS

This chapter details the results from McSharry *et al.* (2003), in which the authors develop a dynamical model with three differential equations that generate synthetic electrocardiograms signals that are similar to a realistic ECG signal.

Before explaining how these equations were developed, it is important to understand some features from the realistic ECG signal. Section 2.1 explains the ECG morphology. Section 2.2 then explains, according to McSharry *et al.* (2003), the dynamical model developed.

2.1 THE ELECTROCARDIOGRAM MORPHOLOGY

The electrocardiogram is a special graph that represents the electrical activity of the heart. Thus, the ECG provides a time-voltage chart of the heartbeat. For many patients, this test is a key component of clinical diagnosis and management in both inpatient and outpatient settings. To record the cardiac electrical currents, conductive electrodes are selectively positioned on the surface of the body. For the standard ECG recording, electrodes are placed on the arms, legs, and chest wall (precordium) (Goldberger *et al.*, 2013; Dubin, 2000).

A single sinus (normal) cycle of the ECG, corresponding to one heartbeat, is traditionally labeled with the letters P, Q, R, S and T on each of its turning points, see Figure 1.

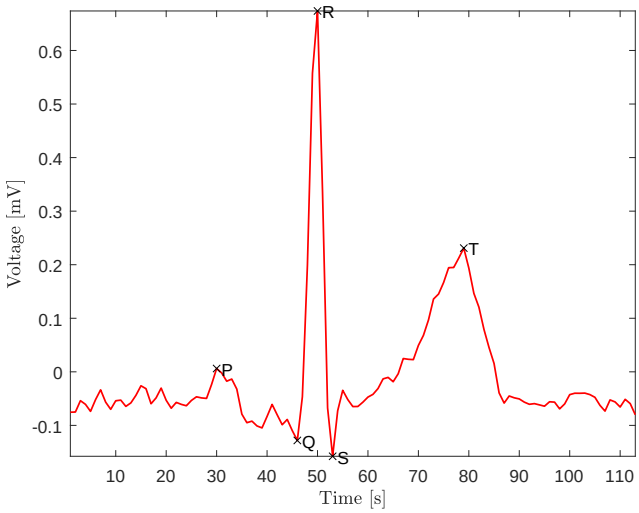
The ECG may be divided into the following sections (Goldberger *et al.*, 2013):

- P-wave: represents atrial depolarization, it is a small positive (or negative) deflection before the QRS complex.
- PQ-interval: the time between the beginning of atrial depolarization and the beginning of ventricular depolarization.
- QRS-complex: the largest-amplitude portion of the ECG, caused by currents generated when the ventricles depolarize prior to their contraction. Although atrial repolarization occurs before ventricular depolarization, the latter waveform, i.e. the QRS-complex, is of much greater amplitude and atrial repolarization is therefore not seen on the ECG.
- QT-interval: the time between the onset of ventricular depolarization and the end of the ventricular repolarization.

- ST-segment: the time between the end of S-wave and the beginning of T-wave. Significantly elevated or depressed amplitudes away from the baseline are often associated with cardiac illness.

- T-wave: ventricular repolarization, whereby the cardiac muscle is prepared for the next cycle of the ECG.

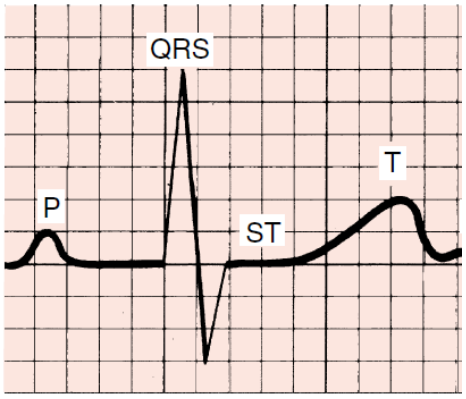
Figure 1 – Morphology of one PQRST-complex of the ECG.



Source: Made by the author.

Figure 2 is an example taken from Goldberger *et al.* (2013), which shows the heart beat morpghology mentioned before. While Figure 1 represents a real eletrocardiogram.

Figure 2 – Morphology of one PQRST-complex of the ECG.

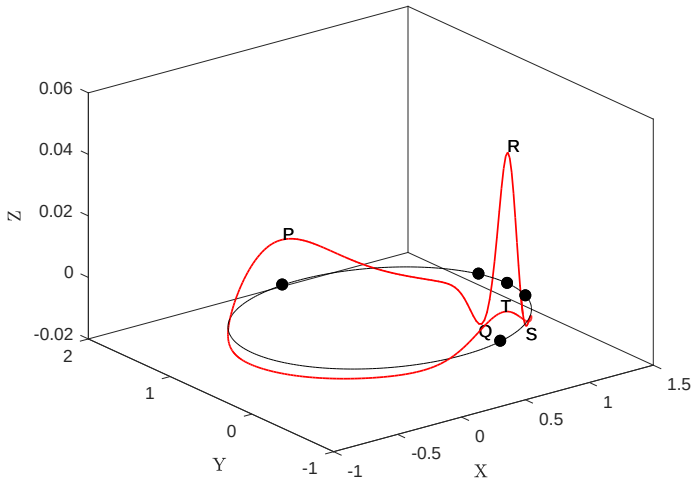


Source: (Goldberger *et al.*, 2013).

2.2 A DYNAMICAL MODEL FOR GENERATING SYNTHETIC ELECTROCARDIOGRAM SIGNALS

The model generates a trajectory in a three-dimensional (3-D) state-space with coordinates (x,y,z), as illustrated in Figure 3. The quasi-periodicity of the ECG is reflected by the movement of the trajectory around an attracting limit cycle of unit radius in the (x,y) plane. Each revolution on this circle corresponds to one RR-interval or heart-beat. Interbeat variation in the ECG is reproduced using the motion of the trajectory in the z-direction. Distinctive points on the ECG, such as P, Q, R, S and T are described by events corresponding to negative and positive attractors/repellor in the z-direction. These events are placed at fixed angles along the unit circle given by θ_P, θ_Q, θ_R, θ_S and θ_T, represented by the dots as in Figure 3. When the trajectory approaches one of these events, it is pushed upwards or downwards away from the limit cycle, and then as is moves away it is pulled back toward the limit cycle.

Figure 3 – Typical trajectory by the dynamical model in the 3-D space given by (x,y,z).



Source: Based on (McSharry *et al.*, 2003)

The dynamical equations of motion are given by a set of three ordinary differential equations

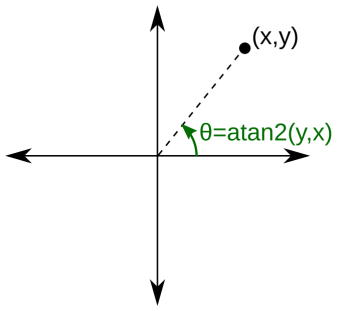
$$\begin{aligned} \dot{x} &= \rho x - \omega y, \\ \dot{y} &= \rho y + \omega x, \\ \dot{z} &= -\sum_{i\in\{P,Q,R,S,T\}} a_i \Delta\theta_i \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right) - (z - z_0). \end{aligned} \tag{1}$$

The $a_i \in \mathbb{R}$, $b_i \in \mathbb{R}$ and $\theta_i \in \mathbb{R}$ terms correspond to the amplitude, width and center parameters of the Gaussian terms of this equation, respectively. And $\rho = 1 - \sqrt{x^2 + y^2}$, $\Delta\theta_i = (\theta - \theta_i) \mod 2\pi$, $\theta = \text{atan2}(y,x)$, the four quadrant arctangent of the real parts of the elements of x and y, with $-\pi \leq \text{atan2}(y,x) \leq \pi$ (see Figure 4), and ω is the angular velocity of the trajectory as it moves around the limit cycle. As seen in Equation (1), each of P, Q, R, S and T-waves of the ECG waveform are modeled with a Gaussian function located at specific angular positions θ_i. Baseline wander was introduced by coupling the baseline value z_0 in (1) to the respiratory frequency f_2 using the following function

$$z_0(t) = A \sin(2\pi f_2 t) \tag{2}$$

in which A = 0.15 mV.

Figure 4 – $\text{atan2}(y, x)$ **returns the angle** θ **between length** r **and the positive x axis, confined to** $[-\pi, \pi]$**.**



Source: Made by the author.

The formula that gives rise to a Gaussian distribution (also known as a normal distribution) is given by

$$f(n) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2}\left(\frac{n-\mu}{\sigma^2}\right), \tag{3}$$

where n is randon variable, μ is the expected value and σ^2 is the variance. Comparing Equation (1) with the third equation of system represented in Equation (1), it is possible to verify that this equation is a sum of Gaussian.

These equations of motion given by system in Equation (1) were integrated numerically using a fourth-order Runge-Kutta method (Press *et al.*, 1992) with a fixed time step $\Delta t = \frac{1}{f_s}$ where f_s is the sampling frequency.

Table 1 specifies the appropriate values for the parameters a_i, b_i e θ_i for the P, Q, R, S, and, T points (McSharry *et al.*, 2003).

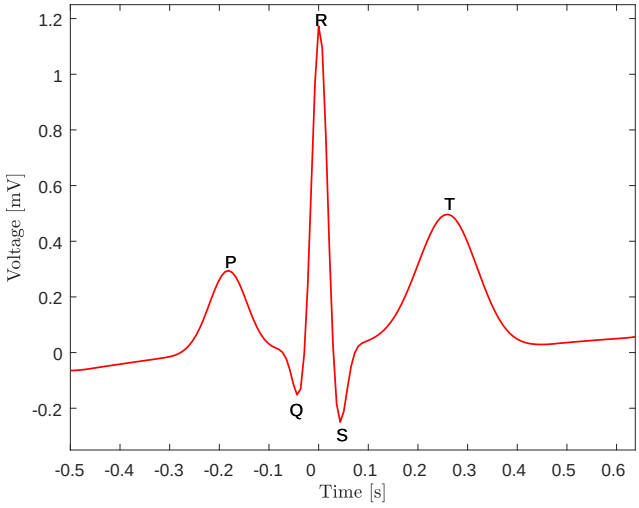
Table 1 – Parameters of the ECG model given by (1).

Index (i)	P	Q	R	S	T
Time (sec)	-0.2	-0.05	0	0.05	0.3
θ_i (radians)	$-\frac{1}{3}\pi$	$-\frac{1}{12}\pi$	0	$\frac{1}{12}\pi$	$\frac{1}{3}\pi$
a_i	1.2	-5.0	30.0	-7.5	0.75
b_i	0.25	0.1	0.1	0.1	0.4

Source: McSharry *et al.* (2003)

A trajectory generated by Equation (1) in three dimensions corresponding to (x,y,z) is illustrated in Figure 3. This illustrates how the positions of the events P, Q, R, S, T act on the trajectory in the z-direction as it processes around the unit circle in the (x,y) plane. The z variable from the 3-D system of equation 1 yields a synthetic ECG with realistic PQRST morphology, illustrated in Figure 5. The similarity between the synthetic ECG and the real ECG may be seen by comparing Figure 5 with Figure 1.

Figure 5 – Morphology of one PQRST-complex of the synthetic ECG.



Source: Based on (McSharry *et al.*, 2003)

2.3 SYNTHETIC CARDIOPATHIES

Besides of generating a healthy synthetic signal, the ordinary equations demonstrated by equation 1 can generate synthetic signals of heart diseases. For this to be possible, its necessary just change the values of Gaussian terms a_i, b_i and θ_i in Table 1 to valid values that are capable of shaping some cardiopathies. Table 2 represents the values of a_i, b_i and θ_i that form the following diseases: sinus bradycardia, sinus tachyardia, ventricular flutter, atrial fibrillation and ventricular tachycardia.

Table 2 – Parameters of the ECG model adapted from (Sayadi *et al.*, 2010).

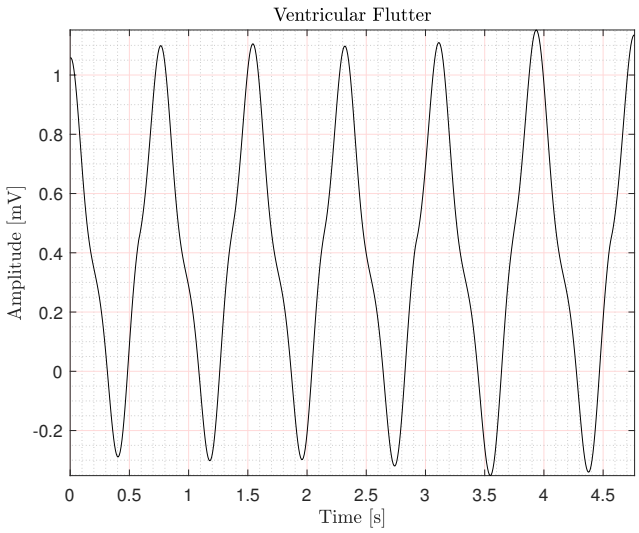
Gaussian kernels	**Sinus bradycardia**	**Sinus tachycardia**	**Ventricular flutter**	**Atrial fibrillation**	**Ventricular tachycardia**
P	(0.75, 0.15, $\frac{-3\pi}{8}$)	(0.75, 0.15, $\frac{-\pi}{6}$)	(0, 0.1, $\frac{-2\pi}{3}$)	(0.7, 0.12, $\frac{-5\pi}{7}$)	(1, 0.2, $\frac{-10\pi}{13}$)
Q	(-1, 0.1, $\frac{-\pi}{13}$)	(-7, 0.1, $\frac{-\pi}{13}$)	(0, 0.1, $\frac{-\pi}{12}$)	(0.6, 0.9, $\frac{-pi}{4}$)	(-12, 0.1, $\frac{-\pi}{3}$)
R	(20, 0.1, 0)	(20, 0.1, 0)	(20, 0.6, $\frac{-\pi}{2}$)	(18, 0.1, 0)	(1, 0.3, 0)
S	(-9.5, 0.1, $\frac{\pi}{15}$)	(-9.5, 0.1, $\frac{\pi}{17}$)	(-20, 0.6, $\frac{\pi}{2}$)	(-0.1, 0.05, $\frac{\pi}{30}$)	(3, 0.4, $\frac{2\pi}{11}$)
T	(0.21, 0.475, $\frac{2\pi}{5}$)	(0.21, 0.475, $\frac{\pi}{2}$)	(0, 0.1, $\frac{5\pi}{8}$)	(0.55, 0.17, $\frac{7\pi}{11}$)	(5, 0.5, $\frac{\pi}{2}$)

2.3.1 Ventricular Flutter

Ventricular flutter, also named ventricular fibrillation, is a completely disorganized ventricular rhythm resulting in immediate cessation of cardiac output and cardiac arrest (Goldberger *et al.*, 2013).

Figures 6 and 7 illustrate a comparison of the synthetic signal and the real signal of sinus bradycardia, respectively.

Figure 6 – Synthetic Ventricular Flutter with 300 bpm.



Source: Made by the author.

Figure 7 – ventricular flutter.

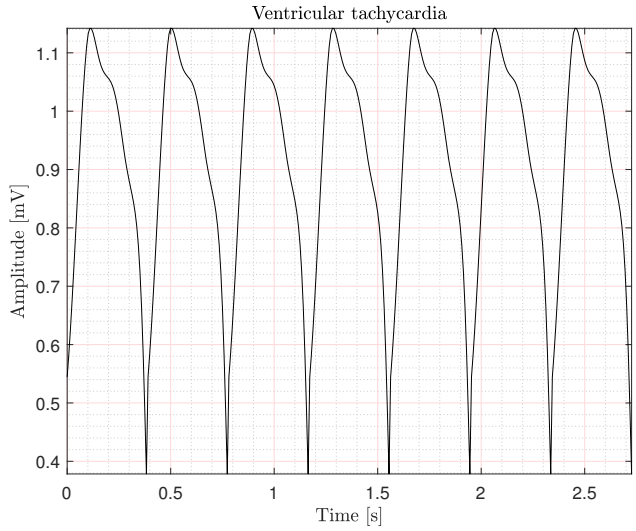
Source: (Goldberger *et al.*, 2013).

2.3.2 Ventricular Tachycardia

Ventricular tachycardia is another cardiopathy that acellerates the hear rate, caused by abnormal electrical signals in the heart ventricles (Goldberger *et al.*, 2013).

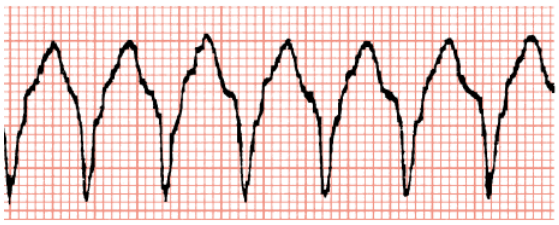
Figures 8 and 9 illustrate a comparison of the synthetic signal and the real signal of sinus bradycardia, respectively.

Figure 8 – Synthetic Ventricular Tachycardia with 100 bpm.



Source: Made by the author.

Figure 9 – Ventricular Tachycardia.



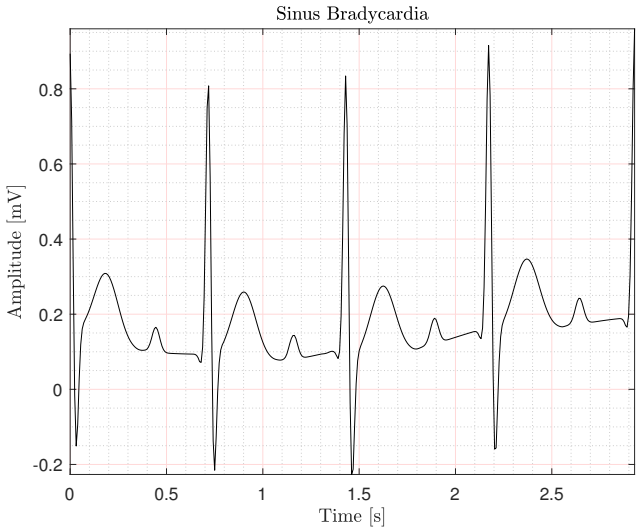
Source: Made by the author.

2.3.3 Sinus Bradycardia

Sinus bradycardia is a heart disease that slows down the heart rate. Typically, the heartbeat ranges from 60 to 100 bpm, if the hear rate is less than 60 bpm it is called sinus bradycardia (Goldberger *et al.*, 2013).

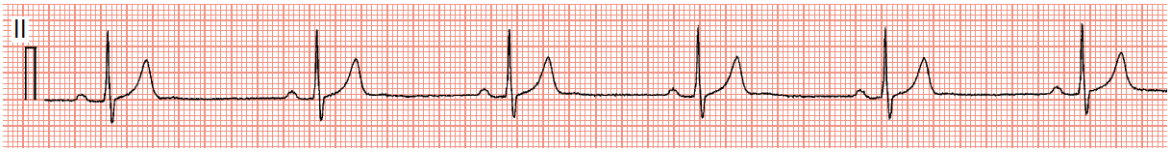
Figures 10 and 11 show a comparison of the synthetic signal and the real signal of sinus bradycardia, respectively.

Figure 10 – Synthetic Sinus Bradycardia with 40 bpm.



Source: Made by the author.

Figure 11 – Sinus Bradycardia.



Source: (Goldberger *et al.*, 2013).

2.3.4 Sinus Tachycardia

Sinus tachycardia is sinus rythm with a heart rate exceeding 100 beats/minute (Goldberger *et al.*, 2013).

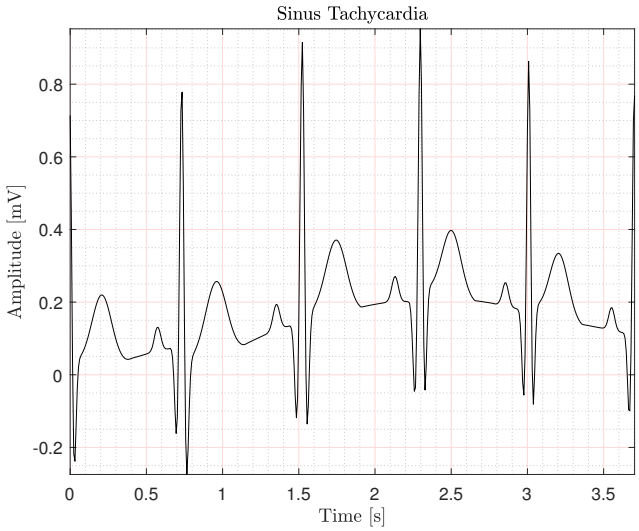
Figures 12 and 13 show a comparison of the synthetic signal and the real signal of sinus bradycardia, respectively.

2.3.5 Atrial Fibrillation

Atrial fibrillation is the most common type of cardiac arrhythmia. This cardiopathy accelerates the heart rate in a skyrocketing way (Goldberger *et al.*, 2013).

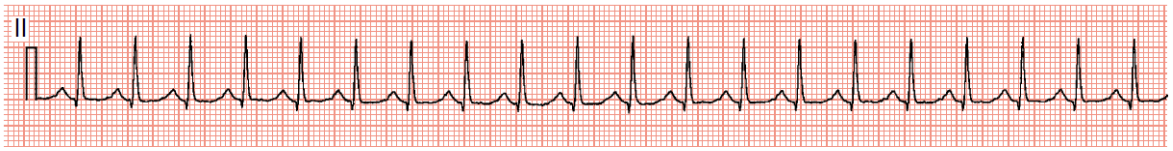
Figures 14 and 15 show a comparison of the synthetic signal and the real signal of sinus bradycardia, respectively.

Figure 12 – Synthetic Sinus Tachycardia 140 bpm.



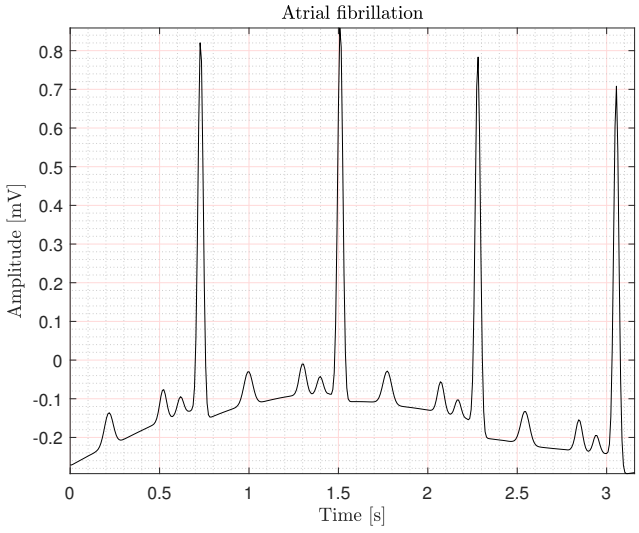
Source: Made by the author.

Figure 13 – Sinus Tachycardia.



Source: (Goldberger *et al.*, 2013).

Figure 14 – Synthetic Atrial Fibrillation with 60 bpm.



Source: Made by the author.

Figure 15 – Atrial Fibrillation.

Source: (Goldberger *et al.*, 2013).

2.4 CHAPTER 2 SYMBOL LIST

a_i	Gaussian amplitude
b_i	Gaussian width
θ_i	Gaussian center
f_2	respiratory frequency
z_0	baseline
$f(n)$	Gaussian distribution
n	randon variable
μ	expected value
σ^2	variance
Δt	fixed time step
f_s	sampling frequency

3 KALMAN FILTER

Modern filter theory began with Wiener's work in the 1940s (Wiener, 1949). His work describes a filter used to produce an estimate of a desired or target random process by linear time-invariant (LTI) filtering of a noisy observation (assuming known stationary signal and noise spectra and additive noise). This filter is known as Wiener filter, that minimizes the mean square error (MSE) between the estimated random process and the desired process (Kailath *et al.*, 2000). However, this filter was seldom used in practice (Sayed, 2008).

In 1960 Rudolf Emil Kalman introduced KF (Kalman, 1960), who published this paper to present solutions for some limitations of the Wiener filter. Basically, the KF is a stochastic filter and a generalization of the Wiener filter (Kay, 1993) .

Considering the same problem that Wiener had dealt with earlier, Kalman considered the noisy measurement to be a discrete time sequence in contrast to a continuous-time signal. He also posed the problem in a state-space setting that accommodated the time-variable MIMO scenario nicely. Engineers, especially in the field of navigation, were quick to see the Kalman technique as a practical solution to some applied filtering problems that were intractable using Wiener methods. Also, the rapid advance in computer technology that occurred in the 1960s certainly contributed to popularizing Kalman filtering as a practical means of separating signal from random noise (Brown; Hwang, 1983).

3.1 THE DISCRETE KALMAN FILTER

This section describes the KF in its original formulation Kalman (1960), and also it is based on the notes of Welch e Bishop (2001).

3.1.1 The Process to be Estimated

The Kalman filter addresses the general problem of trying to estimate the state $\mathbf{x} \in \mathbb{R}^n$ of a discrete-time controlled process that is governed by the linear stochastic difference equation

$$\mathbf{x}_k = A\mathbf{x}_{k-1} + B\mathbf{u}_k + \mathbf{w}_{k-1}, \tag{5}$$

with a measurement $\mathbf{z} \in \mathbb{R}^m$ given by

$$\mathbf{z}_k = H\mathbf{x}_k + \mathbf{v}_k. \tag{6}$$

The $n \times n$ dynamics matrix A in the difference Equation (5) relates the state at the previous time step $k-1$ to the state at the current step k, in the absence of either a driving function or process noise. The $n \times l$ matrix B relates the optional control input $\mathbf{u}_k \in \mathbb{R}^l$ to the state $\mathbf{x}_k$. The $m \times n$ matrix H in the measurement Equation (6) relates the state to the measurement $\mathbf{z}_k$.

The random processes $\mathbf{w}_k$ and $\mathbf{v}_k$ represent the process and measurement noise, respectively. They are assumed to be independent of each other, white, and with normal probability distributions, which the correlation equations can be written as follows:

$$Q_k = \mathbb{E}\left[\mathbf{w}_k \mathbf{w}_k^\intercal\right], \tag{7}$$

$$R_k = \mathbb{E}\left[\mathbf{v}_k \mathbf{v}_k^\intercal\right], \tag{8}$$

where Q_k is the process noise covariance matrix and R_k is the measurement noise covariance matrix. In practice, the process noise covariance Q_k and measurement noise covariance R_k matrices might change with each time step or measurement, however here they are assumed as constant, so they will be represented without the subindex k.

3.1.2 The Computational Origins of the KF

Define $\hat{\mathbf{x}}_k^- \in \mathbb{R}^n$ to be the *a priori* state estimate at step k given knowledge of the process at step $k-1$, and $\hat{\mathbf{x}}_k \in \mathbb{R}^n$ to be the *a posteriori* state estimate at step k given measurement $\mathbf{z}_k$. Then the *a priori* and *a posteriori* estimate errors can be defined as

$$\begin{aligned} \mathbf{e}_k^- &= \mathbf{x}_k - \hat{\mathbf{x}}_k^-, \\ \mathbf{e}_k &= \mathbf{x}_k - \hat{\mathbf{x}}_k. \end{aligned} \tag{9}$$

The *a priori* estimate error covariance matrix is then

$$P_k^- = \mathbb{E}\left[\mathbf{e}_k^- \mathbf{e}_k^{-\intercal}\right], \tag{10}$$

and the *a posteriori* estimate error covariance matrix is

$$P_k = \mathbb{E}\left[\mathbf{e}_k \mathbf{e}_k{}^{\intercal}\right]. \tag{11}$$

In deriving the equations for the KF, in order to find an equation that computes an *a posteriori* state estimate $\hat{\mathbf{x}}_k$ as a linear combination of the *a priori* estimate $\hat{\mathbf{x}}_k^-$ and a weighted difference between an actual measurement $\mathbf{z}_k$ and a measurement prediction $H\hat{\mathbf{x}}_k^-$ as shown in Equation (12)

$$\hat{\mathbf{x}} = \hat{\mathbf{x}}_k^- + K_k\left(\mathbf{z}_k - H\hat{\mathbf{x}}_k^-\right). \tag{12}$$

The difference $\left(\mathbf{z}_k - H\hat{\mathbf{x}}_k^-\right)$ in Equation (12) is called the measurement innovation. The innovation reflects the discrepancy between the predicted measurement $H\hat{x}_k^-$ and the actual measurement z_k. A residual of zero means the two are in complete agreement.

The $n \times m$ matrix K_k in Equation (12) is chosen to be the *gain* or *blending factor* or, even, *Kalman gain* that minimizes the *a posteriori* error covariance equation (11). This minimization can be accomplished by first substituting Equation (12) into the above definition for e_k, substituting that into Equation (11), performing the indicated expectations, taking the derivate of the result with respect to K_k, setting that result equal to zero, and then solving for K_k. One form of the resulting K_k that minimizes Equation (11) is given by

$$K_k = P_k^- H^T\left(HP_k^- H^T + R\right)^{-1}. \tag{13}$$

Looking at Equation (13), it can be seen that as the measurement error covariance R approaches zero, the gain K weights the residual more heavily. Specifically,

$$\lim_{R_k \to 0} K_k = H^{-1}. \tag{14}$$

On the other hand, as the *a priori* estimate error covariance P_k^- approaches zero, the gain K weights the residual less heavily. Noting that this limit presented only can be done if the matrix H is quadratic. Specifically

$$\lim_{P_k^- \to 0} K_k = 0. \tag{15}$$

Another way of thinking about the weighting by K_k is that as the measurement error covariance R approaches zero, the actual measurement $\mathbf{z}_k$ is more reliable, while the predicted

measurement $H\hat{\mathbf{x}}_k^-$ is less reliable. On the other hand, as the *a priori* estimate error covariance P_k^- approaches zero the actual measurement $\mathbf{z}_k$ is less reliable, while the predicted measurement $H\hat{\mathbf{x}}_k^-$ is more reliable.

3.1.3 The Probalistic Origins of the KF

The justification for Equation (12) is based on the probability of the *a priori* estimate $\hat{\mathbf{x}}_k^-$ conditioned on all prior measurement $\mathbf{z}_k$ (Bayes' rule). The Kalman filter maintains the first two moments of the state distribution,

$$\begin{aligned} \mathbb{E}[\mathbf{x}_k] &= \hat{\mathbf{x}}_k, \\ \mathbb{E}[(\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^\intercal] &= P_k. \end{aligned} \tag{16}$$

The *a posteriori* state estimate Equation (12) reflects the mean (the first moment) of the state distribution - it is normally distributed if the conditions of Equation (7) and Equation (8) are met. The *a posteriori* estimate error covariance Equation (11) reflects the variance of the state distribution (the second non-central moment). In other words,

$$\begin{aligned} p\left(\mathbf{x}_k|\mathbf{z}_k\right) &\sim \mathcal{N}\left(E[\mathbf{x}_k], E\left[(\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^\intercal\right]\right) \\ &= \mathcal{N}\left(\hat{\mathbf{x}}_k, P_k\right). \end{aligned} \tag{17}$$

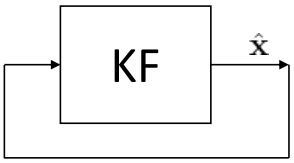
For more information about the probabilistic origins of the Kalman filter, see (Brown; Hwang, 1983).

3.1.4 The Discrete KF Algorithm

The KF estimates a process by using a form of feedback control: the filter estimates the process state at some time and then obtains feedback in the form of noisy measurements (Figure 16). As such, the equations for the KF can be seperated in two steps: *time update* equations and *measurement update* equations. The time update equations are responsible for projecting forward, in time, the current state and error covariance estimates to obtain the *a priori* estimates for the next time step. The measurement update equations are responsible for the feedback - *i.e.* for incorporating a new measurement into the *a priori* estimate to obtain an improved *a posteriori* estimate.

The time update equations can also be thought of as *predictor* equations, while the mea-

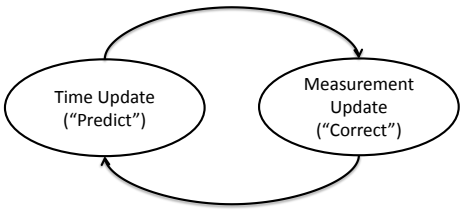
Figure 16 – KF feedback control.



Source: Made by the author.

surement update equations can be thought of as *corrector* equations. Indeed the final estimation algorithm resembles that of a *predictor-corrector* algorithm for solving numerical problems as shown in Figure 17.

Figure 17 – The simplified discrete Kalman filter cycle. The *time update* projects the current state estimate ahead in time. The *measurement update* adjusts the projected estimate by an actual measurement at that time.



Source: Based on (Welch; Bishop, 2001).

The specific equations for the time updates are given by Equations (18) and (19).

$$\hat{\mathbf{x}}_k^- = A\hat{\mathbf{x}}_{k-1}^- + B\mathbf{u}_k, \tag{18}$$

$$P_k^- = AP_{k-1}A^T + Q, \tag{19}$$

and the equations for the measurement updates are given by (20)-(22).

$$K_k = P_k^- H^T \left(HP_k^- H^T + R\right)^{-1}, \tag{20}$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + K_k\left(\mathbf{z}_k - H\hat{\mathbf{x}}_k\right), \tag{21}$$

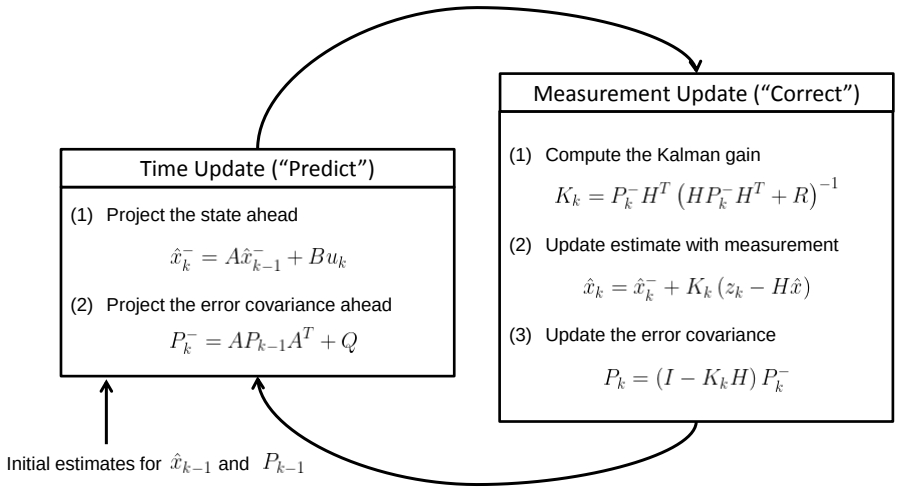
$$P_k = \left(I - K_k H\right) P_k^-. \tag{22}$$

Note how the time update Equations (18) and (19) project the state and covariance estimates forward from time step $k-1$ to step k. A and B are from Equation (5), while Q is given by Equation (7).

The first task during the measurement update is to compute the Kalman gain, K_k. Note that Equation (20) is equal to Equation (13). The next step is to actually measure the process to obtain z_k, and then to generate an *a posteriori* state estimate by incorporating the measurement as in Equation (21). Again Equation (21) is simply Equation (12) repeated here for completeness. The final step is to obtain an *a posteriori* error covariance estimate via Equation (22).

After each time and measurement update pair, the process is repeated with the previous *a posteriori* estimates used to project or predict the new *a priori* estimates. This recursive nature is one of the very appealing features of the Kalman filter. It makes practical implementations much more feasible than, e.g., an implementation of a Wiener filter (Brown; Hwang, 1983) which is designed to operate on all of the data directly for each estimate. The Kalman filter instead recursively conditions the current estimate on all of the past measurements. Figure 18 offers a complete picture of the operation of the filter, combining the high-level diagram of Figure 17 with the equations from equations (18)-(22).

Figure 18 – A complete picture of the operation of the Kalman filter, combining the high-level diagram of Figure 17 with the equations from Equations (18)**-**(22)**.**



Source: Based on (Welch; Bishop, 2001)

3.2 THE EXTENDED KALMAN FILTER

As described in Section 3.1, the Kalman filter addresses the general problem of estimating the state $\mathbf{x}_k \in \mathbb{R}^n$ from a discrete-time controlled process that is governed by a linear stochastic equation. However, in this work, the problem presented has is a non-linear model. So it is important to consider one of the most successful applications of the Kalman filtering, named the EKF, which linearizes the non-linear system around the current mean and covariance (Welch; Bishop, 2001).

3.2.1 The Process to be Estimated

Similarly to a Taylor series, the process around the current estimate can be linearized using partial derivatives of the process and the measurement functions, yielding estimates even in the face of non-linear relationships. Assume that the process has a state vector $\mathbf{x}_k \in \mathbb{R}^n$, and is governed by the non-linear stochastic difference equation

$$\mathbf{x}_k = f_k\left(\mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{w}_{k-1}\right), \tag{23}$$

with measurement $\mathbf{z}_k \in \mathbb{R}^m$ given by

$$\mathbf{z}_k = h_k\left(\mathbf{x}_k, \mathbf{v}_k\right), \tag{24}$$

where the random variables $\mathbf{w}_k$ and $\mathbf{v}_k$ represent the process and measurement noise as in Equations (7) and (8), respectively. In this case the non-linear function f_k in the difference Equation (23) relates the state at the previous time step $k-1$ to the state at the current time step k, including as parameters any input function u_k and the zero-mean process noise $\mathbf{w}_k$. The non-linear function h_k in the measurement Equation (24) relates the state $\mathbf{x}_k$ to the measurement $\mathbf{z}_k$.

There are cases that the functions f_k and h_k varies according to the time k, but in this cases these functions do not varies, so they will be represented as f and h.

The state and measurement vectors can be approximated as

$$\tilde{\mathbf{x}}_k = f\left(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_k, 0\right), \tag{25}$$

and

$$\tilde{\mathbf{z}}_k = h\left(\hat{\mathbf{x}}_k,0\right), \tag{26}$$

where $\hat{\mathbf{x}}_k$ is some *a posteriori* estimate of the state from a previous time step k.

It is important to note that a fundamental drawback of the EKF is that the distributions of the random processes, are no longer normal after undergoing their respective nonlinear transformations. The EKF is simply an *ad hoc* state estimator that only approximates the optimality of Bayes' theorem by linearization.

3.2.2 The Computational Origins of the EKF

In order to estimate a process with a non-linear difference and measurement relationships consider the linearized dynamics

$$\mathbf{x}_k \approx \tilde{\mathbf{x}}_k + A\left(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}\right) + W\mathbf{w}_{k-1}, \tag{27}$$

$$\mathbf{z}_k \approx \tilde{\mathbf{z}}_k + H\left(\mathbf{x}_k - \tilde{\mathbf{x}}_k\right) + V\mathbf{v}_k, \tag{28}$$

where

- $\mathbf{x}_k$ and $\mathbf{z}_k$ are the actual state and measurement vectors;
- $\tilde{\mathbf{x}}_k$ and $\tilde{\mathbf{z}}_k$ are the approximated state and approximated measurement vectors from Equations (25) and (26);
- $\hat{\mathbf{x}}_k$ is an *a posteriori* estimate of the state at step k;
- the random variables $\mathbf{w}_k$ and $\mathbf{v}_k$ represent the process and measurement noise as in Equations (7) and (8), respectively;
- $\dot{A}$ is the Jacobian matrix of the partial derivates of f with respect to $\mathbf{x}$, that is

$$\dot{A}_{[i,j]} = \frac{\partial f_{[i]}}{\partial \mathbf{x}_{[j]}}\left(\hat{x}_{k-1},u_k,0\right), \tag{29}$$

  where $\dot{A}_{[i,j]}$ refers to the (i,j) element of matrix A;

- $\dot{W}$ is the Jacobian matrix of partial derivates of f with respect to $\mathbf{w}$,

$$\dot{W}_{[i,j]} = \frac{\partial f_{[i]}}{\partial \mathbf{w}_{[j]}}\left(\hat{x}_{k-1},u_k,0\right); \tag{30}$$

- $\dot{H}$ is the Jacobian matrix of partial derivatives of h with respect to $\mathbf{x}$,

$$\dot{H}_{[i,j]} = \frac{\partial h_{[i]}}{\partial \mathbf{x}_{[j]}} \left(\tilde{x}_k, 0\right); \tag{31}$$

- $\dot{V}$ is the Jacobian matrix of partial derivatives of h with respect to $\mathbf{v}$,

$$\dot{V}_{[i,j]} = \frac{\partial h_{[i]}}{\partial \mathbf{v}_{[j]}} \left(\tilde{x}_k, 0\right). \tag{32}$$

For simplicity in the notation, it is not used the time step subscript k with the Jacobians $\dot{A}$, $\dot{W}$, $\dot{H}$ and $\dot{V}$, even though they are different at each time step.

Now, it is defined as a new notation for the prediction error,

$$\tilde{\mathbf{e}}_{\mathbf{x}_k} = \mathbf{x}_k - \tilde{\mathbf{x}}_k, \tag{33}$$

and the measurement prediction error,

$$\tilde{\mathbf{e}}_{\mathbf{z}_k} = \mathbf{z}_k - \tilde{\mathbf{z}}_k. \tag{34}$$

In practice one does not have access to $\mathbf{x}_k$ in Equation (33), which is the actual state vector, it is the quantity one is trying to estimate. On the other hand, one does have access to $\mathbf{z}_k$ in Equation (34), which is the actual measurement that one is using to estimate x_k. Using Equations (33) and (34), it is possible to write the governing equations for an error process as

$$\begin{aligned} \tilde{\mathbf{e}}_{\mathbf{x}_k} &\approx \tilde{\mathbf{x}}_k + A\left(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}\right) + W\mathbf{w}_{k-1} - \tilde{\mathbf{x}}_k, \\ \tilde{\mathbf{e}}_{\mathbf{x}_k} &\approx A\left(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}\right) + \varepsilon_k, \end{aligned} \tag{35}$$

$$\begin{aligned} \tilde{\mathbf{e}}_{\mathbf{z}_k} &\approx \tilde{\mathbf{z}}_k + H\left(\mathbf{x}_k - \tilde{\mathbf{x}}_k\right) + V\mathbf{v}_k - \tilde{\mathbf{z}}_k, \\ \tilde{\mathbf{e}}_{\mathbf{z}_k} &\approx H\tilde{\mathbf{e}}_{\mathbf{x}_k} + \eta_k, \end{aligned} \tag{36}$$

where ε_k and η_k represent new independent random variables having zero mean and covariance matrices WQW^T and VRV^T, with Q and R as in Equations (7) and (8), respectively, these new variables are replacing the terms $W\mathbf{w_{k-1}}$ and $V\mathbf{v}_k$, respectively.

Notice that the Equations (35) and (36) are linear and that they closely resemble the difference and measurement Equations (5) and (6) from the discrete Kalman filter. This motivates the use of the actual measurement residual $\tilde{e}_{z_k}$ in the Equation (34) and a second (hypothetical) Kalman filter to estimate the prediction error $\tilde{e}_{x_k}$ given by (35). This estimate, call it $\hat{e}_k$, could then be used along with Equation (33) to obtain the *a posteriori* state estimates for the original non-linear process as

$$\hat{\mathbf{x}}_k = \tilde{\mathbf{x}}_k + \hat{\mathbf{e}}_k. \tag{37}$$

The random variables of Equation (35) and Equation (36) have approximately the following probability distributions (represented by the letter p):

$$\begin{aligned} p(\tilde{\mathbf{e}}_{\mathbf{x}_k}) &\sim \mathcal{N}(0, E[\tilde{\mathbf{e}}_{\mathbf{x}_k}\tilde{\mathbf{e}}_{\mathbf{x}_k}^{\intercal}]), \\ p(\varepsilon_k) &\sim \mathcal{N}(0, WQ_kW^{\intercal}), \\ p(\eta_k) &\sim \mathcal{N}(0, VR_kV^{\intercal}). \end{aligned}$$

Given these approximations, the Kalman filter equation used to estimate $\hat{\mathbf{e}}_k$ could be represented by

$$\hat{\mathbf{e}}_k = K_k\tilde{\mathbf{e}}_{\mathbf{z}_k}. \tag{38}$$

and substituting Equation (38) back into Equation (37) and making use of Equation (34), it can be seen that:

$$\begin{aligned} \hat{\mathbf{x}}_k &= \tilde{\mathbf{x}}_k + K_k\tilde{\mathbf{e}}_{\mathbf{z}_k} \\ &= \tilde{\mathbf{x}}_k + K_k(\mathbf{z}_k - \tilde{\mathbf{z}}_k). \end{aligned} \tag{39}$$

Now Equation (39) can be used for the measurement update in the extended Kalman filter, with $\tilde{\mathbf{x}}_k$ and $\tilde{\mathbf{z}}_k$ coming from Equation (25) and Equation (26), and the Kalman gain K_k coming from Equation (20) with appropriate substitution for the measurement error covariance.

Equations (40)-(44) represent the complete set of EKF equations. Note that $\hat{\mathbf{x}}_k^-$ was substituted for $\tilde{\mathbf{x}}_k$ to remain consistent with the earlier "super minus" *a priori* notation, and that now can be attached the subscript k to the Jacobians $\dot{A}$, $\dot{W}$, $\dot{H}$ and $\dot{V}$, to reinforce the notation that they are different at each time step.

$$\hat{\mathbf{x}}_k^- = f\left(\hat{\mathbf{x}}_{k-1}, u_k, 0\right), \tag{40}$$

$$P_k^- = A_kP_{k-1}A_k^T + W_kQ_{k-1}W_k^{\intercal}. \tag{41}$$

As with the basic discrete Kalman filter, the time update Equations (40) and (41) project the state and covariance estimates from the previous time step $k-1$ to the current time step k.

As with the basic discrete Kalman filter, the measurement update Equations (42)-(44) correct the state and covariance estimates with the measurement $\mathbf{z}_k$.

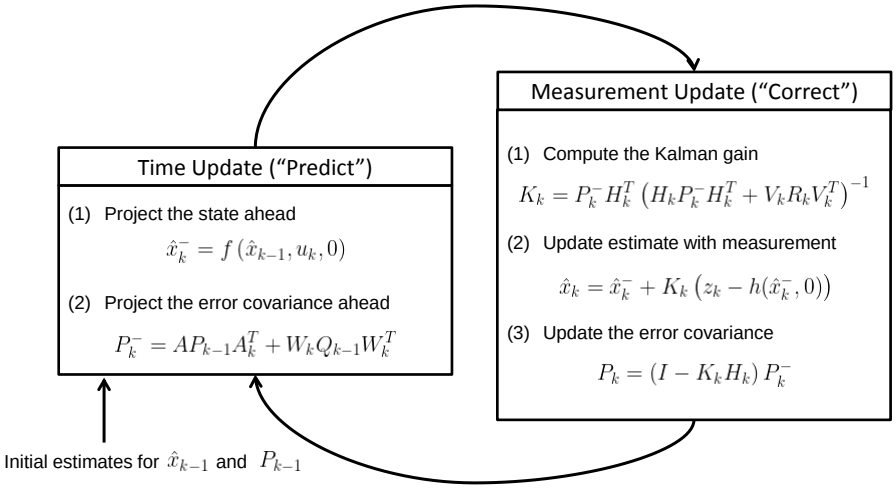
$$K_k = P_k^-H_k^T\left(H_kP_k^-H_k^{\intercal} + V_kR_kV_k^{\intercal}\right)^{-1}, \tag{42}$$

$$\hat{\mathbf{x}}_k = \hat{x}_k^- + K_k \left(z_k - h(\hat{x}_k^-,0) \right), \tag{43}$$

$$P_k = \left(I - K_k H_k \right) P_k^-. \tag{44}$$

Figure 19 offers a complete picture of the operation of the EKF, combining the high-level diagram of Figure 17 with the equations from Equations (40)-(44).

Figure 19 – A complete picture of the operation of the extended Kalman filter, combining the high-level diagram of Figure 17 with the Equations (40)**-**(44)**.**



Source: Based on (Welch; Bishop, 2001).

3.3 CHAPTER 3 SYMBOL LIST

$\mathbf{x}_k$	variable state
$\mathbf{z}_k$	measurement state
$\mathbf{u}_k$	control input
$\mathbf{w}_k$	process noise
$\mathbf{v}_k$	measurement noise
k	time step
A	dynamic matrix
B	matrix relates the control input
H	mtrix relates the state to the mesurement
Q_k	process noise covariance matrix
R_k	measurement noise covariance matrix
$\hat{\mathbf{x}}_k^-$	*a priori* state estimate
$\hat{\mathbf{x}}_k$	*a posteriori* state estimate
$\mathbf{e}_k^-$	*a priori* estimate error
$\mathbf{e}_k$	*a posteriori* estimate error
P_k^-	*a priori* error covariance matrix
P_k	*a posteriori* error covariance matrix
K_k	Kalman gain
f_k	non-linear function
h_k	non-linear function
$\tilde{\mathbf{x}}_k$	approximated state vector
$\tilde{\mathbf{z}}_k$	aproximated measurement vector
$\dot{A}$	Jacobian matrix of the partial derivatives of f respect to $\mathbf{x}$
$\dot{W}$	Jacobian matrix of the partial derivatives of f respect to $\mathbf{w}$
$\dot{H}$	Jacobisn matrix of partial derivatives of h respect to $\mathbf{x}$

$\dot{V}$	Jacobian matrix of partial derivatives of h with respect to $\mathbf{v}$
$\tilde{\mathbf{x}}_{x_k}$	prediction error
$\tilde{\mathbf{e}}_{z_k}$	measurement prediction error
ε_k	independent random variable
η_k	independent random variable

4 METHODOLOGY

This chapter presents the methodology used to implement the EKF. The EKF was chosen because the ECG signal is non-linear. Some previous results of the EKF implementation are also presented. Part of this methodology was based on Sameni *et al.* (2005), Sameni *et al.* (2007).

4.1 EXTENDED KALMAN FILTER IMPLEMENTATION

With this general overview of KF theory and the previously developed dynamical ECG model explained in Chapters 3 and 2, respectively, it is possible to use the synthetic dynamical ECG model within a KF framework. To do so, the dynamic equations in Equation (1) need to be modified. As presented in (Sameni *et al.*, 2005), these equations can be transferred into polar coordinates. Moreover, assuming the z_k state variable in Equation (1) to be in millivolts, b_i's and θ_i's in radians, and time in seconds, it is clear that the a_i's are in mV/(rads $\times$ s). To simplify the dimensions and later relate the model parameters with real ECG recordings, the a_i term in system (1) will be replaced with:

$$a_i = \frac{\alpha_i \omega}{b_i^2} \quad i \in \{P,Q,R,S,T\}, \tag{45}$$

where the α_i are the peak amplitudes of the Gaussian functions used for modeling each of the ECG components, in millivolts. This definition may be verified from (1), by neglecting the baseline wander term $(z - z_0)$ and integrating the $\dot{z}$ equation with respect to t. With these changes, the new form of the dynamic equations in cylindrical coordinates is

$$\begin{aligned} \dot{r} &= r(1-r), \\ \dot{\theta} &= \omega, \\ \dot{z} &= -\sum_{i \in P,Q,R,S,T} \frac{\alpha_i \omega}{b_i^2} \Delta_i \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right) - (z - z_0), \end{aligned} \tag{46}$$

where r and θ are, respectively, the radial and angular state variables in polar coordinates. This new sets of equations have some benefits compared to the original equations in (McSharry *et al.*, 2003). First of all, the polar form is much simpler and its interpretation is straightforward. Accordingly, the first equation will reach the limit cycle $r = 1$ with any initial value of r, but as it further seen the second and third equations of (46) are independent from r. This means

that the first differential equation may be excluded as it does not affect the synthetic ECG (the z state variable). Another benefit of this representation is that the phase parameter θ is an explicit state-variable, noting that this parameter indicates the angular location of the P, Q, R, S and T waves in Table 1. This point is further used in the implementation of the EKF (Sameni *et al.*, 2005). In this case, the simplified dynamic model of the equation (46) in its discrete form, with the assumption of a small sampling period of δ is as follows:

$$\begin{aligned} \theta_{k+1} &= (\theta_k + \omega\theta) \mod 2\pi, \\ z_{k+1} &= -\sum_i \delta \frac{\alpha_i \omega}{b_i^2} \Delta\theta_i \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right) + z_k + \eta, \end{aligned} \tag{47}$$

where $\Delta\theta_i = (\theta_k - \theta_i) \mod 2\pi$, η is a random additive noise that models the inaccuracies of the dynamic model (including the baseline wander), and the summation over i is taken over the number of Gaussian functions (or turning points) used for modeling the shape of the desired ECG. In fact, due to the flexibility of Gaussian mixtures, it is believed that, by using a sufficient number of Gaussian functions, they can be fitted to signals recorded from different ECG leads. However, to illustrate the general filtering framework, it is used only five Gaussians to model the ECG channels containing the P, Q, R, S and T waves.

Henceforth, θ_k and z_k are assumed to be the state variables, and ω, α_i, θ_i, b_i and η are assumed as independent and identically distributed Gaussian random variables considered to be process noises. Following the notation of equation (25), the system state and process noise vectors are defined as follows:

$$\begin{aligned} \mathbf{x}_k &= [\theta_k, z_k]^\intercal, \\ \mathbf{w}_k &= [\alpha_P, ..., \alpha_T, b_P, ..., b_T, \theta_P, ..., \theta_T, \omega, \eta]^\intercal, \end{aligned} \tag{48}$$

and the process noise covariance matrix is given as $Q_k = E[w_k w_k^T]$.

To set up an EKF model based on the nonlinear synthetic model of equation (47), it is necessary to have a linearized version of the model. Consequently, the state-equation of equation (47) requires linearization using the Jacobian matrix of the partial derives shown in equations (25), (26), (29), (30), (31) and (32). By defining

$$\begin{aligned} \theta_{k+1} &= f_1(\theta_k, \omega, k), \\ z_{k+1} &= f_2(\theta_k, z_k, \omega, \alpha_i, \theta_i, b_i, \eta, k), \end{aligned} \tag{49}$$

following equations represent the linearized model with respect to the state variables θ_k and z_k:

$$\begin{aligned}
&\frac{\partial f_1}{\partial z_k} = 0, \\
&\frac{\partial f_1}{\partial \theta_k} = \frac{\partial f_2}{\partial z_k} = 1, \\
&\frac{\partial f_2}{\partial \theta_k} = -\sum_{i \in P,Q,R,S,T} \delta \frac{\alpha_i \omega}{b_i^2} \left[1 - \frac{\Delta\theta_i^2}{b_i^2}\right] \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right).
\end{aligned} \tag{50}$$

Similarly, the linearization of Equation (49) with respect to the process noise components yields in $i \in \{P,Q,R,S,T\}$

$$\begin{aligned}
&\frac{\partial f_1}{\partial \omega} = \delta, \\
&\frac{\partial f_2}{\partial \eta} = 1, \\
&\frac{\partial f_1}{\partial \alpha_i} = \frac{\partial f_1}{\partial b_i} = \frac{\partial F1}{\partial \theta_i} = \frac{\partial F_1}{\partial \eta} = 0, \\
&\frac{\partial f_2}{\partial \alpha_i} = -\delta \frac{\omega \Delta\theta_i}{b_i^2} \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right), \\
&\frac{\partial f_2}{\partial b_i} = 2\delta \frac{\alpha_i \omega \Delta\theta_i}{b_i^3} \left[1 - \frac{\Delta\theta_i^2}{2b_i^2}\right] \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right), \\
&\frac{\partial f_2}{\partial \theta_i} = \delta \frac{\alpha_i \omega}{b_i^2} \left[1 - \frac{\Delta\theta_i^2}{b_i^2}\right] \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right), \\
&\frac{\partial f_2}{\partial \omega} = -\sum_i \delta \frac{\alpha_i \Delta\theta_i}{b_i^2} \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right).
\end{aligned} \tag{51}$$

The noisy ECG recordings are assumed to be observations for the EKF. The relationship between the states and observations of the EKF depends on the location of the electrodes and the origin of the measurement noise. For example, motion artifacts, environmental noise, or bioelectrical artifacts such as electromyography or electrogastric noise, may be assumed as the measurement noises. While the measurement noise can generally contaminate the ECG in a non-linear and non-Gaussian form, the results are based on the assumption of additive Gaussian noise.

In addition to the noisy ECG observations, the phase θ may also be added as a second observation. In fact, by studying the values of Table 1, it is noticed that the R-peak is always assumed to be located at $\theta = 0$ and the ECG contents lying between two consecutive R-peaks are assumed to have a phase between 0 and 2π (or $-\pi$ and π). So by simply detecting the R-peaks an additional observation is achieved.

Hence the phase observations (ϕ_k) and noisy ECG measurements (s_k) may be related to

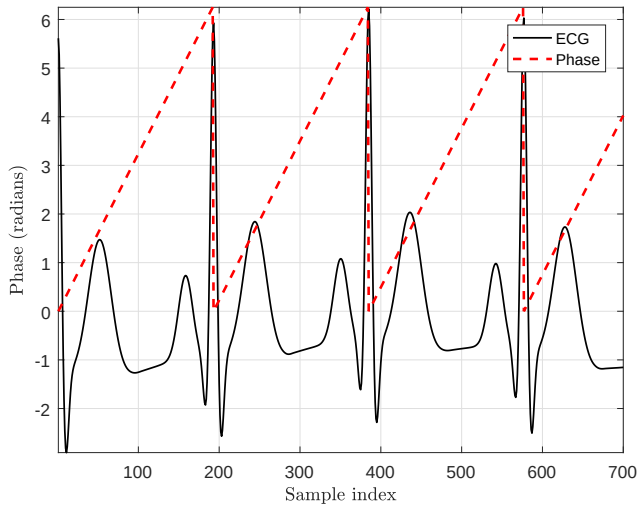
the state vector as follows

$$\begin{bmatrix} \phi_k \\ s_k \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_k \\ z_k \end{bmatrix} + \begin{bmatrix} u_k \\ v_k \end{bmatrix}, \tag{52}$$

where $R_k = E\left[[u_k, v_k]^\intercal [u_k, v_k]\right]$ is the observation noise covariance matrix.

In the context of estimation theory, the variance of the observation noise in equation (52) represents the degree of reliability of a single observation. In other words, when a rather precise measurement of the system states is valid the value of R_k is low, and the Kalman filter gain is adapted such as to rely on the specific measurement. On the other hand, for the epochs with noise where there are no measurements available, the value of R_k is high and the Kalman filter tends to follow its internal dynamics rather than tracking the observations. This point may be used to include additional measurements for the angle θ. In fact, θ_k has a periodic value that starts from $\theta = 0$ at the R-peak and ends at $\theta = 2\pi$ with the next R-peak. Although the only valid phase observation is obtained from the R-peak locations, it is possible to linearly assign a phase value between 0 and 2π to the intermediate samples. This means that even for angles other than $\theta = 0$, it is possible to assign a phase measurement between 0 and 2π to each angle, as illustrated in Figure 20.

Figure 20 – Illustration of the phase assignment approach.



Source: Based on (Sameni *et al.*, 2007)

Before the implementation of the EKF model, it is necessary to have an estimate of the values of the process and measurement noise covariance matrices. Generally, for the 17 noise parameters of equation (49) a 17×17 process noise covariance matrix Q_k should be found, but

if the noises are uncorrelated with each other, the matrix is simplified to a diagonal matrix. The measurement noise covariance matrix R_k has a similar case.

To built the process noise covariance matrix Q_k it will be used the exact values to simulate the normal ECG from Table 1, since the EKF entries are real, the simulated ECG signal is an estimative of the ECG should seem.

The angular frequency ω may be set to $\omega = 2\pi/T$, where T is the RR-interval period in each ECG cycle. A simpler approximation is to use a global ω using the average RR-interval of the whole signal.

The variance of the process noise η should also be estimated. Remembering that η is a parameter that represents the imprecision of the dynamic model, and after some tests it will be considered a value approximate from zero.

From equation (52) it can be observed that u_k is the phase measurement noise. As mentioned before, the phase for each beat is determined from the R-peaks of the signal. A possible noise source for u_k is the sampling error that occurs when the actual R-peak is located between two sample times. By considering that each ECG cycle is equivalent to 2π in the phase domain, u_k would be uniformly distributed in the range of $\pm\omega\delta/2$, where ω is the angular frequency and δ is the sampling period. With this assumption:

$$E\left[u_k^2\right] = (\omega\delta)^2/12. \tag{53}$$

The method used to estimate the variance of the measurement noise v_k is to estimate the noise power from the deviations of the whole signal around the ECG signal, but after some tests it was observated that this value was approximately zero.

4.2 CHAPTER 4 SYMBOL LIST

α_i	peak amplitude of the Gaussian
a_i	Gaussian amplitude
b_i	Gaussian width
θ_i	Gaussian center
z_0	baseline
r	radial state variable
θ	angular state variable
z_k	state variable
ω	angular frequency
η	random additive noise
$\mathbf{x}_k$	system state vector
$\mathbf{w}_k$	process noise vector
ϕ_k	phase observation
s_k	noisy ECG measurement
u_k	measurement phase noise
v_k	measurement noise
R_k	measurement noise covariance matrix
Q_k	process noise covariance matrix
T	RR-interval period in each ECG cycle
δ	small sampling period

5 RESULTS

From the methodology of the work, it should be borne in mind that the construction of the EKF is entirely based on a synthetic healthy ECG signal, with healthy meaning a normal sinus rhythm ECG, based on the data presented in Chapter 2.

The idea of building an EKF based entirely on a synthetic healthy signal would be for the filter to "learn" just how a healthy ECG should behave, because as soon as something different from that healthy or normal ECG appears, the filter would detect an anomaly or error, or be a disease. Thus, important parameters for the EKF such as the error covariance matrix (P_k), the process noise (w_k), the variance matrix of the process noise (Q_k), the measuremente noise ($[u_k,v_k]$) and the variance matrix of the measurement noise (R_k) are defined based on a synthetic normal sinus ECG which were created with 80 bpm and a sampling frequency of 128 Hz, as shown in Figure 21. And the specific values of these parameters are:

$$
\begin{aligned}
P_k &= \begin{pmatrix} 4\pi^2 & 0 \\ 0 & (10\,|\mathbf{x}_k|)^2 \end{pmatrix}, \\
\mathbf{w}_k &= \begin{pmatrix} a_i & b_i & \theta_i & \omega & 0 \end{pmatrix}, \\
Q_k &= \texttt{diag}\begin{pmatrix} (0.1\cdot a_i)^2 & (0.5\cdot b_i)^2 & (0.05\cdot \theta_i)^2 & \omega_{std}^2 & 0 \end{pmatrix}, \\
[u_k,v_k] &= \begin{pmatrix} 0 & 0 \end{pmatrix}, \\
R_k &= \begin{pmatrix} 0.3584\cdot 10^{-3} & 0 \\ 0 & 1\cdot 10^{-6} \end{pmatrix},
\end{aligned}
\tag{54}
$$

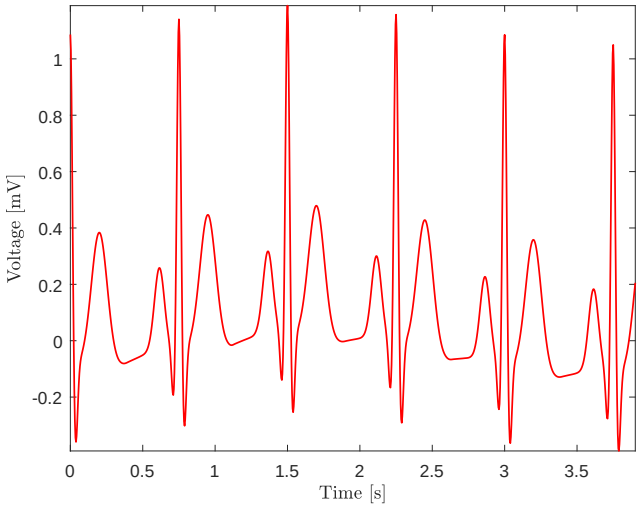
where a_i, b_i, θ_i are values taken from Table 1, ω is the angular frequency with the value of 8.3939, ω_{std} is the heart-rate standard deviation ($\omega_{std} = 0.0667$), and finally `diag` means that the matrix Q_k is a diagonal matrix.

Once set up all the EKF parameters, the next step is to define the EKF states, which in this case are the real signals of healthy ECG with some traces of disease. And so, the EKF will try to estimate the signal.

To validate the work, a signal that contains healthy and diseased samples from a real ECG, i.e., an ECG obtained from a human being, must be estimated as a state. To build this signal, real ECG samples from the MIT database were used.

As there are many heart diseases, it was necessary to create a criterion to choose which diseases would be used to test the effectiveness of EKF. Therefore, observing the existing heart

Figure 21 – Synthetic normal ECG signal with 80 bpm and f_s = 128 Hz.



Source: Made by the author.

diseases, it was noticed that there are two ways to modify the behavior of a healthy ECG (normal sinus rhythm). Thus, this work divided heart disease into two major classes. The first class is composed of those heart diseases that modify the patterning of points P, Q, R, S, and T that a healthy ECG has. Examples of this disease are ventricular flutter and ventricular tachycardia, shown in Figures such and such , respectively. The second class comprises those diseases that have P, Q, R, S, and T points well defined as well as a healthy ECG, but the heart rate is altered, with examples being bradycardia, shown in Figure 11, which slows the heart rate and sinus tachycardia that accelerates the heartbeat, shown in the Figure 13.

Based on this observation and the division of heart diseases into two classes, ventricular flutter was chosen to represent the class of heart diseases that modify the structure of the ECG, and supraventricular tachycardia was chosen to represent the class of heart diseases that modify the rhythm cardiac. Thus, two signals will be created, a first signal containing healthy ECG and ventricular flutter samples, and a second signal containing healthy ECG and supraventricular tachycardia samples.

Once defined which diseases will be used for the functioning of the EKF, it must be defined how the validation of the proposal will be carried out, that is, if the EKF serves as a classifier of heart diseases. Therefore, two analyzes will be carried out: the first analysis consists of a qualitative analysis of the error signal of the state to be estimated with the estimated signal, and the second analysis, already a quantitative analysis, will calculate the energy of the windows.

For the calculation of energy used the Equation (55):

$$Eg = \frac{\mathbf{sig}^2}{wnd}, \tag{55}$$

where $\mathbf{sig}$ represents the sign of the error that consists of the difference between the estimated signal and the signal to be estimated, and wnd represents the window size, that is, the number of samples in that window.

5.1 FIRST ANALYSIS: WHEN THE MORPHOLOGY CHANGES

First, it will be analyzed the behavior of the EKF in relation to changing the ECG structure. So, the Figure 22 shows the signal created and the output of the EKF. The dashed red line represents the signal created with healthy ECG samples and ventricular flutter heart disease, which is the state to be estimated by the EKF, the solid black line is the output (response) of the EKF, the estimated state.

Also, in relation to Figure 23, the blue dashed vertical lines are the separations of the created windows, these windows were created for the analysis of the signal energy. The windows were created according to the healthy and sick samples.

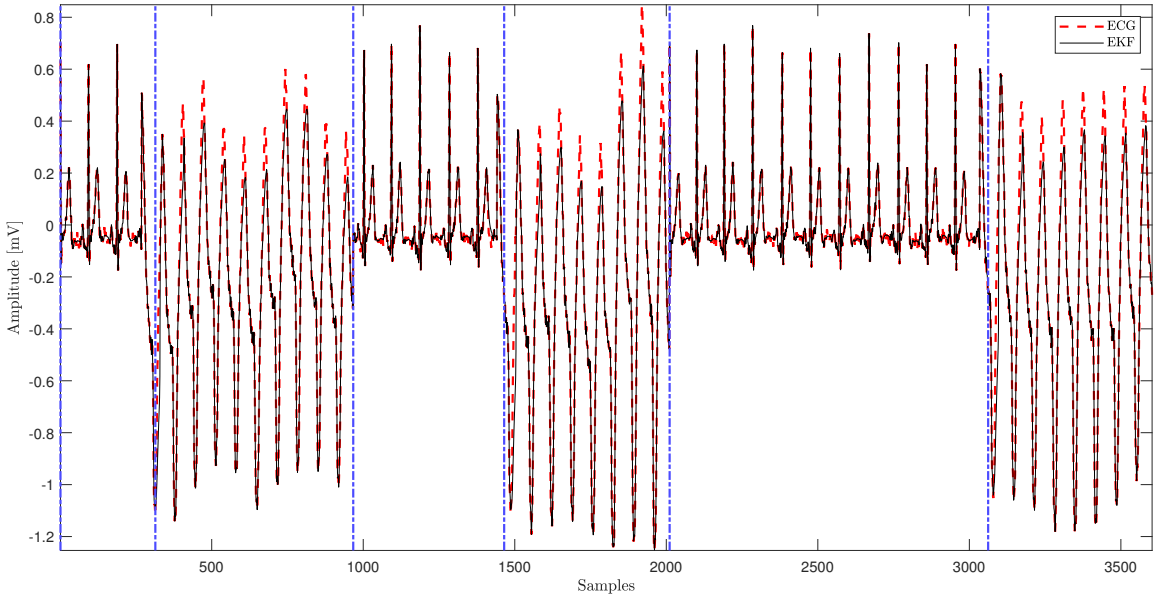
In a visual and comparative analysis, it can be seen that apparently the filter was able to estimate both the behavior of the healthy signal and the behavior of the ventricular flutter. In a closer look, shown in Figure 23, it can be seen that the filter made a very small mistake in estimating the ventricular flutter, so, in other words, the filter was able to estimate the morphology change in a satisfactory way.

However, the chosen classification criteria will be based on the sign of the error, not the state estimated by the EKF itself. As mentioned before, a qualitative analysis of the error signal will be done first, and then a quantitative analysis will be done.

So, Figure 23 shows the error between the sign to be estimated and the EKF output. As can be seen, the windows that indicate the ventricular flutter samples have a greater error when compared to the error presented by the windows that indicate the healthy ECG samples. This error indicates that there is an anomaly in the signal, and this anomaly can be translated as a disease, which in this case is a disease that has totally changed the structure of the ECG.

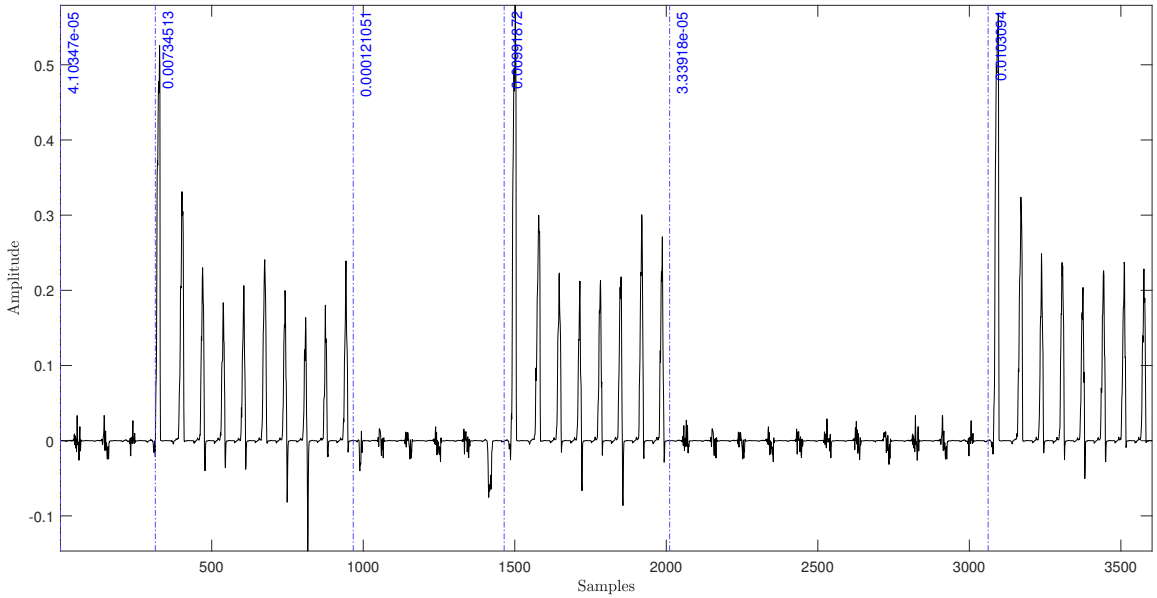
Now, performing a quantitative analysis, the energy of the windows will be calculated accordind to Equation (55). Figure 23 brings the energy values of each window, and Table 3

Figure 22 – EKF Output: ventricular flutter.



Source: Made by the author.

Figure 23 – Error signal of the input in relation to the EKF output: ventricular flutter.



Source: Made by the author.

better illustrates these values.

Table 3 – The value of the ventricular flutter energy windows.

Window	Sample	Energy
1	Normal sinusal rythm 1	$4.10347 \cdot 10^{-5}$
2	Ventricular flutter 1	$7.35513 \cdot 10^{-3}$
3	Normal sinusal rythm 2	$1.21051 \cdot 10^{-5}$
4	Ventricular flutter 2	$9.91072 \cdot 10^{-3}$
5	Normal sinusal rythm 3	$3.33918 \cdot 10^{-5}$
6	Ventricular flutter 3	$1.03094 \cdot 10^{-2}$

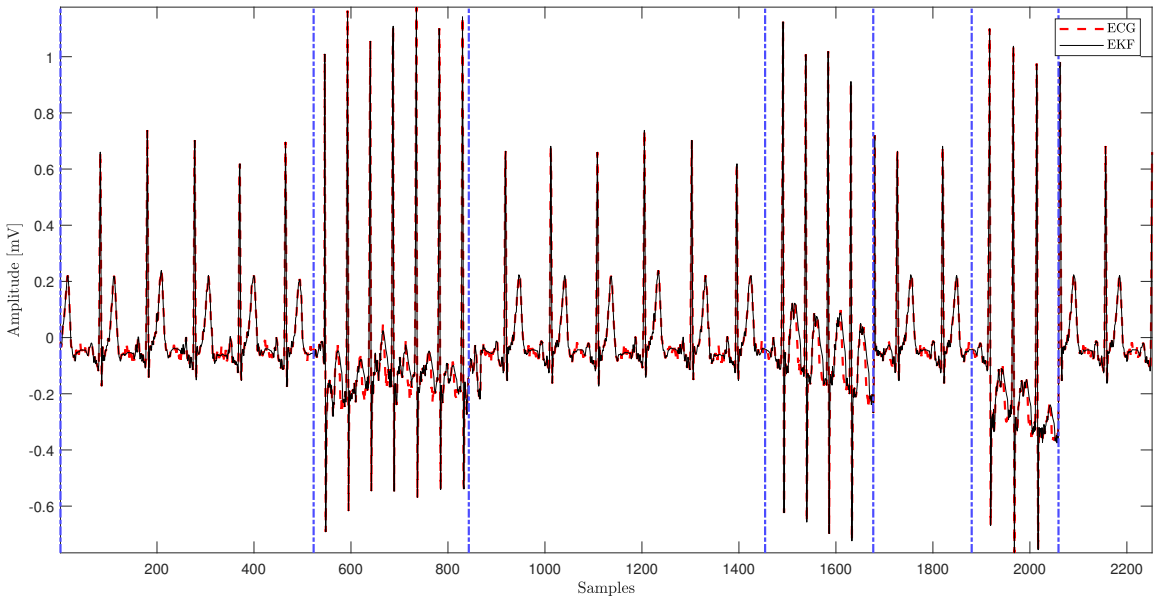
5.2 SECOND ANALYSIS: WHEN THE HEART RYTHM CHANGES

The second analysis to be done will be in relation to diseases that alter heart rate. Then, a signal was created with ECG samples of normal sinus rhythm and supraventricular tachycardia. And just as the first case presented was analyzed, this will be analyzed equally.

So, Figure 24 shows the signal created and the output of the EKF. The dashed red line represents the signal created with healthy ECG and supraventricular tachycardia cardiopathy samples, which is the status to be estimated by the EKF, while the continuous black line is the output (response) of the EKF, the estimated status.

And just like in the first case presented, the blue dashed vertical lines are the separations of the created windows, these windows were created for the analysis of the signal energy. The windows were created according to healthy and sick samples.

Figure 24 – EKF Output: ventricular flutter.

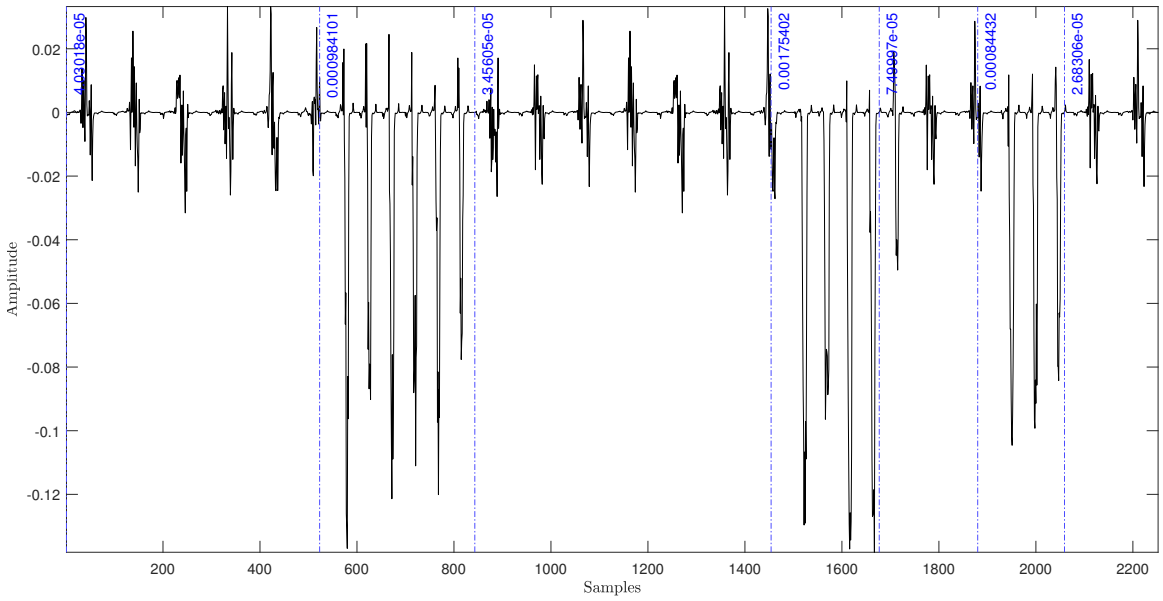


Source: Made by the author.

As can be seen in Figure 24, the EKF was able to better estimate the supraventricular

tachycardia than the ventricular flutter, however the classification analyzes will be based on the error sign corresponding to Figure 25.

Figure 25 – Error signal of the input in relation to the EKF output: supraventricular tachycardia.



Source: Made by the author.

As in the case of ventricular flutter, the qualitative analysis of the error signal of supraventricular tachycardia cardiopathy indicates a greater error in disease samples when compared to errors in healthy samples.

Regarding the quantitative analysis, the energy of the windows will be calculated according to Equation (55). Figure 25 brings the energy values of each window, and Table 4 better illustrates these values.

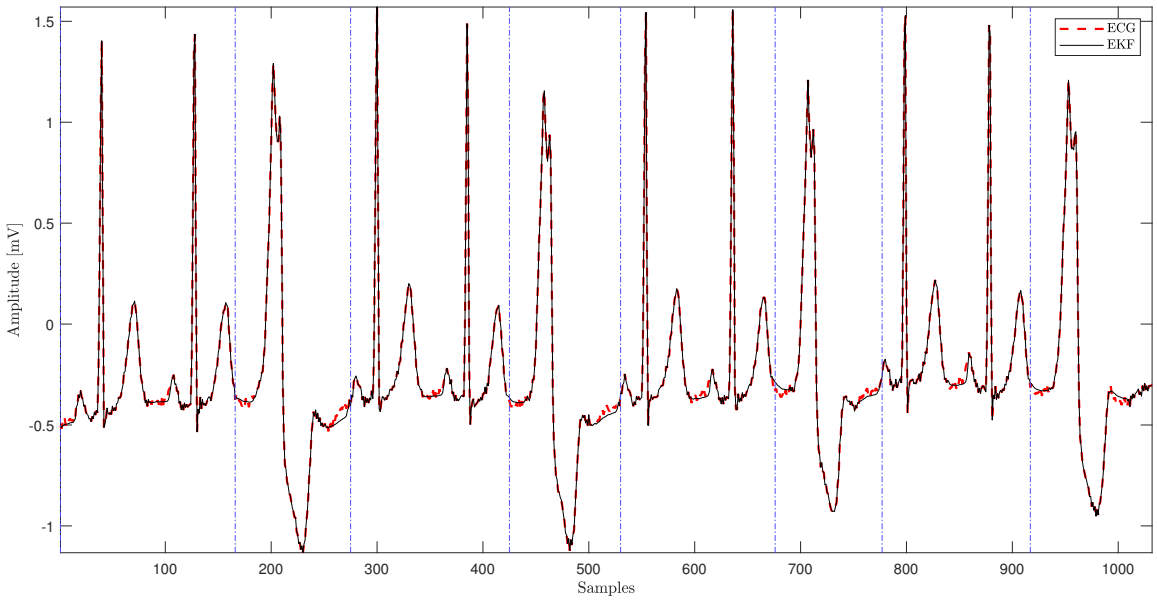
Table 4 – The value of the supraventricular tachycardia energy windows.

Window	Sample	Energy
1	Normal sinusal rythm 1	$4.10348 \cdot 10^{-5}$
2	Supraventricular tachycardia 1	$9.84101 \cdot 10^{-4}$
3	Normal sinusal rythm 2	$3.45605 \cdot 10^{-5}$
4	Supraventricular tachycardia 2	$1.75402 \cdot 10^{-3}$
5	Normal sinusal rythm 3	$7.49997 \cdot 10^{-5}$
6	Supraventricular tachycardia 3	$8.4432 \cdot 10^{-4}$
7	Normal sinusal rythm 4	$2.68306 \cdot 10^{-5}$

5.3 COMPARING THE ENERGIES

In order to compare the energies and verify that KF can be used as a classifier, a third disease was used in the filter, this being ventricular trigeminy. This disease consists of two healthy heartbeats before one sick heartbeat (Goldberger *et al.*, 2013), the red dashed line in the Figure 26 better illustrates this disease, just as the continuous black line illustrates the output of the EKF.

Figure 26 – EKF output: ventricular trigeminy.



Source: Made by the author.

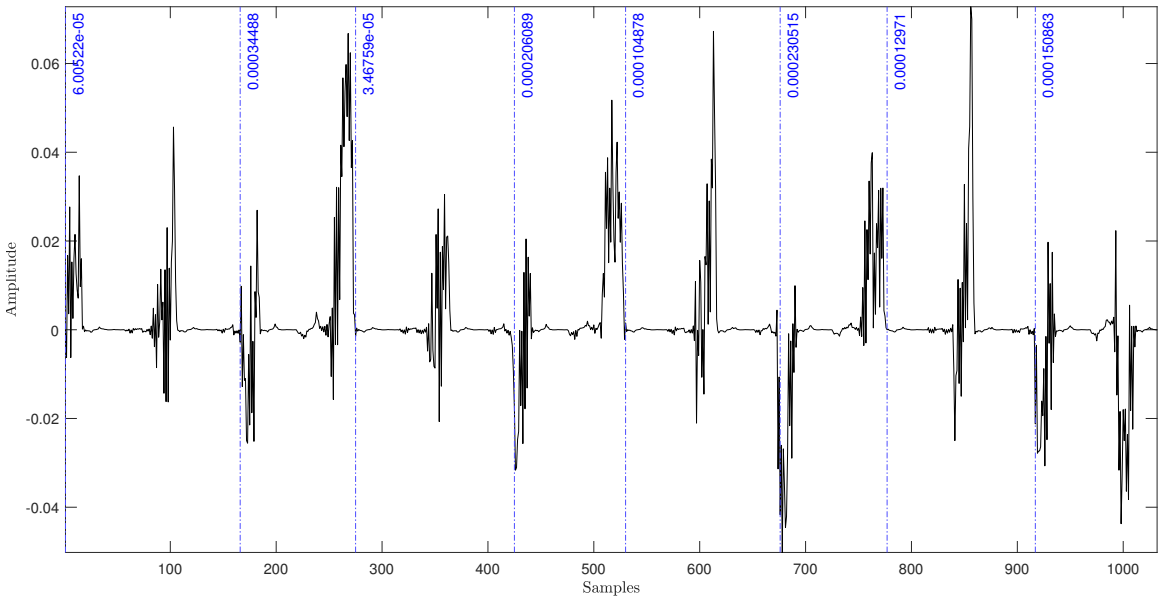
Figure 27 shows the error between the sign to be estimated and the EKF output. As done with the other deseases, windowns were created to calculate the sample energies, the separation of the windows are shown by the dashed blue lines.

Table 5 express better the energies values.

Table 5 – The value of the ventricular trygeminy energy windows.

Window	Sample	Energy
1	Normal sinusal rythim 1	$4.82359 \cdot 10^{-5}$
2	Ventricular trigeminy 1	$3.41465 \cdot 10^{-4}$
3	Normal sinusal rythim 2	$3.20557 \cdot 10^{-5}$
4	Ventricular trigeminy 2	$2.25208 \cdot 10^{-4}$
5	Normal sinusal rythim 3	$7.62638 \cdot 10^{-5}$
6	Ventricular trigeminy 3	$2.60397 \cdot 10^{-4}$
7	Normal sinusal rythim 4	$8.37457 \cdot 10^{-5}$
8	Ventricular trigeminy 4	$1.49192 \cdot 10^{-4}$

Figure 27 – Error signal of the input in relation to the EKF output: ventricular trigeminy.

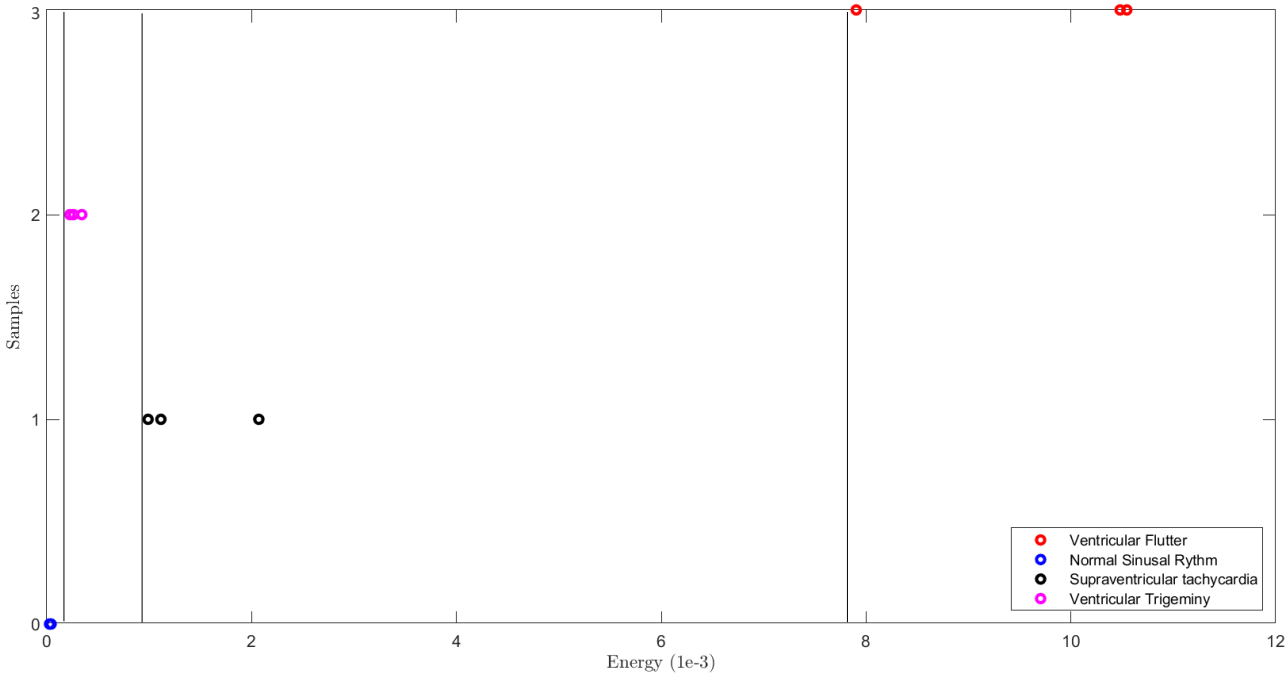


Source: Made by the author.

However, this test was done to build an energy comparison and show the filter's effectiveness. Thus, Figure 28 shows a visual way of comparing the energy of these diseases, axis x brings the values of the energies of each sample, and the energies presented are divided by 10^{-3}, and axis y matches the samples, with 0 being set for normal sinus rhythm samples, 1 for supraventricular tachycardia samples, 2 for ventricular trigeminy samples, and 3 for ventricular flutter samples. The vertical black lines just indicate where the energies begin, in order to demonstrate that they are diferent, although the sinus normal rythm and the ventricular trigeminy seem to have the same energy, but the black line proves they do not have the same energy.

Through this comparison, it is possible to see the difference in energies, and the possibility of using them as a classification criterion. But the focus is to see through the energies that it is possible to use the EKF as a heart disease classifier.

Figure 28 – Comparing the energies.



Source: Made by the author.

5.4 CHAPTER 5 SYMBOL LIST

P_k	error covariance matrix
$\mathbf{w}_k$	process noise
a_i	Gaussian amplitude
b_i	Gaussian width
θ_i	Gaussian center
ω	angular frequency
$\mathbf{x}_k$	estimate state-ECG signal
u_k	measurement phase noise
v_k	measurement noise
R_k	measurement noise covariance matrix
Q_k	process noise covariance matrix
ω_{std}	heart-rate standard deviation
f_s	frequency sampling
`diag`	diagonal matrix
E_g	energy
sig	sign of the error
wnd	window size

6 CONCLUSION

The objective proposed in this study was to evaluate whether FK can be used as a classifier of heart disease or if it can somehow help in these classifications. As the objective is a study of a new and different technique from the use of KF, the results presented were observed in a way that allows concluding whether the filter can be used as a classifier or not, and from this answer, ways of improvement will be suggested technique that should be performed in future works.

From this objective, the work was developed as follows: first, the concept of synthetic ECG and what is a KF and its algorithms were presented.

With the definitions of the KF equations and algorithm, it was noticed that this filter is a state estimator that finds the best solution or answer for a given state that you want to obtain an indirect measure or obstructed by noise, in the case of this work, the state to be estimated is a human's ECG. Throughout the theoretical development of the FK, it was seen that this filter is an optimal estimator for linear differential equations, and the human ECG does not have a linear behavior, so it was necessary to use an update of the FK, this being the EKF that is specific for nonlinear systems and equations. The EKF has the same equations and algorithm as the normal FK, however, the only difference is that the nonlinear equations are linear by some method, in the case of this work it was used as a Jacobian linearization form, before starting to execute the algorithm.

After defining the EKF, the next step would be to define the differential equations that would represent the human ECG, so Chapter 2 of this dissertation summarized the work presented by (McSharry *et al.*, 2003), in which in this work the authors developed a system with three differential equations capable of simulating a Synthetic ECG in a realistic way of a human being. From these equations it was possible to develop an algorithm for the EKF and validate the objective of this research.

For the validation of the objective, that is, to obtain results, the following technique was chosen: The EKF should estimate, as a state, a signal constructed with healthy samples and sick samples from a real ECG, that is, an ECG obtained from being human. As there are several heart diseases, a criterion was created to choose which diseases would be used to test the effectiveness of the EKF, thus, heart diseases were divided into two large classes: the first class is composed of those heart diseases that modify the standardization of points P, Q, R, S, e, T that a healthy ECG

has, the ventricular flutter being chosen to represent this class, and the second class is composed of those diseases that have points P, Q, R, S, e, T well defined as well as a healthy ECG, but the cardiac rhythm is altered, being chosen the supraventricular tachycardia to represent this class.

The developed EKF was based entirely on a healthy synthetic ECG, that is, normal sinus rhythm, as the expectation was that the filter would only “learn” to estimate a healthy ECG signal, so when the ECG changed to a heart disease the EFK would detect some error or anomaly, and that would indicate an illness.

Two techniques were chosen to assess whether EKF can be used as a classifier. The first technique was based on a qualitative analysis of the error signal of the state to be estimated with the estimated signal, and the second technique was based on a quantitative analysis, the energy of the windows will be calculated.

Through the qualitative analysis it was observed that comparing the two obtained errors, the morphology error and the heart rate error, it can be seen that the morphology error is much greater (when considering the amplitude module) than the heart rate error, in this case this means that the ventricular flutter could be a health issue more serious than the supraventricular tachycardia, but this affirmation only can be done by the health professional. A possible explanation for the result presented earlier is that the EKF was built from a healthy signal structure. Arrhythmia disease (represented by supraventricular tachycardia) has the same morphology as a normal sinus ECG, while the flutter "destroys"the standardized morphology of the ECG signal. Perhaps this is why the ventricular flutter error was greater when compared to the supraventricular tachycardia error.

Through the quantitative analysis of the energies of the error windows, presented by Tables 3 and 4, it was noticed that the energies of the ventricular flutter samples are much higher than the energies of the healthy samples, and the healthy samples are considered to be practically zero. The energies of the supraventricular tachycardia samples are not as significant as the energies of the ventricular flutter, but they are different energies from the healthy samples and also quite different from the ventricular flutter samples, indicating the possibility of classification through the energy of the signal window of the error.

Finally, the research developed is the first step towards a new technique to use KF, in this case, to recognize it as a heart disease classifier. From the results obtained, it is possible to use it as a classifier or complement a classification technique. Now the next step is to improve KF to implement this proposed technique.

As a suggestion for improvements and future work, for example, the EKF showed to be a better option because it estimates non-linear signal, and the ECG is non-linear, however along the work, this filter demonstrated to be very sensible when small modifications are made to the developed algorithm, like changes to sampling frequency. Maybe this occurred because of the jacobian matrices. A possible solution for this is to use the Unscented Kalman filter (UKF), which instead of using the first-order linearization of the nonlinear system in its formulation (as the EKF does), the UKF uses a deterministic sampling approach, by using the unscented transformation (UT), that is a method for calculating the statistics of a random variable which undergoes a nonlinear transformation (Haykin, 2001).

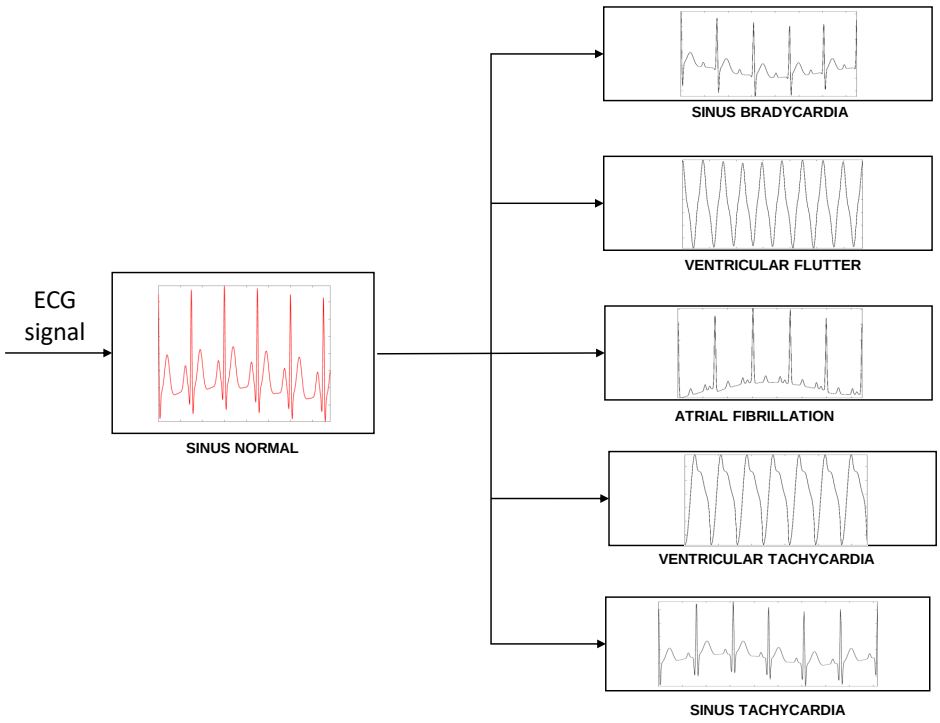
Another ideia is to improve the classification method like using the synthetic diseases presented before as a classifier criterion. For example, following the logic of Figure 29, first, there is an ECG signal that will pass over a Kalman filter that contains information about the behavior of a healthy synthetic ECG (as shown in the methodology of this work). If that first filter does not detect any anomalies or modifications, the signal is expected to be entirely healthy. If something different is detected, that same signal will be used in other Kalman filters that will be built from the information of diseased synthetic signals presented in Table 2. That filter that shows the smallest error for heart disease, will indicate which disease it would be.

But, as mentioned before, this work is an assessment of the possibility of using Kalman filter as a classifier. These suggestions will be used as future works to improve the method preseted in this work.

So, summarizing, the main ideas for futures works are:

- Using the Unscented Kalman filter;
- Using the synthetic diseases as classifier criterion;

Figure 29 – Classifier example using the synthetic diseases as criterion.



Source: Made by the author.

REFERENCES

Almalchy, M. T.; Ciobanu, V.; Popescu, N. Noise removal from ECG signal based on filtering techniques. *In*: **2019 22nd International Conference on Control Systems and Computer Science (CSCS)**, 2019. p. 176–181.

Blackburn, Henry; Keys, Ancel; Simonson, Ernedt; Rautaharju, Pentti; Punsar, Sven. The electrocardiogram in population studies: A classification system. **Circulation**, v. 21, n. 6, p. 1160–1175, 1960.

Brown, Robert Grover; Hwang, Patrick Y. C. **Introduction to Random Signals and Applied Kalman Filtering: with Matlab Exercises**. Hoboken, New Jersey: John Wiley and Sons, 1983.

Cardenas, O. A.; Flores Nava, L. M.; Castañeda, F. G.; Moreno Cadenas, J. A. ECG arrhythmia classification based on fuzzy cognitive maps. *In*: **2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)**, 2019. p. 1–4.

Clifford, Gari; McSharry, Patrick. A realistic coupled nonlinear artificial ECG, BP, and respiratory signal generator for assessing noise performance of biomedical signal processing algorithms. **Proceedings of SPIE - The International Society for Optical Engineering**, v. 5467, p. 290–301, 2004.

Clifford, Gari; Shoeb, A; McSharry, P. E.; Janz, BA. Model-based filtering, compression and classification of the ECG. **International Journal of Bioelectromagnetism**, v. 7, p. 158–161, 2005.

Dolinský, Pavol; Andras, Imrich; Michaeli, Linus; Grimaldi, Domenico. Model for generating simple synthetic ECG signals. **Acta Electrotechnica et Informatica**, v. 18, p. 3–8, 2018.

Dubin, Dale. **Rapid Interpretation of EKG's**. Tampa, Florida: Cover Pub Co, 2000.

Goldberger, Ary L.; Goldberger, Zachary D.; Shvilkin, Alexei. **Goldberger's Clinical Electrocardiography: A Simplified Approach**. Philadelphia, PA: Elsevier Saunders, 2013.

GUTIERREZ-GNECCHI, Jose Antonio; MORFIN-MAGANA, Rodrigo; LORIAS-ESPINOZA, Daniel; TELLEZ-ANGUIANO, Adriana del Carmen; REYES-ARCHUNDIA, Enrique; MENDEZ-PATINO, Arturo; CASTANEDA-MIRANDA, Rodrigo. DSP-based arrhythmia classification using wavelet transform and probabilistic neural network. **Biomedical Signal Processing and Control**, v. 32, p. 44–56, 2017.

Haykin, Simon. **Kalman Filtering and Neural Networks**. New York, USA: John Wiley and Sons, 2001.

Kailath, Thomas; Sayed, Ali H.; Hassibi, Babak. **Linear Estimation**. Upper Saddle River, New Jersey: Prentice-Hall, 2000.

Kalman, R. E. A new approach to linear filtering and predction problems. **Transaction of the ASME-Journal of Basic Engineering**, p. 35–45, 1960.

Kay, Steven M. **Fundamentals of Statistical Signal Processing: Estimation Theory**. Upper Saddle River, New Jersey: Prentice Hall Signal Processing Series, 1993. v. 1.

Kovács, Péter. ECG signal generator based on geometrical features. **Annales Universitatis Scientiarum Budapestinensis de Rolando Eotvos Nominatae. Sectio Computatorica**, v. 37, 2012.

Kubicek, J.; Penhaker, M.; Kahankova, R. Design of a synthetic ECG signal based on the Fourier series. *In*: **2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)**. New Delhi, India, 2014. p. 1881–1885.

McSharry, P. E.; Clifford, G. D.; Tarassenko, L.; Smith, L. A. A dynamical model for generating synthetic electrocardiogram signals. **IEEE Transactions on Biomedical Engineering**, v. 50, n. 3, p. 289–294, 2003.

Mneimneh, M. A.; Yaz, E. E.; Johnson, M. T.; Povinelli, R. J. An adaptive Kalman filter for removing baseline wandering in ECG signals. *In*: **2006 Computers in Cardiology**. Valencia, Spain, 2006. p. 253–256.

Moody, G. B.; Mark, R. G. The impact of the MIT-BIH arrhythmia database. **IEEE Engineering in Medicine and Biology Magazine**, v. 20, n. 3, p. 45–50, 2001.

Press, W. H; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. **Numerical Recipes in C**. Cambrige, Uk: Cambrige University Press, 1992.

Sameni, Reza; Clifford, Gari; Jutten, Christian; Shamsollahi, Mohammad. Multichannel ECG and noise modeling: Application to maternal and fetal ECG signals. **Journal on Advances in Signal Processing**, 2007.

Sameni, R.; Shamsollahi, M. B.; Jutten, C.; Babaie-Zade, M. Filtering noisy ECG signals using the extended Kalman filter based on a modified dynamic ECG model. *In*: **Computers in Cardiology, 2005**. Lyon, France, 2005. p. 1017–1020.

Sameni, R.; Shamsollahi, M. B.; Jutten, C.; Clifford, G. D. A nonlinear Bayesian filtering framework for ECG denoising. **IEEE Transactions on Biomedical Engineering**, v. 54, n. 12, p. 2172–2185, 2007.

Sayadi, Omid; Shamsollahi, Mohammad B; Clifford, Gari D. Synthetic ECG generation and Bayesian filtering using a Gaussian wave-based dynamical model. **Institute of Physics and Engineering in Medicine**, v. 31, p. 1309–1329, 2010.

Sayed, Ali H. **Adaptive Filters**. Hoboken, New Jersey: John Wiley and Sons, 2008.

TAYEL, Mazhar B.; ELTRASS, Ahmed S.; AMMAR, Abeer I. A new multi-stage combined kernel filtering approach for ECG noise removal. **Journal of Electrocardiology**, v. 51, n. 2, p. 265–275, 2018.

Ting, Chee-Ming; Salleh, S. ECG based personal identification using extended Kalman filter. *In*: **10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010)**. Kuala Lumpur, Malaysia, 2010. p. 774–777.

Vullings, R.; de Vries, B.; Bergmans, J. W. M. An adaptive Kalman filter for ECG signal enhancement. **IEEE Transactions on Biomedical Engineering**, v. 58, n. 4, p. 1094–1103, 2011.

Welch, Greg; Bishop, Gary. **An Introduction to the Kalman Filter**. Univertsity of North Carolina at Chapel Hill, 2001.

WHO. **Cardiovascular Diseases (CVDs)**. 2017.

Wiener, Norbert. **Extrapolation, Interpolation and Smoothing of Stacionary Time Series**. New York, Wiley, 1949. v. 1.