The Art of Heart Rate Variability

Driver Fatigue Application

Master's thesis in Biomedical Engineering

FABIO FORCOLIN

The Art of Heart Rate Variability

Drive Fatigue Application

FABIO FORCOLIN

This thesis has been written during the Erasmus exchange program between Chalmers Tekniska Högskola and Universitá degli studi di Padova, during the second semester of the Academic Year 2014/2015.

The study have been carried out at SAFER - Vehicle and Traffic Safety Center at Chalmers, under the supervision of Post-doc Rubén Buendia and Prof. Bengt-Arne Sjöqvist, professor at the department of Signals and Systems in Chalmers. The supervisor at Universitá degli Studi di Padova has been Prof. Alfredo Ruggeri, Prof. of Biomedical Instrumentation for the Mater's Degreen in Biomedical Engineering.

The Art of Heart Rate Variability
Driver fatigue Application
FABIO FORCOLIN
Department of Signals and Systems
Chalmers University of Technology
Università degli Studi di Padova

Abstract

A substantial percentage of accidents are caused by drivers falling asleep at wheel. Among the different ways of detecting sleepiness at wheel, physiological measurements start changing at an earlier stage of fatigue what is crucial for accident avoidance. This motivated to study how heart rate and its variability (HRV) relate to drivers sleepiness. The study was conducted on a population of around 80 drivers that drove 3 times (morning, afternoon, night) for about 90 minutes, with subjective sleepiness evaluations every 5 minutes, leading to over 3500 epochs to analyse. In order to derive HRV indices, outlier detection and spectral transformation need to be applied to the data. Different technics are available for these purposes with no consensus about their suitability. In this thesis, the most relevant HRV indices were derived using the main technics for outliers detection and spectral transformation. Two methods were used for outliers detection, one based on a heart beat interval differing in more than 30% the average of the 4 beats before and the other consisting in an interval deviating 5 std from mean. Regarding spectral transformation, DFT, AR-model and LS Periodogram were considered. The agreement between methods was evaluated using Bland-Altman plots and Student T-Tests. The number of samples classified as outliers and removed is, in average, 0.278% for the percentage methods and 0.0648% for the standard deviation method. Outlier detection methods showed to have a low influence on the value of the HRV indices calculated. Spectral transformation methods showed, instead, to have much higher consequences on the HRV indices. The outlier detection method does not have an influence on the capability of different HRV indices of discriminating between morning, afternoon and night. The spectral transformation method does not show a significant influence either. However LS-periodogram performed slightly worst in this regards.

Keywords: Heart rate variability, Driver fatigue, outliers detection, spectral transformation, KSS.

Acknowledgements

I would like to express my gratitude to my supervisor, Postdoc. Rubén Buendia, for his effort, patience, help and knowledge that added considerably to my experience. I appreciate his wide knowledge in statistical methods for data analysis, his precious help for writing the report and all the time dedicated to me during these months for completing the thesis and obtaining all the results. He provided me materials, technical support, directions and became essential part of my work. I am also thankful to him and my examiner, Professor Bengt-Arne Sjöqvist, for proposing me such an attractive study, which was perfectly fitting my interests. It is thank to them if I started being more and more passionate about safety and human behaviour.

A special thank goes to Cristina, who put me in contact with Rubén and for her support during the year. My gratitude goes also to my friends and colleagues in Chalmers and and SAFER, for the time dedicated to me for discussions, exchanges of knowledge, experiences and the funny moments.

Thanks also to my family for supporting me in all my decisions, in particular during the last year. My greatest thank goes to Alessandra, whose support was fundamental during the year, who encouraged me continuing and to do my best and I apologize for the time stolen to her.

Last but not least, I would thank Autoliv who provided the dataset to analyse and without which this project would have not been possible.

<div align="right">

Fabio Forcolin, Gothenburg, June 2015

</div>

Nomenclature

- RR or NN: R-to-R or Normal-to-Normal. Interval (in seconds) between two R peaks.
- HRV: Heart Rate Variability.
- HR: Heart Rate
- ECG: ElectroCardioGram.
- EEG: ElectroEncephaloGram.
- CAN BUS: Controller Area Network
- ANS: Autonomic Nervous System.
- CNS: Central Nervous System.
- PNS: Peripheral Nervous System.
- PN: Parasympathetic Nervous System.
- SN: Sympathetic Nervous System.
- KSS: Karolinska Sleepiness Scale.
- EBS: Estimated Breath Cycle.
- LS: Lomb-Scargle periodogram.
- FFT/DFT/FT: Fast/Discrete Fourier Transform.
- AR: AutoRegressive model.
- SDNN: Standard Deviation of NN intervals.
- RMSSD: Root Mean Square of Successive Differences.
- pNN50%: percentage of NN interval that differs more than 50 ms.
- mHR: mean HR.
- stdHR: HR standard deviation.
- RRTri: Triangular Index.
- TINN: Triangular Interpolated index.
- VLF: Very Low Frequency.
- LF: Low Frequency.
- HF: High Frequency.
- n.u.: normalized units.
- SD1, SD2: length of minor and major axis of Poincarè plot.
- BA: Bland-Altman
- BP: Boxplot
- CI: Confidence interval

Contents

List of Figures

List of Tables

List of Tables

1
Background

1.1 Driver fatigue

Driver sleepiness is one of the main causes of road accidents. One of the factors that influences sleepiness is *driver fatigue*: the transitory period between awake and asleep that, if not interrupted, can lead to sleep [1]. From 2009 to 2013, concerning the crashes caused by drowsiness, in 7% of the cases the occupants reported only minor injuries and in 13% of them the occupants needed to be hospitalized. In for 21% of the cases the person was killed [2]. Driver fatigue accounts for 20-35% of serious accidents, leading to a large number of deaths (almost 1550 in 100000 crashes according to NTHSA) and serious injuries.

Different studies for understanding the causes of fatigue and its consequences were conducted in different parts of the world, with different results mainly due to different road traffic nature and different ways of conducting the experiment (people, simulator, real drive...). Even if the results are not identical, they converge to the same fact that sleepiness in drivers increases sensibly the possibilities of crash and thus its consequences.

Comparisons between different driving hours [3] or different ages [4], asses that driving at night increases driver fatigue and thus the chances of falling asleep at the wheel. Among the most exposed individuals to driver fatigue are young drivers and professional shift workers. Beside night driving, other factors that have been demonstrated increasing driver fatigue are low speeds and long driving tasks [3], conclusion never achieved when simulators were chosen for the experiments. The same study found also that sleepiness in drivers appears to be related to a reduced speed and a leftward position of the vehicle on the road: the speed reduction is probably due to a redirection of driver's attention to more important tasks.

1.2 Subjective measurements of sleepiness

Different tests have been proposed to assess the degree of sleepiness and they base on subjective perceptions of the person examined. Tests commonly used are Epworth Sleepiness Scale (ESS), Stanford Sleepiness Scale (SSS), Oxford SLEep Resistance (OSLER) and Karolinska Sleepiness Scale (KSS) [5], [6]. All those scales differ mainly in the range of values and all record the sleepiness from subjective answers about the feeling of sleepiness of the person under test.

To find a correlation between the measurements and the real need of sleeping, a feedback from the driver is needed and it should be given in real-time. Therefore, a

KSS	Meaning
1	Extremely alert
2	Very alert
3	Alert
4	Rather alert
5	Neither alert nor sleepy
6	Some signs of sleepiness
7	Sleepy, but no effort to keep awake
8	Sleepy, some effort to keep awake
9	Very sleepy, great effort to keep awake, fighting sleep

Table 1.1: KSS values and their meanings.

passenger in the right front seat asks periodically how the driver feels and registers the result, e.g., according to the KSS [7]. Since driver behaviour should not be altered during the task, every interaction with humans or activities different from driving (e.g. listening to music, drinking coffee, texting...) are forbidden. However, there are no other ways to have a real-time feedback and thus this approach could bias the results.

For this study, the KSS was used. KSS is a 9-point self-reported scale that measures the subject's fatigue. It spans from "extremely alert" to "extremely sleepy-fighting sleep" [8]. Values and their meaning are reported in Table 1.1 [9].

1.3 Heart and Heart Rate Variability

Heart is one of the vital organs and it is located in the center of the chest cavity (over the hypocondriac-epigastric quadrant), more precisely in the anterior mediastinum between the two lungs, behind the sternum and the costal cartilages. Its main function is to pump the blood through all the body parts in order to sprinkle tissues and organs with oxygenated blood and to lead the de-oxygenated blood to the lungs. Heart works like a pump: by a series of contractions (systole and diastole) of the ventricles and opening and closure of valves, blood enters and exits the heart continuously. Heart is an involuntary organ, which means it is not controlled consciously by the person itself but by its autonomic nervous system (ANS). The contraction of the heart is mediated by the pacemaker cells (modified cardiomyocytes) that are able to autonomously generate rhythmic action potentials[10].

1.3.1 Potential conduction

The signal generated by the pacemaker cells propagates through a long path from its generation to death. Figure 1.1 represents it schematically.

The signal arises from the right atrium, specifically in the sinoatrial node (SA). This signal stimulates the contraction of the atria with consequent blood ejection
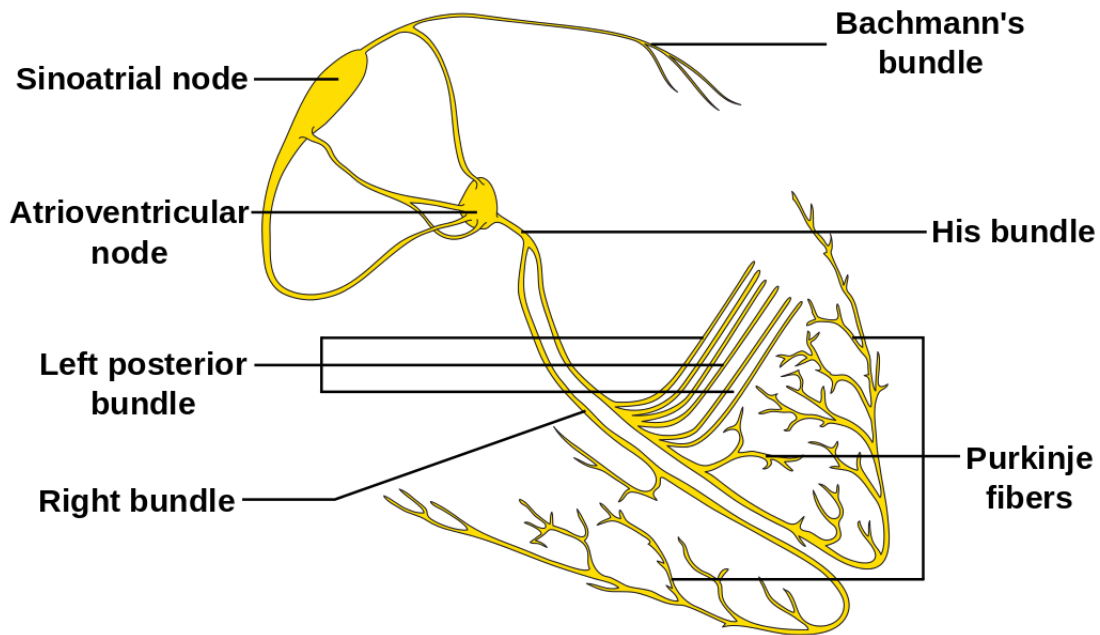
Figure 1.1: Heart potential conduction path [11].

and rapidly moves to the atrioventricular node where it is delayed in time to let the blood fills the ventricula. The signal then propagates to the Purkinje fibers by means of the right and left His bundle, reaching the left and right side of the heart and the endocardium at the apex of the heart as well. The signal ends in the ventricular epicardium. The shape of the signal, during its propagation, is schematized in Figure 1.2[12].

It can happen that the heart sometimes produces beats that differ from the normal ones in shape or duration. Among the different abnormal beats there are ectopic beats. Originating outside of the sinoatrial node, those beats cause disturbances in the cardiac rhythm. Those can be symptoms of pathology but happen also to healthy subjects and they are common in adult people but rare in young. Ectopic beats lead to extra or skipped heartbeats and they are not dangerous, at least in healthy patients [13]. The most common ectopic beats are Premature Ventricular Contraction (PVC) and Premature Atrial Contraction (PAC). An example of PAC is shown in Figure 1.3.

1.3.2 Heart Rate Variability

Heart Rate Variability (HRV) is defined as the variation in time interval between heartbeats. The reference method to estimate HR is the ECG. When using ECG, the beat-to-beat signal is also called "RR signal" because it refers to the interval between two consecutive R peaks (Figures 1.4 and 1.5). The occurrence of an R peak causes blood ejection and is called systole [15]. Excluding ectopic beats, the HR is controlled by the autonomic nervous system (ANS). Therefore, HRV analysis provides information about the ANS' activity. By means of its innervation, ANS can increase or decrease heart's contraction frequency and strength. Some studies state that a greater variability in heart rate means that heart can keep up better
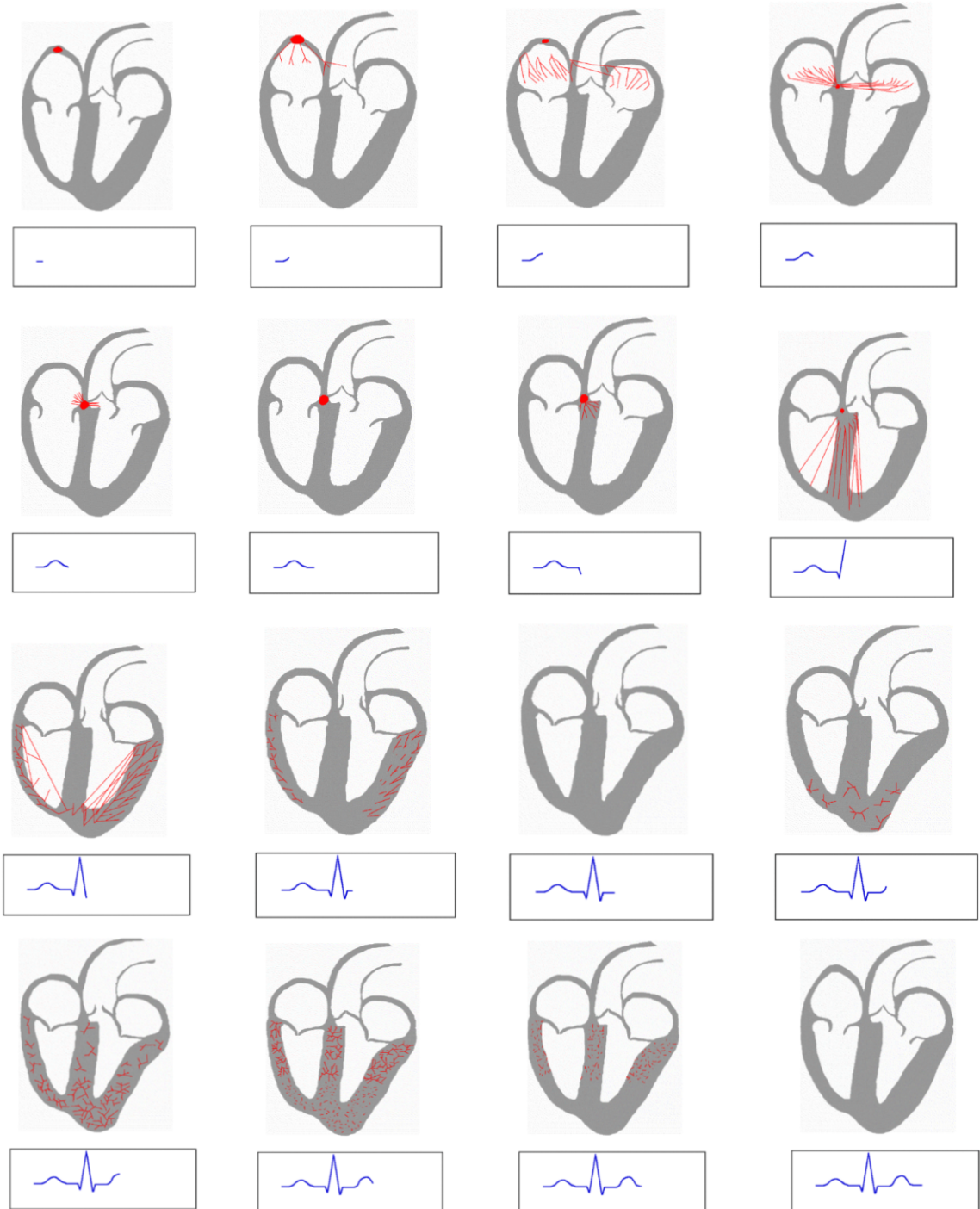
Figure 1.2: Signal shape during propagation [12].
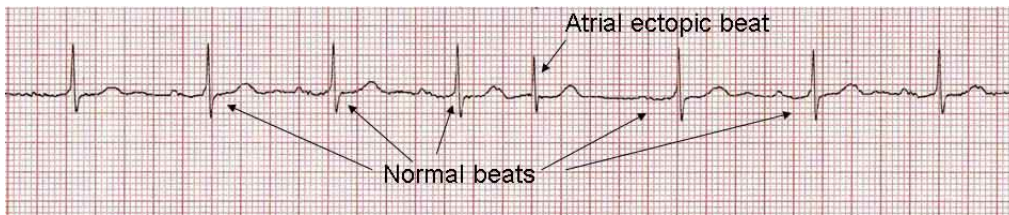
Figure 1.3: Atrial ectopic beat [14].

with changes [16].

Studies on HRV started in 1965 with Hon and Lee that found a more robust and predictable way to diagnose fetal distress in the alterations of the heart beat interval if compared to the heart rate [17]. That study, together with Ewing in the '70s and others later, contributed to the understanding of the correlation between the HRV and the autonomic nervous system. The availability of cheaper, new, digital, robust and powerful ECGs pushed lots of researchers to refine the previous knowledge and lead to important results like the correlation between HRV and mortality after an acute myocardial infarction[18].
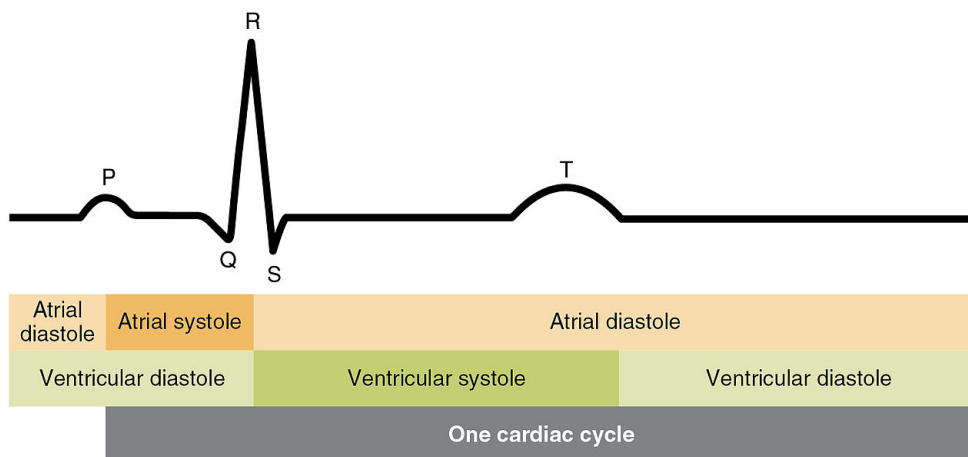
Figure 1.4: Cardiac cycle [19].

Despite the large amount of studies performed, HRV variation meanings are still not totally known yet. The sinoatrial node is responsible for the heart contraction, and the variations in its activity are due the sympathetic and parasympathetic nervous system and the humoral factors[1]. In turn, variations in the activity of the nervous system are given by different factors like thermoregulation, stress, physical activity, meals and, of more interest for this study, the sleep-wake cycle[15].

The ECG signal is composed by a set of different frequencies in the range 0-0.4Hz as in Figure 1.6. The signal is usually divided in three bands: very low frequency (below 0.04 Hz), low frequency (between 0.04Hz and 0.15Hz) and high frequency (between 0.15Hz and 0.4Hz). The origins of the peak in the VLF band are not well known, but scientists think they are linked to the chemoreceptors[2] on the heart. The

---

[1]Factors transported by the circulatory system like hormones and humoral immunity factors.
[2]Sensory receptor that transduces a chemical signal into an action potential.
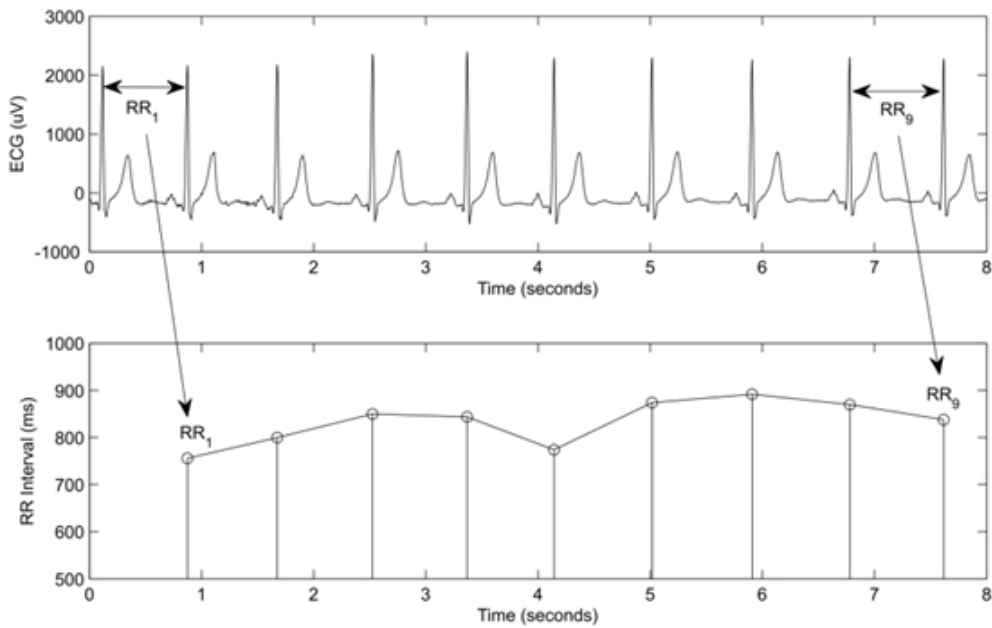
Figure 1.5: RR (HRV) signal [20].

spectral power of the LF band, instead, is related to the activity of the baroreflexes[3] and represents the sympathetic and parasympathetic activity. The last band, HF, is related to the respiration and represents the parasympathetic activity only [21]. The two frequencies of the two branches of the nervous system overlap and this problem becomes more evident at low respiratory rates since parasympathetic activity shifts into lower frequencies and overlaps the frequency interval where the sympathetic region is defined [22]. This sort of overlapping can lead to misleading results and thus it is safer to study the ratio of those frequencies instead of their absolute value. HRV is linked with the sleep-wake cycle and thus it is interesting to find a correlation between the incoming drowsiness and variability in the heart rate.

There are three main methods to analyze the heart rate variability: time domain (linear and non linear like Poincaré), frequency domain and geometric [18]. Time domain methods are the simplest to perform and include the extraction of different features:

- SDNN: standard deviation of the NN intervals (intervals between adjacent QRS complexes resulting from sinus node depolarization)
- Mean NN interval
- Mean heart rate
- Difference between longest and shortest NN interval
- Difference between night and day heart rate
- RMSSD: square root of the mean of the squares of the successive differences between adjacent NNs
- NN50: number of pairs of successive NNs that differ by more than 50 ms
- pNN50: proportion of NN50 divided by total number of NNs
- SD1, SD2: major and minor axis of the ellipse that fits the Poincarè plot

---

[3]Mechanism that helps maintaining blood pressure at a constant level.
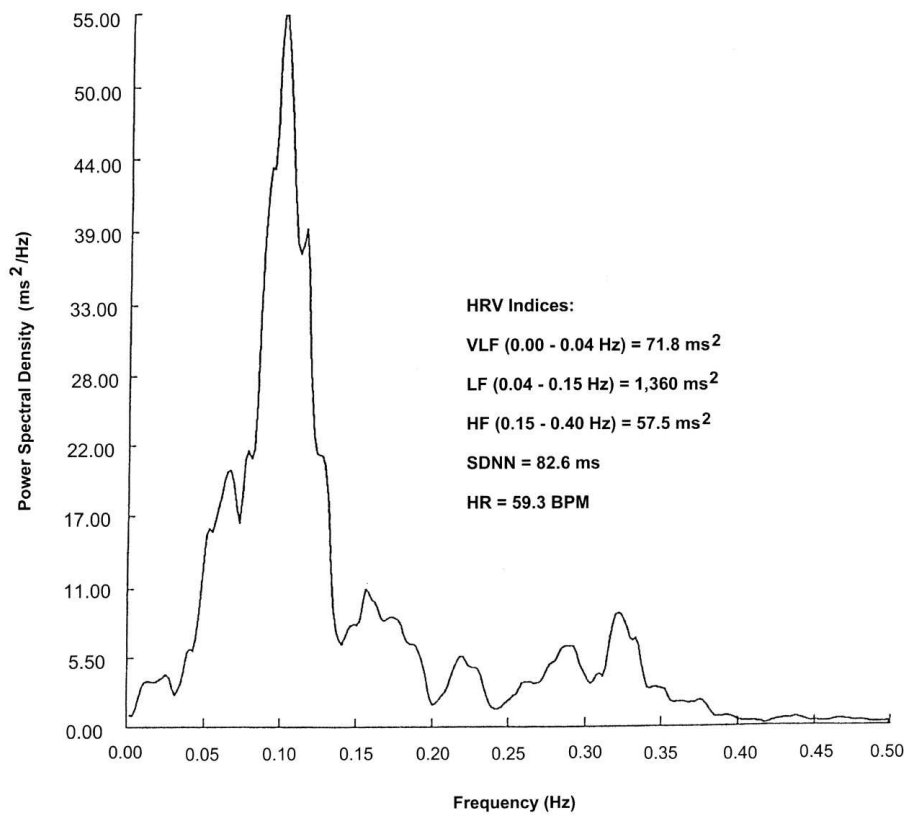
Figure 1.6: Power Spectral Density of an ECG and frequency bands.

- EBC: Estimated Breath Cycle, the range within a moving window of a given time duration within the study period

The HR spectrum is often divided in three frequency bands for frequency domain analysis. This bands are:

- Very low frequency (VLF): from 0.0033 to 0.04Hz. Less interesting of the other bands since its meaning is not clear.
- Low frequency (LF): from 0.04Hz to 0.15Hz
- High frequency (HF): from 0.15Hz to 0.4Hz

Several methods have been proposed for the time to frequency domain transformation. They are generally divided in parametric and nonparametric, each one with pros and cons. Nonparametric methods, like discrete Fourier transform (DFT), are simpler and faster compared to the others. The parametric ones, instead, give smoother results, are easier to post-process and work better with a low number of samples (typical for experimental data). They need to be validated after being implemented. If parametric methods are an application of interest, it is possible to choose between LS periodogram, wavelet transform and autoregressive models. In particular, LS periodogram plays a special role in HRV analysis since, unlike the other methods, it does not require an evenly sampled data set.

Concerning the geometric methods, the series of NN intervals is converted into a geometric pattern and by a formula it is possible to get information about the geometric/graphic properties of the pattern. The most known geometric method is the triangular index that is defined as the integral of the density distribution normalized on its maximum.

1.4 The nervous system

The nervous system is the organ of the body depicted to coordinate voluntary and involuntary actions and transmits signals to the body's organs. Figure 1.7 represents it schematically. In mammals, the nervous system is divided in Central Nervous System (CNS) and Peripheral Nervous System (PNS). The first one consists of two important parts, the brain and the spinal cord, meanwhile the other one consists of the nerves that connect the different parts of the body. The PNS can be divided by the function that it performs and specifically these are:

- Motor neurons: enable the movement
- Autonomic Nervous System, which regulates the body's unconscious actions
- Enteric nervous system: regulates the gastrointestinal function

The autonomic nervous system (ANS) is divided in two groups: sympathetic and parasympathetic, which usually work as antagonists. ANS controls vital parts of the body like the heart and the lungs. With the term controlling it is meant an increase or decrease of performance and functionality, for example, increasing or decreasing heart rate, breath rate or increasing/decreasing contractile force. In particular, the sympathetic one acts in response to stress, causing, for example, an increase in HR by increasing the firing rate of pacemaker cells. The parasympathetic, instead, works as the opposite, decreasing the pacemaker cells' fire rate.

Parasympathetic division   Sympathetic division

Brain

Constricts pupil — Eye — Dilates pupil

Stimulates saliva production — Salivary glands — Inhibits saliva production

Constricts bronchi — Lung — Dilates bronchi

Slows heart — Heart — Accelerates heart

Adrenal gland

Spinal cord — Liver — Stimulates epinephrine and norepi- nephrine release

Stomach — Stimulates glucose release

Stimulates stomach, pancreas, and intestines — Pancreas

Intestines — Inhibits stomach, pancreas, and intestines

Stimulates urination — Bladder — Inhibits urination

Promotes erection of genitals — Genitalia — Promotes ejac- lation and vaginal contractions

Figure 1.7: The Autonomic Nervous System.

1.4.1 Parasympathetic nervous system

The parasympathetic nervous system, represented by the left side of Figure 1.7, is responsible for the "rest-and-digest" or "feed-and-breed" activities and it prevails on the sympathetic fibers for what it concerns salivary and lacrimal glands and erectile tissue [23]. An activation of the PN usually, but not always, means a relaxation or decrease in activity of the function of the organ which it is controlling. It prepares the body for sleeping, resting, digesting and relaxing. Fibers of the PN arise from the CNS and include cranial (facial, oculomotory, vagus), spinal and pelvic nerves. From the vagus nerve, parasympathetic fibers go to the vital organs like heart, lungs, intestine and stomach. It exploits several functions, of which the most known are miosis, contraction of lungs and reduction of their volume, decrease in blood pressure and frequency, and greater enteric mobility. Interesting for this study is that the parasympathetic nervous system decreases the heart rate and therefore modifies its variability. Its function is mediated by ganglia which uses different chemicals. By means of the M2 muscarinic receptors, the PN acts to bring the heart back to normal after the action of the sympathetic nervous system (slowing down the heart rate, reducing contractile force). Exploiting the M3 muscarinic receptors, it can control instead the lungs, causing bronchoconstriction. The PN signals are characterized by frequencies that lie around $0.0 - 0.4$ Hz but concentrated mostly in the range $0.15 - 0.4$Hz [24], that is referred to as high frequency (HF) [23]. Studies [25],[26],[27], show that its activity increases in those situations in which the body is subject to fatigue, e.g. sleepiness.

1.4.2 Sympathetic nervous system

Working as an antagonist of the PN, the SN constitutes the other branch of the nervous system (right side of Figure 1.7). Its primary process is to regulate the "fight-or-flight" response but it is also constantly active at a basic level to maintain homeostasis. It prepares the body to physical activity. The nerves of the sympathetic nervous system arise from the middle of the spinal cord. The functions of the SN are mostly the opposite of the one processed by the PN, but sometimes they cooperate for the same function: everything is up to which receptor each fiber activates/deactivates. For example, it inhibits the stomach, intestine and pancreas instead of stimulating it. The SN dilates pupils, inhibits peristalsis, activates sweat secretion and, more interesting for this study, increases heart rate and contractile force and dilates bronchioles [28]. The activity of the SN is reflected in all the frequencies but it is more concentrated in the LF band, 0.04 and 0.15 Hz [24].

According to literature the parasympathetic nervous system, which is reflected in the high frequency range, overcomes the sympathetic during sleepiness. Thus it's reasonable to think of a decrease in the low frequency power spectrum and also in the ratio between LF and HF power. An increase, instead, in the sympathetic activity (driver awake) means an increase mainly (but not only) in the low frequencies spectrum and a decrease in the parasympathetic activity (which would in turn decrease the high frequency component). Therefore, it is the ratio between LF and HF that represents better the variation.

1.5 Statistical methods

The main statistical method used throughout the thesis is the student T-test of association. A t-test is a statistical hypothesis test. It is usually applied to compare whether two groups (data sets) belong to the same population or not. For instance, one could test some measurements in patients who received a drug or a placebo to see if the cure had any effect. In this study, the association between the different HRV parameters and the KSS will be tested by means of the Student T-Test. The two sample t-test is used but, for a complete understanding, the one-sample t-test will be explained first [29], [30], [31].

1.5.1 One-sample t-test

Outcome of the t-test, is the rejection or not of a null hypothesis, denoted with H_0, under the assumption "Data are normally distributed".

Presupposing that one wants to test if a certain number of n samples with mean $\bar{X}_1$ and standard deviation s originate from a population of mean μ_{H_0}. The null hypothesis is H_0: "True population mean score is μ_{H_0}" and the alternative hypothesis is thus H_a: "True population mean score is not μ_{H_0}". The significance level is usually set to $\alpha = 0.05$: this means that we are sure of this hypothesis at 95%. It is interesting to notice that, the bigger the difference between μ_{H_0} and $\bar{X}_1$ the greater the probability to reject the null hypothesis. To assess if the means are "far enough" in a statistic way, the sample means need be to be standardized and the t-value calculated according to eq. 1.1:

$$t = \frac{\bar{X} - \mu_{H_0}}{\hat{\sigma}_{\bar{X}}} \tag{1.1}$$

where with $\hat{\sigma}_{\bar{X}}$ the sample standard deviation error mean is denoted. Its value is not known, thus it has to be estimated with eq. 1.2:

$$\hat{\sigma}_{\bar{X}} = \frac{\hat{\sigma}}{\sqrt{n}} \tag{1.2}$$

where $\hat{\sigma}$ denotes the sample's standard deviation.

The t-value can thus be calculated and compared with the critical value of t, from the $t - table$. The t-table gives the percentage points of the Student t-distribution (Figure 1.8), which is a function of only the degrees of freedom df, eq. 1.3.

$$df = N_1 + N_2 - 2 \tag{1.3}$$

N_1 and N_2 are the number of samples.

In the table, according to the degrees of freedom df and the significance level α, the confidence interval CI is obtained. If the $t - value$ is included in the CI the null hypothesis is accepted, otherwise it is rejected.
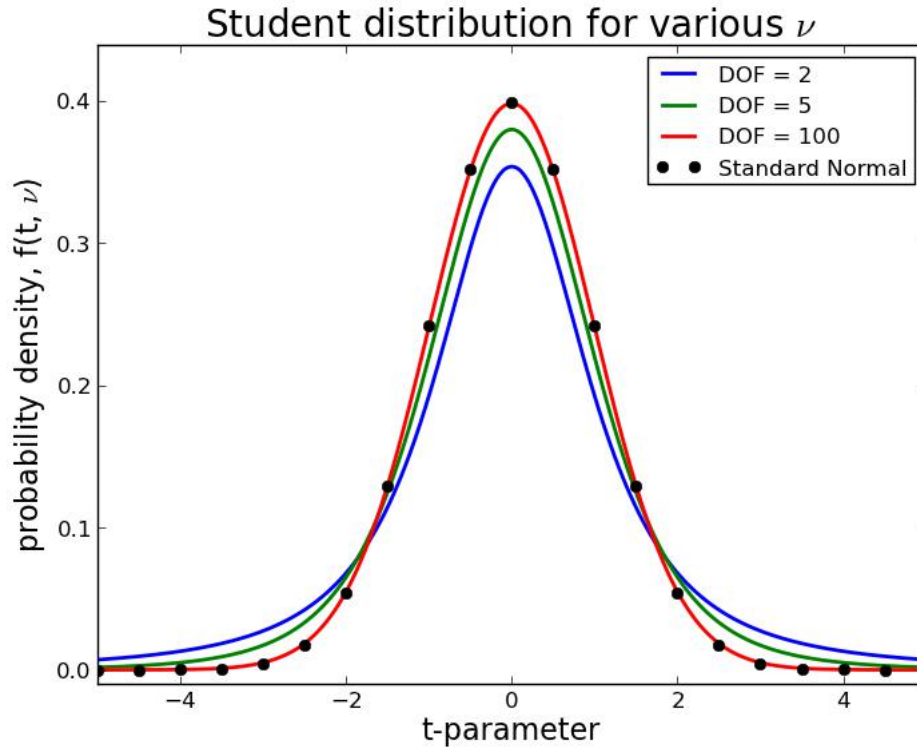
Figure 1.8: Student's distribution as a function of the degrees of freedom *DOF*.

1.5.2 Two-sample t-test

The test works on the difference between the means of the two samples. It is almost the same as the one-sample t-test with the following differences. In this case, the null hypothesis is "The means of the two data sets are equal". The standard deviation error becomes now:

$$\hat{\sigma}_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}} \tag{1.4}$$

Equation 1.4 is only valid with a large number of samples. If that is not the case, variations for this formula exist. The $t - value$ of eq. 1.1 becomes now:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\hat{\sigma}_{\bar{X}_1 - \bar{X}_2}} \tag{1.5}$$

The acceptance or rejection of the null hypothesis is the same as it is for the one-sample t-test.

1.6 Methods' agreement

Bland-Altman (BA) analysis is a process for assessing agreement between measurement methods. The method is also known as Tukey mean-difference plot [32] but popularized in medical statistics by Bland and Altman. The BA process is based on the idea that any two methods used for measuring the same parameter, should

have high correlation when a set of samples are chosen such that the property to be determined varies considerably. The horizontal axis of the plot represents the mean of each pair of measurements and the vertical axis the distance (difference) between them. Thus, each pair is represented on the plot as a point $p = (\frac{x_1+x_2}{2}, x_1 - x_2)$. In a BA plot, data of interest are the bias (distance of the mean of differences from zero), the variance (95% limit of agreement, $\pm 2SD$) and the magnitude of the measures. Bias is a measure of the accuracy: a high bias means a low accuracy. The variance is a measure of the precision: high variance means low precision. Accuracy is the closeness of some measurements to the correct or accepted value of the measured quantity. Precision is the closeness of a set of measurements to each other [33]. The optimal condition is high accuracy and precision together, which is translated in low bias (close to zero) and low variance. Standard deviation and bias only, give an absolute information that is more complete if it refers to the magnitude. For instance, a bias of 2 ± 1 gives rise to very different conclusion if the magnitude is, e.g. 3 or 80.

A Bland-Altman assessment for agreement was used to compare the different methods for the transformation from time to frequency domain and for the two outliers removal methods. Figure 1.9 shows an example of Bland-Altman plot: the ratio between LF and HF power computed using the AR model is compared for the two methods for detecting outliers, pct and std. It is possible to see that the magnitude is close to 1, with a bias that is almost zero with low variance. Thus, the two methods, std and pct, lead to very similar results when dealing with the AR model. The p-value computed with the t-Test gives us an estimate of how reliable those values are.
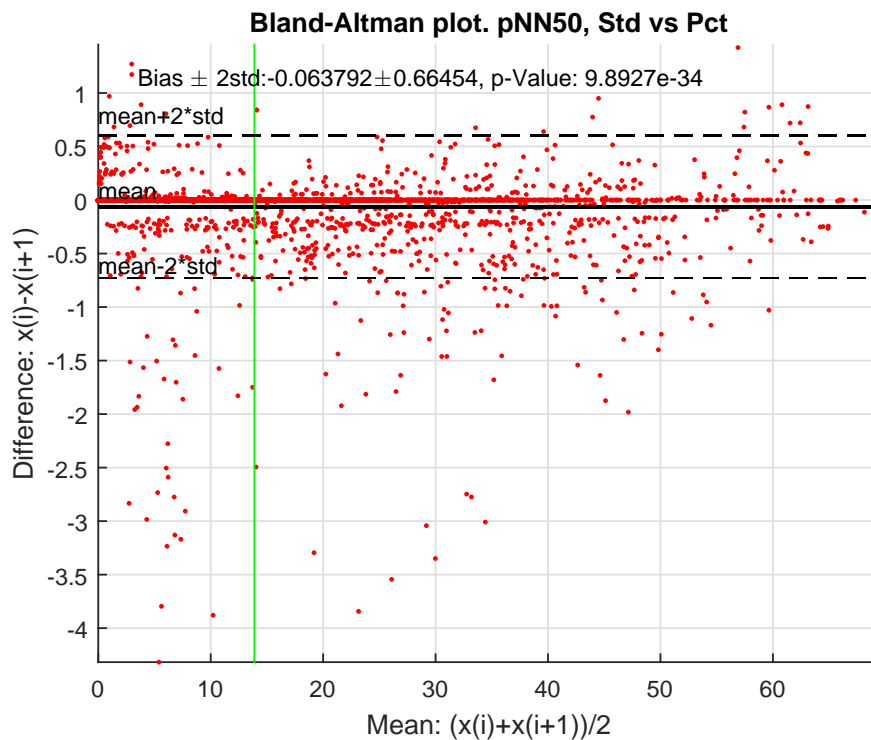

Figure 1.9: Bland-Altman Plot.

1.7 HRV association to Driver Fatigue

Drowsy drivers are potentially dangerous as drunk drivers are. Several studies were conducted in the previous years to find a correlation between drowsiness and heart rate variability for minimizing accident risks while driving. It is valuable to look at the different results they have achieved, the way of conducting the experiments and their limitations. In this way it will be possible to understand which are the best features to extract and which kind of results to expect. Moreover, it will be useful for performing the analysis in the most efficient way and without repeating the same errors. Each study, for how accurate it can be, is biased by some uncontrolled factors or assumptions that are taken. Every study analyzed was conducted in a specific way: some with real driving tasks, other with the help of driving simulators, some during night and others in the morning, for longer or shorter periods of time, with professional or normal drivers. Moreover, they differ also in the measured quantity: some recorded vehicle measures (e.g. steering wheel movements), others used behavioral or subjective measures and, probably the most reliable, physiological measures. Each of those measurement methods has its advantages and limitations as resumed in Table 1.2.

The study of Mahachandra et al. [34] deals with the HRV as an indicator of sleepiness. It was conducted on 16 participants with at least one year of experience at the wheel. They had to drive for 3.5 hours in a simulator representing a closed loop circuit. Their EEG and heart rate were recorded continuously during the drive. The drivers performed the task in normal and physiological conditions (e.g. not awake from 24 hours as in other studies). The promising results presented are in accordance with the literature and in general with all the studies analyzed: a reduction in the performance of HRV due to sleepiness, a higher peak in the high frequency band and a correlation between the decrement in RMSSD (28%) and the parasympathetic activity were found. Also the SD1 of the Poincaré plot from fractal method of HRV showed a decrement of 27% as an indicator of sleepiness. As in other studies ([35] [36][37] [25]), a driving simulator was utilized to conduct the experiment. Simulators are chosen for their cheaper costs, efficiency, experimental control and safety. However, it is not recommendable to perform studies with simulators since results could be biased: the driver does not perceive the same risk and thus has a different behavior. Nevertheless, the newest and advanced simulators (like the motion-based ones that provide a view close to 360° and an extensive moving base) have been proved to have low influence in the behavioral changes of the driver [25].

Patel et al. [26], tried to find the best artificial intelligence based system able to detect fatigue in drivers, analyzing the HRV in a pre-recorded set of 12 ECGs. Using the Tompson algorithm, thresholding to reduce artifacts and transformations of the signal from time to frequency domain, they implemented a neural network with a precision of 90% (tested on only six data sets). The results show that LF/HF can be an indicator of fatigue: it decreased with increasing fatigue and thus a decline in driving control and performance of the driver. Moreover, the study assesses an accuracy of 90% with ANN. The only limitation of this study is the low number of subjects.

Yang et al. [36], instead, developed a dynamic Bayesian network to extract and

Ref.	Measures	Parameters	Advantages	Limitations
[38] [39]	Subjective measures	Questionnaire	Subjective	Not possible in real time
[40] [41]	Vehicle based measures	Deviation from the lane position; Loss of control over the steering wheel movements	Nonintrusive	Unreliable
[42] [43]	Behavioral measures	Yawning; Eye closure and blink; Head pose	Non intrusive; Easy of use	Light conditioning, background
[44] [45]	Physiological measures	Statistical and energy features derived from EEG, EoG, EEG	Reliable; Accurate	Intrusive

Table 1.2: Advanges and limitations of different measurement methods

classify features from an ECG (52 subjects) as LF, HF and their ratio, to recognize drivers fatigue. The ECG is first smoothed with a low-pass filter and then transformed by the FFT and analyzed. The study had a significantly statistic number of subjects and concluded according to the study of Patel.

A study conducted by Michail [46], included 21 subjects driving in real road traffic deprived of sleeping for 24 hours. Each subject drove 1 hour with EEG and ECG connected to him. In eight cases the experiment had to be stopped because the level of sleepiness was too high and thus unsafe to continue. The main parameters taken into account for what it concerns HRV were its LF and HF power and the ratio between them. In particular, they found a decrease in LF/HF during sleepiness between 7 and 42%. Compared to different studies, it has a reasonable number of subjects and it was performed on a real road. On the other hand, the conditions of the subjects can not be considered physiological and thus the results lose their statistical meaning.

The study of Gang [47], seems to show very good results. His implementation of the support vector machine classifier (leave-one-out method) and wavelet analysis for detecting a correlation between HRV and sleepiness had an accuracy of 95%. Conventional LF/HF ratios had an accuracy of only 68%. In particular, they found an interesting correlation between the entropy of the heart rate variability and the incoming sleepiness. However, looking more into detail, those results appear strange and not so meaningful since the correlation was found only for the VLF band: this band of frequencies is not totally explained by the scientists and its origins are mostly unknown (see VLF in section 1.4).

Also other groups of research , e.g. [35], have concluded finding a strict variation in the PSD of the different frequency bands between alert and sleepy state. In particular, Masahiro [37] found an increase in HRV when a decrease in the mental workload was occurring.

2

Introduction

Driving is a complex task that requires to be constantly aware of what it is happening on the road. Drivers are really good at avoiding threats, taking actions when required like braking, steering and changing direction. Sometimes, instead, threats are very difficult to avoid because they do not depend on the driver, e.g. an incoming vehicle from a blind spot, or are unpredictable. Among the last category, it lies drowsiness: one of the causes of accidents and fatalities among the drivers that leads every year to severe injuries and also to death. Numerous studies have demonstrated that a substantial percentage of accidents are caused by drivers falling asleep at the wheel, especially at night. NHTSA (US National Highway Traffic Safety Administration), has estimated 100000 vehicle crashes due to drowsiness in 2012, of which 1550 deaths, 71000 injuries and 12.5 billion $ in monetary loss.

Scientists have studied different ways for trying to predict fatigue, basing their hypothesis on three types of measures: vehicle-based, behavioral and physiological. Among the vehicle-based measures it is possible to find steering and acceleration patterns or deviations from lane position [25]. Behavioral measures, instead, record and analyze visual signs like yawning or eye blinking. Using capacitive foils on the car's roof, it is also possible to measure the position and movements of the head [48]. Physiological measures are obtained, instead, by the analysis of the correlation between physiological signals and driver fatigue. Physiological signals can be electrocardiograms (ECG), electromyograms (EMG), electroencephalograms (EEG) or electrooculograms (EoG) [25].

Subjective parameters to evaluate sleepiness like KSS (see 1.2) were used in various studies leading to KSS values ranging between 7-9 when drowsiness was occurring [3]. It is of interest to study the possibility to detect those situations that are not predictable for the human in an earlier stage [25]. Among the different ways to detect sleepiness at the wheel, physiological measurements are the ones which start changing at an earlier stage what is crucial for accident avoidance. It seems very promising to study how heart rate and its variability relate to drivers' different levels of sleepiness, extracting particular features from the heart rate signal, in the time and the frequency domain.

Most of the previous studies are inconclusive or inconsistent because of the low number of subjects, the inaccurate method of analysis or the incorrect environment (e.g. simulator instead of real road). Conversely, the high quality data and large data set used in this study enables conclusive results. This study is conducted on a population of 80 drivers that drove 3 times (morning, afternoon, night) for 80 to 90 minutes, with relative KSS evaluations every 5 minutes, leading to over 3500 epochs to analyse. When analysing HRV indices some choices need to be

done. The ones having a highest effect on the values of the parameters are the outlier detection method and the spectral transformation methods in the case of the frequency domain. Among the outliers detection methods, two offer better performance. They are the percentage method and the standard deviation method [49]. In the same way for spectral transformation, the most commonly used methods are Fourier Transform (FT), Autoregressive Model (AR-model) and Lomb Scargle (LS) periodogram. As both choices have been reported having a significant influence on the HRV indices values and there is no consensus about their suitability, there is a need of assessment their influence.

The goal of this thesis is to determine dependency of the HRV indices values on the methods utilized for outlier detection and spectral transformation. Agreement between the different methods tested will be evaluated. Moreover how the different methods allow separation of time of the day based on HRV will be tested as indirect indicator of what methods are most suitable for driver fatigue detection.

The thesis begins with a brief background on the most important topics for understanding the thesis and the results. It includes an introduction to driver fatigue and sleepiness and the subjective methods used for its measurement. It continues with the functioning of the heart, signal propagation and how the autonomic nervous system has influence on it. In the end, the statistical method used for understanding which features can be meaningful for the goal are explained. Successively, the methods and materials used for processing the data and extracting the features are presented in details. Results, conclusions and future improvements are presented in the last chapters.

3

Materials and Methods

3.1 Participants, instrumentation and data collection

The data set included 80 drivers selected randomly. Only healthy men and women, sustaining a valid driving license and between 18 and 65 years old were included. Each driver performed three drives in a day: one in the morning around 08:00-09:00, one in the afternoon around 14:00 and one in the night around 21:00-23:00. The drives were lasting 80 to 90 minutes each. During the drive, a Holter recorded the ECG with a sampling frequency of 256 Hz. Every five minutes, the driver assistant recorded the KSS by asking how the driver was feeling. The drivers were asked not to speak or interact with people or devices in the car (radio, cellphone...) except for the experimenter in the right front seat for recording the KSS (refer to 1.2). Visual signs, like yawning, were also recorded but they are not of interest for this study.

3.2 Data pre-processing

Each file provided in the dataset(successively called "drive") is organized as a structure with the following fields:

- GPS data
- CAN BUS data: vehicle speed, yaw rate, acceleration, steering, brake, engine...
- Annotation: KSS, driver actions, drowsy warning, attend warning, drowsy accept
- ECG, Respiration signal (only for some drives), start and end time of the drive
- Time vector for every parameter

Between all the parameters mentioned above, only ECG, KSS and time vector were used. From each drive, the ECG signal and the time were used for the computation of the features. A visual inspection on some random data-sets was done to check the quality of the signal (e.g. noise, artifacts) and its characteristics. The ECGs had an amplitude varying from -0.4 V up to 1.8 V. The frequency of sampling used, 256 Hz, provided a high-enough quality signal. Noise in the signal, like random fluctuations, was present only in the P, Q S, and T waves. Thus it was not affecting the successive R peaks detection. Some periods were blank in the ECG probably because drivers moved, but this did not happen during the actual recording time.

3.2.1 Peak detection and HRV computation

Every drive is analyzed independently from the others. For each of them, the ECG signal is divided in epochs (sweeps) of 5 minutes, corresponding to the occurrences of the KSS in time. Five minutes epochs find confirmation in different studies like [50] and were chosen despite hindering the analysis of VLF because the meaning of this frequency band is unclear [26]. Moreover, the KSS values are registered every 5 minutes, thus they give information on the 5 minutes prior to their recording.

For every sweep, the HRV signal had to be computed. To do so, it is necessary to measure the length of each QRS complex. QRS complexes are difficult to detect due to their shape, their physiological variability and the type of noise that can be present in the signal. Thus the distance in time between two R peaks is detected since it is the same [51]. The R peaks detection is done using the Pan-Tompkins algorithm. An example of peak detection is shown in Figure 3.1.
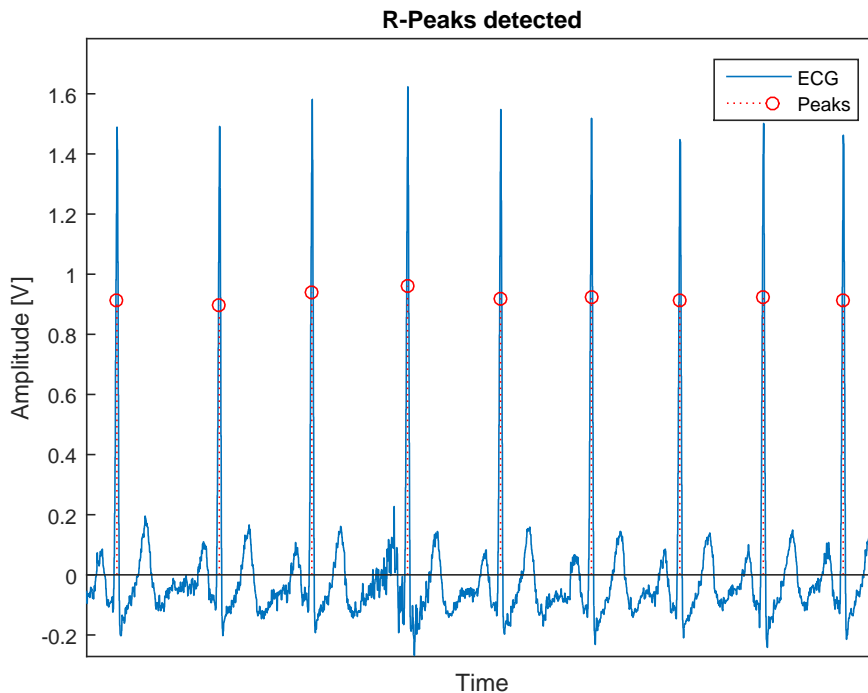
Figure 3.1: R peaks detected in a part of ECG with Pan-Tompkins algorithm.

After obtaining all the R peaks, the time interval between two consecutive peaks is calculated. Each of the resulting points is also related to the time at which the second of the two peaks has occurred. For example, if two peaks occurred at times t_1 and t_2, the resulting point will be $p = (t_2 - t_1, t_2)$. An example of HRV signal is shown in Figure 3.2.

3.2.2 Outliers detection and processing

The HRV signal presented some abnormalities like very high peaks. The correspondent pairs of R-peaks that generated the spikes were thus checked. Those peaks

Figure 3.2: HRV signal in a 5 minutes sweep.

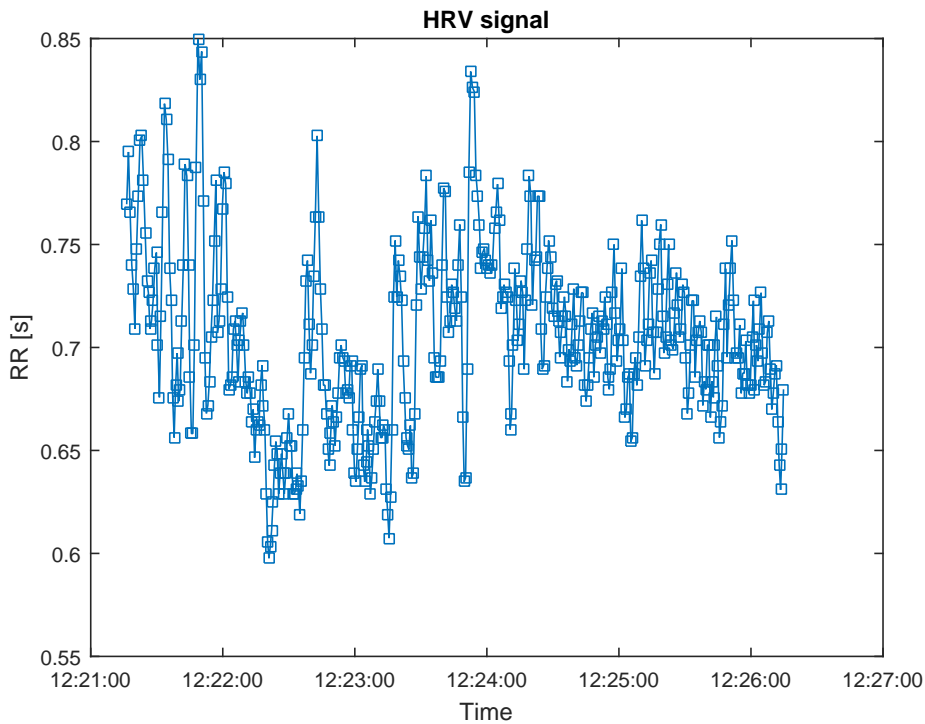were corresponding, in the ECG signal, to delayed, missing or ectopic beats (Ectopic beats: section 1.3). Even though ectopic beats are physiological in nature, they need to be removed since they are not controlled by the ANS. Therefore, an algorithm for detecting and correcting those beats is required.

Various methods have been presented in literature. For this study, two methods were adopted according to study [49]. The first one, successively referred as pctChg, classifies an RR interval as outlier if its value differs more than 30% from the mean of the previous four accepted intervals. With the second method, called stdMethod, all the points whose values differ more than five standard deviations from the mean of all RR intervals, are classified as outliers. Both methods were implemented and compared. Figures 3.3 and 3.4 show an example of outliers detected (red crosses) in a sweep of HRV signal.

As it can be seen in the two figures, the two methods don't give always the same results. Samples that are detected as outliers with the first method are not necessarily marked as outliers by the other. It is also possible that when one method detects some outliers the other detects none. The reason is that pctChg is sensible to the number of points that the peak is made of and the number of previous points used for calculating the mean. If the peak is high but several points go gradually increasing to that peak, the variation in percentage can be less than 30% and thus no outliers are detected. This fact is not present when considering the stdMethod. The outliers detection has been validated checking whether the detected outliers were corresponding to abnormalities or not. Looking closer in the ECG in their correspondence, it is possible to see that some of them actually are ectopic beats.
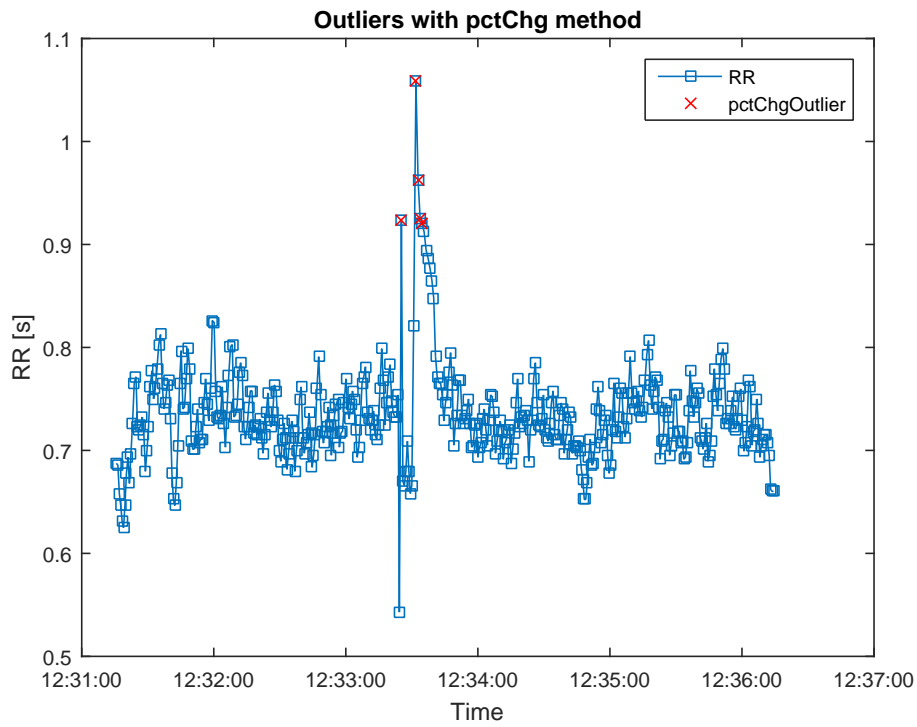
Figure 3.3: Outliers (red crosses) with pctChg method with threshold 30%.

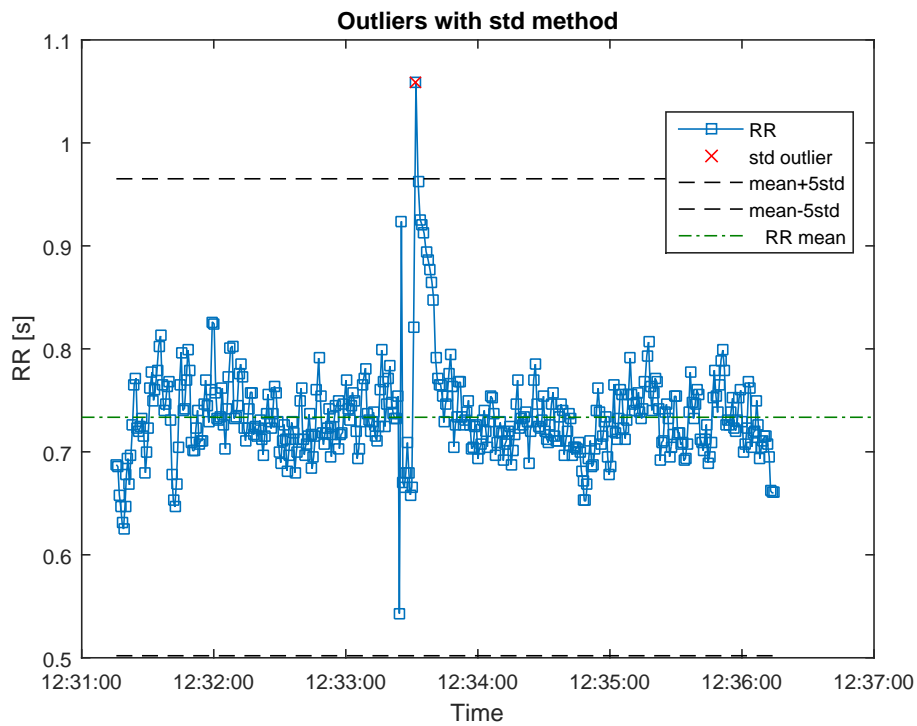**Figure 3.4:** Outliers (red crosses) with stdMethod.

Figure 3.5 clearly displays the detection of an ectopic beat.
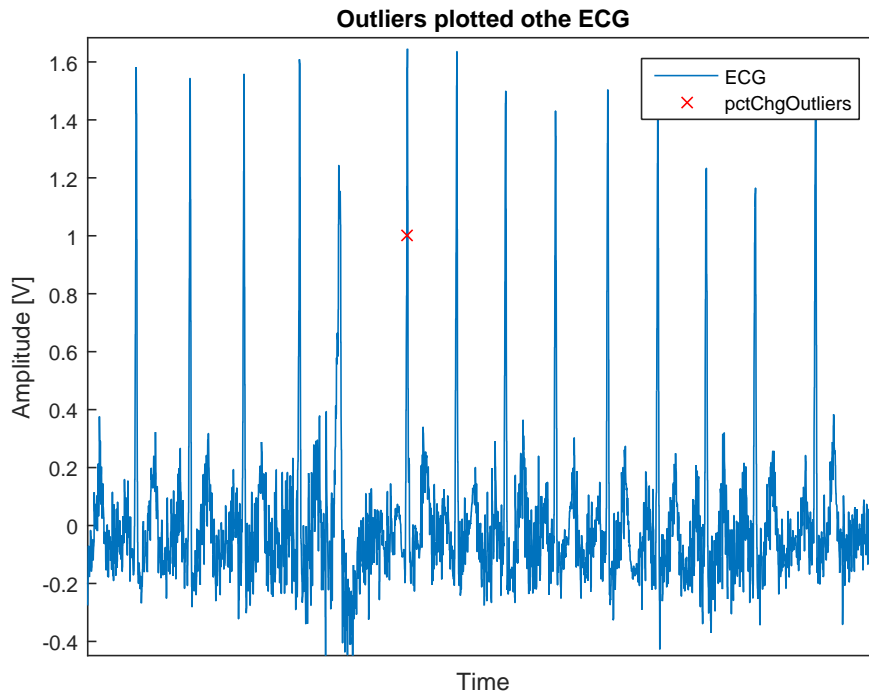

Outliers plotted othe ECG

Figure 3.5: Ectopic beat detected by the pctChg algorithm.

The samples detected and classified as outliers need to be processed. Several methods for processing outliers, for example removal, interpolation or replacement exist. Each technique can be done in different ways. For instance, if one uses interpolation, it can be linear, quadratic or cubic. Removal was recommended over interpolation in [52]. Nevertheless [49] determined than the influence of the outlier removal method is non-significant. It is clear that, regardless of the choice, errors will be introduced, such as increases of the low or high frequency component, false shapes and trends [52] [53]. For this study, simple removal of outliers has been chosen according to the recommendation in [52]. Figure 3.6 shows the HRV signal after the removal of the outliers detected by the pctChg method.

3.3 Data processing

The analysis of the data comes in four different domains: time domain, frequency domain, geometrical and non-linear. In each section, the methods in Table 3.1 are discussed and some examples are given. A complete list of features and their explanation can be found in Table 3.1.

3.3.1 Time domain

The simplest analysis to perform is the one in the time domain since it does not require transformations of the data. The features that have been extracted, according
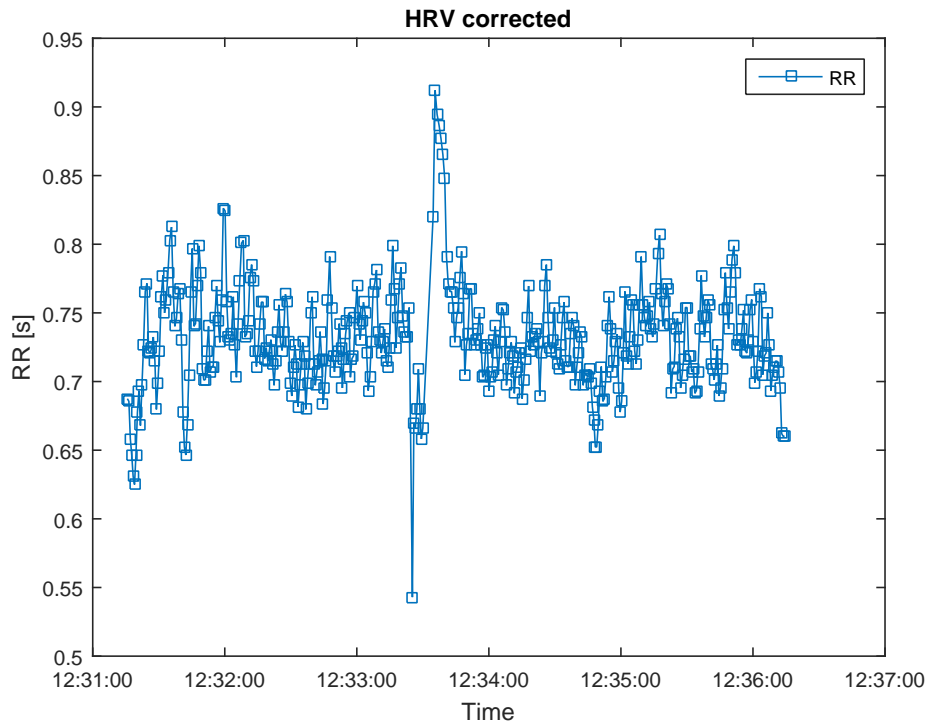
Figure 3.6: HRV signal after removal of pctChg's outliers.

Domain	Feature	Symbol	Physiological meaning
Time	Mean heart rate	mHR	
	Standard deviation of Heart Rate	stdHR	HR variation
	Root mean square of successive differences	RMSSD	Short term component
	Standard deviation of NN intervals	SDNN	Estimate of overall HRV
	Percentage of NN interval that differs more than 50 ms	pNN50%	Short term component
Frequency	Low frequency power	LFP	Simpathetic activity
	High frequency power	HFP	Parasympathetic activity
	Ratio between LFP and HPF	LHRatio	Simpathovagal balance
	Low frequency power normalized	LFP n.u.	
	High frequency power normalized	HFP n.u.	
Geometrical	Triangular index	RRtri	Estimate of overall HRV
	Triangular Interpolation Index	TINN	
Non-Linear	Poincaré	Poin	SD1: Short, SD2: Long term variability, Ratio: randomness in HRV

Table 3.1: Features, symbol and physiological meaning.

to recommendation in [50], are explained here.

Firstly HR and its standard deviation were calculated. Afterwards, the analysis is extended to RMSSD, defined as the square root of the mean of the sum of the squares of differences between adjacent NN intervals. Last parameter is pNN50%, the number of pairs of adjacent NN intervals differing by more than 50 ms in the entire recording divided by the total number of all NN intervals. Three variants are possible: counting all such NN intervals pairs or only pairs in which either the first or the second interval is longer. The first of the three approaches has been chosen.

3.3.2 Frequency domain

Particular attention needs to be dedicated to the analysis in the frequency domain for calculating the power spectrum of the HRV signal computed in the previous steps. First of all, there exists a large number of methods that can be used to perform this analysis. There are several methods for spectral transformation: Fourier Transform (DFT), Autoregressive Model (AR) and Lomb-Scargle periodogram (LS). All the three methods have been implemented to compare their performance.

DFT and AR-model methods need an evenly sampled dataset. Thus, since RR is an unevenly sampled signal it needs to be interpolated and resampled. This produces an additional error. LS can work with unevenly sampled data representing this an advantage [54] [55].

The Fourier spectrum can be calculated in different ways, either by squaring the FFT transformed signal or by computing the DFT of the autocorrelation of the signal. The second approach has been chosen. In this way, the spectrum is sensitive to the order of autocorrelation used. Autocorrelation is estimated using eq. 3.1 (N = number of samples, m = order of autocorrelation).

$$\hat{R}_x(m) = \frac{1}{N-m} \sum_{n=0}^{N-m-1} x(n)x(n+m) \tag{3.1}$$

Since this estimator is non-polarized, it converges to the true value (eq.3.2).

$$E[\hat{R}_x(m)] = R_x(m) \tag{3.2}$$

It is also consistent, thus its variance sets to zero when N goes to infinite, but it increases with m when N is a finite number (eq. 3.3).

$$var[\hat{R}_x(m)] \approx \frac{N}{(N-m)^2}[...] \tag{3.3}$$

A good trade-off between computational efficiency and quality of the spectrum has been obtained with an order $m = 40$. Higher orders give a spectrum made of more points, thus smoother and that can catch better the variations in frequency. On the other hand, it follows more noise and artifacts as well. It is possible to see an example of spectrum in Figure 3.7, where the green and red vertical lines demarcate the VLF, LF and HF bands.

Concerning the autoregressive model, it leads to a smoother spectrum if compared to the DFT method, but since it is a parametric method it depends on the order
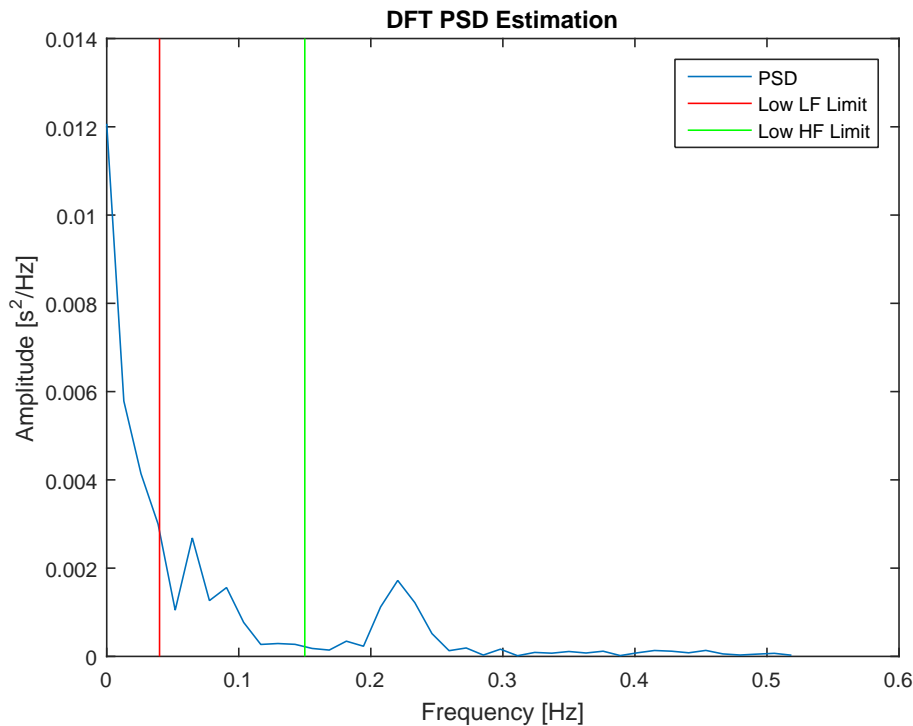
Figure 3.7: DFT: Power Spectral Density. Red and green vertical lines delimit VLF, LF and HF frequency bands.

used. Different orders have been tried. Akaike creterion (AIC), Final prediction Error (FPE) and Parzen index were all giving similar results for the orders between 10 and 20. Thus, after a trial on few drives, order $p = 16$ has been chosen agreeing to recommendation of [56]. Similar consideration made for the DFT about using higher orders are valid. An example of PSD estimation with AR(16) model is shown in Figure 3.8.

The last method used was the Lomb-Scargle periodogram. As already mentioned, it is the most robust of the three to missing or unevenly sampled data. An oversampling factor of 4, as recommended in Matlab's documentation has been chosen [57]. Example of PSD estimation with this method is provided in Figure 3.9.

The three plots are similar in terms of shape and values.

After obtaining the PSD with the three methods, a simple integration in the three frequency bands leads to the features of interest: LFP, HFP, LHRatio, LFn.u. and HFn.u.. The three methods are compared in Figure 3.10 for the power in LFn.u. and HFn.u.

3.4 Geometric and non-linear index

The RR (HRV) signal can be converted to a histogram. Exploiting the geometrical properties of the shape of the histogram, it is possible to extract the so-called geometric indexes. Geometrical methods are robust to the analytic quality of the series of NN intervals (low sensitivity to outliers) but they need a reasonable number of
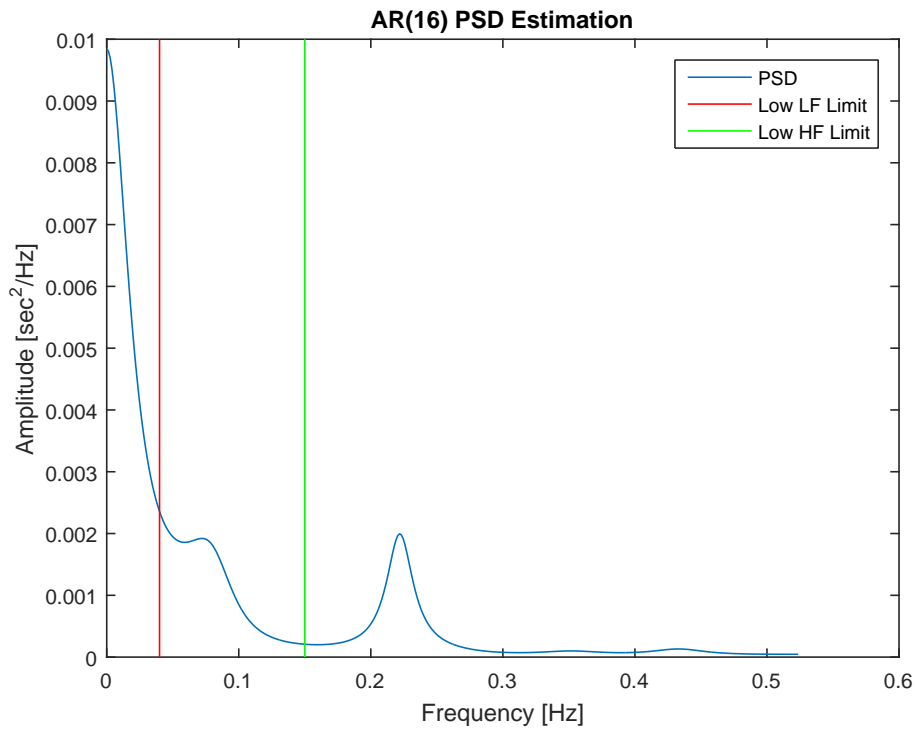
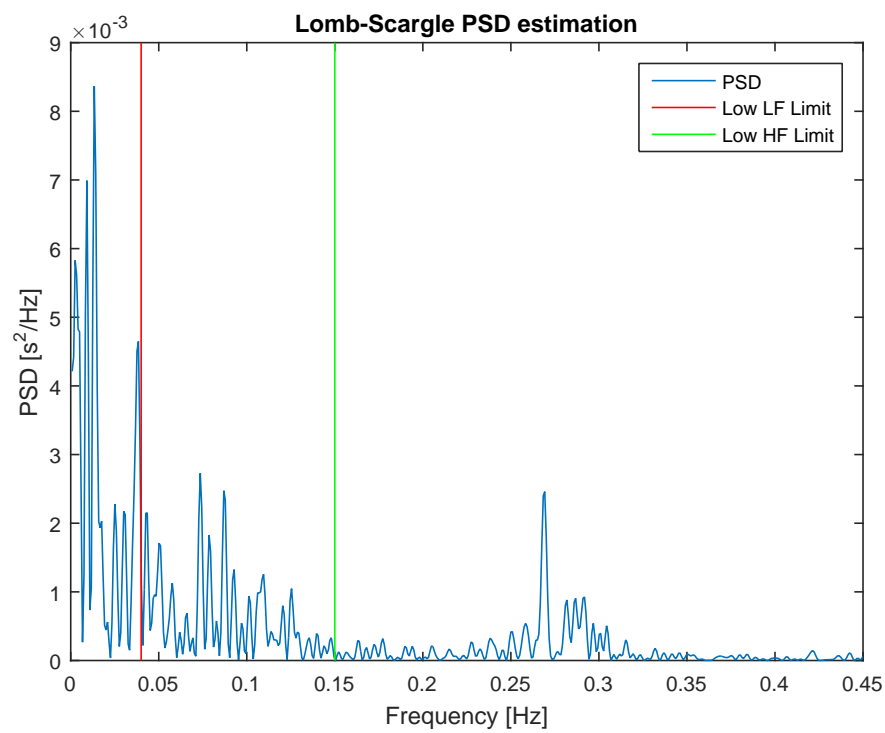**Figure 3.8:** AR: Power Spectral Density. Red and green vertical lines delimit VLF, LF and HF frequency bands.

Figure 3.9: Lomb-Scargle: Power Spectral Density. Red and green vertical lines delimit VLF, LF and HF frequency bands.
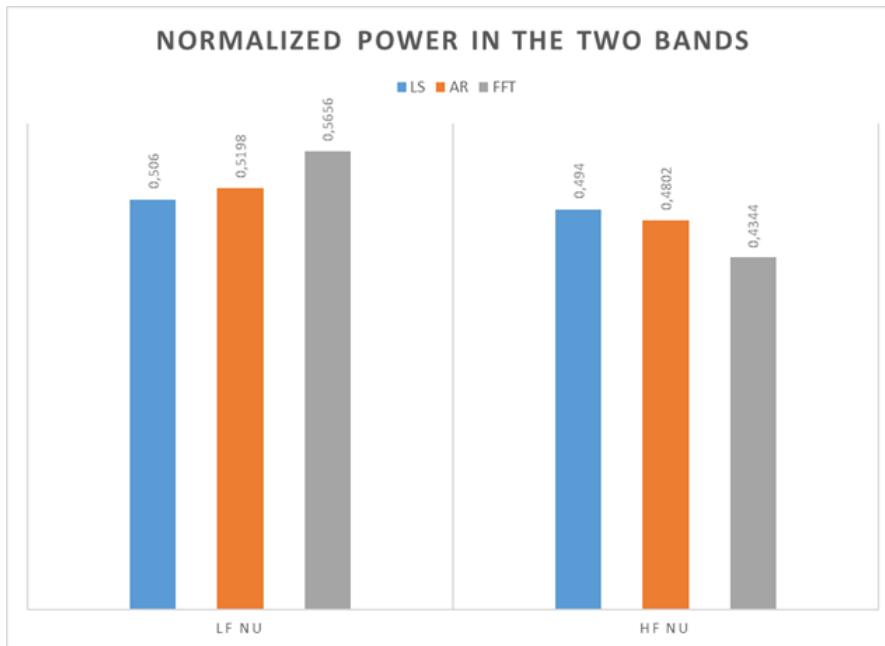
Figure 3.10: Examples of power (n.u.) in LF and HF band for the three methods

NN intervals to construct the geometric pattern [18].

To study the variations in the short and long term, the triangular index (RRTri) and the Triangular interpolation index (TINN) have been calculated. RRTri corresponds to the ratio between the number of RR intervals and the maximum height of the histogram and gives an estimation of the overall HRV. A bin width corresponding to $1/f_{sampling}$ was used [18]. TINN, instead, is the distribution baseline width measured as a triangular base approximating the distribution of all RR intervals. To determine the triangle that best approximates the histogram, the LS method was used [58]. An example of Triangular Interpolation of NN interval is shown in Figure 3.11. When the triangle that best fits the histogram is found, its base is returned as the feature.

Concerning the logarithmic index, what it is calculated is the best value of the parameter k for the curve $y = Ae^{-k*d}$ that fits the histogram of the absolute difference between successive intervals.

The only non-linear method considered is the Poincarè plot. Non-linear features are interesting since the HRV regulation is probably non-linear [58]. The Poincaré plot is the plot of the RR_n interval against the RR_{n+1} one. All the points of the plot which lie on the line $y = x$ are representing no variations between an RR interval and the consecutive. This because they have identical coordinates and thus there are no variations in the HR: a point would be $p = (RR_n, RR_{n+1}) = (RR_n, RR_n)$. Points that lie over this line represent an increase in the HR since RR_{n+1} is greater than RR_n, points under this line represent, similarly, a decrease in the HR. Statistically, in the Poincaré plot for a healthy person, each point over the line of identity would find a correspondent point under it [59].

From the distributions of the points on the plot, it is possible to fit an ellipse from which it is possible to obtain the parameters SD1, SD2, and their ratio. The center of the ellipse corresponds to the average of the RR intervals. SD1 is calculated as
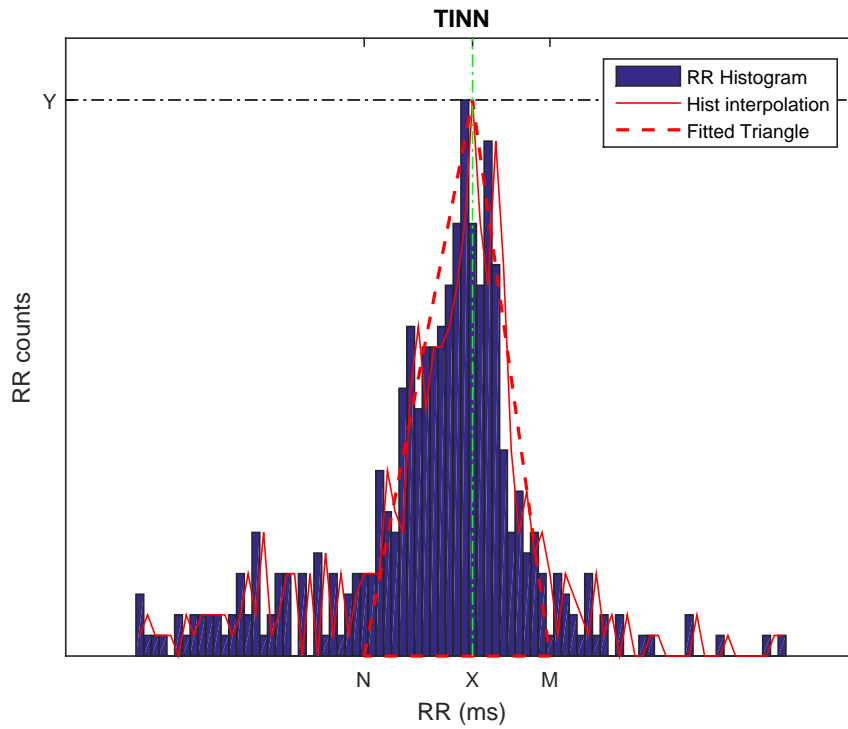
Figure 3.11: TINN: Triangular Interpolation of NN intervals.
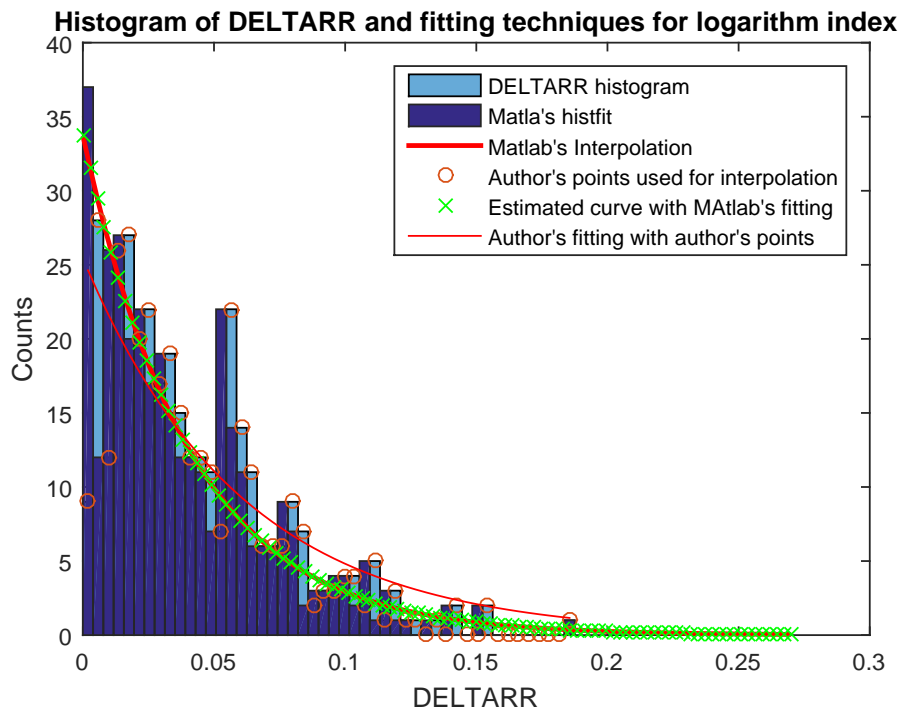

Figure 3.12: Logarithmic index.

the distance of the points to the diagonal $y = x$, SD2 the standard deviation of the distance of the points to the line $y = -x + mean(RR)$. SD1 and SD2 represent the short and long term variability respectively. Their ratio is the ratio between the short and long term variation of RR intervals [58]. In Figure 3.13 it is possible to see an example of a Poincaré plot. Since there were multiple occurrences of some pairs (RR_n, RR_{n+1}), the color is used to represent the number of occurrences.
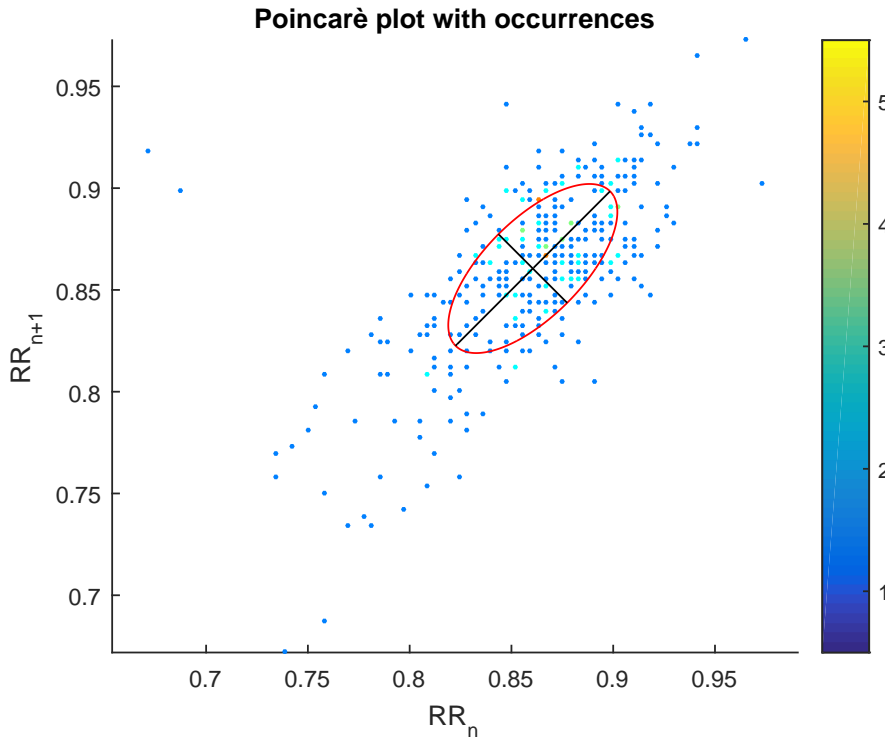

Figure 3.13: Poincaré plot. Length of the two semi-axis of the fitted ellipse are the features of interests.

A scheme of the algorithm is presented in Figure 3.14.

3.5 Statistical tests

Once all the features are collected for all the drivers and all the drives, a t-Test is performed to find a correlation between the HRV features and driver's fatigue. The Student t-Test is used to test whether the means of two populations are statistically significant or not. Initially, the project was set to obtain this result using the KSS values recorded during the drive. Since those values where not provided, it was decided to substitute the KSS with the time at which the drive was performed. Thus, the test is done dividing the features in three sets: the first one corresponds to the morning, the second one to the afternoon, the third one to the night.
T-Test and Bland-Altman analysis (see section 1.6) are computed also for comparing the agreement of two methods. When it is applied to agreement between two methods, the t-Test tells whether the bias is statistically significant or not. For that it also needs to be paired. A first analysis has been done for the outliers detection
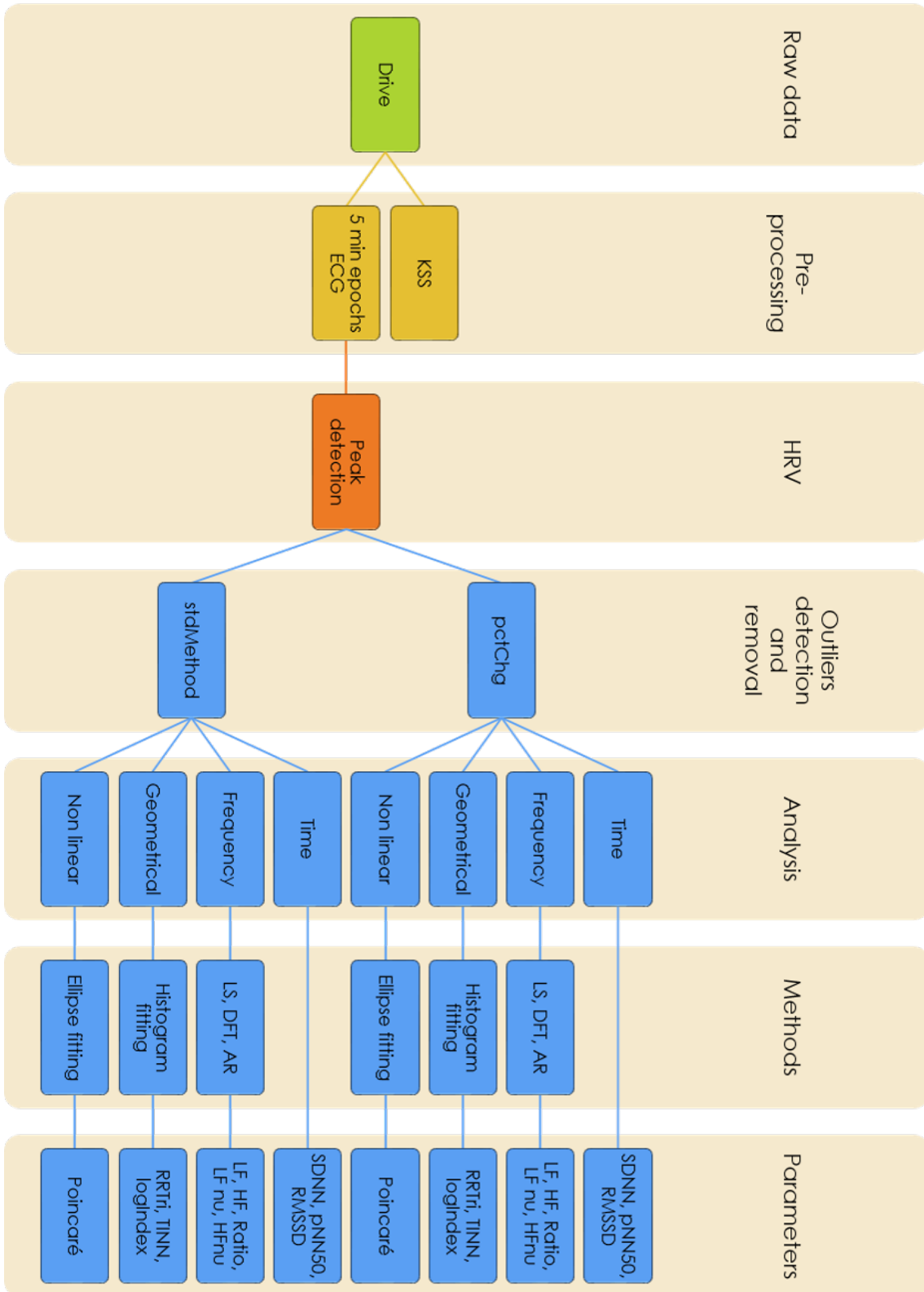
Figure 3.14: Flowchart for computing the features.

methods pctChg and stdMethod for a given feature. All the features were selected. For HR, SDNN, pNN50%, TINN, SD1, SD2, Ratio the BA plot is shown, meanwhile the remaining are summarized in Figures 4.1, 4.3, 4.4 and 4.2. A second t-Test and BA analysis were computed for comparing the different methods for the frequency domain analysis: AR vs FFT, AR vs LS and FFT vs LS, for the stdMethod. Same for pctChg. If two drives (driveX with std and driveX with pct) have different lengths in term of sweeps, the full drive is discarded. Thus we compare drives with the same length. This, in most cases, means that no epochs were removed. It can also mean than an epoch was discarded by both methods. As a last possibility, it may happen that in one drive one method removed one epoch and the other a different one leading to an error. However, since the percentage of outliers was generally much smaller than 5%, the risk of doing this error is minimum.

Figure 3.15: Flowchart for the statistical information

4

Results

In this chapter, the results achieved will be presented structured in two sections by means of tables and plots. The first one shows ho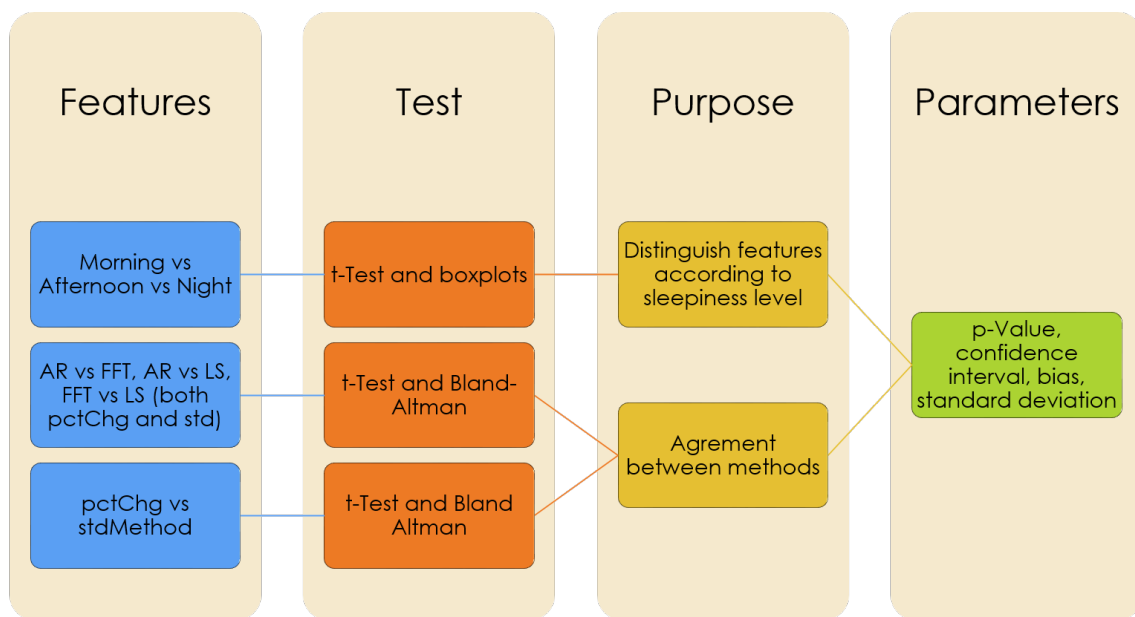w choices and assumptions in HRV processing affect the results. The second part deals with the correlation between the HRV features extracted and driver fatigue.

4.1 Influence of choices on the results: agreement between methods

In this section, the consequences of choosing an outliers detection method rather than another are evaluated. Successively, the differences in the values of the resulting features in the frequency domain due to the adopted spectral transformation will be presented. The tables and plots are intended to assess the agreement between different ways of obtaining the same results.

4.1.1 Outliers detection methods

The first results presented concern the outliers detection methods. The number of samples classified as outliers and removed is, in average, 0.278% for the percentage methods and 0.0648% for the standard deviation method. The two methods have discarded, respectively, 21 and 2 sweeps in total, due the large number of outliers present (greater than 5%). In addition, it is interesting to see if the choice of using one method or another affects the results and how: the two methods can remove different outliers and also in different numbers. Thus, t-Test and Bland-Altman plots between two sets of features processed with both outliers detection methods are shown. The features selected are Heart Rate, SDNN, pNN50%, TINN, SD1 and Ratio (evaluated for the three methods it was computed). In each plot, bias and variance are taken into account with respect to the magnitude of the values. All the HRV indices for the Bland-Altman plot are summarized at bias and confidence interval. For an explanation of the Bland-Altman plot, refer to section 1.6.

It is possible to see, in Figure 4.1 and 4.1a, the influence on the results due to the different choice for the outliers detection methods, for different features. Each point in the plot has two dimensions: one is the mean between the value given by the two methods for a specific feature, the other is their difference. For example, if HR calculated choosing the pctChg method assumes a values of 70 and with stdMethod assumes value 72, the point on the BA plot will be $p = (-2, 71)$. In each plot is shown the straight line represents the systematic error (bias), which asses the mean

(a) PctChg vs Std: Ratio AR

(b) PctChg vs Std: Ratio LS

(c) PctChg vs Std: Ratio FFT

(d) PctChg vs Std: pNN50%

(e) PctChg vs Std: SDNN
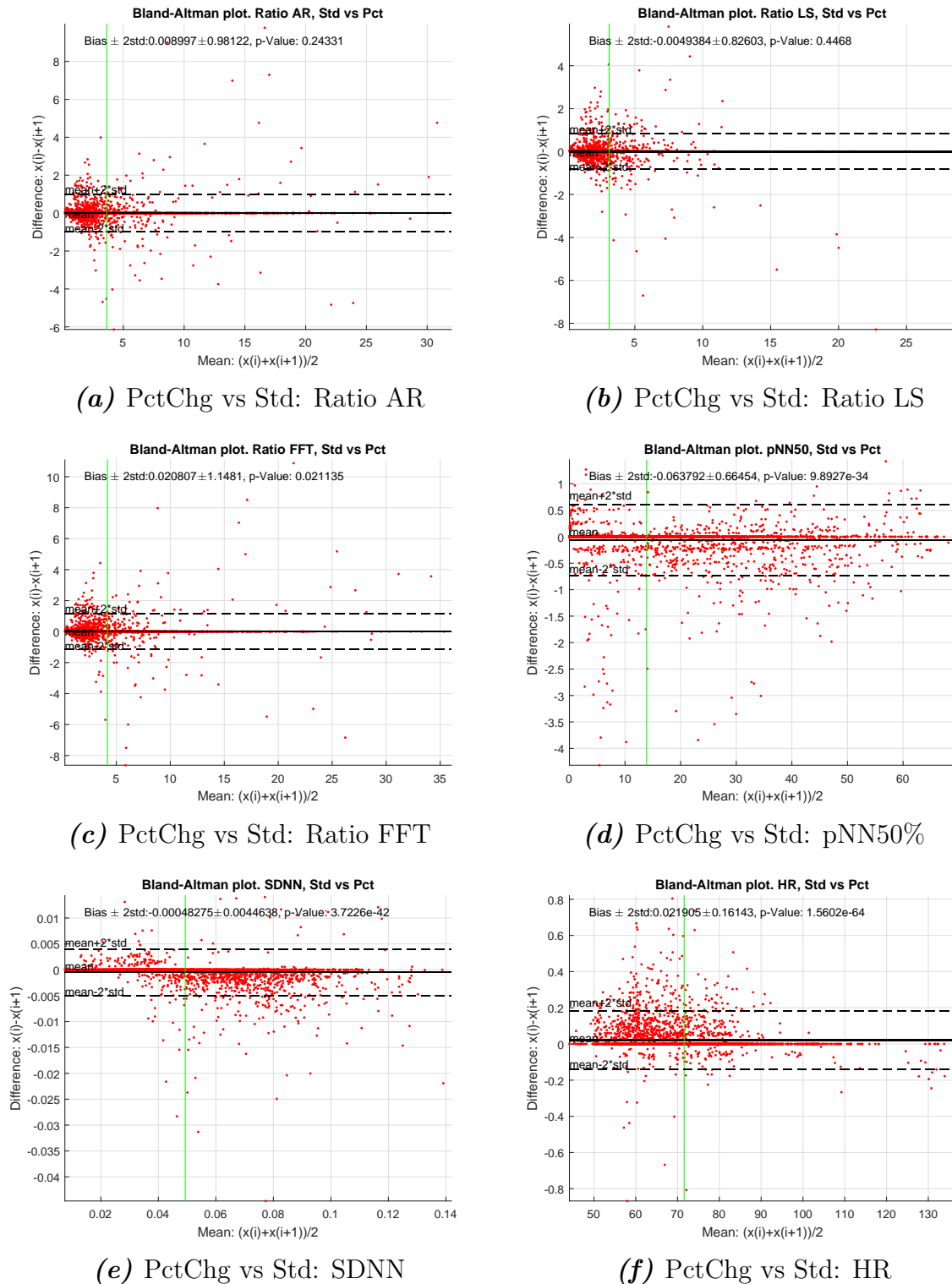
(f) PctChg vs Std: HR

Figure 4.1: Influence on results due to the chosen outlier detection method.

difference between the two sets of values. The confidence interval CI (dashed line), also referred as random error, corresponds at two stand deviations. Thus, 96% of the values are, in principle, contained between the two dashed lines. Bias and CI
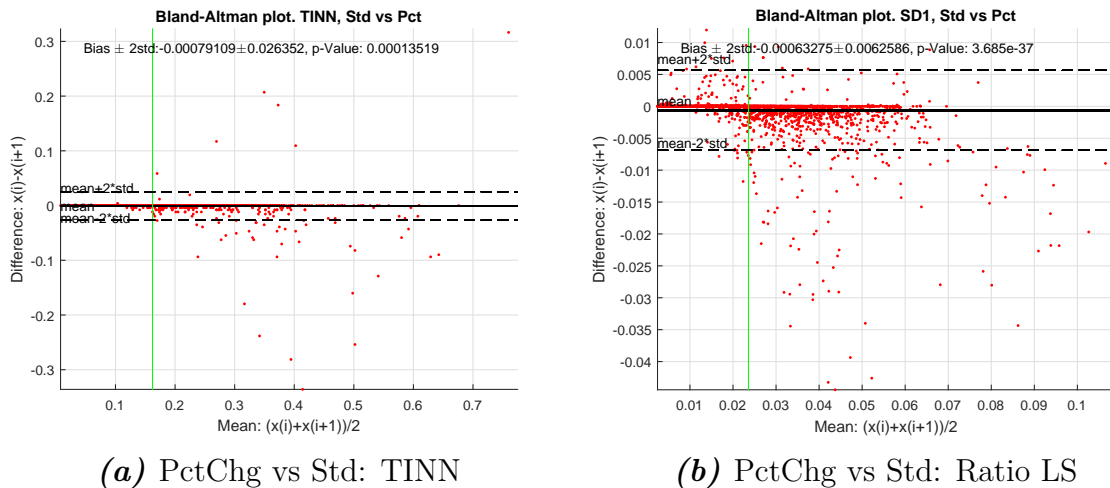
(a) PctChg vs Std: TINN

(b) PctChg vs Std: Ratio LS

Figure 4.2: Influence on results due to the chosen outlier detection method.

have to be compared to the mean magnitude of the measure, which correspond to the vertical line.

4.1.2 Spectral transformation methods

Even though the three time-to-frequency domain transformation methods are in principle equivalent, they give slightly different results due to errors, non-verified hypothesis and assumptions. The differences in the resulting features computed with those methods are thus evaluated. The comparison is done for the feature Ratio, in pairs for all the methods: AR and FFT, AR and LS, FFT and LS. The plots are shown in Figure 4.3.

4.2 Association between HRV and driver fatigue

Finding the difference in the values of the HRV indices for morning, afternoon and night was considered as an indirect indicator of the capability of a feature to discriminate between a driver suffering fatigue or not. This way, if the indices obtained using a certain outlier detection or spectral transformation method would show higher differences than others, it may be more suitable for our application.

The following box plots show how the features are distributed between morning, afternoon, and night. Each box plot shows the median, the first and third quartile, the minimum and the maximum value. The notches around each median give a rough idea on how significantly different the medians are: if the notches do not overlap, the medians are different at the 5% significance level. If a point is larger than the 25th and 75th percentiles, it is marked as outlier and represented by a red cross. The interval corresponds to approximately $\pm 2.7\sigma$ and 99.3% coverage when the data are normally distributed. The plotted whisker extends to the adjacent value, which is the most extreme data value that is not an outlier [60].

Due to their large number, not all the features are shown in the plots. However, their same parameters visible in the plots above, are summarized in Table 4.1-4.4.

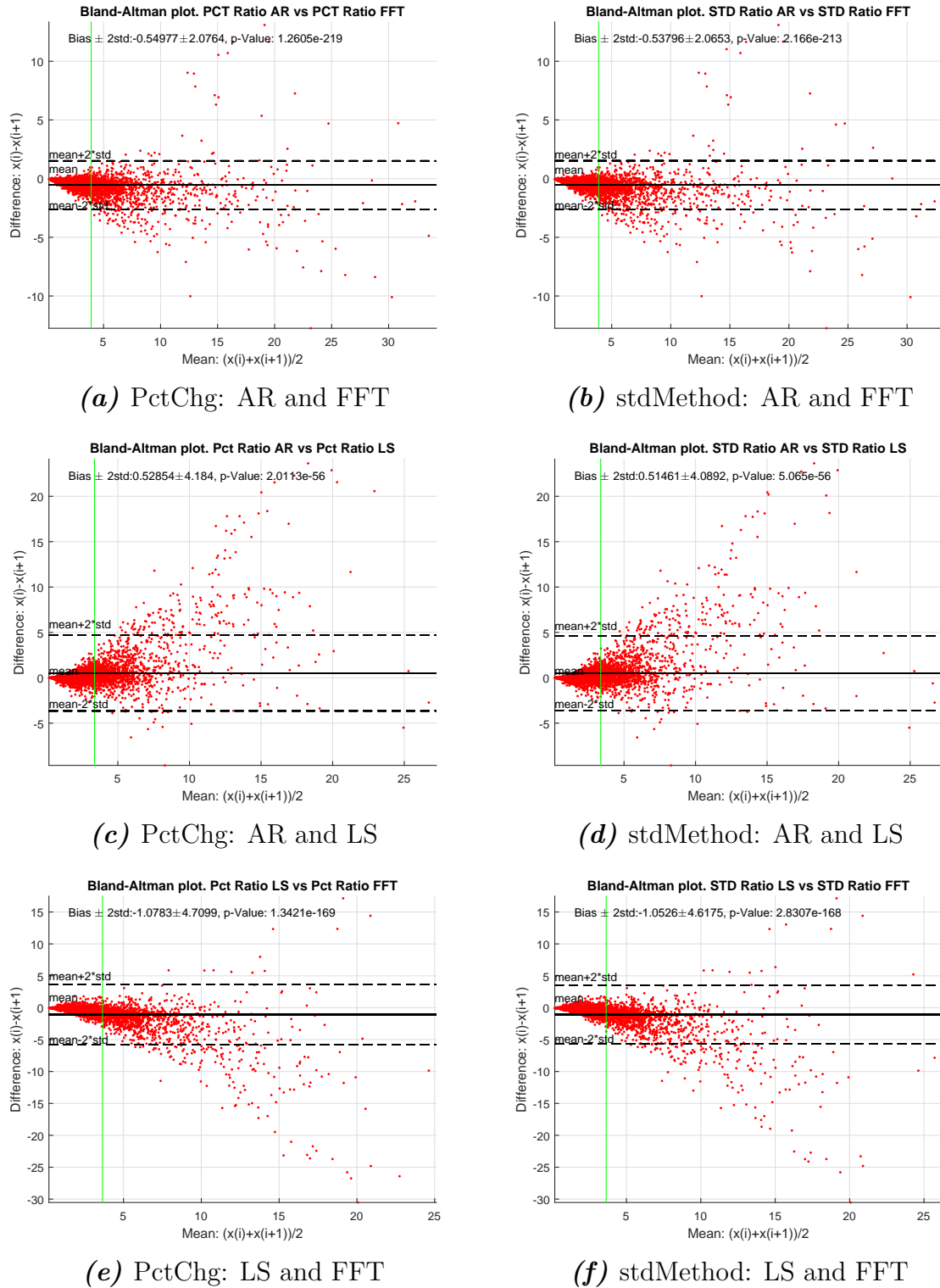*(a)* PctChg: AR and FFT

(b) stdMethod: AR and FFT

(c) PctChg: AR and LS

(d) stdMethod: AR and LS

(e) PctChg: LS and FFT

(f) stdMethod: LS and FFT

Figure 4.3: Influence on results due to the chosen spectral transformation method.

	Feature		t-test	Bland-Altman
			p-value	Bias, 2std, magnitude
AR	HF	Std vs Pct	2,23E-33	2,86E-05 3,00E-04 4,72E-04
	HFnu	Std vs Pct	5,39E-02	1,04E-03 6,90E-02 3,01E-01
	LF	Std vs Pct	1,06E-63	4,78E-05 3,55E-04 1,00E-03
	LFnu	Std vs Pct	5,39E-02	-1,04E-03 6,90E-02 6,99E-01
	Ratio	Std vs Pct	2,43E-01	9,00E-03 9,81E-01 3,63E+00
FFT	HF	Std vs Pct	5,39E-24	2,52E-05 3,16E-04 5,02E-04
	HFnu	Std vs Pct	1,52E-04	1,96E-03 6,57E-02 2,70E-01
	LF	Std vs Pct	3,51E-47	4,02E-05 3,50E-04 1,24E-03
	LFnu	Std vs Pct	1,52E-04	-1,96E-03 6,57E-02 7,30E-01
	Ratio	Std vs Pct	2,11E-02	2,08E-02 1,15E+00 4,18E+00
LS	HF	Std vs Pct	9,44E-02	8,95E-04 6,81E-02 3,08E-01
	HFnu	Std vs Pct	2,45E-52	4,37E-05 3,60E-04 1,02E-03
	LF	Std vs Pct	9,44E-02	-8,95E-04 6,81E-02 6,92E-01
	LFnu	Std vs Pct	3,74E-37	-8,94E-04 8,84E-03 3,33E-02
	Ratio	Std vs Pct	4,47E-01	-4,94E-03 8,26E-01 3,11E+00

Table 4.1: Table of the p-values, bias and CI for the frequency domain features.

	Feature	t-Test	Bland Altman
		p-value	Bias, 2std, magnitude
HR	Std vs Pct	1,56E-64	2,19E-02 1,61E-01 7,16E+01
logIndex	Std vs Pct	5,61E-16	2,18E-05 3,42E-04 4,77E-04
pNN50	Std vs Pct	7,17E-69	-5,22E-02 3,71E-01 1,93E+01
RMSSD	Std vs Pct	3,69E-37	-6,33E-04 6,26E-03 2,36E-02
RRTri	Std vs Pct	9,15E-30	-4,20E-04 4,68E-03 6,52E-02
SD1	Std vs Pct	3,72E-42	-4,83E-04 4,46E-03 4,93E-02
SD2	Std vs Pct	1,35E-04	-7,91E-04 2,64E-02 1,62E-01
SDNN	Std vs Pct	3,10E-04	-2,11E-02 7,42E-01 3,24E+01
TINN	Std vs Pct	9,89E-34	-6,38E-02 6,65E-01 1,39E+01

Table 4.2: Table of the p-values, bias and CI for the time, geometrical and non-linear domain features.

	Feature		t-Test	Bland Altman
			p-value	Bias, 2std, magnitude
HF	Pct	AR vs FFT	5,71E-125	-3,16E-05 1,63E-04 4,74E-04
	Std	AR vs FFT	3,58E-202	-2,82E-05 1,12E-04 5,01E-04
	Pct	AR vs LS	2,09E-05	-8,21E-06 2,45E-04 4,62E-04
	Std	AR vs LS	8,68E-05	-1,42E-06 2,38E-04 4,87E-04
	Pct	LS vs FFT	1,60E-40	-2,34E-05 2,21E-04 4,78E-04
	Std	LS vs FFT	2,95E-38	-2,68E-05 2,61E-04 5,01E-04
	Pct	AR vs FFT	2,43E-01	3,09E-02 5,23E-02 2,85E-01
	Std	AR vs FFT	2,45E-52	3,00E-02 5,19E-02 2,86E-01
	Pct	AR vs LS	4,08E-23	-7,08E-03 9,05E-02 3,04E-01
	Std	AR vs LS	4,14E-22	-6,93E-03 9,07E-02 3,05E-01
	Pct	LS vs FFT	4,53E-05	3,80E-02 7,93E-02 2,88E-01
	Std	LS vs FFT	2,35E-05	3,69E-02 8,13E-02 2,90E-01

Table 4.3: Table of the p-values, bias and CI for the frequency domain features.

	Feature		t-Test	Bland Altman
			p-value	Bias, 2std, magnitude
	Pct	AR vs FFT	9,87E-05	-2,40E-04 5,69E-04 1,10E-03
	Std	AR vs FFt	4,53E-05	-2,32E-04 5,44E-04 1,14E-03
LF	Pct	AR vs LS	1,32E-23	-2,43E-05 3,06E-04 9,91E-04
	Std	AR vs LS	3,67E-18	-2,01E-05 2,94E-04 1,04E-03
	Pct	LS vs FFT	0,00E+00	-2,15E-04 5,03E-04 1,11E-03
	Std	LS vs FFT	0,00E+00	-2,12E-04 5,15E-04 1,15E-03
	Pct	Ar vs FFT	0,00E+00	-3,09E-02 5,23E-02 7,15E-01
	Std	AR vs FFt	0,00E+00	-3,00E-02 5,19E-02 7,14E-01
LFnu	Pct	AR vs LS	4,08E-23	7,08E-03 9,05E-02 6,96E-01
	Std	AR vs LS	4,14E-22	6,93E-03 9,07E-02 6,95E-01
	Pct	LS vs FFT	0,00E+00	-3,80E-02 7,93E-02 7,12E-01
	Std	LS vs FFT	0,00E+00	-3,69E-02 8,13E-02 7,10E-01
	Pct	AR vs FFT	1,26E-219	-5,50E-01 2,08E+00 3,91E+00
	Std	AR vs FFT	2,17E-213	-5,38E-01 2,07E+00 3,90E+00
	Pct	AR vs LS	2,01E-56	5,29E-01 4,18E+00 3,37E+00
	Std	AR vs LS	5,07E-56	5,15E-01 4,09E+00 3,37E+00
	Pct	LS vs FFT	1,34E-169	-1,08E+00 4,71E+00 3,65E+00
	Std	LS vs FFT	2,83E-168	-1,05E+00 4,62E+00 3,64E+00

Table 4.4: Table of the p-values, bias and CI for the frequency domain features.

(a) HR, pctChg method

(b) HR, stdMethod

(c) SDNN, pctChg method

(d) SDNN, stdMethod

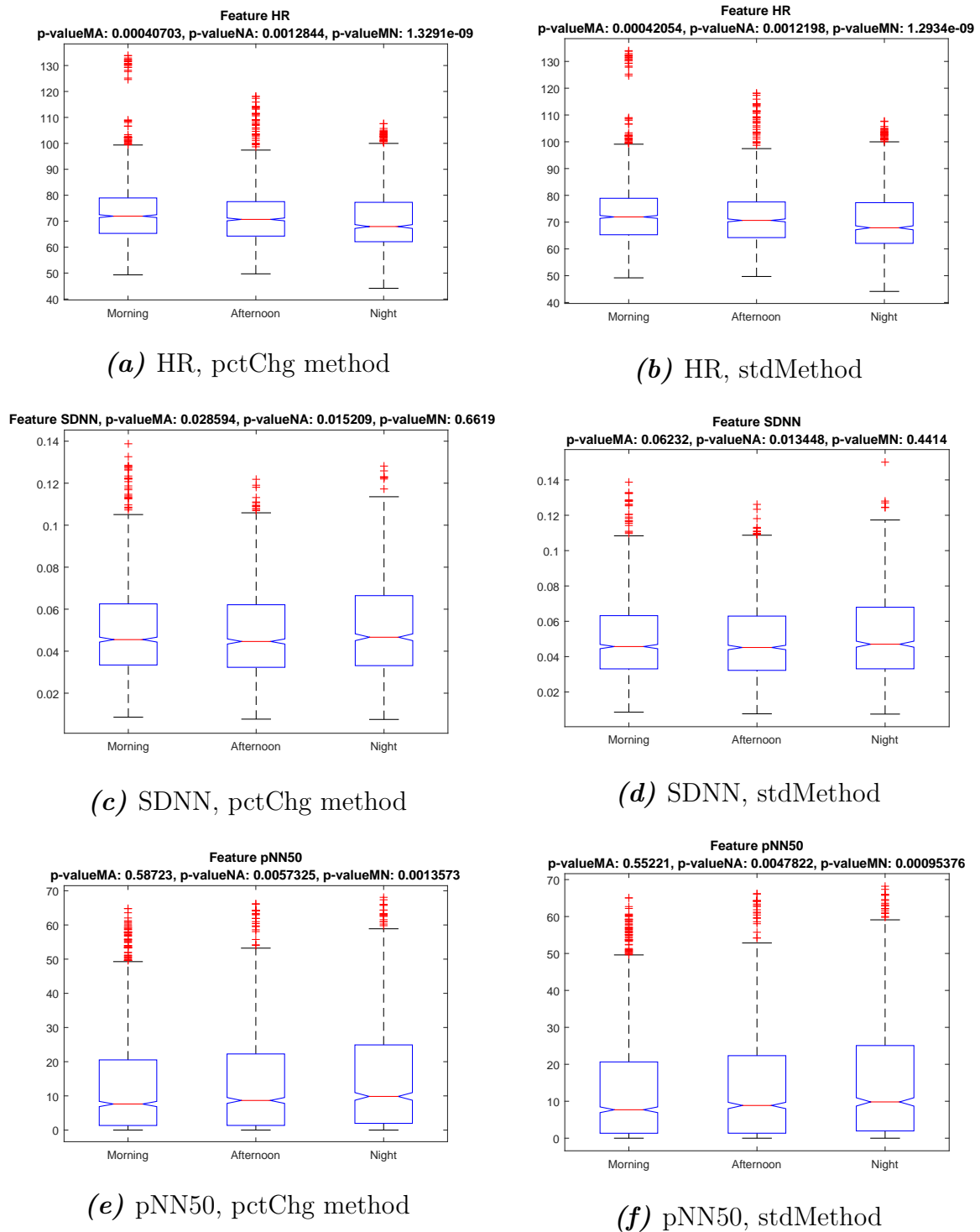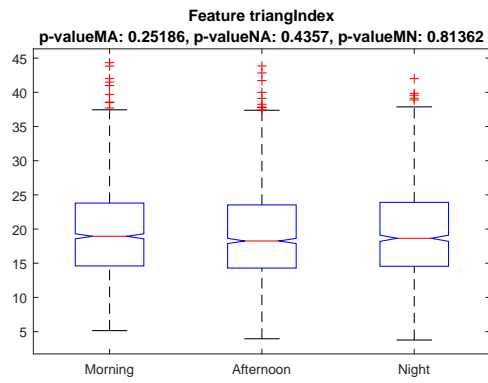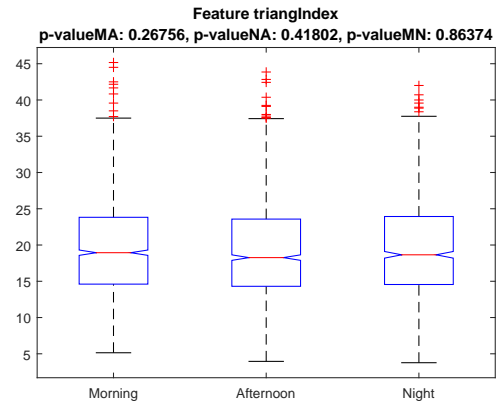**(e)** pNN50, pctChg method

(f) pNN50, stdMethod

Figure 4.4: Boxplots for different features in the time domain, comparing morning, afternoon and night drives. Each feature is presented for the two outliers detection methods.
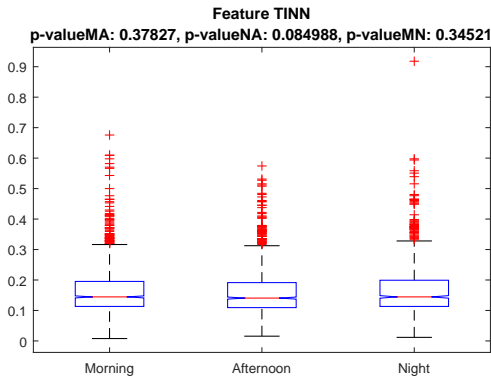
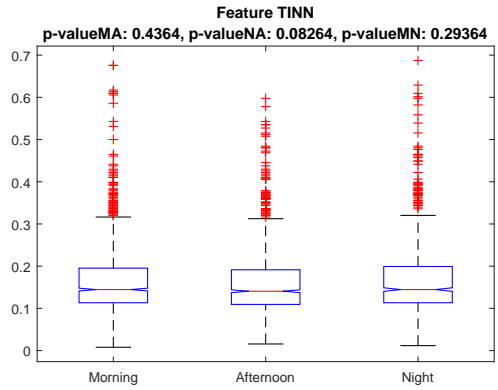*(a)* RRTri, pctChg method

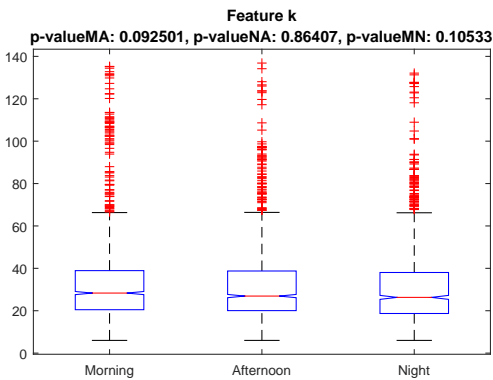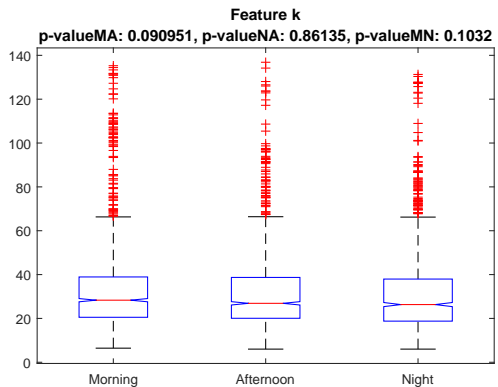*(b)* RRTri, stdMethod
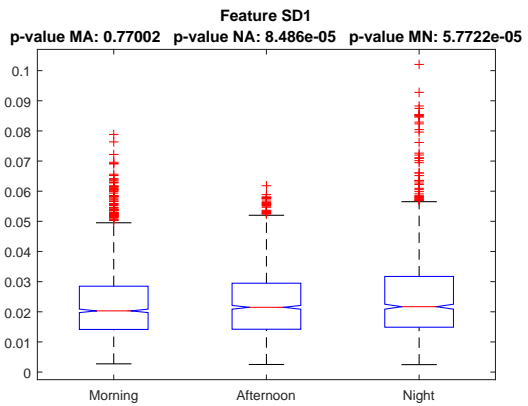
(c) TINN, pctChg method

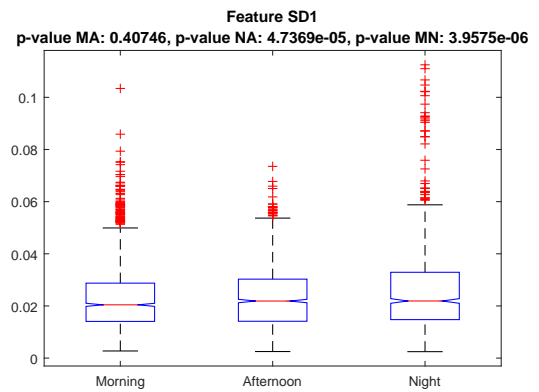*(d)* TINN, stdMethod

(e) k, pctChg method

(f) k, stdMethod

Figure 4.5: Boxplots for different features in the time domain, comparing morning, afternoon and night drives. Each feature is presented for the two outliers detection methods.

(a) SD1, pctChg method

(b) SD1, stdMethod

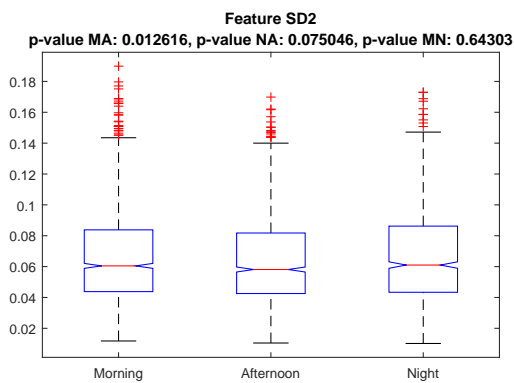*(c)* SD2, pctChg method

(d) SD2, stdMethod

Figure 4.6: Boxplots for different features in the non-linear domain, comparing morning, afternoon and night drives. Each feature is presented for the two outliers detection methods.
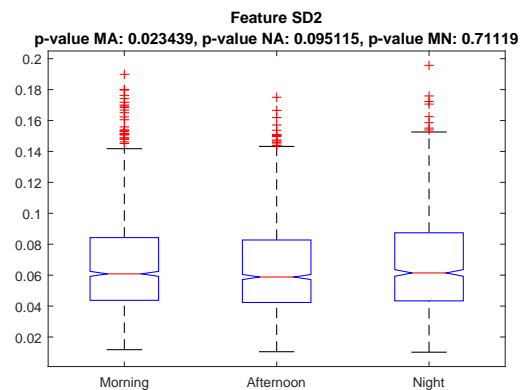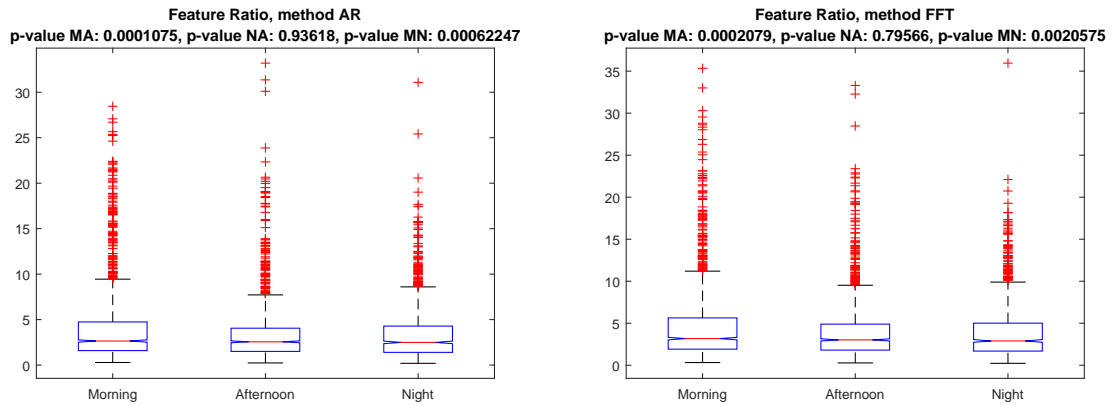
(a) Ratio AR

(b) Ratio FFT
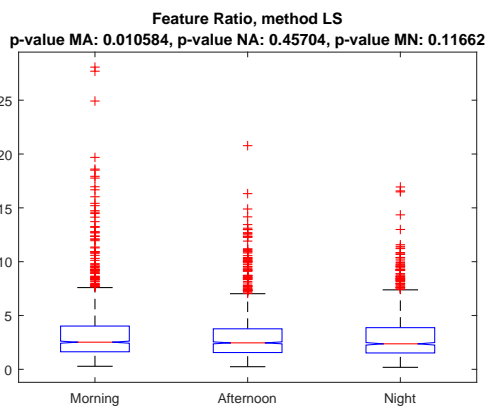
(c) Ratio LS

Figure 4.7: Boxplots for the feature Ratio for the three transformation methods, comparing morning, afternoon and night drives.
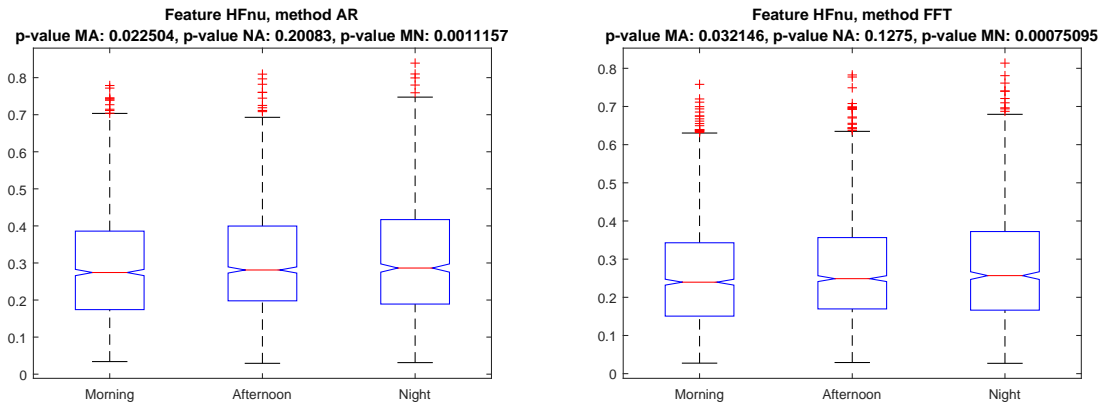
(a) HFnu AR

(b) HFnu FFT

(c) HFnu LS

Figure 4.8: Boxplots for the feature HFnu for the three transformation methods, comparing morning, afternoon and night drives.

(a) LFnu AR

(b) LFnu FFT

(c) LFnu LS

Figure 4.9: Boxplots for the feature LFnu for the three transformation methods, comparing morning, afternoon and night drives.

5

Discussions

In this chapter, the results presented in the previous chapter will be discussed to answer the research questions.

5.1 Comparison between outliers detection methods
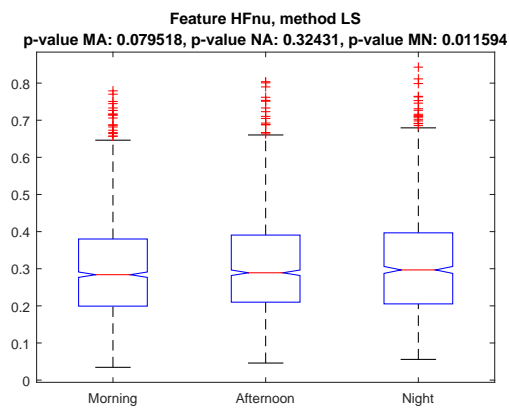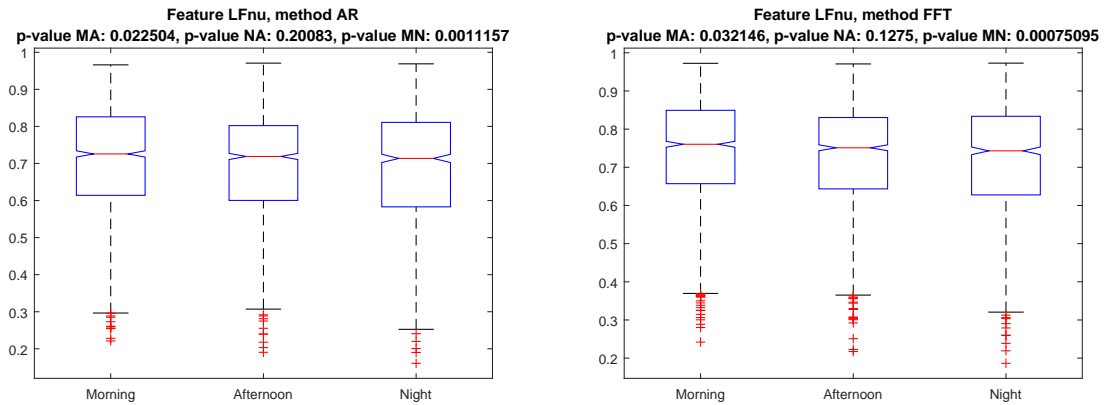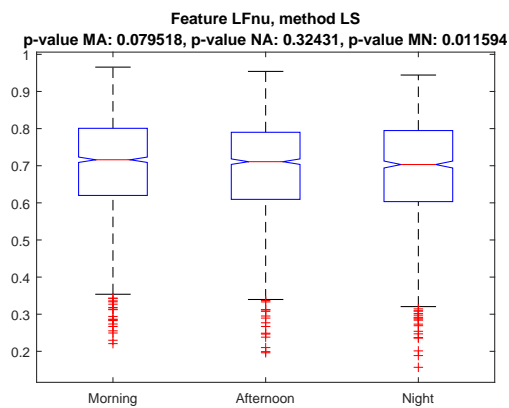
The number of sweeps discarded and outliers detected by the two outliers detection methods is significantly different. The pctChg method has marked and removed 0.278% of the samples and discarded 21 sweeps. The stdMethod has found, instead, 0.0648% of outliers and removed only two epochs. Thus, it is possible to say that the stdMethod is more conservative than the pctChg method. This is strictly related to the threshold adopted (30% variation and 5std), which were chosen according to recommendation in [49].

From the analysis of Figures 4.1 and 4.2 is possible to infer whether choosing an outlier detection method or another is useful or not for the results.

Frequency domain figures are shown in Figures 4.1a. Attending Figure 4.1a, it is represented the difference between the feature Ratio, calculated with the autoregressive model, after removing the outliers with the pctChg method, and with the stdMethod. From the picture, it is to compare the bias (systematic error, straight horizontal line), the confidence intervals at two standard deviations (random error, dashed lines) and the mean of both methods average (vertical line). It is possible to notice that most of the points are spread along a line whose Difference value is very close to zero and to the bias, whose value 0.008997 is sensibly small. The value of the CI has to be compared with the mean magnitude of the measure. Since it is close to 4 and the CI has a value of $\pm$ 0.98122, approximately 95% of the results given by the two methods has a variation within 25%, but most of them very close to zero. The variation is significant but still not too big to clearly mark the results as different. The p-value returned by the Student t-Test is around 24.3%, which means that the Bias is not statistically significant. Similar conclusion can be drawn for the same feature, calculated with Lomb-Scargle and Fourier Transform. The latter shows, among the three, the biggest, but still small, difference (Figure 4.1b and Figure 4.1c. The Bias are generally very low as well as the variance in HRV indices in the frequency domain. Thus, it can be concluded that the outlier detection method does not significantly influence the values of HRV indices in the frequency domain.

Regarding the time, geometric and non-linear domain features, Figure 4.1d, 4.1e,

4.1f, 4.2a and 4.2b, it is possible to see that the confidence interval, if compared to the magnitude of the mean, is very small. Figure 4.1d assess a variation of ±0.664 with respect to a magnitude of 14. Thus the variation is below 5% and thus small enough to infer that the choice of the outliers detection method has almost no influence on the results. The p-value is extremely small too, but this is due to a large dataset rather than a high bias. This way, the Bias is around 200 times lower than the mean of the magnitudes what is consistent among all time domain features.

From the analysis it is thus possible to state that choosing pctChg or stdMethod for detecting outliers has non-significant influence on the outcoming results. This result is less evident for the features related to the frequency domain than the others.

5.2 Comparison between different time-to-frequency domain transformation methods

In Figures 4.3 are shown the effects, on the results, of the different choice for the time-to-frequency domain transformation method. Only the feature Ratio is shown, but the following discussion can be similarly applied to the other features in the frequency domain, which are summarized in the table in Tables 4.1, 4.3 and 4.4. An analysis similar to the one in the previous section has been done. A detailed examination of Figure 4.3a reveals that, contrary to what it was happening for the outliers detection methods, the points of the BA plot are less aligned on the zero and more spread as a cloud. The value assumed by bias is also bigger, around 0.5 in absolute value instead of 0.008. Thus the mean difference between the two methods is bigger. Concerning instead the CI, they are approximately 50% of the magnitude mean when comparing AR and FFT and approximately 100% when comparing LS to any other (Figure 4.3d, 4.3e and 4.3f). The variation is big enough for saying that the choice of the transformation method is relevant for our application. The fact that AR and FFT has higher agreement among them than compared to LS may happen because, unlike the LS method, they are affected by unevenly sampled data. Moreover, they show a marked increasing/decreasing trend that is not present when comparing FFT with AR. Moreover AR is also sensitive to the order chosen for the model. The results point at the time to frequency transformation methods having a high influence over the HRV indices values, especially when referring to LS.

5.3 Capability of HRV indices of indirect fatigue detection

The box plots of Figures 4.7-4.9, together with their p-values (MA = morning-afternoon, NA = night-afternoon, MN = morning-night), show the capability of the calculated HRV indices of discriminating between morning, afternoon and night drives. This can be considered an indirect indicator of driver sleepiness. In particular, in Figure 4.7, it is possible to infer that the p-values for the feature Ratio calculated with the pctChg method, is almost identical whether the transforma-

tion method chosen is the FFT, the AR-model or the LS-periodogram. The LS-periodogram performed slightly worse than the other two methods as highlighted by its higher p-values. Moreover, the feature Ratio seems to perform better when discriminating fatigue between morning and afternoon, and between morning and night since those p-values are sensibly lower than 5%.

Similar conclusions can be drawn for the feature HFnu and LFnu calculated with the three transformation methods (Figure 4.8 and 4.9). Differently from the the feature Ratio and LFnu, the normalized high frequency shows a higher discriminative potential since the lower overlapping between the boxes in the plot. The effect is more evident with the AR-model, as shown in Figure 4.8a. Also in this case, the LS-periodogram has slightly worse performances.

The very similar p-values of the time domain features calculated with the pctChg method and the stdMethod, point out that there are no appreciable effects on the results adopting one outlier detection method or the other, as shown in Figures 4.4. Among the time domain features, HR is the one that allows a better discrimination between morning, afternoon and night since the lower overlapping between the boxes.

Similarly to the time domain features, the geometrical and the non-linear (Figure 4.5 and Figure 4.6) are not affected by the outliers detection method used.

5. Discussions

6
Conclusions

The dependency of the HRV indices values on the methods utilized for outlier detection and spectral transformation was tested on this thesis using data recorded on drivers. Outlier detection methods showed to have non-significant influence on the value of the HRV indices calculated. Spectral transformation methods showed, instead, to have a substantial influence on the HRV indices.

Moreover, how the different methods allow separation of time of the day based on HRV was tested as indirect indicator of what methods are most suitable for driver fatigue detection. The outlier detection method chosen did not show an influence on the capability of the different HRV indices of discriminating between morning, afternoon and night. The spectral transformation methods did not show a significant influence either. However, between the three transformation methods, the LS-periodogram performed slightly worse in this regards.

6. Conclusions

References

[1] S. K. Lal *et al.*, "A critical review of the phychophysiology of driver fatigue", *Biological Psychology*, vol. 55, pp. 173–194, 2001.

[2] "Prevalence of motor vehicle crashes involving drowsy drivers, united states, 2009 - 2013", *Foundation for Traffic Safety*, 2014.

[3] D. Sandebrg *et al.*, "The characteristics of sleepiness during real drive at night - a study of driving performance, physiology and subjective experience",

[4] A. J. Filtness and L. A. Reyner, "Driver sleepiness-comparisons between young and older man during a monotonous afternoon simulated drive", *Biological Psychology*, vol. 89, pp. 580–583, 2012.

[5] *Slubjective measures of sleepiness.* [Online]. Available: `http://www.sleeprate.com/faq_article/subjective-measures-of-sleepiness/`.

[6] *Quantifying sleepiness.* [Online]. Available: `http://www.uptodate.com/contents/quantifying-sleepiness`.

[7] T. Åkerstedt *et al.*, "Having to stop driving at night because of dangerous sleepiness - awarness, physiology and behaviour", vol. 22, pp. 380–388, 2013.

[8] K. Kaida *et al.*, "Validation of the karolinska sleepiness scale against performance and eeg variables", *Clinical Neurophysiology*, vol. 117, pp. 1574–1581, 2006.

[9] T. Akerstedt *et al.*, "Subjective and objective sleepiness in the active individual", vol. 52, pp. 29–37, 1990.

[10] *Heart.* [Online]. Available: `http://en.wikipedia.org/wiki/Heart`.

[11] *Potential conduction.* [Online]. Available: `http:\\tmedweb.tulane.edu\pharmwiki\doku.php\cardiac\electrophysiology`.

[12] *Electrical conduction system.* [Online]. Available: `http://en.wikipedia.org/wiki/Electrical_conduction_system_of_the_heart`.

[13] G. A. Ng, "Treating patients with ventricular ectopic beats", *Heart*, pp. 1707–1712, 2006.

[14] *Ectopics.* [Online]. Available: `http://www.rbain.org.uk/Ectopics.html`.

[15] *Heart rate variability.* [Online]. Available: `http://en.wikipedia.org/wiki/Heart_rate_variability`.

[16] J. Schesser, *Hrv analysis.*

[17] Hon and Lee, "Electronic evaluation of the fetal heart rate. viii. patterns preceding fetal death, further observations.", *American journal of obstetrics and gynecology*, vol. 87, 1963.

[18] *Heart rate variability: standards of measurement, physiological interpretation, and clinical use.* [Online]. Available: `http://circ.ahajournals.org/content/93/5/1043.full`.

References

[19] *Cardiac cycle.* [Online]. Available: `http://commons.wikimedia.org/wiki/File:2028_Cardiac_Cycle_vs_Electrocardiogram.jpg`.

[20] *What is hrv.* [Online]. Available: `http://www.megaemg.com/knowledge/heart-rate-variability-hrv/`.

[21] R. J. Leor-Librach *et al.*, "A common origin of the very low frequency hear rate and blood pressure variability - a new insight into an old debate", *Autonomic Neuroscience: Basic & Clinical*, vol. 96, pp. 140–148, 2002.

[22] B. A. Leor-Librach *et al.*, "Effect of respiration in heart rate variability (hrv) analysis", *Engineering in Medicine and Biology Society*, pp. 1776–1779, 2006.

[23] *Parasympathetic nervous system.* [Online]. Available: `http://en.wikipedia.org/wiki/Parasympathetic_nervous_system`.

[24] U. Rayendra *et al.*, "Heart rate variability: a review", vol. 44, pp. 1031–1051, 2006.

[25] A. Sahaydhas *et al.*, "Detecting driver drowsiness based on sensors: a review", vol. 12, pp. 16 937–16 953, 2012.

[26] M. Patel *et al.*, "Applying neural network analysis on heart rate variability data to access driver fatigue", vol. 38, pp. 7235–7242, 2011.

[27] K. Saroj *et al.*, "A critical review of the psycophysiology of driver fatigue", vol. 55, pp. 173–194, 2001.

[28] *Sympathetic nervous system.* [Online]. Available: `http://en.wikipedia.org/wiki/Sympathetic_nervous_system`.

[29] *Making sense of the two-sample t-test.* [Online]. Available: `http://www.isixsigma.com/tools-templates/hypothesis-testing/making-sense-two-sample-t-test/`.

[30] *The t-test.* [Online]. Available: `http://www.socialresearchmethods.net/kb/stat_t.php`.

[31] *Student's t-test.* [Online]. Available: `http://en.wikipedia.org/wiki/Student%27s_t-test`.

[32] *Bland-altman plot.* [Online]. Available: `http://en.wikipedia.org/wiki/Bland%E2%80%93Altman_plot`.

[33] *Precision and accuracy.* [Online]. Available: `http://www.ncsu.edu/labwrite/Experimental%20Design/accuracyprecision.htm`.

[34] M. Mahachandra, "Sensitivity of heart rate variability as indicator of driver sleepiness.", *SEANES*, 2012.

[35] W. C. Liang *et al.*, "Changes in physiological parameters induced by indoor similated driving: effect of lower body exercise at mid-term break", vol. 9, pp. 6913–6933, 2009.

[36] G. Yang *et al.*, "A driver fatigue recognition model based on information fusion and dynamic bayesian network", *Information science*, vol. 180, pp. 1924–1954, 2010.

[37] M. Miyaji, "Method of drowsy state detection for driver monitoring function", *International Journal of Information and Electronics Engineering*, vol. 4, 2014.

[38] Ingre *et al.*, "Subjectivesleepiness, simulated driving performance and blink duration: examining individual differences.", vol. 15, pp. 47–53, 2006.

[39] Sommer *et al.*, *Biosignal based discrimination between slight and strong driver hypovigilance by support-vector machines.* Springer, 2010, vol. 67, pp. 177–187.

[40] C. Liu *et al.*, "Predicting driver drowsiness using vehicle measures: recent insights and future challenges", vol. 40, pp. 239–245, 2009.

[41] M. Vitaterna *et al.*, "Overview of circadian rhythms", vol. 25, pp. 85–93, 2001.

[42] P. Philip *et al.*, "Fatigue sleep restriction and driving performance", vol. 37, pp. 473–478, 2005.

[43] L. Hartley *et al.*, "Review of fatigue detection and prediction technologies", 2000.

[44] X. Yu *et al.*, "Real-time nointrusive detection of driver drowsiness", 2009.

[45] *Seeingmachines driver state sensor.* [Online]. Available: `http://www.seeingmachines.com/product/dss/`.

[46] E. Michail, "Eeg and hrv markers of sleepiness and loss of control during car driving", *Engineering in medicine and biology society*, pp. 2566–2569, 2008.

[47] G. Li *et al.*, "Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier", vol. 13, pp. 16 494–16 511, 2013.

[48] T. Akerstedt *et al.*, "Sleepiness at wheel. white paper.", *ASFA*, 2012.

[49] K. J. Kemper *et al.*, "Heart rate variability: impact of differences in outlier identification and management strategies on common measures in three clinical populations", *Pediatric research*, vol. 62, pp. 580–583, 2007.

[50] T. force of the european society of cardiology, the north american society of pacing, and electrophysiology, "Standard of measurements, physiologial interpretation and clinical use", *European heart journal*, pp. 354–381, 1996.

[51] Pancand and Tompkins, "A real-time qrs detection algorithm", *Biological Psychology*, vol. BME-32, pp. 230–236, 1985.

[52] M. A. Peltola, "Role of detecting r-r intervals in the analysis of heart rate variability", 2012.

[53] G. D. Cliffor *et al.*, "Quantifying errors in spectral estimates of hrv due tobeat replacement and resampling", *IEEE*, vol. 52, pp. 530–638, 2005.

[54] P. Laguna *et al.*, "Power spectral density of unevenly sampled data by least-square analysis: performance and application to heart rate signals", vol. 45, pp. 698–715, 1998.

[55] G. G. Clifford, "Signal processing methods for heart rate varibility", PhD thesis, Unversity of Oxford, St. Corss College, 2002.

[56] A. P. R. Anita Boardmann Fernando Soares Schlindwein and A. Leite, "A study on the optimum order of autoregressive model for heart rate variability", *Physiological Measurements*, vol. 23, pp. 325–336, 2002.

[57] *Plomb.* [Online]. Available: `http://se.mathworks.com/help/signal/ref/plomb.html?refresh=true`.

[58] L. C. M. Vanderlei *et al.*, "Geometric indexes of heart rate variability in obese and eutrophic children",

[59] J. Piskorski *et al.*, "Geometry of the poincaré plot of rr intervals and its asymmetry in healthy adults.", *Physiological measurements*, pp. 278–300, 2007.

[60] *Boxplot.* [Online]. Available: `http://se.mathworks.com/help/stats/boxplot.html`.