Abstract

Recent years have seen a rapid expansion of consumer sleep monitoring devices, using a wide variety of cheap and accessible sensor technology. However, these products tend to be seriously lacking in terms of validation, clinical or at all, and have very limited medicinal usefulness. On the other side, the gold standard of sleep medicine is polysomnography, a highly comprehensive and resource intensive method that sacrifices any and all practicality for its high accuracy. The essential motivation behind this thesis is a desire to unite these two worlds through extensive validation and multivariate data modelling techniques.

This thesis focuses on building a predictive model for sleep apnea, a sleep disorder characterized by cessation of or periods of shallow breathing during sleep. The work consists of accessing and exporting anonymized polysomnography recordings done at St. Olav's hospital between 2012 and 2015, as well as reconfiguring the data to a more well-structured format. A selection of preprocessing and data reduction techniques are described and implemented, and multivariate analysis and modelling is performed on several different configurations of the data. Using a subset of 10 sensors from the polysomnography data and Partial Least Squares Regression, a model with good predictive ability for a characteristic of sleep apnea is achieved, that also performs well under validation.

However, more data would have to be included before any proper conclusions can be drawn. We therefore end the work by proposing an outline for an extensive research project involving both polysomnography scored by trained experts and a selection of commercial wearable physiological monitors, which might eventually lead to a commercial at-home sleep monitor with a validated clinical ability to recognize respiratory disturbances in its users.

Sammendrag

I nyere år har det vært en eksplosjon på markedet for søvnovervåkende konsumerprodukter, som bruker et vidt spekter av billig og letttilgjengelig sensorteknologi. Disse produktene har imidlertid en tendens til å mangle validering, og er per i dag i svært liten grad nyttige innen søvnmedisin. På den andre siden er gullstandard i søvnevaluering, polysomnografi, en svært omfattende og resurskrevende undersøkelse, som ofrer all lettvinthet på nøyaktighetens alter. Den grunnleggende motivasjonen bak dette arbeidet er å legge grunnlaget for en bro mellom disse to verndene, gjennom nøyaktighet, validering, og multivariate modelleringsteknikker.

Dette arbeidet fokuserer på å bygge en prediktiv modell for søvnapné, en søvnforstyrrelse som kjennetegnes av gjentatte pustestopp om natten (og svært ofte høylytt snorking). Arbeidet består av å få tilgang til samt eksportere polysomnografidata fra St. Olav's Hospital, i tillegg til å rekonfigurere denne dataen til et mer strukturert format. Et utvalg av preprosesseringsteknikker forklares og implementeres, og multivariat analyse og modellering utføres på flere forskjellige datakonfigurasjoner. Gjennom Partial Least Squares Regression av en undergruppe bestående av 10 sensorer, blir en modell med gode prediktive og validerte egenskaper for en søvnapnékarakteristikk oppnådd.

Det vil imidlertid være nødvendig å inkludere mer data før noen definitive konklusjoner kan bli nådd. Vi avslutter derfor arbeidet med å foreslå en skisse for et større forskningsprosjekt som involverer både profesjonelt tolket polysomnografi og et per nå udefinert utvalg kommersielt tilgjengelige 'wearable' søvnmonitorprodukt, som til slutt kan lede til et søvnovervåkende konsumerprodukt med validert klinisk evne til å gjenkjenne respirasjonsfortyrrelser under søvn hos sine brukere.

Preface

This thesis concludes a Master's degree in Engineering Cybernetics from the dept. of Engineering Cybernetics at NTNU.

The long-term goal of this research would be to use the results and experience to develop scientifically sound easy to use commercial at-home sleep evaluation tools. Collaborating partners in this larger project includes St. Olav's Hospital and Novelda AS, whom I worked for in summer (full time) and autumn (part time) 2015. They also provided the funds for my data storage in this project, and a course in polysomnygraphy interpretation attended by me in Bergen, Norway, February 2016.

I will disclose that the idea for this project occurred during an ordinary lunch at Novelda, from some casual smalltalk about the data processing tools (or lack of such) used in polysomnygraphy interpretation, and the lack of an adequate alternative. Being a young and arrogant fledgling cyberneticist, I felt at once vaguely offended and like I could do something to improve the situation. Now, much older and wiser (an entire year later), I realize that the task might have been a bit more complicated than I initially anticipated. This work is by no means complete, and I would like to keep carrying this project forwards.

This report presents work carried out during spring/autumn 2016. It assumes that the reader has a fair understanding of linear algebra, but not necessarily in-depth knowledge of multivariate statistics and modelling, nor significant experience with the field of sleep medicine.

The reader must also feel assured that the irony of spending late nights and loosing copious amount of sleep on a topic such as this is not lost on me.

Trondheim, December 1, 2016

Hanne Siri A. Heglum

Acknowledgments

I would like to thank my supervisor Prof. Tor Onshus, for your excellent guidance and infinite patience.

Without Prof. Harald Martens, this work would not have happened. Your enthusiasm, knowledge, and experience is truly inspirational.

Dr. Morten Engstrøm, who helped me gain access to everything I needed at St. Olav's, and who had so much faith in me. Thank you.

I must also thank Ellen Beate Hove. You have much more faith in me than I do, and I would not have gotten through all these years without your help.

In the same breath, I would like to thank a different Ellen, who saved my life a couple of years ago.

Less seriously, I will thank Sirius the Cat. My Big Beautiful Boy, too cute for this world.

Lastly, to all my friends and family. Thank you for putting up with me.

H.S.A.H.

Contents

1

Introduction

Laugh and the world laughs with you; snore, and you sleep alone!

– Anthony Burgess

1.1 Background and motivation

Anyone able to read this will be at least partially aware of the importance of sleep - it critical for human health both physically and mentally, and plays a pivotal role in our quality of life. Knowledge of the existence and anguish of sleep disorders is nothing new - records dating back to ancient times discuss insomnia and its possible remedies[20]. However, factors that potentiate the development of sleep disorders is closely related to the cultural zeitgeist, and growing evidence suggests that the technological advancements that have afforded modern society such unprecedented conveniences also may be taking their toll on our quality of sleep[28]. At that, the ambitious technologist may think that if technological advances are causing us trouble, then they should also be able to help. Yet the current gold standard of sleep study is a lumbering beast, more lightweight options are seriously undervalidated, and in general there is a real need for inovation in this field[27].

When sleep disorders are suspected, the current gold standard for sleep evaluation is a comprehensive sleep study, called a polysomnography (PSG)[9]. This is a highly resource intensive procedure, and it is poorly suited both for population studies and in cases where the presence of

large ammounts of body-mounted sensors is undesireable or impossible. The data processing and interpretation is commonly done manually, which, while accurate, is highly time consuming. Less comprehensive screening methods do exist, but these either tend to suffer from serious accuracy challenges or be similar to PSG in their interpretation processes[31].

Recent years has seen a rapid expansion of consumer sleep monitoring devices, using a variety of sensors to track movement as well as cardiac and respiratory physiology. However, finding relevant information in this data is a difficult task, and while many data mining techniques for wearable sensors in health monitoring have been explored[3], in the case of sleep monitoring, the market interest and availability of commercial products tend to be outpacing the rigorous validation needed to assess the potential use of these devices in clinical and research settings[27].

The amount of cheap, high-quality data available today is huge compared to when the gold standard methods were developed. Several individual suppliers of at-home wearable sleep monitors have performed double registrations with PSG for comparison and validation, but so far the focus has been exclusively on healthy patients, and the scope far too small[27].

The background of this project is the idea that clinical sleep-evaluation need options that are more efficient, using less cumbersome equipment and more efficient data processing than the current clinical alternatives and better validated than the current consumer products[27].

1.2 Objective

This project will focus on sleep apnea, a sleep disorder characterized by cessation of or periods of shallow breathing during sleep, described in more detail in Section 2.3.

While consumer products for sleep evaluation tends to start with a technology, develop a sleep monitor, and then (sometimes) validate against gold standard[27], this project will approach the problem in reverse order. The project will begin by using PSG recordings as a base-

line, with scored results available for development and validation. Through preprocessing and multivariate analysis tools, we will attempt to build a predictive model for apnea characteristics.

If it turns out that the necessary model parameters are one of the many that now can be recorded with simple wireless sensors available commercially today or in the near future[1], the results of this examination may in turn be used to choose which sensors and data processing methods to combine to provide more simple and lightweight integrated sleep evaluation products with real clinical applicability, than anything that is currently available.

1.3 Limitations

This thesis will only focus on analysis of respiratory issues during sleep. This excludes the problem of separating sleep from wake, as well as other sleep disorders not related to respiration. Only data from adult patients is included. We will not focus on the characteristics of different sleep stages, nor pay much attention to popular parameters such as sleep latency, sleep effiency, or the hypnogram. The data set itself has several limitations, described in detail in Section 7.2.

1.4 Structure of the report

This thesis is structured as follows: Chapter 2 begins with a brief historical perspective on sleep science, before describing some technical aspects of sleep medicine today. We then define and describe sleep apnea, and the specific equipment relevant to thesis. The theory and interpretation of the multivariate analysis tools used is described in Chapter 3. Chapter 4 describes the process of aquiring data, and transforming it to a usable format, and Chapter 5 describes the preprocessing steps taken to prepare the data for the analysis that is described and discussed in Chapter 6. Chapter 7 provides discussion of the methods used, as well as a suggestion for how this work could continue. Chapter 8 provides a summary and conclusion.

[1] say, any combination of respiration curves, heart rate, body movement, and blood oxygenation[7] [23]

1.5 Software used in this thesis

MATLAB

Preprocessing, preliminary visualization, and data infrastructure and management has been handled using MATLAB version R2016a.

The Unscrambler®X

The Unscrambler®X version 10.3 has been used for the multivariate modelling.

DOMINO

DOMINO Software (part of the SOMNOScreen™System provided by SOMNOmedics) was used for extracting and formatting the PSG data.

2

An Introduction to Sleep

This chapter begins by summarizing the history of sleep medicine in Section 2.1, before describing some of the technical aspects and challenges of sleep medicine today in Section 2.2. Sleep apnea, the sleep disorder most relevant to the work of this thesis, is described in Section 2.3, and Section 2.4 describes the equipment used in this thesis.

2.1 A brief history of sleep medicine

More has been learned about sleep in the last 60 years than in the past 6000.

– Alan Hobson, *Sleep*

The importance of sleep is obvious to any living human, and yet the nature of the phenomena has confounded philosophers and physicians for milennia. Sleep medicine and science today relies on modern sensors and computer technology, but it is important to apreciate the roots of our current understanding. This section will very briefly summarize the relevant history.

The foundations of sleep science can be traced back to ancient civilizations across the globe[20]. Sleep has been both feared and revered; dream interpretation has been an integral part of civilizations and religions[30], and in both ancient Rome and Greece the similarity between sleep and death has been emphasized; Homer wrote in the *Illiad* "Sleep and death, who are twin brothers", and Ovid (43 B.c.-A.D. 17) in *Amores II*, "What else is sleep but the image of chill

death?".

Greek philosophers wrote about the nature of sleep - Alcmaeon of Croton (approx. 500 BCE) wrote about sleep as analgous to the tide, that it is “produced by the withdrawal of the blood away from the surface of the body to the larger (‘blood-flowing’) vessels and that we awake when the blood diffuses throughout the body again”[25], and Aristotle wrote an entire opus (*On Sleep and Sleeplessness*) dedicated to the nature and dichotomy of sleep and wake[2]. Chinese and Egyptian texts reveal at least some knowledge of sleep disorders (primarily insomnia), and discuss their treatments[20].

The first references to a more "modern" sleep science occur, perhaps unsurprisingly, during the 17th century. Thomas Willis, considered to be one of the fathers of neurology, refers in his book *The Practise of Physick* (1692) to sleepiness and insomnia, as well as descriptions of narcolepsy and restless leg syndrome. He also recognized the effect of coffee on sleep[17]. Descartes proposed "the hydraulic model" for sleep[6]. Throughout the 18th century biological rhythms gain recognition, and during the 19th century several competing theories on the nature of sleep are proposed. However, during this time the majority of focus and curiosity are geared towards explaining the mystery of sleep itself, with almost no attention being payed to the treatment of sleep disorders. The 19th century shows more practical advances, including the first medications produced specifically to induce sleep, as well as the first clinical descriptions of insomnia and narcolepsy[20].

Nevertheless, there is little controversy in claiming that the true birth of sleep science as it is known today began with the development of the polysomnogram. Hans Berger invented the electroencephalogram (EEG) and recorded the first human EEG in 1924[16]. Following this, two groups (Harvey, Hobart, Loomis, Davis, and Davis at Harvard University, and Blake, Gerard, and Kleitman at the University of Chicago) used the EEG to preform sleep-related research during the 1930’s, examining the brain wave changes during sleep. They were able to distinguish distinct and different stages in the brain activity during sleep, a foundation for the sleep staging that is still the basis of sleep studies today.

In 1953 Kleitman and his student Eugene Aserinsky were the first to recognize the Rapid Eye Movement (REM) sleep stage, a discovery that proved hugely controversial at the time[1]. They developed the electroocculogram to better characterize the eye movements during REM sleep, and discovered the connection between the REM sleep phase and vivid dreaming[20].

The true dawn of modern sleep science followed in 1968, with the publishing of the fundamental document *A Manual of Standardized Terminology, Techniques, and Scoring System for Sleep Stages of Human Subjects* by Allan Rechtschaffen and Anthony Kales[26]. This publication contained the first formal definition of the polysomnogram, defining a sleep epoch as 30 s and clarifying the stages of sleep (wakefulness, movement time, stages 1-4 [NREM sleep], and REM sleep). A separate manual for the scoring of sleep in children was published in 1971.

The term 'polysomnography' (PSG) was proposed by Holland, Dement, and Raynal in 1974, to describe the recording, analysis and interpretation of multiple simultaneous physiological parameters during sleep. As a tool, PSG has been considered essential in the formulation of diagnoses for sleep disorder patients and in the enhancement of understanding of both normal sleep and its disorders ever since[14].

All in all, the 20th century saw an explosion in the activity in sleep science. The Mean Sleep Latency Test for evaluating daytime sleepiness was developed in the early 1980s, with formal guidelines published in 1986, knowledge of circadian rhythms were substantially expanded through extensive studies for the duration of the century, and sleep disorders including Obstructive Sleep Apnea (OSA), Restless Leg Syndrome (PLM - Periodic Leg Movement), and Narcolepsy saw their first formal definitions, along with an advent of medicinal and non-medicinal treatment options, compiled in the first Diagnostic Classification of Sleep and Arousal Disorders in 1979[20].

[1] REM sleep is characterized by chaotic low frequency EEG, and apart from the characteristic eye movements can be almost indistinguishable from wakefulness. This was completely contrary to the reigning belief that sleep brain activity were defined by low frequency high amplitude waves

The first comprehensive classification of disorders of sleep and arousal was developed by the American Academy of Sleep Medicine (AASM) in association with the European Sleep Research Society, the Japanese Society of Sleep Research, and the Latin American Sleep Society in 1990 and later revised in 1997. AASM also developed the Manual for the Scoring of Sleep and Associated Events in 2007, revised in 2012 in order to address a changed digital information landscape[4]. This manual contains rules for the interpretation and classification of sleep phenomena from polysomnography data, and is to this day the definitive reference for the evaluation of PSG.

2.2 Sleep medicine today

The current state of sleep evaluation standard procedure is summarized in Chapter 1 of the Atlas of Sleep Medicine [5]:

> The most important initial step in the evaluation of a patient with sleep complaints is a detailed and focused in-clinic interview and physical examination by an experienced provider, which helps generate a differential diagnosis. In selected patients the diagnostic polysomnogram (PSG) then plays a pivotal role in confirming the clinician's suspicions and helps guide further management.

2.2.1 Polysomnography

The term 'Polysomnogaphy' was proposed by Holland, Dement, and Raynal in 1974, to describe the recording, analysis, and interpretation of multiple, simultaneous physiological parameter[14]. It has since been considered the 'gold standard' in sleep evaluation, and has been described as being the single most important laboratory technique for the assessment of sleep and its disorders[5].

With roots extending to Rechtschaffen and Kales[26] and beyond, a large part of PSG still involves the scoring of sleep stages based on EEG activity, eye movements (EOG), and muscle activation recordings (EMG). PSG today will also involve the registration of a wide array of

respiration-related variables, including oronasal airflow, snoring, expansion and contraction of thorax and abdomen, blood oxygen saturation, and electrocardiography (ECG). It is also not uncommon to include sleeping position and leg movements, and if the situation calls for it even more channels could be added, including but not limited to esophageal pressure or concurrent video recordings[9].

The procedure is complex and thorough, and should be performed by a trained technologist with an understanding of what questions the study seeks to answer, in addition to access to a full medical and psychiatric history of the patient. Sophisticated knowledge and understanding is needed to understand how different aspects may affect the study, how to make protocol adjustments where necessary, anticipate difficulties, and guarantee that the most pertinent data are collected[14].

No matter the skill of the technologist, the sources of error are numerous: environmental noise (including electromagnetic interference), unstable electrodes and registration electronics, or simply patient movements during sleep just to name a few. In some sleep labs the patient stays overnight for the procedure, however in Norway the most common practice is for the patient to be mounted with the equipment in the afternoon, before going home with the rig in place to sleep as normal as possible in their own beds[9]. Both solutions have benefits - in-lab recording allows for greater control over potential physical disturbances (electrodes falling off, etc.), but may have a greater impact on sleep than the at-home alternative[19].

Nevertheless, polysomnography will always be highly resource intensive, time consuming, and expensive. The amount of equipment that the patient has to deal with is significant, and has to be mounted for many hours. The application of sensors is in itself time consuming, and the interpretation of data after the fact is laborious and cumbersome. As such, PSG providers do not always have the resources to provide more than a single night of PSG to the patient, which is problematic due to the so-called 'First Night Effect', a phenomena of adaption difficulties that has been known since 1966[12]. Furthermore, more recent studies have indicated that this first-night effect may in fact last for more than one night[13] and indeed, intuitively it is hard to

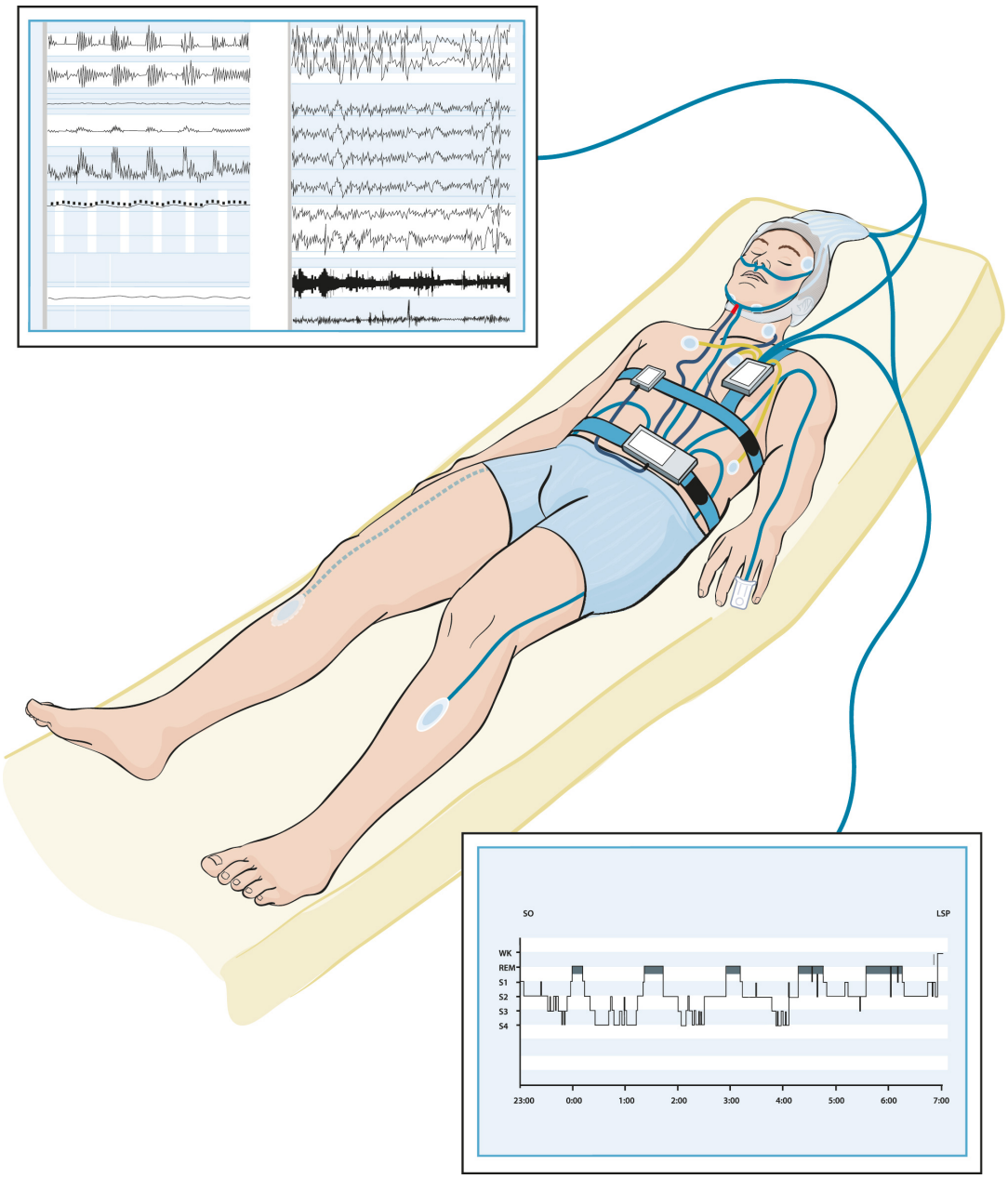


Figure 2.1: Common placement of sensory equipment during polysomnography. Top panel left shows 5 minutes of respiratory recordings, right thirty seconds of EEG, EOG, and EMG. Lower panel shows a somnogram. Figure from [9], illustration ©Illumedic

imagine sleep *not* being affected by wearing such a substantial rig of sensors on-body.

Polysomnography: The interpretation process

When data has been collected, an experienced technician or physician will analyze the data to evaluate length and quality of sleep. The AASM Manual for the Scoring of Sleep and Associated Events[4] is commonly considered the fundamental document of sleep scoring, and its standards are widely adopted. It contains lists of parameters to be reported by PSG, technical and digital specifications for PSG equipment, as well as visual rules for the scoring and interpretation of the data.

The analysis is done through a manual epoch-by-epoch review. The analyst divides these epochs into one of five sleep stages[2], W (Wakefulness), N (non-REM) 1, 2, and 3[3], and REM (Rapid Eye Movement), according to the visual scoring rules given in [4]. She also notes temporary arousals and other artifacts during sleep , as well as respiratory events including various types of sleep apnea (see section REF). Main statistics and features of the result are summarized in a PSG report, which among other things typically includes a hypnogram (plot of sleep stages over time), sleep efficiency and latency statistics (related to the ratio of sleep time versus time spent in bed), and the apnea-hypopnea index number AHI (see REF).

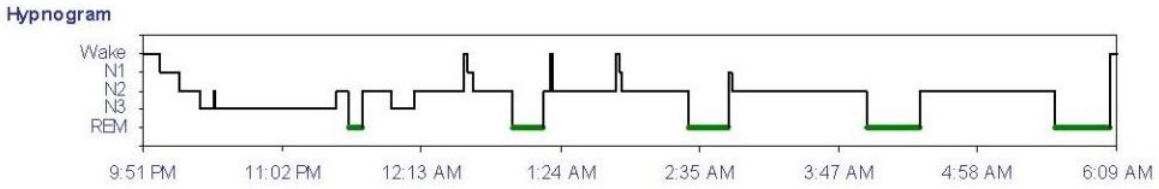


Figure 2.2: A typical hypnogram.

Historically, thirty second epoch lengths was used because of the common $10mms^{-1}$ paper speed for analog polysomnography. The standard has remained after digital systems took over, and today it is not uncommon to discuss the effect of screen size and quality on the accuracy of

[2] Sleep staging has long been considered to be a fundamental part of sleep science and medicine. However, since this thesis primarily deals with respiratory events and data, the sleep stages will not be given much attention here.

[3] Some also include a fourth stage of NREM, N4

the scoring. It is claimed that compressed displays of greater than 30s screen for EEG 'should be avoided'[14], and even experienced scorers will spend significant time going through an entire night of data. For slower signals such as oxygen saturation and respiratory signals, 2-5 minutes per screen is recommended.

Automated scoring- or interpretation methods have been developed, but their accuracy and reliability have thus far not been enough to make them anything more than an aid to clinicians at best, to be controlled by an experienced scorer[9].

Physiological Signals Recorded with PSG

The following section contains very brief descriptions of the most common physiological signals recorded during a PSG. There are multiple possible sensor technologies being used, and multiple possible arrays. This section will focus on the ones used by St. Olav's hospital, from whom we have gathered the data used in this thesis.

EEG, electroencephalography, is the recording of the spontaneous electrical activity of the brain over a period of time, recorded by an array of (typically noninvasive) electrodes placed on the scalp. Diagnostic information is usually extracted from the spectral content of the signals (commonly divided into frequency bands), and sometimes from specific visually recognizable signal shapes.

EOG, electrooculography, measures the corneo-retinal standing potential between the front and the back of the human eye, typically recorded by electrodes placed above and below the eye, or to the left and right. In PSG, EOG is instrumental in recognizing the rapid eye movements that characterizes both wakefulness and the rapid eye movement (REM) sleep stage, as well as the slow eye movements characteristic of a person moving from wakefulness to sleep.

EMG, electromyography, records the electrical activity produced by skeletal muscles. In PSG, four electrodes are used to record muscle tension in the body as well as monitor leg movements during sleep. Two electrodes are typically placed on the chin above and below the jaw line, as

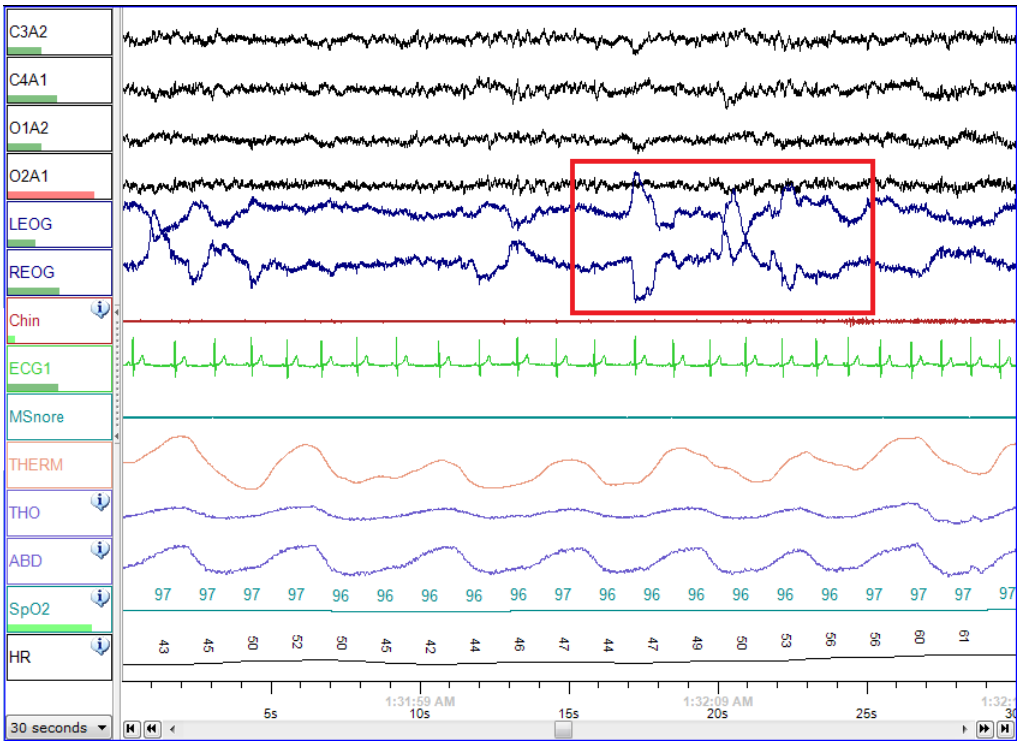


Figure 2.3: One epoch of a PSG recording. The large eye movement highlighted by the red box is a characteristic of REM sleep, but also of wakefullness.

well as one on each leg.

ECG, electrocardiography, records the electrical activity of the heart using electrodes placed on the skin. PSG typically uses two or three electrodes, whereas a typical ECG would use ten.

Pulse Oximetry measures the blood oxygen saturation, typically using a photoplethysmograph (an optical sensor) clipped either on a finger or an earlobe.

Pulse Oximetry Plethysmography measures the change in the volume of arterial blood with each pulse beat, also using a photoplethysmograph located in the same clip.

Respiratory Inductive Plethysmography (RIP) evaluates pulmonary ventilation (movement of air in and out of the lungs) by measuring the expansion and contraction of the chest and abdominal wall.

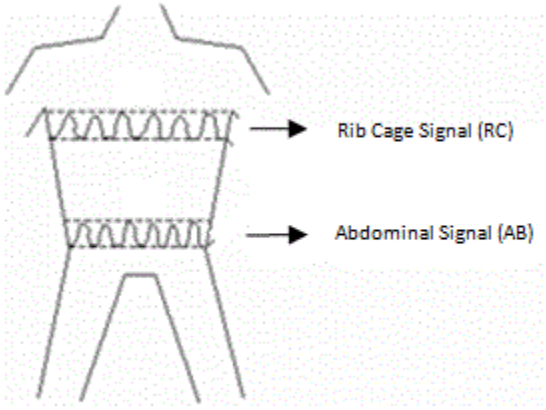


Figure 2.4: Respiratory Inductive Plethysmography (RIP) uses parallel belts to evaluate respiration. Figure from Wikipedia.

Flow of air during breathing is measured using pressure transducers (and/or a thermocouple), fitted in or near the nostrils.

Snoring may be recorded using a sound probe on the neck.

2.2.2 Current alternatives to Polysomnography

In addition to being resource intensive and expensive, there are cases such as examination of children or very sick patients, where the presence of large amounts of body-mounted sensors is undesirable or impossible. As a tool PSG is also uniquely poorly suited to population studies or long-term monitoring of patients. For all of these reasons and more, reasonable and clinically applicable alternatives to PSG are in demand.

Respiratory polygraphy (RPG) is a less comprehensive method used in cases where respiratory disturbances during sleep is strongly suspected as the cause of a patient's difficulties. It commonly records airflow, nasal pressure, respiratory effort, and oxygen desaturation. However, albeit more lightweight than PSG, RPG still requires substantial body-mounted sensors, and it requires a similar interpretation process[10].

Simpler methods, such as wrist-mounted motion sensors (actigraphy) and patient-written

sleep diaries, are options in cases where PSG and RPG are unnecessary or impossible, and when long-term monitoring of individual patients is desirable. However, sleep diary data is encumbered with subjectivity and recall challenges. Actigraphy data are, while objective, sparse and of suboptimal accuracy[31].

There are currently no consumer devices on the market which can reliably screen for or indicate sleep apnea. Analysis of single channels of data have shown promise for quantifying sleep disordered breathing in certain settings, as well as distinguishing central from obstructive apnea, but whether a combination of clinical information and a consumer wearable device could provide risk stratification for sleep apnea remains a testable hypothesis[27].

2.3 Sleep Apnea

Sleep apnoea is a typically chronic sleep disorder characterized by cessation of or periods of shallow breathing during sleep. These pauses can last from a few seconds to several minutes, and may occur thirty times or more in a single hour[24]. In most cases, the sleeper himself will not be aware of these cessations, as they typically do not lead to a full awakening, and even when they do the cause may not be obvious to the patient. They are however accompanied by a number of other symptoms.

Sleep apnea is associated with symptoms including extremely loud snoring, persisten daytime sleepiness, long sleep latency (i.e. trouble falling asleep), trouble staying asleep, and dry mouth and headaches in the morning. Left untreated, sleep apnea has been shown to have serious and life-shortening consequences, including but not limited to heart disease, high blood pressure, diabetes, weight gain, stroke, chronic pain, depression, accidents caused by daytime sleepiness, and more.[1][15]

Sleep apnea is also estimated to be very common, affecting an estimated one out of six norwegians between the ages 30 and 65[15]. And while risk factors involve being male, overweight, and above 30, anyone at any age can be affected, even children[1].

In spite of all this, it is estimated that a majority of those affected remain undiagnosed and untreated. Dedicated sleep clinics often have long waiting lists and capacity trouble, and the condition cannot be detected by a routine examination by a general practitioner.[1][15]

2.3.1 Types of sleep apnea

The AASM Manual for the Scoring of Sleep and Associated Events[4] contains scoring rules for five primary types of respiratory disturbances in adults:

Obstructive Sleep Apnea (OSA) is the most common type of apnea, arising from a physical blockage of the airway. Typically, soft tissue at the rear of the throat collapses backwards to block the airway, resulting in no air entering the lung when the sleeper expands the chest to inhale[1]. In PSG, OSA is characterised by a continued effort to breath in the absence of airflow.

Central Sleep Apnea (CA) is characterized by a cessation of airflow accompanied by a absent inspiratory airflow for the entire period of absent flow. In other words, during a central apnea episode, the airway is not blocked, but the brain fails to signal the muscles to breathe.

Mixed Sleep Apnea (MA) is a combination of obstructive and mixed apnea.

Hypopnea events are periods of overly shallow breathing that do not meet the strict apnea criteria. These can also be categorized as obstructive, central, or mixed, but doing so is optional and not common.

Cheyne-Stokes Breathing is not a type of apnea in itself, but rather a pattern of consecutive apneas separated by a crescendo and decrescendo change in breathing amplitude, in cycles usually taking 40 seconds to two minutes.

2.3.2 Diagnosing sleep apnea: The Apnea-Hypopnea Index

Sleep apnea is typically diagnosed from PSG (or RPG) data, using the visual rules found in [4]. Chapter VIII Part 1 of this manual, Respiratory Rules for Adults, is included as Appendix **??**.

In general, an apnea is scored as a cessation of airflow by a ≥90% amplitude drop in a flow

sensor lasting longer than 10 seconds. The effort sensors then characterize the type of apnea event. A hypopnea is scored similarly, but from a ≥30% amplitude drop, accompanied by a ≥3% oxygen desaturation from the pre-event baseline.

During the scoring, the analyst will recognize apnea episodes according to these rules. She will mark them, and they will form the basis for the Apnea-Hypopnea Index, which is simply the number of apnea events divided by the number of hours of sleep. AHI is categorized as follows:

- Normal: 0-4
- Mild sleep apnea: 5-14
- Moderate sleep apnea: 15-29
- Severe sleep apnea: 30 or more

2.4 Equipment Used in this thesis

This thesis uses PSG data aquired from St. Olav's hostpital patients. This data has been recorded using SOMNOscreen™plus polysomnography equipment supplied by Somnomedics[29], with DOMINO software for analysis and scoring. The box itself contains a light sensor, body position detection, movement artefact detection, internal thoracic or abdominal effort detection as well as a pulse oximeter for recording SpO2, pulse rate, and pulse waveform. Standard external sensors include a miniaturized headbox with referential inputs for EEG/EOG, EMG, and ECG, and their respective electrodes/sensors, thermistors for recording nasal and oral flow, microphone for recording snoring sound, nasal cannula pressure sensor for additional flow and snore recording, effort sensors for thoracic and abdominal effort recording (St.Olav's uses RIP Belts), and an activity sensor.

The SOMNOscreen™plus employs active filtering, with adjustable highpass and lowpass filters (see Appendix [BLAH], St.Olav's global PSG filter settings) as well as bandpass filters at 50Hz for all sensors. Sampling rates vary from 4-512 Hz, and all sensors use 16 bit ADC. Further tech-

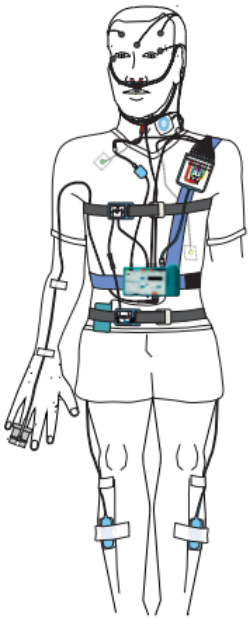


Figure 2.5: Somnomedics SOMNOScreen™plus PSG equipment configuration. Figure from [29]

nical specifications of each sensor are trade secrets which have not been aquired for this project.

The exact sensor configurations used by the hospital for each recording vary slightly - all use the PSG configurations, but the number of EEG and EMG electrodes vary slightly. Since the focus of this thesis is on sensors other than EEG and EMG, this is not regarded to be a problem. The hospital has not kept records of exactly which box has been used for each recording, so it is not possible to retroactively see which exact sensor has been used for each data set.

The Domino software performs a wide array of automatic analyses. If these are used at all, they function as an aid to the scorer, and will be corrected by the manual scoring process[9].

3

Multivariate analysis

This chapter summarizes relevant information from [21], [11], and [32] pertaining to the multivariate analysis tools used in this thesis.

3.1 The concept of latent variables

The fundamental hypothesis of latent variable methods is that the information contained in a matrix $\mathbf{X} \in \mathbb{R}^{MxN}$ of M objects[1] and N variables is in fact concentrated in $A \ll N$ underlying ('latent') variables, often referred to as 'scores'[21].

$$(t_1, ..., t_A)' = h_1[(x_1, .., x_N)'] \tag{3.1.1}$$

The by far most common latent variable methods use linear modelling, so that

$$\mathbf{T} = \mathbf{XV} \tag{3.1.2}$$

with $\mathbf{V}$ some linear transformation (the determination of which characterizes the method). By identifying the underlying factors $\mathbf{T}$, we can build a compressed model of $\mathbf{X}$, which will simplify statistical modelling by (often drastically) reducing the number of necessary model parameters to be estimated in an $\mathbf{X}-\mathbf{Y}$ regression. Moreover, the latent variables can be a powerful aid in visualizing the $\mathbf{X}$ data itself, by giving reduced-dimensional 'windows' into the high-

[1] *objects* are sometimes also called *observations*

dimensional **X**-space.

For example, in a data set of physical measurements (height, weight, shoe size etc.) taken from multiple people, all the measurements would be correlated because they'd depend on overall body size. This correlation would manifest itself in a single latent variable. In a data set of temperature variations from multiple places over one year, one latent variable could indicate north-south configuration, while another could be related to height above sea level, or separate coastal from inland locations, etc.

Of course, with plentiful causal knowledge of the system, one could choose a sophisticated mathematical model to use and regress for parameters at leisure. Challenges arise when the system is not already well known. This thesis will focus on bilinear methods, a flexible class of methods that require very little prior system knowledge. These 'soft modelling' tools estimate the latent variables directly from the data itself, requiring minimal a priori inferences from the analyst, but instead gives pragmatic data compression that when used interactively with good graphics can give good predictive ability *and* good causal insight at the same time[21].

3.2 Bilinear Modelling

The bilinear models considered in this thesis are projection methods, and have simple geometric interpretations as projections of the **X**-matrix (M points in an N dimensional space) onto an A-dimensional hyperplane; in other words, the bilinear model factors are intended to *map* the non-random structure of the data in a way that reveals the underlying phenomena causing the variations. The phenomena can then be studied graphically in the resulting compressed model, and hopefully interpreted in a meaningful way[21].

In a bilinear model, data set **X** is modelled as

$$\mathbf{X} = \mathbf{T}\mathbf{P}' + \mathbf{E} \tag{3.2.1}$$

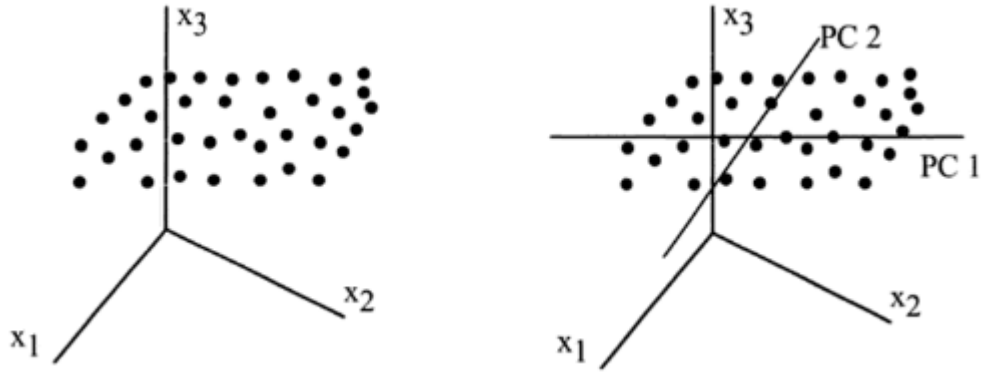


Figure 3.1: Geometric representation of PCA. To the left, a quasi-planar cloud of data points is plotted in three dimensions. To the right, observe how the two principal components in the directions of maximum variance can be used to plot the data in two dimensions, with little loss of information. Figure from [11]

The scores and loadings $\mathbf{TP}'$ represent the *structured part* of the data, while the residual $\mathbf{E}$ represents the unique variation not explained by the A-factor bilinear structure.

The loading matrix $\mathbf{P} \in R^{MxA}$ defines a basis for the A-dimensional space in which we want to map the data - each column p_i, or *loading*, defines a direction in the space. The loadings are used to examine relationships among *variables* in a data set.

The scores matrix $\mathbf{T} \in R^{NxA}$ contains the coordinates of the projection of each object in $\mathbf{X}$ onto the loadings vectors - each column $\mathbf{T}_i$, or *score*, can be interpreted as an approximation of an object in the data set in terms of the corresponding loading. The scores are thus used to examine relationships among the *objects*.

The distance from a point in the original coordinates to its position in the projection formed by the PCA are called the *object residuals*. If these distances are large, it implies that the fit is not good, and thus the model does not represent the data very well.

The products $t_i p_i$ are called bilinear factors, or components. The number of factors A, also known as the dimension of the model, is determined by how many such factors we chose to include, and determines the *model complexity*. The maximum number of factors is the smallest of $N-1$ or M, but usually we will use much fewer than the maximum. In many cases, a large

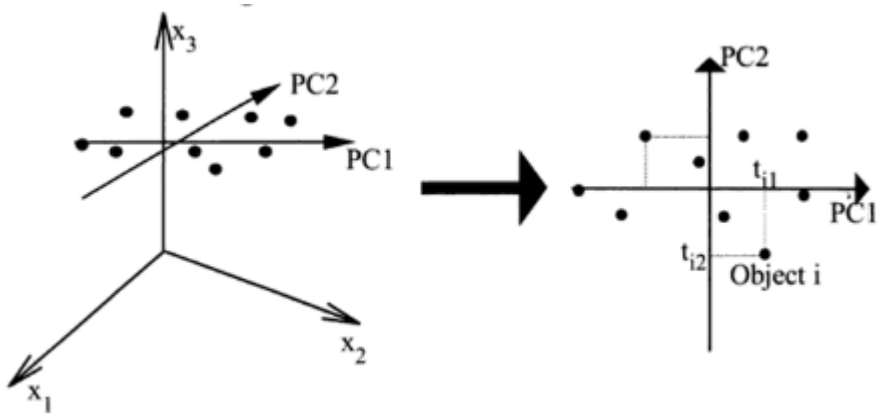


Figure 3.2: The scores map the original data points into the PC coordinate space. The loadings define the direction of the PC vectors in relation to the original coordinates. Figure from [11]

number of variables can be represented by a much smaller number of PCs, which allows us to *see* structures in data whose original dimensionality would be much too high for direct observation, albeit only as approximate projections.

Determining the desired number of components A is of high importance - too many can lead to overfitting (including factors that in reality are the products of pure noise in the data) and unnecessarily difficult interpretation, while too few might leave important phenomena unmodeled and ignored (see Section 3.7).

3.3 Pretreatment

It is common to preprocess the **X**- (and, in the case of regression, **Y**-) variables to a suitable form before initiating analysis. The most common pretreatment strategies are described in most texts detailing bilinear methods, including [21], [32], and [18], and are summarized here.

3.3.1 Centering

Centering is done by subtracting the averages of all variables in **X** and **Y**. This removes variable offset, and increases ease of interpretation and numerical stability, while otherwise leaving the interpretation results unchanged[32].

3.3.2 Auto-scaling

The goal of multivariate analysis is to reveal the relationships between variables. This is difficult if there are significant differences in magnitudes - the larger variables will tend to dominate the analysis. Intuitively, if you are comparing a mouse to an elephant, the elephant will tend to dominate the picture unless you either reduce the size of the elephant or increase the size of the mouse. In essence, we must equalize the obvious before we can observe the interesting.

The standard approach to auto-scaling is to scale each variable to unit variance by dividing it by its own standard deviation, and centering them by subtracting their averages. This leaves all variables expressed in standard units (i.e. zero mean and unit standard deviation), and is a simple, useful, but conservative way of giving every variable an equal chance to contribute to the modelling[21][32].

3.3.3 Weighting

Auto-scaling is simple and often efficient, but should not be used blindly. Martens and Næs[21] chapter 7 (page 320) exemplifies some pitfalls of the process; if, for instance, 99 variables are of one kind and one single variable is of a different kind, standardization through auto-scaling will give a high probability of reaching the noise in the 99 before the 100th has been modelled. We also risk the amplification of meaningless variables - if, for instance, one variable is nearly constant except for a small amount of random noise, the auto-scaling procedure described above will serve only to amplify this noise to the same a priori importance as an informative variable with a much better signal/noise ratio.

Weighting, also known as manual scaling or just *scaling*, is used to increase or decrease the influence of specific variables or objects on the resulting model. One simply multiplies a column (variable) or row (object) in the $\mathbf{X}$ matrix with some constant weight to either increase or decrease its relative magnitude and thus its relative importance, with the choice of weight done on the basis of a priori knowledge or experience. With an appropriate scaling, one can choose to downweight particularly noisy or unimportant variables or objects, increase the influence of

variables or objects known or suspected to have greater importance than others, or, in the case of regression, focus the model towards more important **Y** variables.

3.3.4 Variable selection

Variable selection involves removing a subset of the available variables, usually because they are believed to be unimportant or irrelevant in the current setting. This will often improve the prediction ability of a regression, but *should not* be used carelessly. Removing too much *will* lead to unrealistically positive results - after all, if all one cares about is a perfect linear regression, one may simply remove all but two data points!

3.4 Principal Component Analysis

Principal Component Analysis (PCA) is a bilinear latent variable method, often used in conjunction with graphical tools to make certain data sets easier to explore and visualize[2]. Intuitively, it can be thought of as fitting an A-dimensional ellipsoid to the N-dimensional cloud of M data points (recall that N is the number of variables and M the number of objects). The axes of the ellipsoid are the principal components (PCs), and each represent a direction of maximum variance within the data set, under the constraint that the PCs must be mutually orthagonal and thus *uncorrelated* with each other.

Formally, the first loading $\mathbf{p_1}$ is defined as the unit vector that maximizes the empirical variance of $\mathbf{p_1'X}$[3], i.e.

$$\begin{aligned} \max_{p_1} \quad & \mathbf{p_1} = \mathbf{p_1'X'Xp_1} = \mathbf{t_1't_1} \\ \text{subject to} \quad & ||\mathbf{p_1}|| = 1 \end{aligned} \tag{3.4.1}$$

with $\mathbf{t_{1(i)}} = \mathbf{x_{(i)}p_1}$ being the score that corresponds to the loading $\mathbf{p_1}$, and $\mathbf{t_1p_1'}$ the first principal component. We know from fundamental linear algebra that for a symmetric matrix such as $\mathbf{X'X}$, the maximum possible value occurs when the loading $\mathbf{p_1}$ is the unit eigenvector corresponding

[2] PCA is most commonly employed and most useful in cases where the data is suspected to be collinear, or in other words, when we expect the variables to be correlated

[3] After X has been properly preprocessed.

to the largest eigenvalue of the correlation matrix $\mathbf{X}'\mathbf{X}$.

Further components are found by first subtracting the preceding principal components from X;

$$\hat{\mathbf{X}}_\mathbf{k} = \mathbf{X} - \sum_{i=1}^{k-1} \mathbf{t_i}\mathbf{p}_\mathbf{i}' \tag{3.4.2}$$

and then repeating the optimization problem for this new data matrix,

$$\begin{aligned} \max_{p_k} \quad & \mathbf{p_k} = \mathbf{p}_\mathbf{k}'\hat{\mathbf{X}}_\mathbf{k}'\hat{\mathbf{X}}_\mathbf{k}\mathbf{p_k} = \mathbf{t}_\mathbf{k}'\mathbf{t_k} \\ \text{subject to} \quad & ||\mathbf{p_k}|| = 1 \end{aligned} \tag{3.4.3}$$

It can be shown that this procedure gives the remaining eigenvalues of $\mathbf{X}'\mathbf{X}$ in decreasing order, and the procedure is repeated until the desired number A principal components is obtained.

If we let $A = N$ and extract all eigenvalues of the correlation matrix $\mathbf{X}'\mathbf{X}$ (the smallest of which may be close to or at zero), then $\mathbf{X}$ can be written as exactly $\mathbf{X} = \mathbf{TP}'$, i.e. the scores and loadings model $\mathbf{X}$ exactly and the residual matrix $\mathbf{E}$ is zero. This representation of $\mathbf{X}$ can quite easily be re-written to form exactly the Singular Value Decomposition (SVD) of $\mathbf{X}$. PCA can thus also be obtained from SVD by extracting the elements corresponding to the A largest eigenvalues. With $A < N$, the data matrix $\mathbf{X}$ is *approximated* by $\mathbf{TP}'$, forming the model $\mathbf{X} = \mathbf{TP}' + \mathbf{E}$[21].

3.5 Partial Least Squares Regression

It is possible to use PCA for prediction, by regressing $\mathbf{Y}$ on the scores $\mathbf{T}$ of $\mathbf{X}$ (Principal Component Regression, PCR). However, there may be major eigenvectors in $\mathbf{X}$ that have no relevance for modelling $\mathbf{Y}$. The Partial Least Squares Regression (PLSR) is another bilinear method, which uses the $\mathbf{Y}$ variables actively during the decomposition of $\mathbf{X}$ - in other words, the model is based on the latent structures present in and connecting both data sets.

PLSR finds A $\mathbf{X}$-scores that at both model $\mathbf{X}$ and work as predictors of $\mathbf{Y}$. In other words, we assume that both $\mathbf{X}$ and $\mathbf{Y}$ can be modelled by the same latent variables. The following explana-

tion is a summary of information found in [32]:

The **X**-scores $\mathbf{T} \in \mathbb{R}^{MxA}$ are estimated as linear combinations of the original variables $\mathbf{x_k}$ with coefficients called 'weights', $\mathbf{W} \in \mathbb{R}^{NxA}$, and $\mathbf{W}^* = \mathbf{W}(\mathbf{P^T W})^{-1}$, where **P** are the **X**-loadings, transforms **X** to **T**:

$$\mathbf{T} = \mathbf{XW}(\mathbf{P^T W})^{-1} = \mathbf{XW}^* \tag{3.5.1}$$

The **X**-scores are designed to model **X**, i.e. minimize the **X**-residuals in 3.5.2. If we have multivariate **Y**, we will also find **Y**-scores and **Y**-loadings **U** and **C** to model **Y**, also designed to minimize the residuals **G** in equation 3.5.3.

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} \tag{3.5.2}$$

$$\mathbf{Y} = \mathbf{UC}' + \mathbf{G} \tag{3.5.3}$$

So far, this is the same as PCA. However, in PLSR we are looking for the latent variables in **X** that are most correlated with the latent variables in **Y**. Therefore, the **X**-scores are designed to have the second property that they also are good predictors of **Y**:

$$\mathbf{Y} = \mathbf{TC}' + \mathbf{F} \tag{3.5.4}$$

Equations 3.5.1 and 3.5.4 combine to form the PLSR model,

$$\mathbf{Y} = \mathbf{XB} + \mathbf{F} \tag{3.5.5}$$

where $\mathbf{B} = \mathbf{W}^*\mathbf{C}' = \mathbf{W}(\mathbf{P^T W})^{-1}\mathbf{C}'$ are called the 'PLSR coefficients'.

As with PCA, it can be shown that the PLSR problem reduces to an eigenvalue problem. We quote [32]: "The first weight vector $\mathbf{w_1}$ is the first eigenvector of the combined variance-covariance matrix $\mathbf{X'YY'X}$, and the following weight vectors (component a) are eigenvectors to the deflated versions of the same matrix, i.e., $\mathbf{Z'_a YYZ'_a}$, where $\mathbf{Z_a} = \mathbf{Z_{a-1}} - \mathbf{T_{a-1}P'_{a-1}}$. Similarly, the first score vector ($\mathbf{t}_1$ is an eigenvector to $\mathbf{XX'YY'}$, and later X-score vectors ($\mathbf{t}_a$) are eigenvectors of $\mathbf{Z_a Z'_a YY'}$"

3.5.1 The NIPALS algorithm

The NIPALS ("Nonlinear Iterative Partial Least Squares") algorithm developed by H. Wold is one of the most prolific methods for calculating bilinear models, and is the one used in this thesis. It was first developed for PCA in 1966, and later adapted for PLSR. Today, it exists in several versions, including PLS1 and PLS2, for calculating PLS with one or many Y variables respectively. For implementation details, pseudocode, and thorough discussion, the reader is referred to [11].

3.6 Interpreting the graphics

Both PCA and PLSR can be used to generate a large number of plots of different types. *The Unscrambler®X* generates 16 classes of graphics from one PLSR, most with multiple sub-plots and a wide array of options. This section will only demonstrate some of the most frequently used, focusing specifically on the ones that are included for illustration later in this thesis. The exception is the explained variance plot, which is related to validation, and therefore described in section 3.7.3. However, while performing analysis it is important to also have an understanding of the ones not mentioned in this text, so for a more thorough introduction, the reader is referred to [11].

3.6.1 The Score Plot

A score plot is a two- or three-dimensional plot of the objects in the data matrix mapped by any pair or trio of principal components, and can be viewed as 'windows' into the PC-space. The score plot is a powerful visual aid in interpreting the calibration model, used for outlier identification, recognition of trends, group identification and more.

In a score plot, each dot represents one object. When interpreting the plot, we may mark these dots either with their name, separate them according to category, or connect them with lines.

For example, in section 6.1, we wish to interpret PCA of PSG from a single patient, organized

in a time series format[4]. Plotting the scores of our analysis over the first two PCs, we get the result shown in figure 3.3. This is not immediately easy to interpret, however by coloring the apnea time instances, figure 3.4 clearly shows that apnea objects tend towards the negative along PC1, and normal breathing towards the left. No a priori information of apnea was included in this particular calibration, yet PC1 still lies in this particular direction. PC1 is the most dominant component, and carries 44% of the total variance, so we have an indication that apnea is an important latent variable in the PSG data of this patient.

In a similar manner, we will later see that PC2 appears to carry a connection between pulse and blood oxygen saturation.

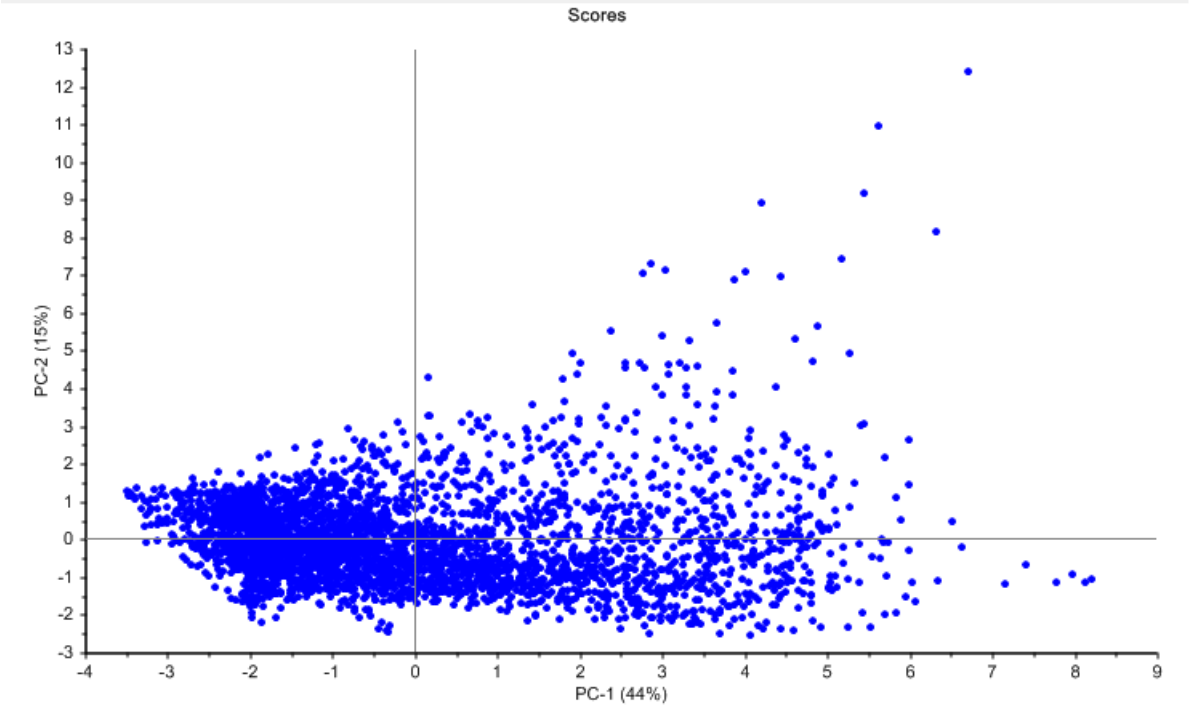


Figure 3.3: Scores for PCA of patient 2407, PCs 1 and 2

[4]This analysis will of course be discussed in more detail in section 6.1. It is included here only as relevant illustration of the concepts in question.

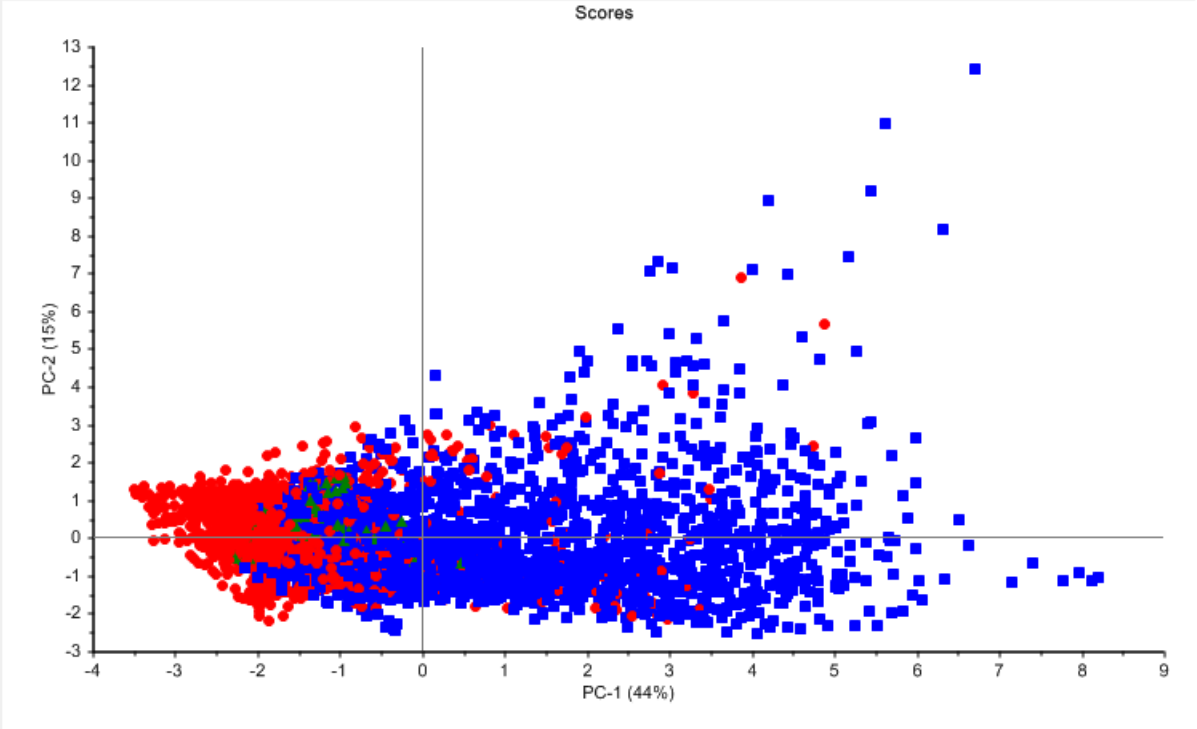


Figure 3.4: Scores for PCA of patient 2407, PCs 1 and 2. Red dots indicate obstructive apnea, green hypopnea, and blue regular respiration

3.6.2 The Loadings Plot

The loadings plot concerns itself with illustrating relationships between the *variables*; specifically, how much each variable contributes to each PC. In the correlation loadings plot, a value of zero in relation to a PC indicates no contribution from that variable, while a value close to one indicates very high importance. Furthermore, variables that appear clustered in the loadings plot will have a high amount of correlation.

The loadings plot in figure 3.5 is taken from the same analysis as the scores plot in the previous section. The clustered variables along the PC1 axis can be interpreted as highly correlated as well as of high importance to PC1. The Pleth-variable is also important to PC1, but has a negative correlation (i.e. negative coefficients in the linear combinations) with the PC1-relevant variables. From the scores plot interpretation in the previous chapter we already suspect that PC1 may have a relation with apnea, and we are now able to single out the variables that appear to have the highest importance in relation to that particular latent variable.

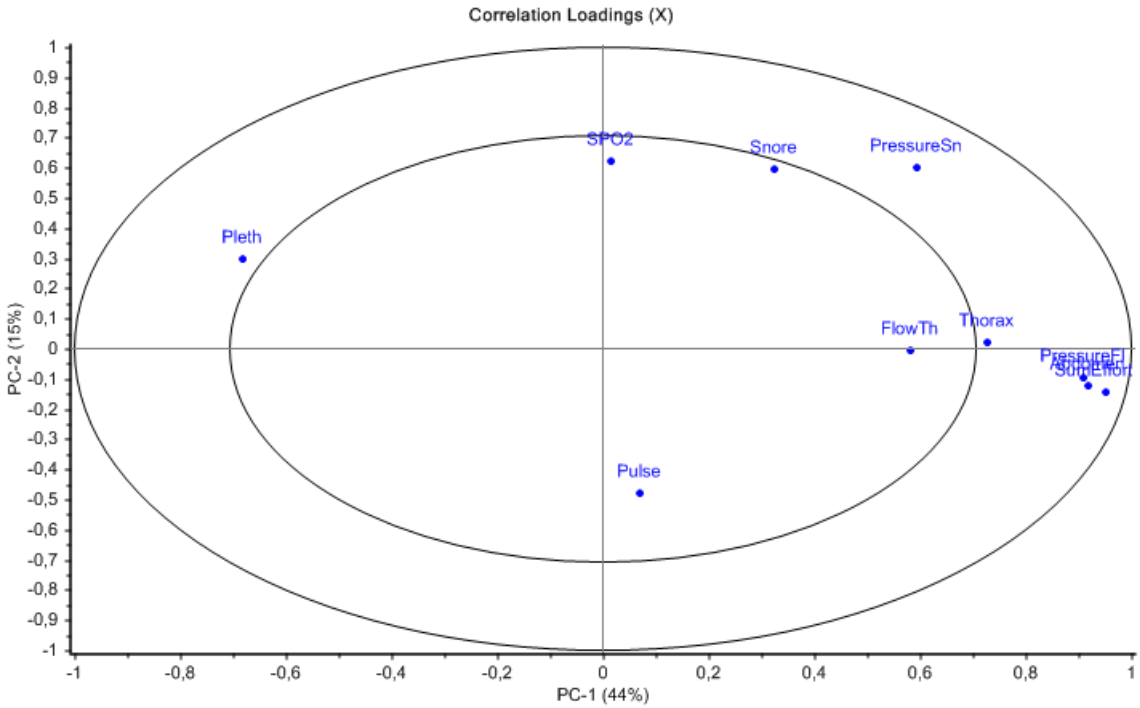


Figure 3.5: Loadings for PCA of patient 2407, PCs 1 and 2. The cluster to the right contains the Pressure Flow, Abdomen, and Sum Effort sensors

This example also illustrates that the scores and loadings plots are complimentary, and of most value when used in conjunction.

In PLSR, we must also consider the loading weights **W**, which represent the effective loadings that are the most directly connected to building the regression relationship between **X** and **Y**. While the loadings and the loading weights often end up quite similar, the difference between them tells us how much the guidance from **Y** has affected our decomposition of **X**.

PLS also has the set of **Y**-loadings, **C**. These, as well as the weights, are used to interpret relationships between the **X** and **Y** variables, and to decipher patterns in the scores related to these.

3.6.3 The influence plot

The final plot we will consider in this thesis is the influence plot, which is a plot of residual variance against leverage. High residual variance means poor model fit, while high leverage means having a large effect on the model. Therefore, samples in the upper right corner (large contribution to the model and high residual variance) are potentially dangerous outliers[11].

Adding more PCs decreases the residual variance, and even outliers will eventually fit in the model. However, the model now concentrates on describing the variations due to these few different samples instead of modeling the variations in the whole data set. In the influence plot, this can be seen as outliers migrating from the upper left to the lower right corner, as illustrated in figure 3.6.3.

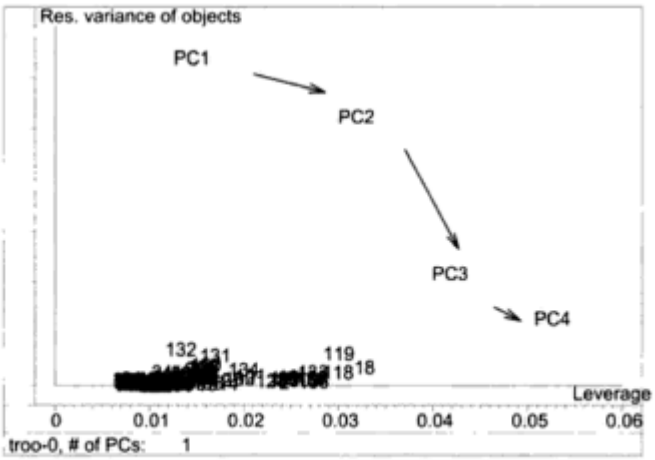


Figure 3.6: Behaviour of outliers in an influence plot. Figure from [11]

3.7 Validation

Although other contexts exists, validation for the purpose of this thesis refers to a measure of the prediction performance of a multivariate model [5]. Esbensen and Guyot[11] describes the purpose of this type of validation as two-fold:

[5] In section 6.1, validation is a measure of *modeling* rather than *prediction performance*, as we do not have any predicted variables in this section. The concepts are very similar, so modeling performance validation will not be treated separately.

> First of all, validation is absolutely essential in order to make sure that the model will work in the future for new, similar data sets [...]. Secondly, validation is often also used in order to find the optimal dimensionality of a multivariate model (**X**,**Y**), i.e. to avoid either overfitting or underfitting[11].

In essence, validation is simply the act of testing the multivariate model on data that was not used to produce it.

3.7.1 Independent test set

Independent test set validation is often considered the 'most ideal' validation method. It requires access to two separate and independent but comparable data sets, and uses one set to calibrate the model and another *only* for validation. In this thesis, we do not have access to two such sets, and while we conceivably could split our samples into two sets and use one for calibration, the other for validation[6], our data sets are not large enough to afford this. For this reason, independent test set validation will not be described in detail here. For more information, the reader is referred to [11].

3.7.2 Cross validation

Cross validation involves partitioning the calibration data set into K segments, and performing the calibration K times, each time omitting one of the segments. The omitted data is predicted by the resulting model, and the difference between the predicted and the true modeled values for the omitted samples are summed and averaged to give the validation results and statistics.

In *segmented cross validation*, each segment k contains multiple objects, for example 10% of the total samples. In *full cross validation*, also known as Leave-One-Out (LOO) validation, each segment contains only one single object. Segmented cross validation is preferable when there is a relative abundance of samples, but not enough to build a proper independent training set (or we do not know how to choose which samples to choose for the set), or when full cross

[6] This would actually be a 'test set switch' situation, which is a form of cross validation[11]

validation would be too time consuming (if each segment is 10% of the total samples, then full cross validation takes 10 times longer than segmented validation).

3.7.3 The explained variance

The **explained variance** tells us about the goodness of fit i.e. how well the model fits the data, as well as its predictive ability, and can be used to determine how many PCs to use.

Let $z_{(i,j)}$ be any variable (prediction or predicted, i.e. x or y), and $\bar{z}_i$ the respective mean (zero for mean-centered data).

We first feed the original data through the model to get the predicted values $\hat{z}_{cal(i,j)}$, and then calculate the model residuals $e_{i,j}$ by subtracting the predicted from the true values. We then apply the model to the test set, get the validation predictions $\hat{z}_{val(i,j)}$ and the validation residuals $f_{i,j}$ accordingly:

$$e_{i,j} = z_{(i,j)} - \hat{z}_{cal(i,j)} \tag{3.7.1}$$

$$f_{i,j} = z_{(i,j)} - \hat{z}_{val(i,j)} \tag{3.7.2}$$

We calculate the *calibration residual sum of squares* SS_{res}, the *predicted residual error sum of squares* $PRESS$, and the *total sum of squares* SS_{tot} as follows:

$$SS_{res} = \sum_i \sum_j e_{i,j}^2 \tag{3.7.3}$$

$$PRESS = \sum_i \sum_j f_{i,j}^2 \tag{3.7.4}$$

$$SS_{tot} = \sum_i \sum_j (z_{i,j} - \bar{z}_i)^2 \tag{3.7.5}$$

We then calculate the Root Mean Square Error of Calibration, called RMSEC[7] or R^2, and the Root Mean Square Error of Prediction, called RMSEP or Q^2:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{3.7.6}$$

$$Q^2 = 1 - \frac{PRESS}{SS_{tot}} \tag{3.7.7}$$

If these statistics are zero, it means that the residual is equal to the original data, i.e. nothing is explained by the model. If they are 1, it means that the residual is zero, and so the entire original data is perfectly modelled and 100% of the variation is explained.

It is common practice to calculate both R^2 and Q^2 at each step of the calibration, i.e. obtain a value after each component we add to the model. We may then plot these values against the number of components. This is the **explained variance plot**.

Note that by the definition of the multivariate methods, PCs are found to minimize residuals at each step. As such, the modeling error will always decrease as more components are added. However, this *does not* mean that its predictive ability increases! A good model will describe only *systematic variations*, not the random variation due to noise. Including too many components makes the model too detailed; it is *overfitted*.

The validation explained variance will typically increase up to a point, before either stagnating or decreasing again. When choosing the number of components to include, the point at which the validation explained variance fails to improve should be considered the upper limit.

[7] Also known as the coefficient of determination

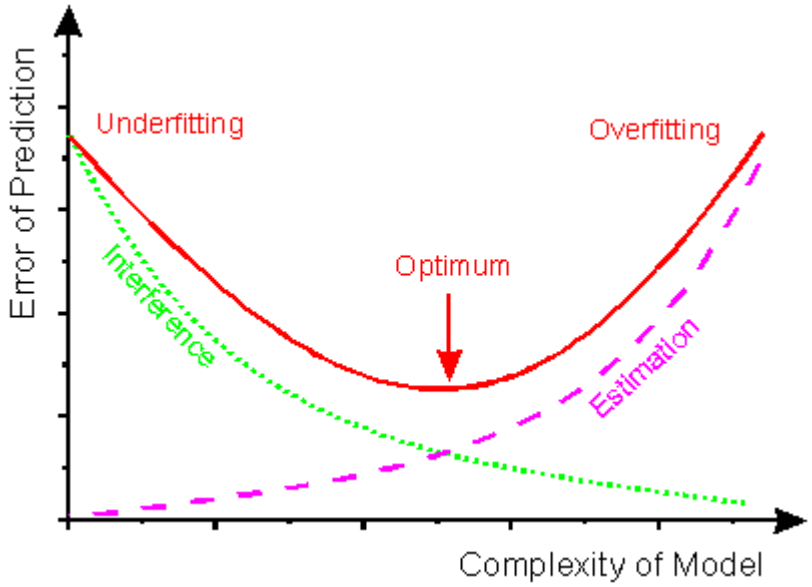


Figure 3.7: The V-shaped empirical prediction error curve: The optimal number of components is that which achieves a balance between minimizing both prediction error and statistical uncertainty. Figure from [8]

4

Data Aquisition

Figure 4.1 summarizes the steps involved in aqcuiring data and preparing it for use. The following chapter will describe each step in detail.

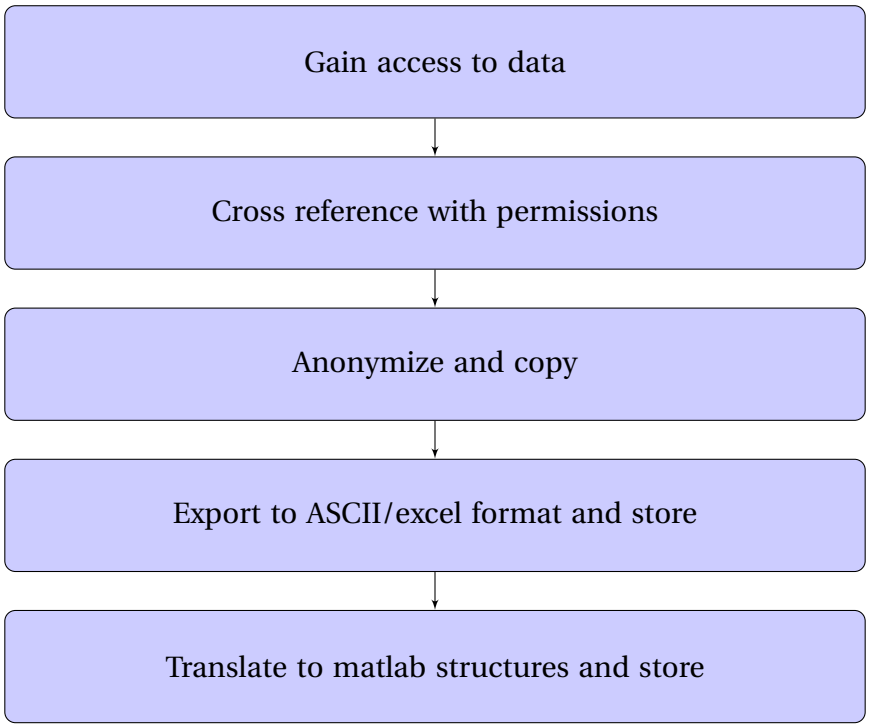


Figure 4.1: Data acquisition

4.1 Accessing St.Olav's patient database

To initiate this study, it was necessary to gain access to quality sleep data. This thesis was proposed as part of the output of a project with the primary goal of adapting the Novelda Xethru® Ultra Wideband Radar into a reliable tool for sleep analysis, and in December 2015 the project filed an application with the Regional Ethics Committee (REK) to gain access to St. Olav's hospital archives of PSG data. The application was evaluated in a committee meeting January 20th 2016, and on February 10th of the same year the project received a letter granting limited access for the duration of 2016, provided that all data be anonymized, and informed consent acquired from every patient involved. An information letter satisfying the demands set by REK was written by Novelda, and Novelda provided the funds to have these printed en masse, shipped with return envelopes enclosed, and all handling fees. The majority of the logistics were handled by the present author. The letters were printed and packed by IndegaardANDSvenil AS. Envelope addressing had to be handled manually, as the patient list could not be sent to the external company.

The complete patient list of eligible PSG recordings, containing some 1700 entries, became available to the project on February 26th. However, patient privacy concerns dictated that permission had to be explicitly acquired from the hospital before this list could be viewed by any external personnel, including project members. Since no St.Olavs employee was available to handle the logistics of addressing consent form envelopes and data anonymization, it became necessary to establish a temporary visitation position at the hospital for the present author.

On March 1st 2016, Dr. Morten Engstrøm filed an application with the St.Olavs administration office, requesting for limited visitation rights at the hospital, including signing of NDAs, a key card, and a computer user at the hospital. The bureaucracy involved was somewhat substantial, but once it was complete it became possible to print the patient address list on stickers and attach these to the packaged envelopes.

The first batch of response forms was received in the first week of April, with more coming in

steadily after that. In total, the project received approximately 800 signed consent forms. These were then opened, catalogued, and cross-referenced against the complete patient list, in order to attain a list of patients eligible for the project.

Once this list was attained, it was necessary to locate the correct recordings in the hospital database. The database had not been designed with this application in mind, and it quickly became apparent that manually searching for the patients was out of the question. A MATLAB procedure was developed to cross-reference the folder structure with the consent list, and generate a batch file to copy the data in question to a safe-storage folder. From there, the recordings had to be manually renamed to anonymous patient ID's, using the DOMINO Somnologica software, before they could be made available to this thesis, Novelda, and the rest of the project collaborators.

Unfortunately for this project, St.Olavs had used a different software and hardware before 2012. The dubious comparability of data before and after the change compiled with the added logistical difficulties of including a search through the less easily parsed naming conventions of the old software led to the decision to discount the earlier recordings and use only data from after 2012. Furthermore, in the data after 2012 a prevalence of spelling errors in patient names adversely affected the success rate of the automated searching method. In the end, the database search had a somewhat disheartening hit rate, leaving the project with a database of 186 anonymous PSG recordings. This data was kept on external hard drives provided by Novelda.

4.2 Data Extraction and Formatting

To facilitate analysis, the anonymous PSG data had to be extracted from DOMINO proprietary files to a universal format. DOMINO supports raw and analysis data export as .txt and EDF+ file formats. For our purpose, .txt was chosen, and each anonymized patient was manually exported.

This project chose to export as much information as possible, including all raw sensor data in addition to manual and automatic analysis results, as well as the excel results-document containing 2508 fields of patient information and statistical summaries. A hierarchical folder structure was created to store the exported data.

Raw data files

The exported ASCII-files for raw sensor data is structured with 7 lines of header information followed by integer data separated by line shifts. The following is an example, from an Abdomen effort sensor file:

```
Signal Type: Effort_DC_Type
Start Time: 29.09.2014 20:00:04
Sample Rate: 32
Length: 1424512
Unit:

Data:
94
86
...
```

Although 'Unit' is included as a header line, it is not filled in for any of the acquired data sets and is thus useless.

Analysis files

The analysis files contain raw PSG results, and consists of 30 files per patient. These are structured with five lines of header information, followed by time and duration information of specific events. If no event has occurred or the analysis has not been preformed, the file is empty following the header. The following is an example of a FlowEvents analysis file:

```
Signal ID: FlowD\flow
Start Time: 17.03.2016 22:00:04
Unit: s
Signal Type: Impuls

23:18:15,284-23:18:36,358; 21;Hypopnea
23:27:14,890-23:27:34,422; 20;Hypopnea
00:43:49,038-00:44:01,256; 12;Obstructive Apnea
...
```

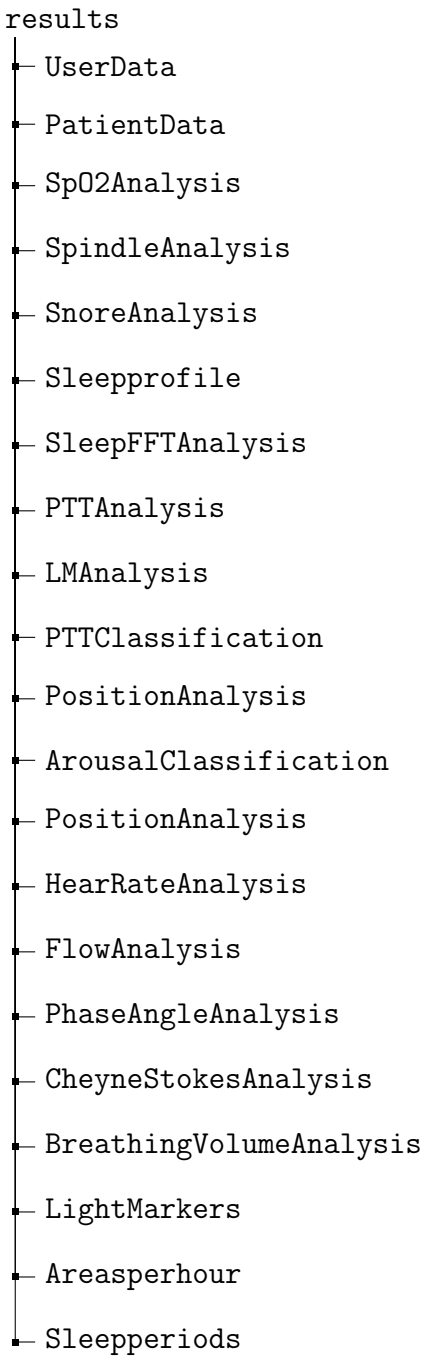
To allow for plotting and analysis of these files, MATLAB code was written to translate the events into numerical states and stack the information as timeseries of length equaling the total duration of the recording. 1 Hz resolution was chosen for simplicity.

The analysis file of greatest interest to this thesis is the FlowEvents file. In addition to the three apnea types and hypopnea, it contains information of Flow Limitation(FloL), indicating a limitation of flow not otherwise scored as an apnea, as well as Body Event (BodE), a disturbance caused by a large body movement disrupting the recordings.

We also note the Sleepprofile file, which contains all 30 second epochs of the recording with their identified sleep stage (Wake, N1, N2, N3, N4, or REM), in other words, the hypnogram.

Results file

The results-excel sheet contains a large number of scalar parameters, primarily numerical summaries of information found in the analysis-files, sorted into the categories listed as follows:



The actual information within each of the categories tends to be rather sparse, owing to not all analysis methods having been used for all patients.

In this thesis, most attention is afforded the FlowAnalysis category, which is where we find information about the patient's AHI, as well as the total durations, types, and number of apnea he experiences during the recording. We also chose to include some miscellaneous parameters

from SpO2 analysis and LM (leg movement) analysis[1], simply in case they would prove relevant.

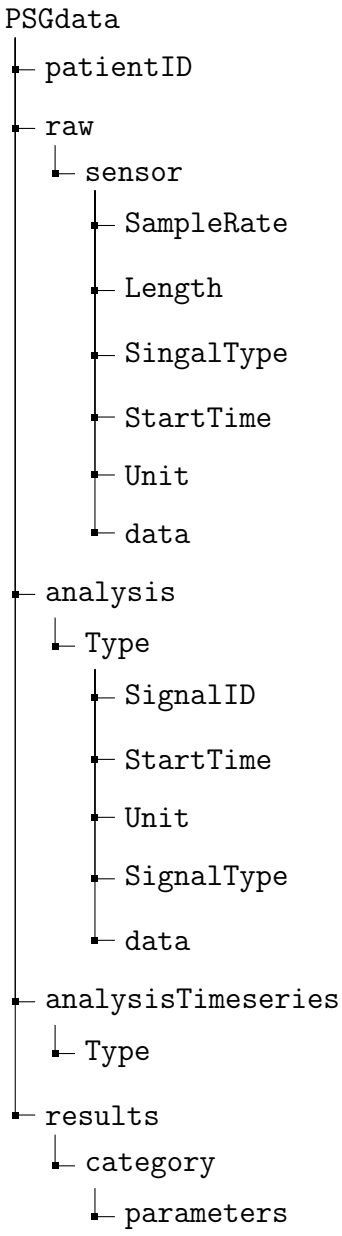
The following PSG output parameters are used in this thesis:

AHI	Apnea-Hypopnea Index
PLM	Periodic Leg Movement Index
DSI	Desaturation Index
DSA	Desaturation Average
DSL	Desautraion Length

The PLM and DSI have been calculated in a manner similar to the AHI - simply the average number of occurrences of the phenomena in question per hour. The DSA is the average size of all desaturations, and DSL is their average length.

MATLAB scripts were written to transform the ASCII files and the excel results into an easily readable .mat struct format, organized as follows:

[1] Periodic Leg Movement is a sleep disorder characterized by simple, repetitive muscle movements during sleep. It has not been treated in detail in this thesis, as it has not been the focus of the work. This index was included in analysis only in case of it proving to be influential.



4.2.1 Problems with the exporting

Consistency issues

The files were not consistently named (some had an 'Abdom'-file, others an 'Abdomen', some had 'Accu', etc). This had to be rectified.

More seriously, the recordings have not consistently used the same equipment, and do not

all use the same number of sensors. As a result, some recordings have fewer data files than others. For consistency, only those containing all the sensors in the chosen subset (see section 6.2.1) could be included. This led to discarding around 40 data sets.

Human error

The exporting itself was done manually, and some human error did occur during the process, only to be discovered rather much later. This lead to the loss of another 5 data sets.

5

Data Preprocessing

Proper pretreatment of data is instrumental for the multivariate analysis methods to work as intended. It is also important to visualize and examine data before analysis, both to aid in the choice of strategy, and to discover problems.

Since the optimal way to pretreat the data is unknown, this project chose an exploratory approach to the preprocessing procedures, including a number of options and carefully storing and labelling the data after each step. Figure 5.1 illustrates the procedure. Analysis can thus be preformed after each and every step, allowing for comparisons and assessment of the pretreatment options in addition to the data analysis itself.

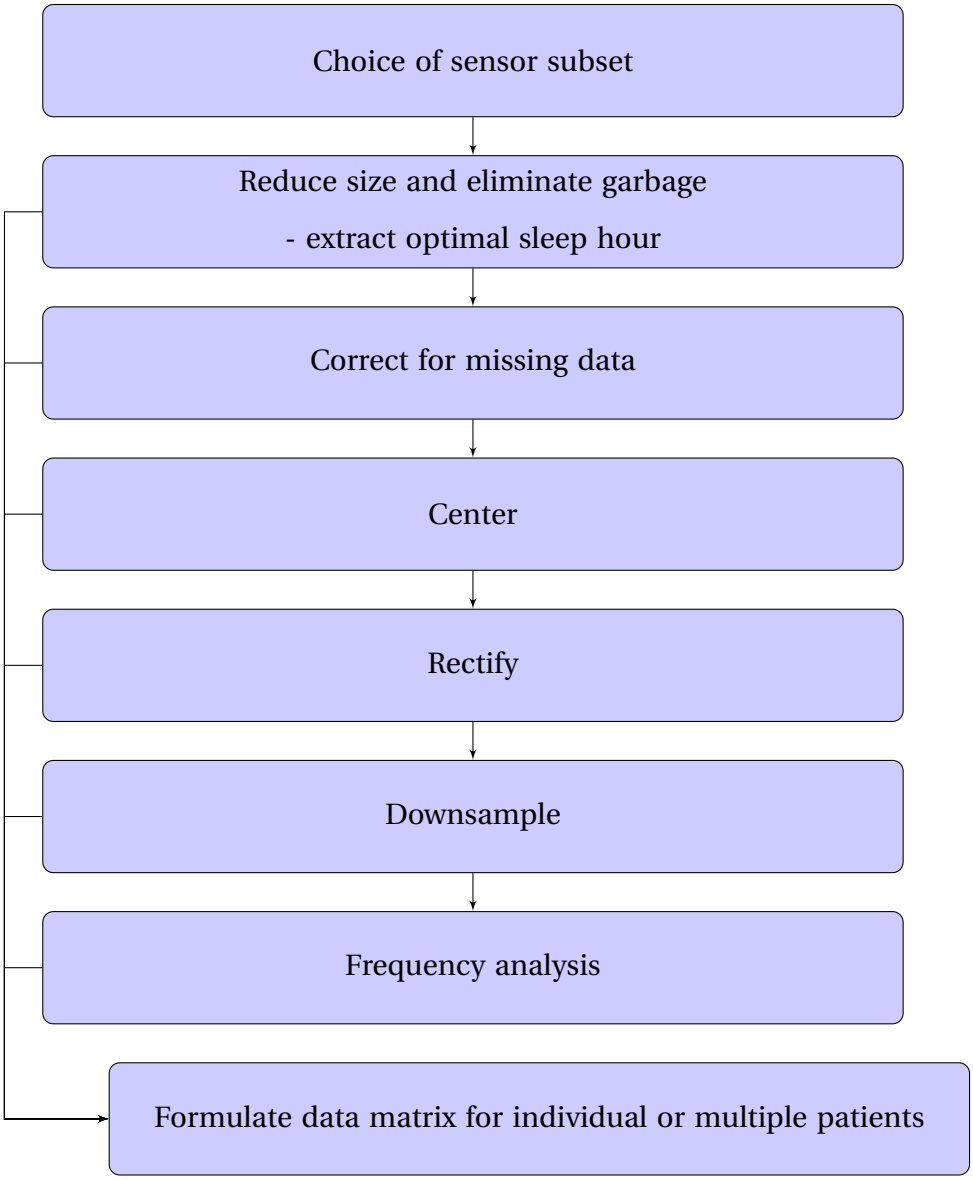


Figure 5.1: Preprocessing procedure summary

5.1 Choice of Sensor Subset

At least in the initial phase, project is less concerned with the fine details of high-resolution EEG and sleep staging, and more with the larger features of respiration, movement, and effort. As such, a subset of the data was chosen for visual examination, with the thought that other sensors could be included on a later stage. The chosen subset includes 10 sensors in total; all sensors directly related to respiration and flow, as well as those mentioned in the diagnostic rules

for sleep apnea. This includes abdominal and thoracic effort recorded with RIP belts (Abdomen and Thorax), effort as recorded with the main box (SumEffort), snore recorded with both microphone (Snore) and nasal cannula (Pressure Snore), pulse, SPO2 saturation and plethysmography recorded through the pulse oximeter (Pulse, SPO2 and Pleth in the figure), flow recorded using both thermistor (FlowTh) and nasal cannula pressure sensor (Pressure Flow). Figure 5.2 shows these sensors during one hour of sleep for a healthy patient.

From the large array of analysis data available from the extracted PSG, the ones chosen for further examination included the sleep profile, illustrating the sleep stages and periods of wakefulness throughout the night, and the flow events, showing apneas of all types, flow limitation events (periods of reduced flow where no apnea has been scored), and body events (artefacts). For some of the later analysis, SPO2 events (desaturations) were also included.

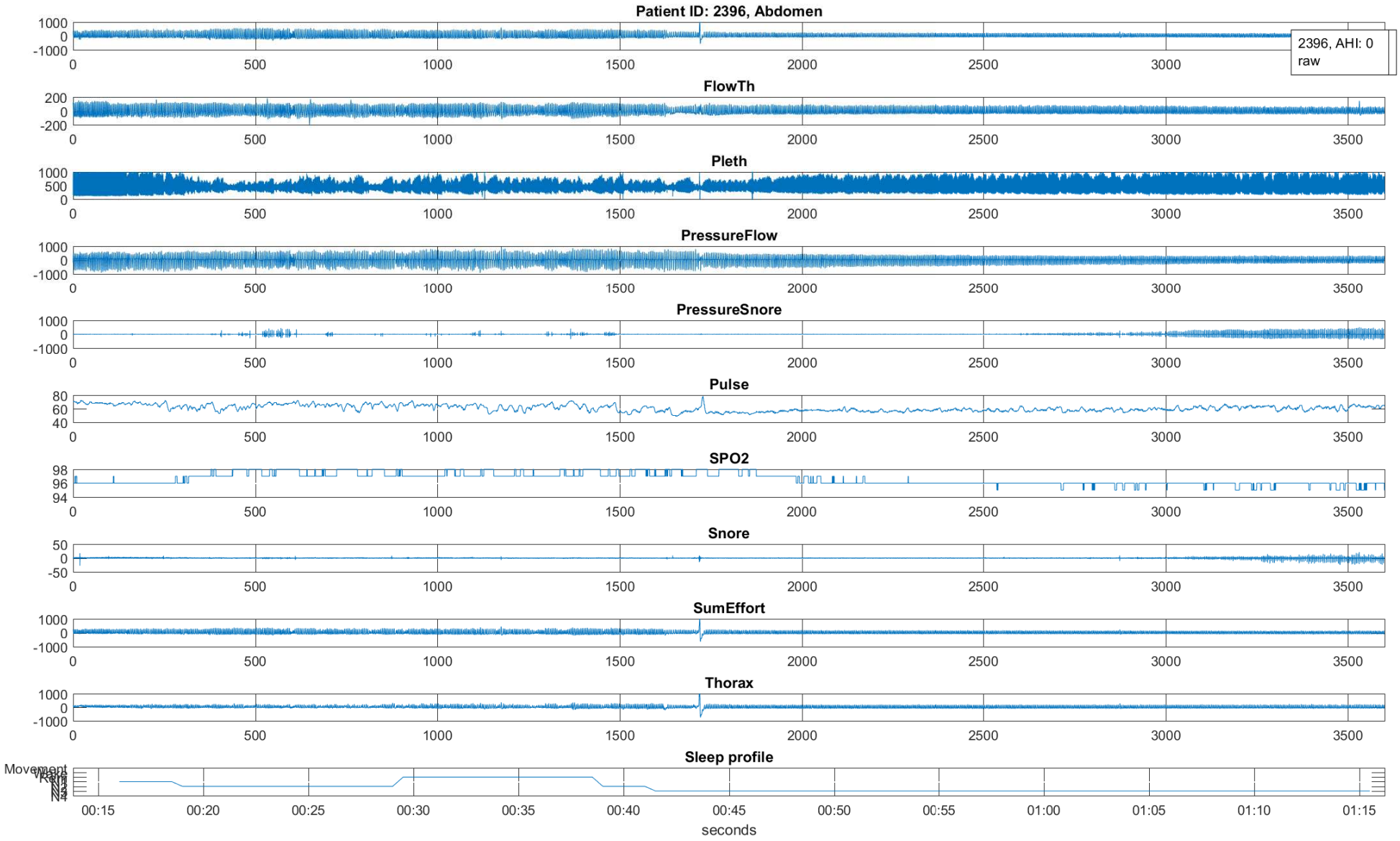


Figure 5.2: One hour of sleep for patient 2396, showing the chosen sensor subset. Sleep profile included for reference

5.2 Choice of 'optimal sleep hour'

While examining the data, it was obvious that not all patients slept steadily throughout the night. Some spend a long time before falling asleep, some wake early, and some wake up several times throughout the night. Although parameters such as sleep latency and sleep efficiency are interesting for future and more extensive work, as well as being frequently cited in clinical practice, considering them was deemed beyond the scope of this thesis. As such, it was desirable to consider only times during which the patients sleep steadily. The manually scored sleep profile analysis data was used to extract the hour of 'best sleep' - either the first hour of constant sleep, or the hour during which the patient has fewest seconds of wakefulness.

5.3 Preliminary visual examination

With a subset of the sensors extracted and considering only one hour from each night, it was possible to preform visual examination of data summaries without being overwhelmed. This is done both to discover visually evident trends to aid further analysis, and to discover problems in the data which must be addressed before continuing.

Figure 5.2 depicts one hour of steady sleep for a patient with no recognized sleep apnea. On the contrast, Figure 5.3 shows a patient with high AHI. The immediate observation made when comparing these two sets of data are that the apnea set has a much wider variation in its amplitude, whereas the healthy patient appears to have a more stationary sensor output. The pattern of rises and falls is the most obvious in the effort-sensors on thorax and abdomen and the flow and pressure flow, as well as the plethysmography, in which the pattern is inverted compared to the others.

To illustrate even more clearly, we use the Flow Events analysis data to mark all the times during this one hour when patient 2407 experiences a respiratory disturbance (Figure 5.4). Patient 2407 in particular exhibits an example of Chenye-Stokes breathing, lasting for the entire duration of the hour. The pattern is incredibly easy to see. Looking at a slightly less obvious case in patient 4171, the periodic pattern is missing, but we can still observe a much greater amount

of fluctuation in sensor amplitude and much more erratic behavior than in the healthy patient in Figure 5.2.

From doing similar visual inspections of the rest of the data, the following observation is made: *While there are large individual differences, high-AHI data sets tend to have much greater fluctuations in sensor amplitudes, whereas low- or no-AHI patients are much closer to stationary, during one hour of steady sleep.*

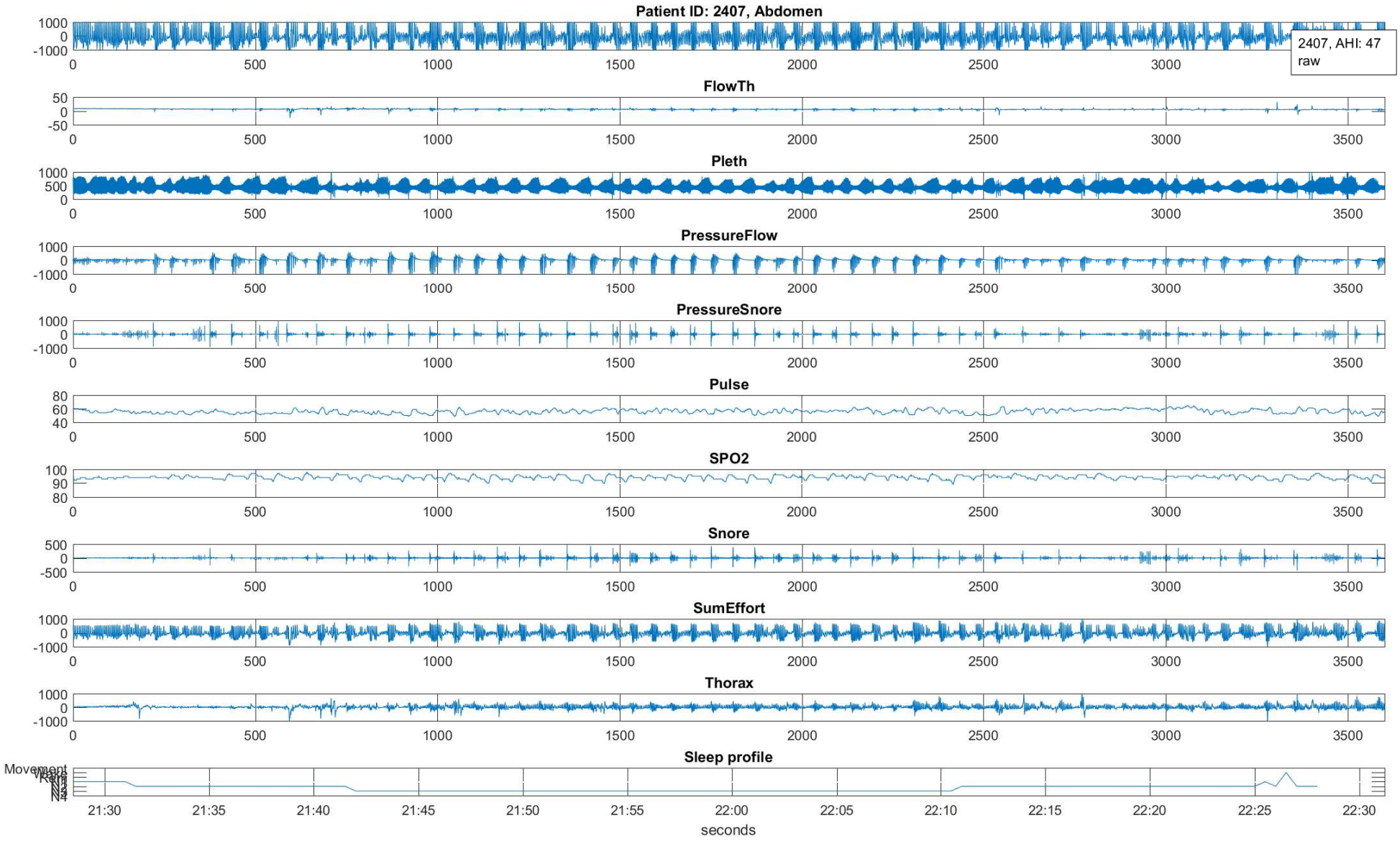


Figure 5.3: One hour of sleep for patient 2407, a patient with high AHI and visible Cheyne-Stokes breathing

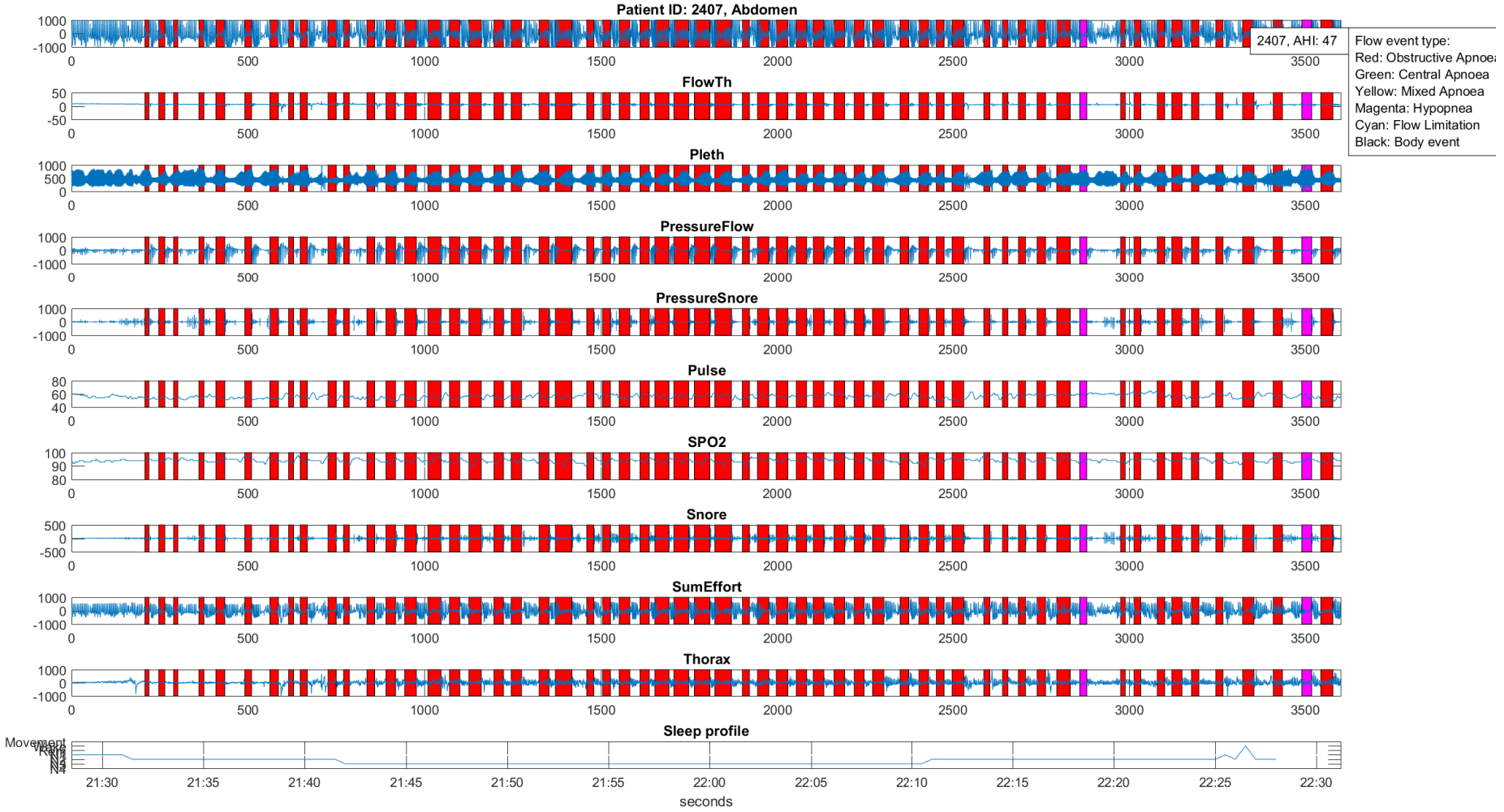


Figure 5.4: One hour of sleep for patient 2407, flow events marked

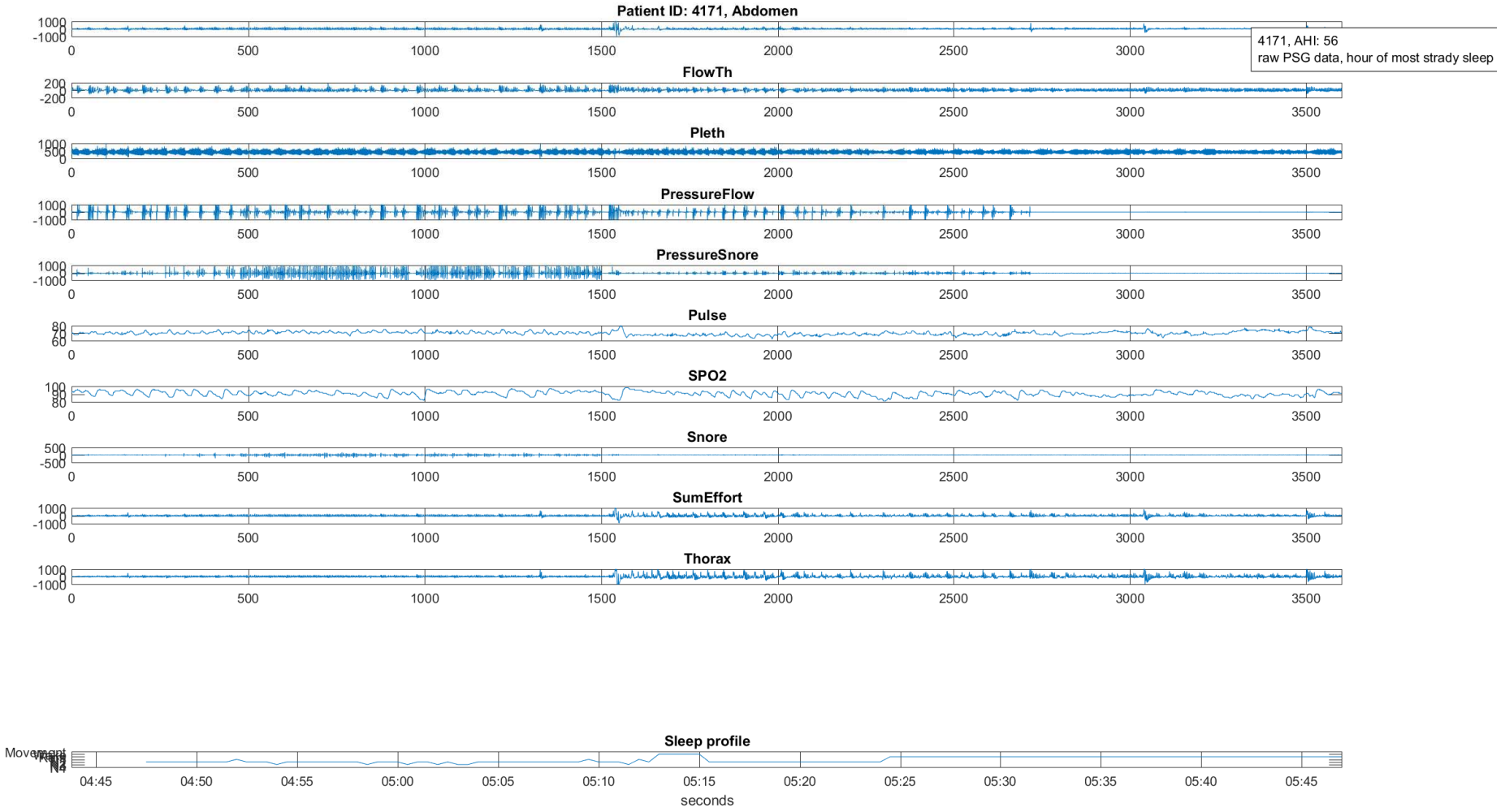


Figure 5.5: One hour of sleep for patient 4171, a patient with high AHI but without signs of Chenye-Stokes breathing

5.4 Missing data-correction

A common problem during PSG is when patient movement completely or partially dislodges a sensor during sleep. Figure 5.6 shows signs of such an event - at approximately 21:55, a sudden and large shift in the pulse oximeter pleth is followed by a significant alteration in the sensor behavior. At the same time, we see the Pulse and SPO2 frequently falling to zero - obviously impossible behavior indicating a fault. It is clear that this patient has partially dislodged the pulse oximeter finger clip.

Similar behavior is observed in a few other data sets - it appears that the finger clip pulse oximeter is particularly prone to dislodging, much more so than the other sensors in the subset.

The pulse and SPO2 data have much lower sampling rates than the plethysomnograph (4Hz versus 128Hz). For this reason, it is decided that the partial data obtained after a dislodgement for the first two may still be used, while the plethysomnograph would be much more seriously corrupted by such an event, and should be discarded. As such, all zero values in Pulse and SPO2 are replaced by NaN in Matlab, and all Pleth discarded provided there are more than 20% zero values of pulse and SPO2 in any given minute.

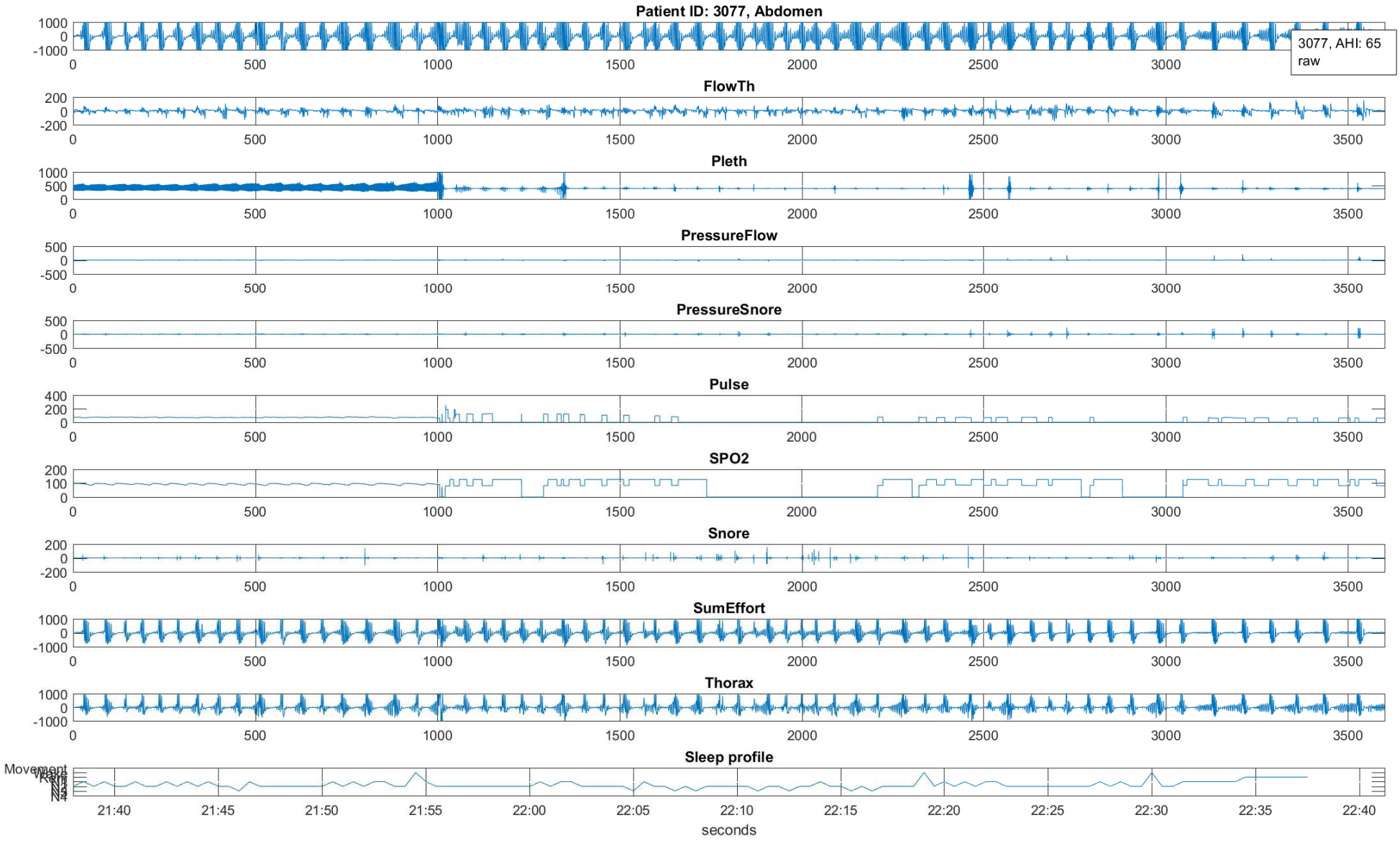


Figure 5.6: Patient 3077 shows signs of an artefact partially dislodging the pulse oximeter at 21:55

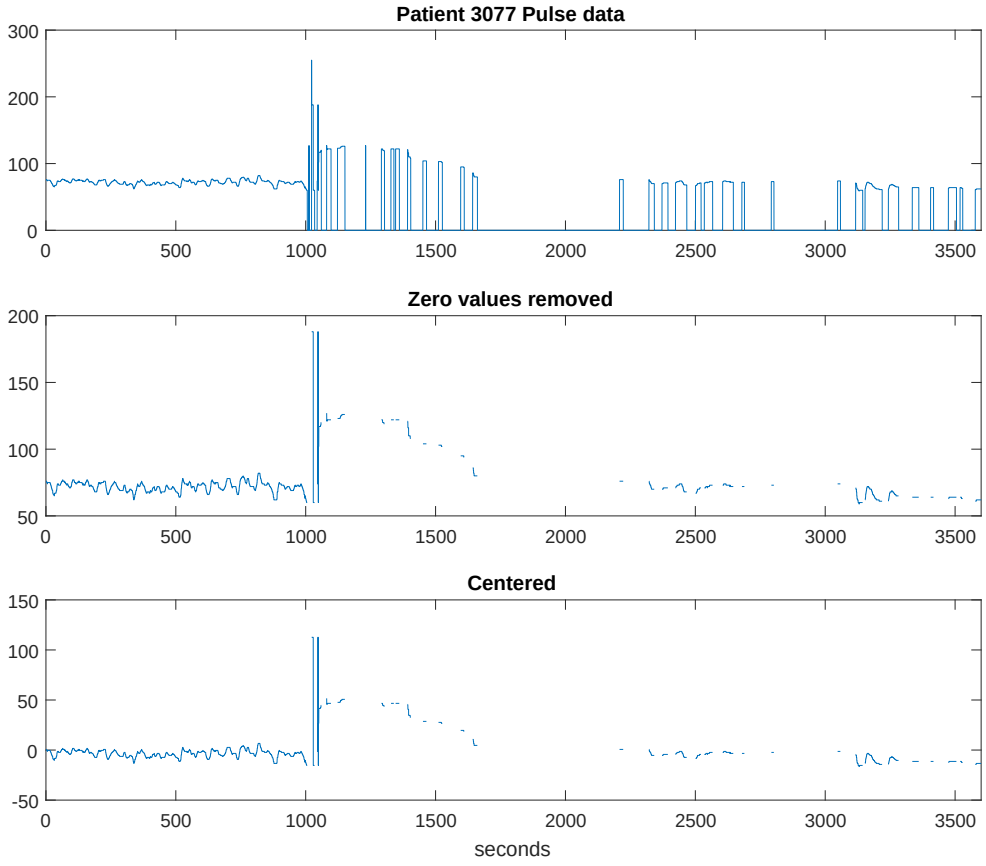


Figure 5.7: Patient 3077 pulse data with first preliminary processing

5.5 Centering, Rectification, and Downsampling

Apnea events are by definition greater than ten seconds in length. We would like to examine these low-frequency features to the exclusion of high frequency components.

The relevant sensors are first centered at zero by subtracting their mean, and then rectified to their absolute values, leaving single sided waves or bursts as seen in Figure 5.8. With apnea events being more than ten seconds in length, a resolution of 1Hz is chosen, as it is both simple and well within the Nyquist frequency of the expected components. To reduce the resolution, each second of data is averaged, and this mean taken to be the new data point. The result is shown in Figure 5.9. To smooth the resulting signal, a 2Hz lowpass filter is then applied, resulting in the final preprocessed data in Figure 5.10.

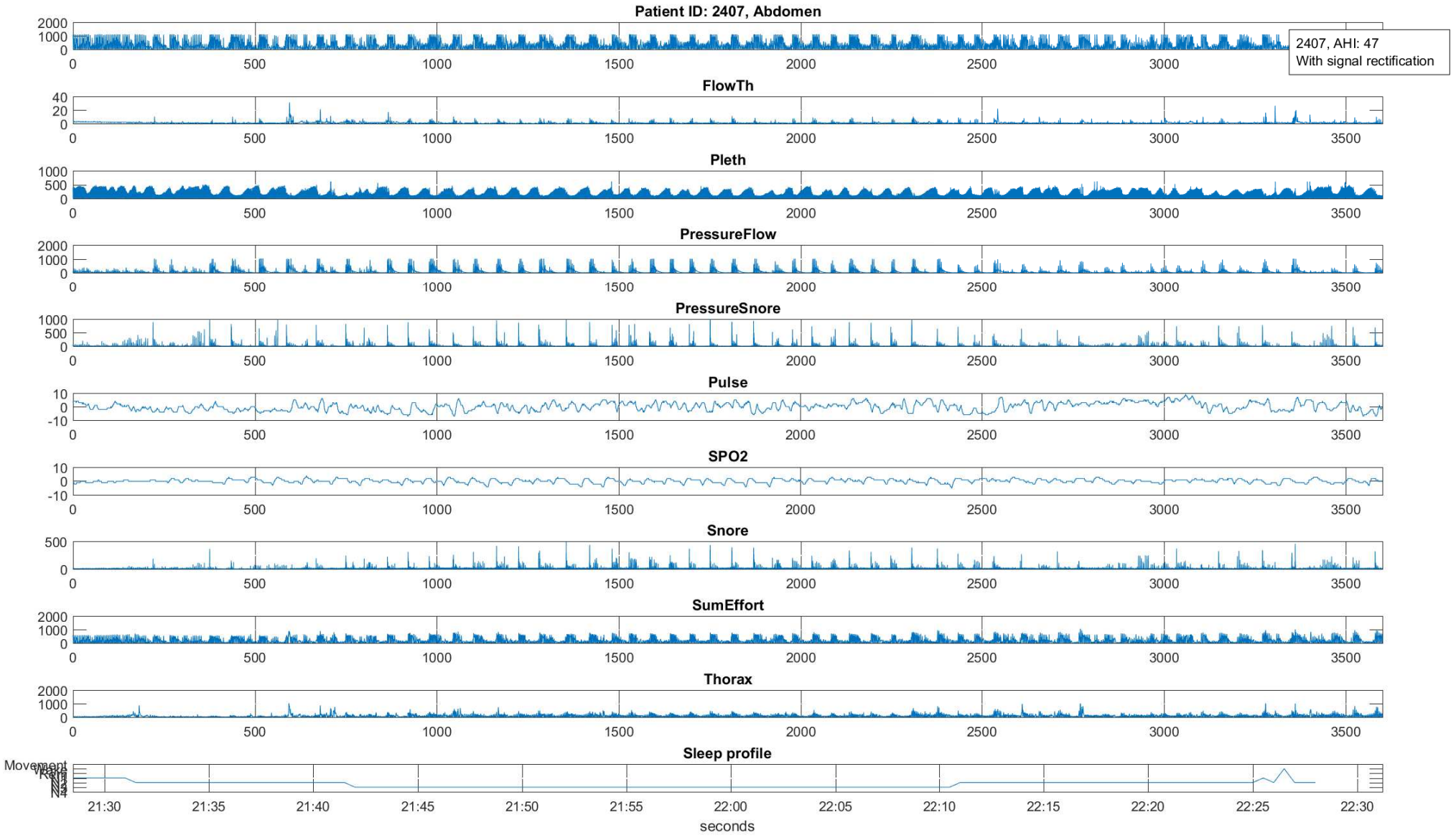


Figure 5.8: Patient 2407, data centered at 0 and relevant sensors rectified

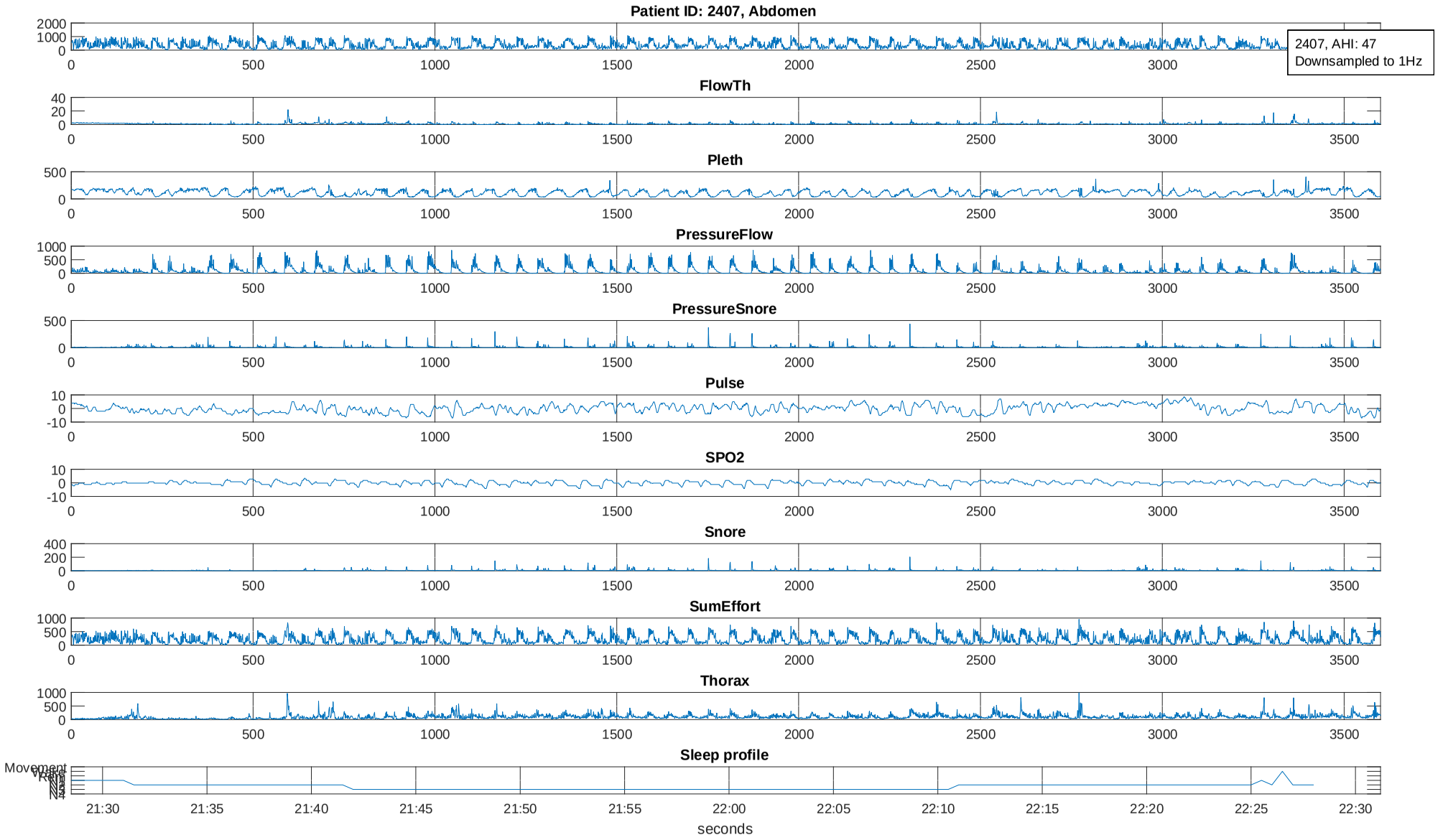


Figure 5.9: Patient 2407, centered, rectified, and downsampled to 1Hz

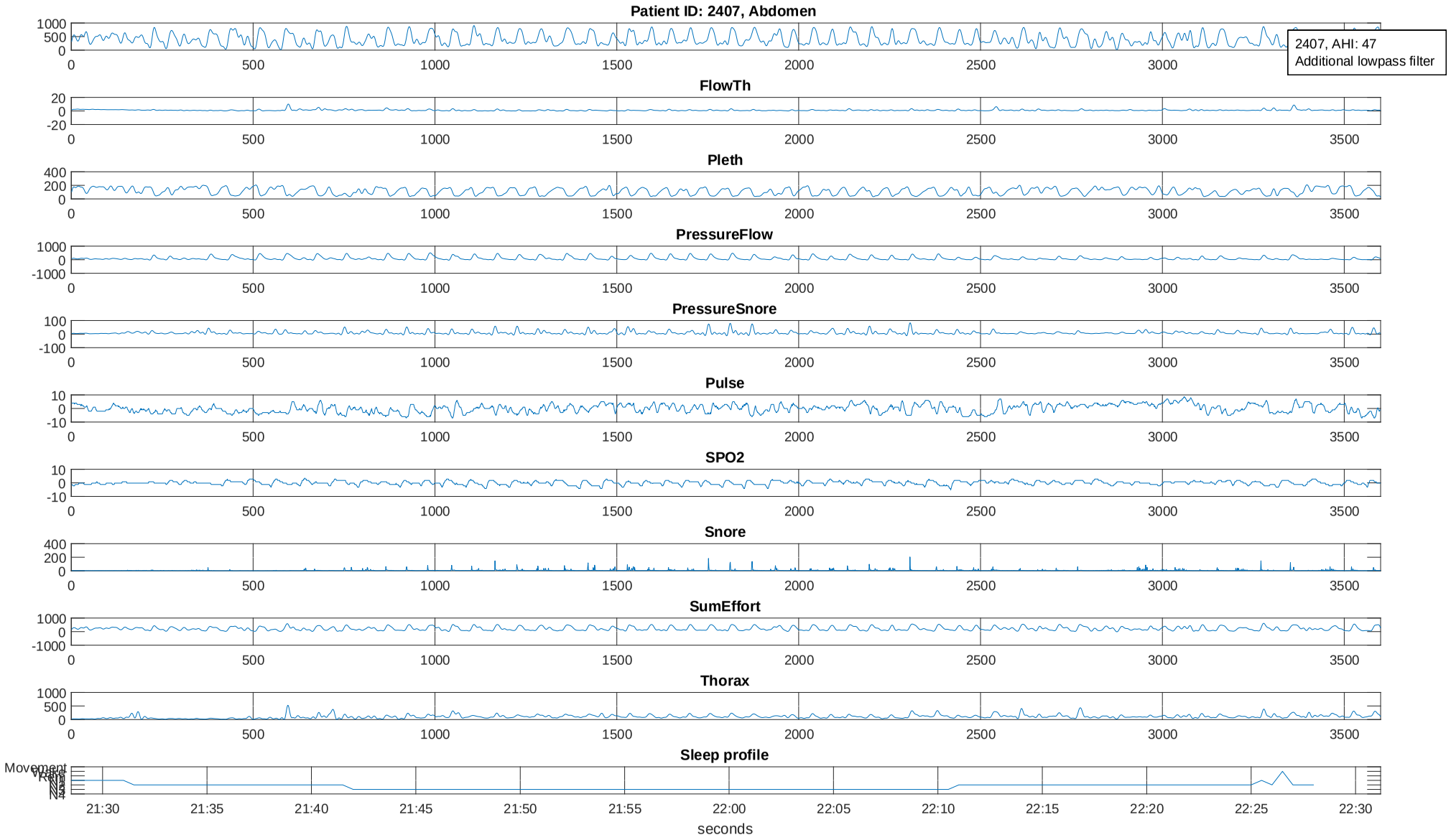


Figure 5.10: Patient 2407, centered, rectified, downsampled to 1Hz, filtered with 2Hz lowpass

5.6 Concatenation to data matrices

Both PCA and PLSR work on data matrices, and the appropriate construction and composition of these are an important first part of the analysis. There is no universal 'best' way to do this, so we will explore a few different approaches in this project. Specifically,

- Preform PCA on individual patients in a timeseries format
- Summarize multiple patients for PLS Regression on diagnostic information

5.6.1 Data matrices for individual patient time series data

For individual patients, timeseries data was stacked in a vertical format, with each point in time represented by one row in the matrix. In order to do this, downsampling was a necessity, as each matrix column needed to be of equal length. 1 Hz was chosen as an adjustable baseline. The FlowEvents analysis time series was included in the matrix, as a potential Y for regression, or as a point marker in PCA. The structure can be seen in Table 5.1.

FlowEvents timeseries	**Sensor 1**	**Sensor 2**	**Sensor 3**	**...**
Event during t1	Amplitude t1	Amplitude t1	Amplitude t1	...
Event during t2	Amplitude t2	Amplitude t2	Amplitude t2	...
...				

Table 5.1: Standard single-patient matrix formulation

5.6.2 Matricized multi-patient summaries

For the comparison of multiple patients, it was desirable to treat each patient, not each time instance, as an object in the analysis. To this end, it was necessary to find some way to summarize each patient recording with all relevant sensor data and associated diagnostic parameters on a single row in the data matrix. The general structure can be seen in table 5.2.

PatientID1	Y-variables	X-variables: Sensor data summaries
PatientID2	Y-variables	X-variables: Sensor data summaries
...	...	...

Table 5.2: Standard multi-patient matrix formulation. The patient IDs are, of course, simply labels, not included in the main data matrix.

The Y-variables

A number of Y-values were included in the matrix construction. *The Unscrambler® X* allows for easy inclusion, exclusion, weighting or downweighting of variables in the analysis, so in this way the analyst will be able to easily pick and choose from a number of potentially important parameters to examine.

From the results-file, the project chose to include AHI, PLM, DSI, DSA, and DSL (see section 4.2). Since these diagnostic values are based on the full-night recordings, and since this project primarily focuses on the single optimal sleep hour, it was also desirable to include information pertinent only to that one hour. As such, the Y-variables also used the FlowEvents and Sleep-stages analysis files to include the total sum of seconds of each flow event type and each sleep stage, as well as the total number of seconds with any flow event present, during the hour in question. The following acronyms all indicate 'sum of seconds' of the described types:

FlowE	Sum of all flow events
ASum	Sum of all apnea events
OSA	Obstructive apnea
CA	Central apnea
MixA	Mixed apnea
HypA	Hypopnea
FloL	Flow Limitation
BodE	Body event

The Y-variables were thus structured as follows:

Results-file parameters	Flow-event parameters	Sleep-stage parameters

With each of the three parameter types breaking down into the following scalar values:

Results-file parameters:	AHI	PLM	DSI	DSA	DSL

Flow-events parameters:	FlowE	OSA	CA	MixA	HypA	ASum	FloL	BodE

Sleep-stage parameters:	Wake	N1	N2	N3	N4	REM

5.6.3 The X-variables: Single-line sensor data summaries

The challenge of this section is to find a way of presenting all the information pertaining to one patient in one single row of a larger matrix, while preserving the most important features of the original data. This thesis explored two possible approaches, histograms, and frequency transformations.

Histograms

Histograms were chosen as a data summary method because it will allow investigation of sensor amplitudes, in particular *amount of amplitude variation*, but also because of their simplicity.

First, the number of bins for each sensor must be decided. In the finished data matrix, we need all object rows to be of equal length. Furthermore, some sensors have greater amount of variation than others, and some higher resolution than others. In light of this, it might feel natural to have different bin sizes for different sensors. However, this would place an inherently unequal weight on the sensors in the analysis, which is undesirable at this stage. As such we would rather like to make a compromise, using a fairly large bin size and accepting that the resulting matrix will be relatively sparse for the sensors with less variation. In other words, we choose that the number of bins for all sensors, all patients, will be the same.

The MATLAB `histogram` function, when given just data, uses an automatic binning alorithm that returns bins with a uniform width, chosen to cover the range of elements in the data and reveal the underlying shape of the distribution[22]. Naturally, this suggestion will not be the same

for same sensor, different patients. However, we will use it to guide our choice.

We first allow the MATLAB histogram algorithm suggest the number of bins necessary for each sensor, for all patients we want to include in the finished matrix. Then, we choose the number of bins for each sensor by taking the mean of all the suggestions from MATLAB plus two standard deviations of the suggestions, rounded up. We believe this choice will achieve an adequate level of detail, without 'overdoing it'.

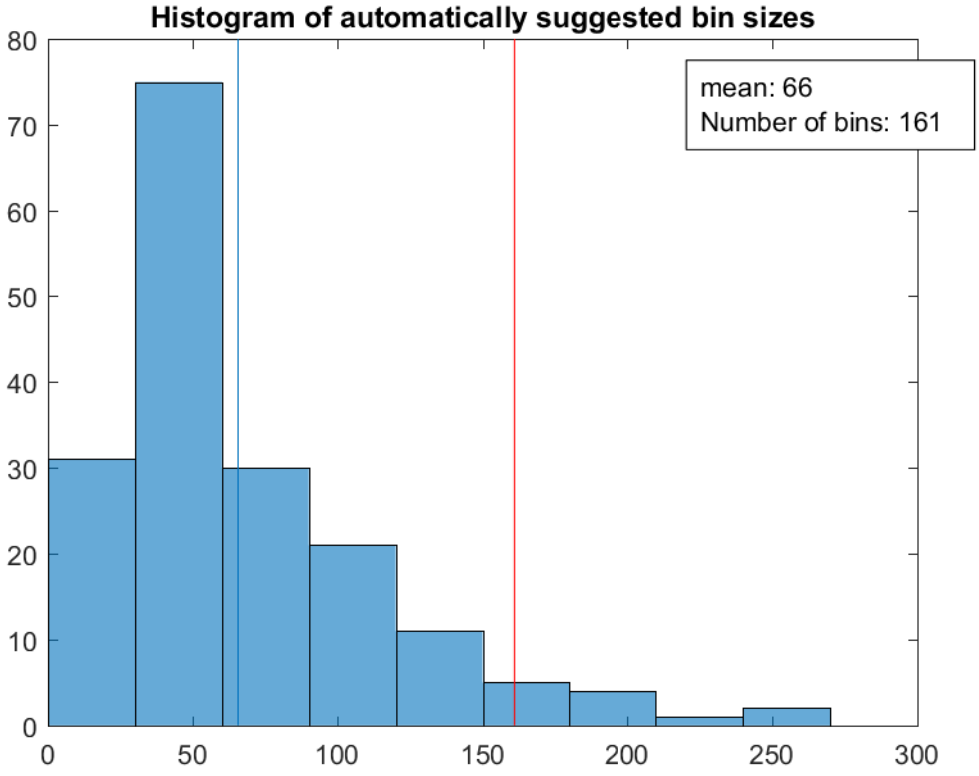


Figure 5.11: Bin number selection for a subset of patients. Blue line is mean, red is selected bin number

Comparability dictates that we use the same edges vector for all same sensors. However, we do not limit different sensors to equal edges. To choose the outer limits, we examine the property `BinLimits` for the generated histograms and choose the extrema for each sensor. We then generate the edges vectors using `linspace`.

With the standardized parameters (bin number and edges vectors) chosen, we can generate sensor histograms for each patient, and include them as rows in the data matrix as illustrated in Table 5.2. We also choose to normalize the histograms at this stage, to increase comparability. The final histogram matrix has a size of $(NumberOfPatients) \times (Yvars_number + (Sensors \times NumBins))$, with a structure as shown in Table 5.3:

PatientID	**Y-variables**	**Sensor 1 edges**	**Sensor 2 edges**	**...**
PatientID1	Y-variables 1	Histogram sensor 1	Histogram sensor 2	...
PatientID2	Y-variables 2	Histogram sensor 1	Histogram sensor 2	...
...	...	...	...	...

Table 5.3: Multi-patient histogram matrix formulation.

It is interesting to examine some of the histograms in greater detail, to evaluate our procedure. Returning to patient 2407, we can see how this procedure results for the abdomen sensor (Figure 5.12) and the pulse sensor (Figure 5.13). We see that the standardized histograms cover approximately the same area and preserve the approximate shape, with a drop in column height more significant for data that originally is more widely distributed.

We may also examine the histograms of some apnea-free data sets. Consider the abdomen sensor of patient 0389, in Figure 5.14. It is clearly much less widely distributed than the high apnea case, and while this naturally proves nothing it lends some credence to our hypothesis that amplitude distribution may be related to apnea.

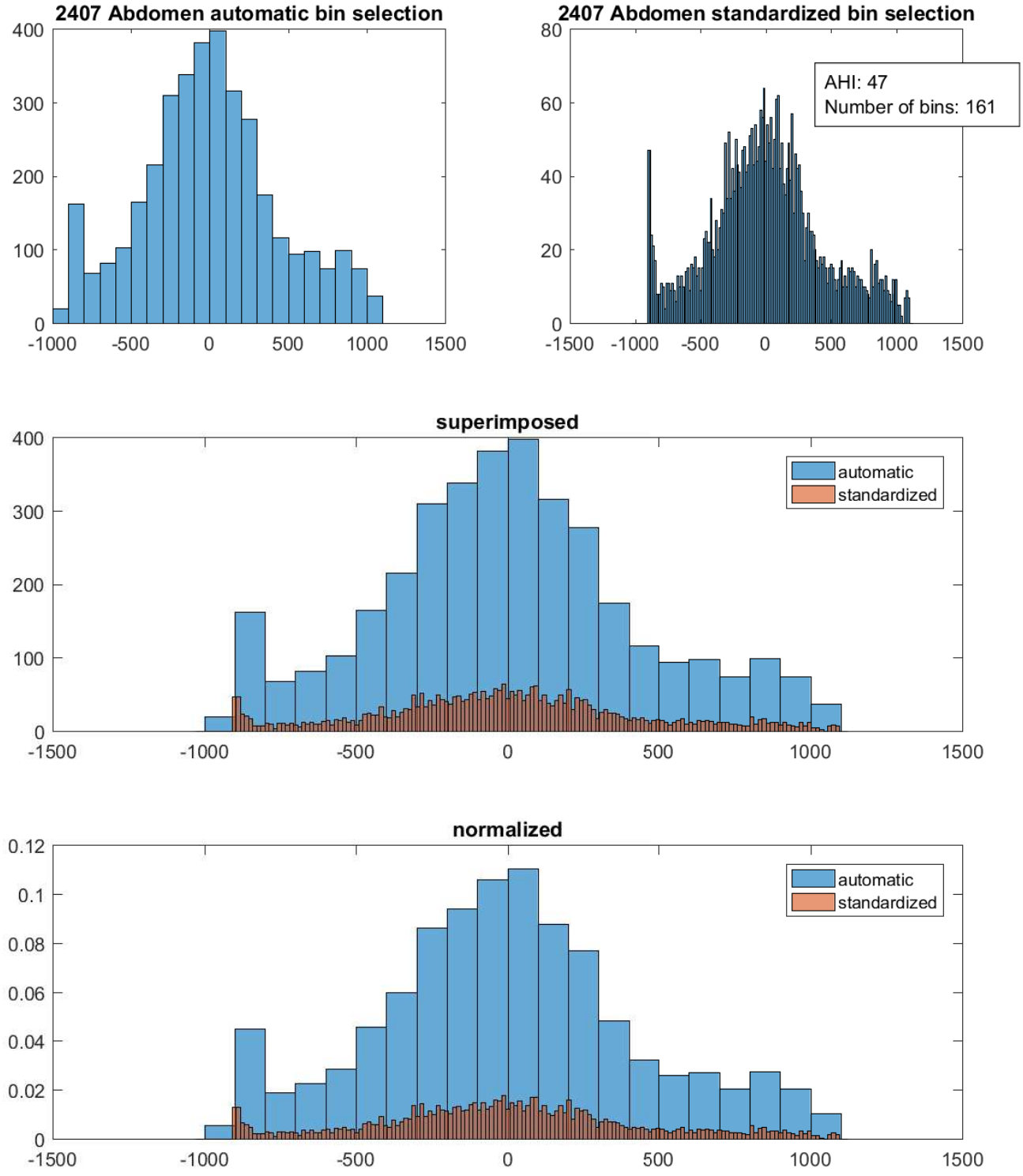


Figure 5.12: Histograms for the abdomen sensor of patient 2407. The drop in column height for the widely distributed data is expected, as the data gets dispersed into a much greater number of bins

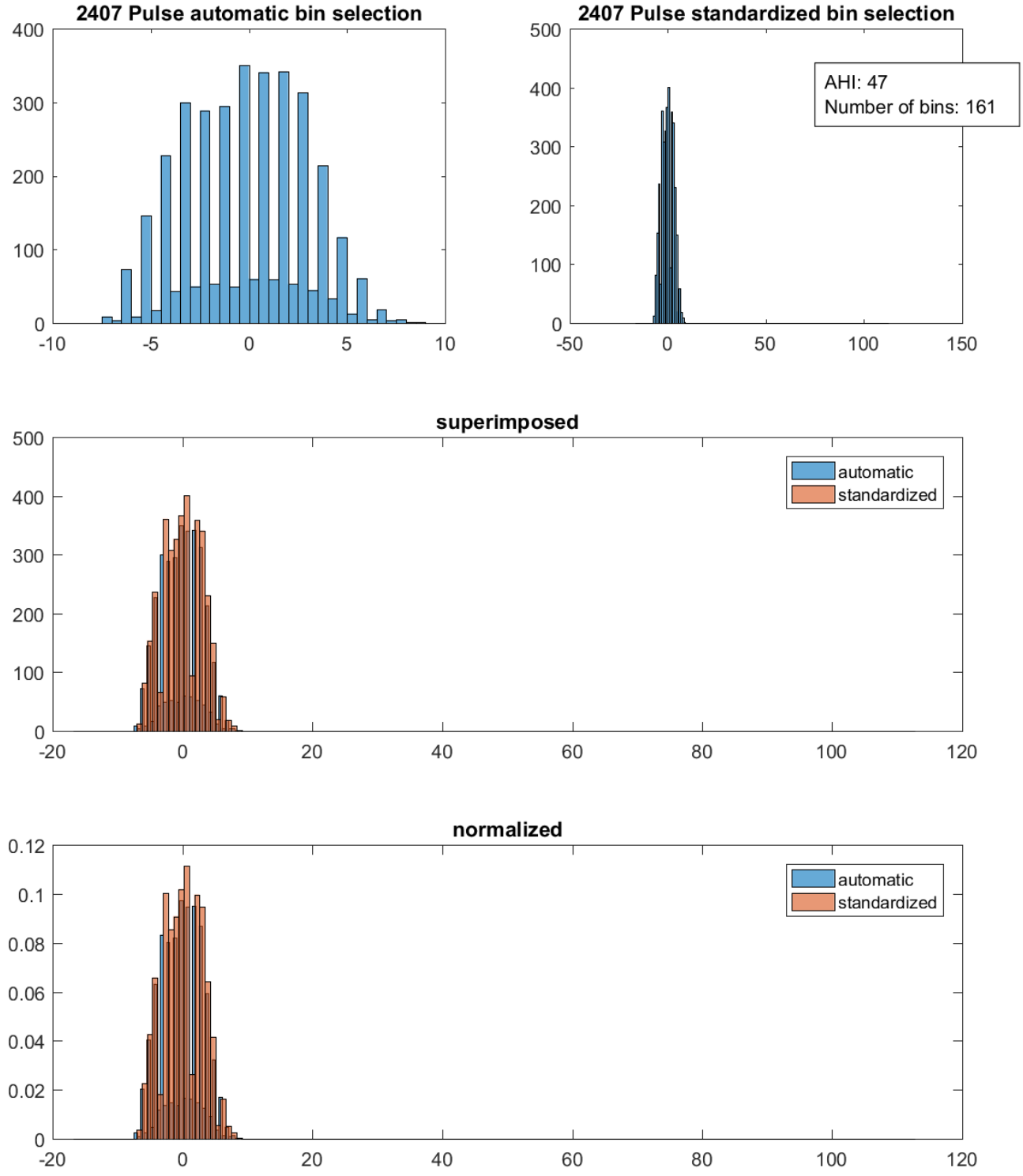


Figure 5.13: Histograms for the pulse sensor of patient 2407

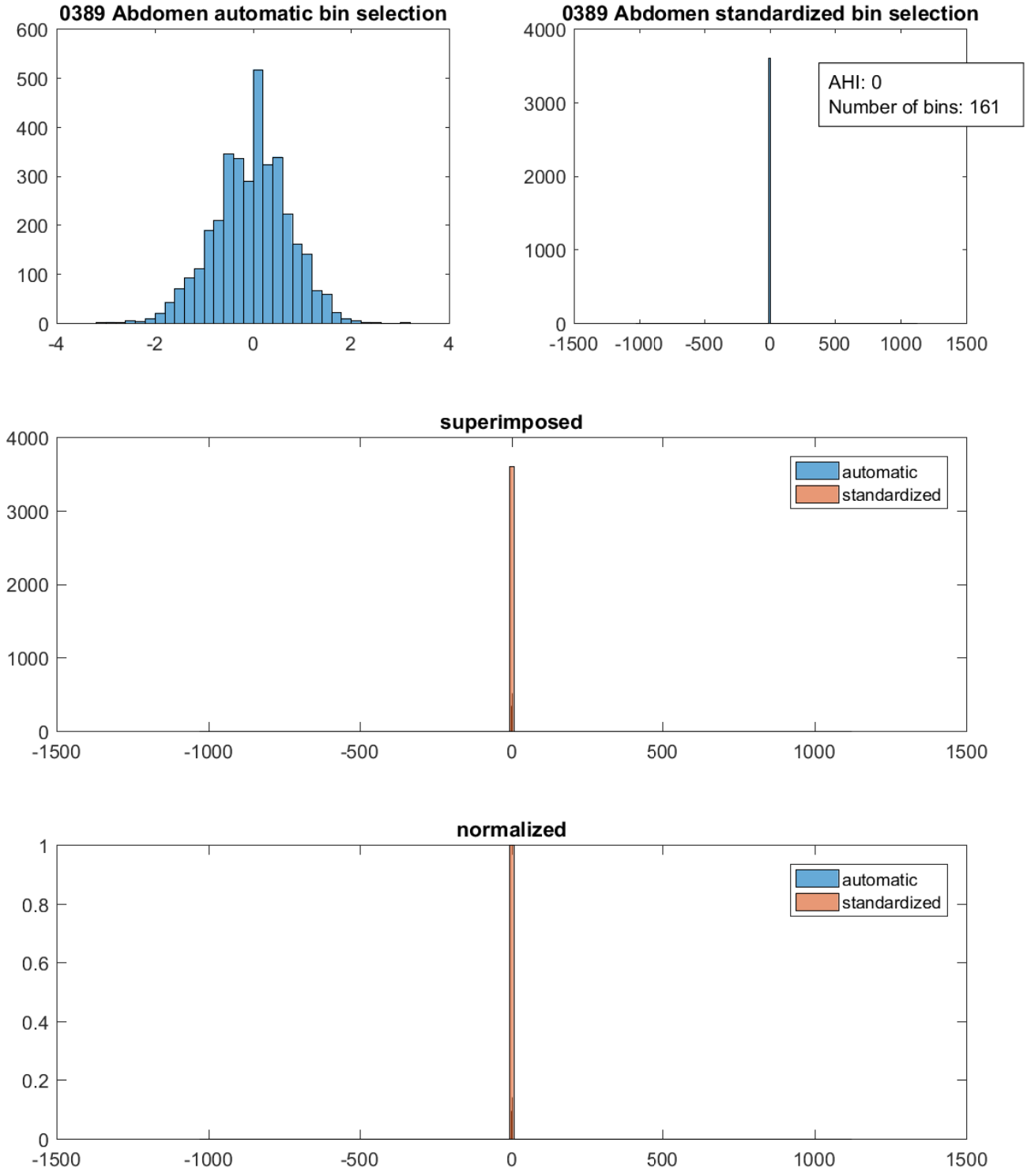


Figure 5.14: Histograms for the abdomen sensor of patient 0389

Frequency-transformations and Power Spectral Density

We wish to examine the frequency components and spectral power of the signals. While the Fast Fourier Transform might not be the most suitable method to use with this data (see DISCUSSION PART), it is well understood and widely used, and therefore worthy of attention.

Using N-point `fft` in MATLAB, with N chosen as the next power of 2 larger than the signal length, we construct matrices with frequency bins as variables. We may then choose to reduce the resolution of the result by sorting the resulting frequencies into a smaller number of bins spanning the same range. To reduce the effect of endpoint discontinuities, we chose to apply a Hanning window. The resulting matrix shape is illustrated in Table 5.4.

PatientID	**Y-variables**	**Sensor 1 bins**	**Sensor 2 bins**	**...**
PatientID1	Y-variables 1	Binned FFT sensor 1	Binned FFT sensor 2	...
PatientID2	Y-variables 2	Binned FFT sensor 1	Binned FFT sensor 2	...
...	...	...	...	...

Table 5.4: Multi-patient FFT matrix formulation

We also construct matrices using the Power Spectral Density of the signals, computed as the absolute magnitude of the fourier transform squared.

PatientID	**Y-variables**	**Sensor 1 bins**	**Sensor 2 bins**	**...**
PatientID1	Y-variables 1	Binned PSD sensor 1	Binned PSD sensor 2	...
PatientID2	Y-variables 2	Binned PSD sensor 1	Binned PSD sensor 2	...
...	...	...	...	...

Table 5.5: Multi-patient PSD matrix formulation

6

Multivariate Analysis of PSG data

In this chapter, multivariate analysis of PSG data is performed. Section 6.1 analyzes individual patients using PCA, first by looking at one patient in detail, before describing some general trends. In Section 6.2, we attempt to build predictive models of apnea by performing PLSR on multiple patients. We first consider a subset of the data in Section 6.2.1, and compare the results to using raw data in Section 6.2.2. In Section 6.2.3 we build a predictive model using the histogram-formulated data matrices, and in 6.2.4, we attempt to do the same using the frequency-formulated data.

6.1 Analyzing the 'optimal sleep hour' for individual patients

We perform PCA of individual patients with data arranged in time-series format as shown in section 5.6.1. The goal here is not to create a predictive model, but rather to visualize the multivariate data in a different format.

Patient 2407 in detail

To start, we consider again patient 2407. We remember that this particular patient has an AHI of 47, and clearly visible Cheyne-Stokes patterned breathing.

We perform a PCA using the 10 sensor subset defined in Section 6.2.1, on data that has been centered, rectified, downsampled, and filtered. We use the NIPALS algorithm, full validation,

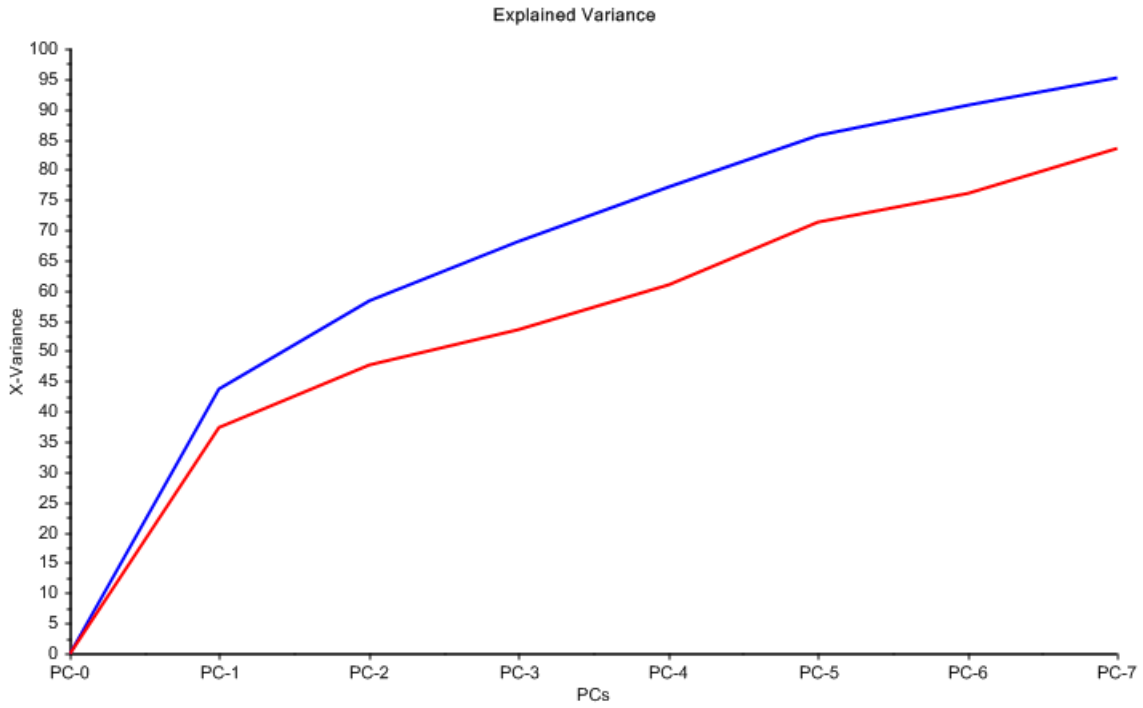


Figure 6.1: Explained variance of PCA of patient 2407, seven principal components included

and standard deviation auto-scaling. The scores of the principal component for a time series PCA such as this each correspond to a time instance in the data series, mapped in the new coordinate systems formed by our principal components.

The explained variance of this analysis is not stellar (Figure 6.1), yet the first principal component explains about 40 percent with good validation, indicating that the analysis certainly finds some meaningful structures in the data.

As we mentioned in Section 3.6, we see from the loadings in Figure 6.2 that the effort-sensors on thorax and abdomen appear strongly correlated with the flow and the pressure flow, and inversely correlated with the plethysmography, along the first principal component axis. This principal component is catching the pattern we observed when visually examining the data set in Section 5.3 - as pressure and flow sensors decreased, the plethysmograph increased accordingly and vice versa.

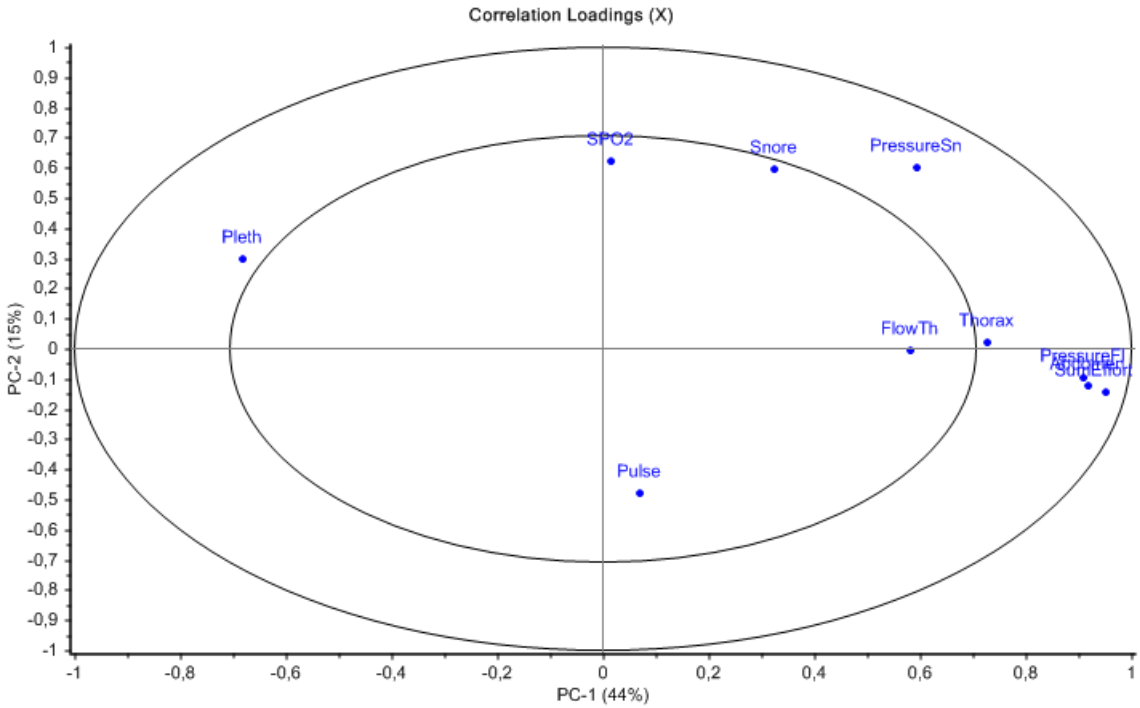


Figure 6.2: Loadings for PCA of patient 2407, PCs 1 and 2. The cluster to the right contains the Pressure Flow, Abdomen, and Sum Effort sensors

The colored scores plot in Figure 6.2 illuminates the situation further. We have not used any information about apnea in building this model for this person, yet it appears that the variation caught by PC1 correlates quite strongly with apnea events, with obstructive apnea time instances falling to the left, regular breathing to the right, and a small smattering of hypopnea around the center-left.

Going back to the loadings, along the second principal component, we see an inverse relationship between SpO2 and Pulse. As oxygen saturation falls, the pulse will increase to rectify it. Figure 6.4 also illustrates this connection. Interestingly, the third principal component appears to be modelling the same phenomena, but in the opposite direction, as seen in Figure 6.5.

While the explained variance continues to increase as we go towards including all seven principal components, we are unable to identify any further clear interpretations of the components. However, it is clear that for this data set, the first three principal components are physically meaningful and explain significant phenomena in the data.

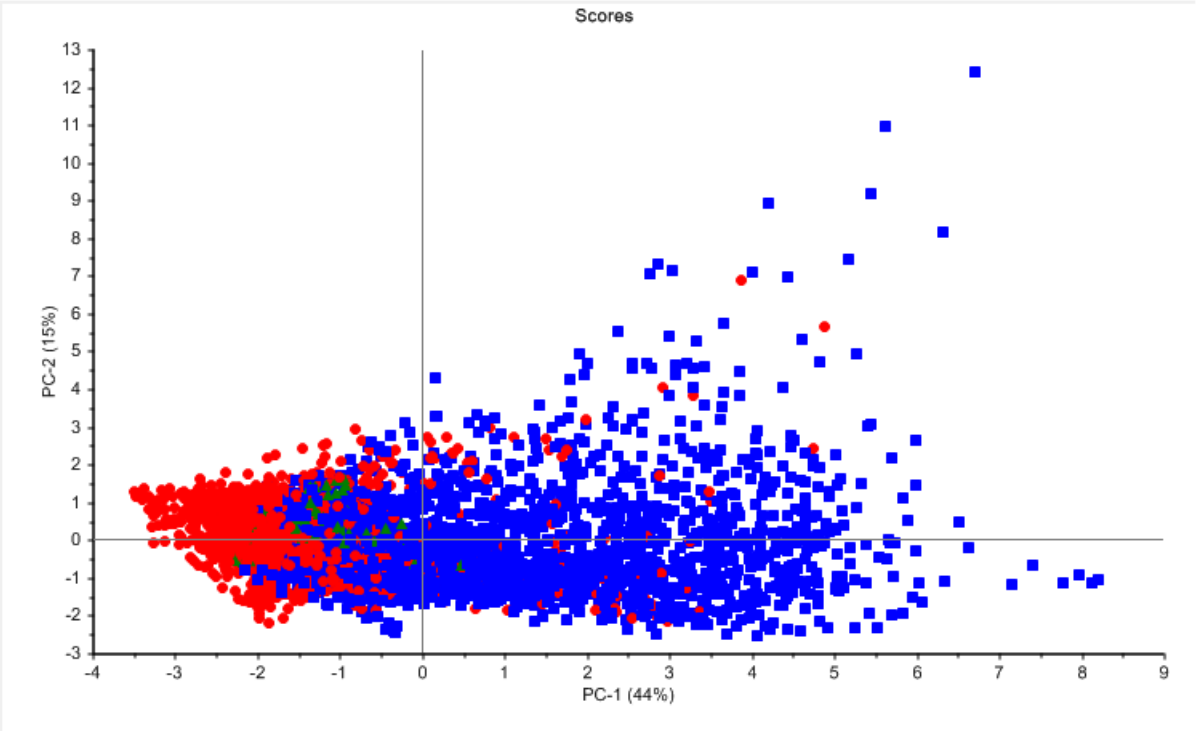


Figure 6.3: Scores for PCA of patient 2407, PCs 1 and 2. Red dots indicate obstructive apnea, green hypopnea, and blue regular respiration

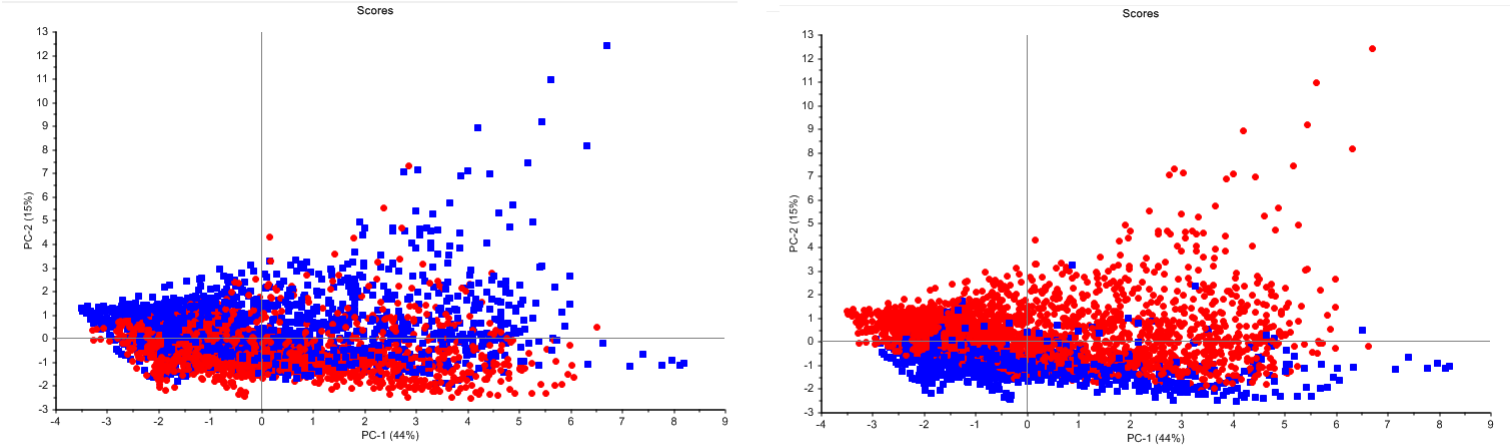


Figure 6.4: Loadings for PCA of patient 2407, PCs 1 and 2, illustrating the relationship between pulse and SPO2. To the left, red indicates pulse above the mean. On the left, red indicates oxygen saturation above the mean.

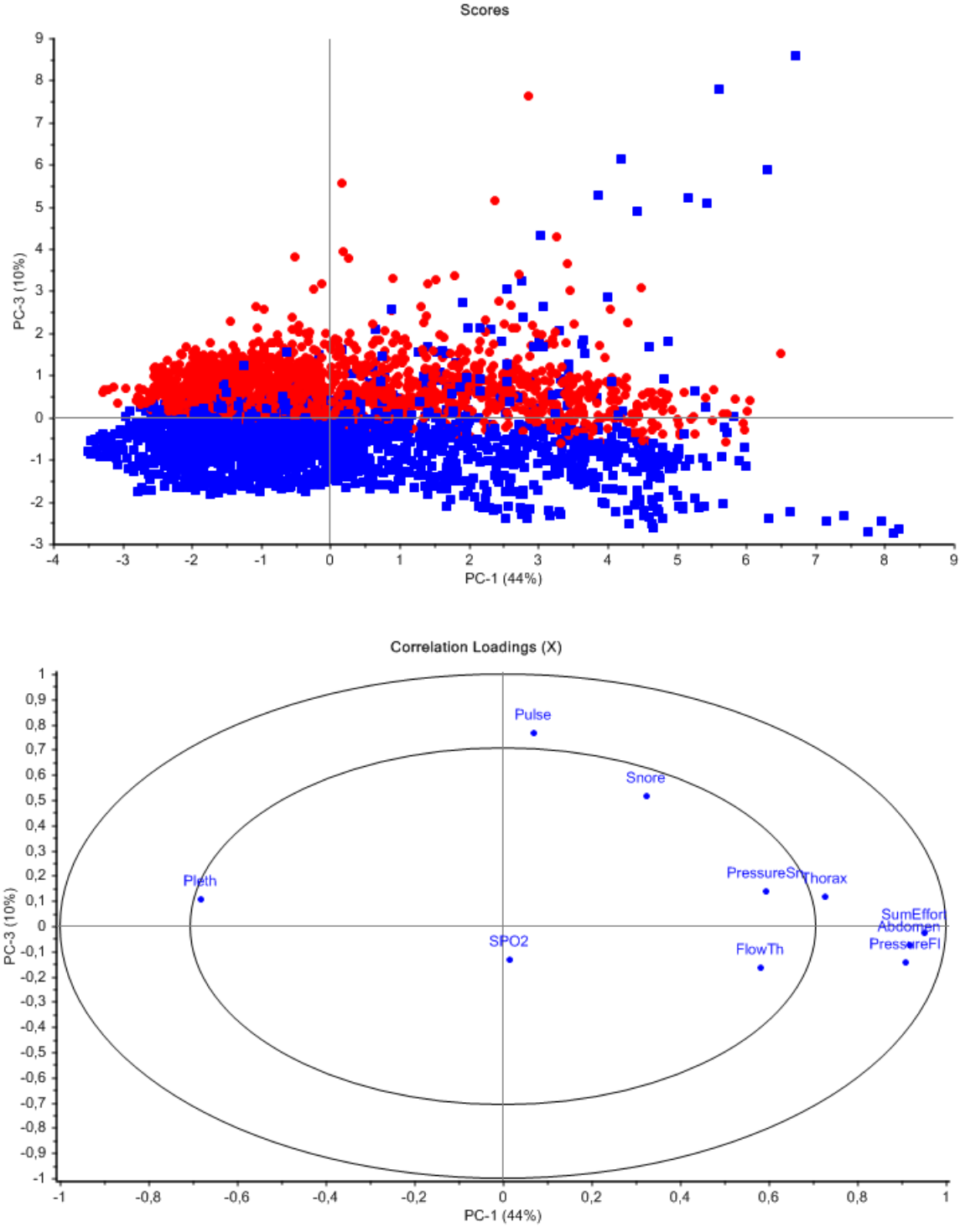


Figure 6.5: Scores and loadings for PCA of patient 2407, PCs 1 and 3. In the scores plot, red indicates oxygen saturation above the mean.

Single patient results

Preforming the same analysis on several other patients with varying degrees of apnea, we observe similar patterns to those seen in patient 2407. In Figure 6.6, we see the pc1pc2 scores and loadings of patient 1921, a patient with an AHI of 52, but without Cheyne-Stokes breathing and with more hypopnea than OSA. We observe that less overall variance is explained by PC1 (28%, as opposed to 44% for patient 2407), but the trend of apnea events going left along the PC1 axis remains. The configuration of pc1pc2 loadings is also similar. We analyze five other high-apnea cases in a similar way, and this trend is evident in all of them (with varying degrees of explained variance).

In Figure 6.7, we see AHI of patient 2396, who has zero signs of apnea and thus normal breathing throughout the hour we are looking at. The loadings configuration is comparable to the other we have looked at, but the scores are distinctly different in form. In the plot, we have chosen to draw lines between consecutive seconds, and color groups in ten-minute intervals. Interestingly, the plot shows evident grouping in time, indicating that characteristics of the patients breathing slowly changes as he sleeps.

Now, since each of these models are developed individually for distinctly different patients, there is no way to really generalize any trends or conclusions, nor predict the apnea of one from the model of another. Several healthy patients also exhibit a flat structure in the pc1pc2 scores similar to that of the high-apnea cases, so without knowledge of the analysis results there is no way to tell one from the other. So why have we done this?

Firstly, the PCA offers a lower-dimensional window into the data. The validation of the explained variance of the PCAs performed here is in general quite high, so with more effort perhaps this decomposition could be used to reduce the amount of data that must be studied in order to form conclusions.

More importantly, this analysis has served as a proof of concept of sorts for our preprocessing methods from Figure 5.1. While our methods are highly reductive, decreasing the resolution

of some sensors by as much as 128 to one, the fact that apnea still tends to show up along the very first principal component lends credence to our hypothesis that the characteristics of apnea are preserved through our data reduction.

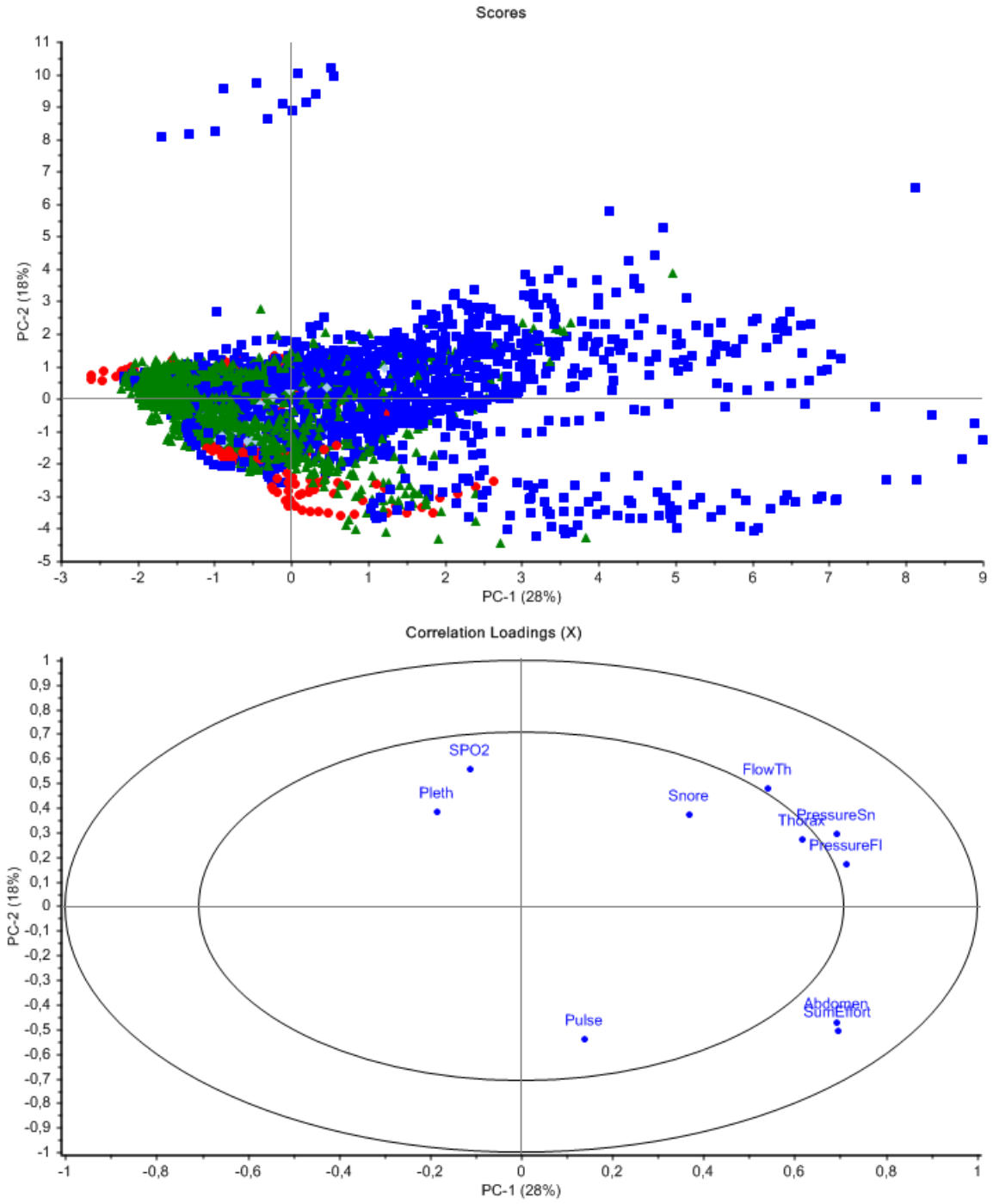


Figure 6.6: Scores and loadings for PCA of patient 1921 (AHI = 52), PCs 1 and 2. Red is OSA, green hypopnea, and blue normal breathing.

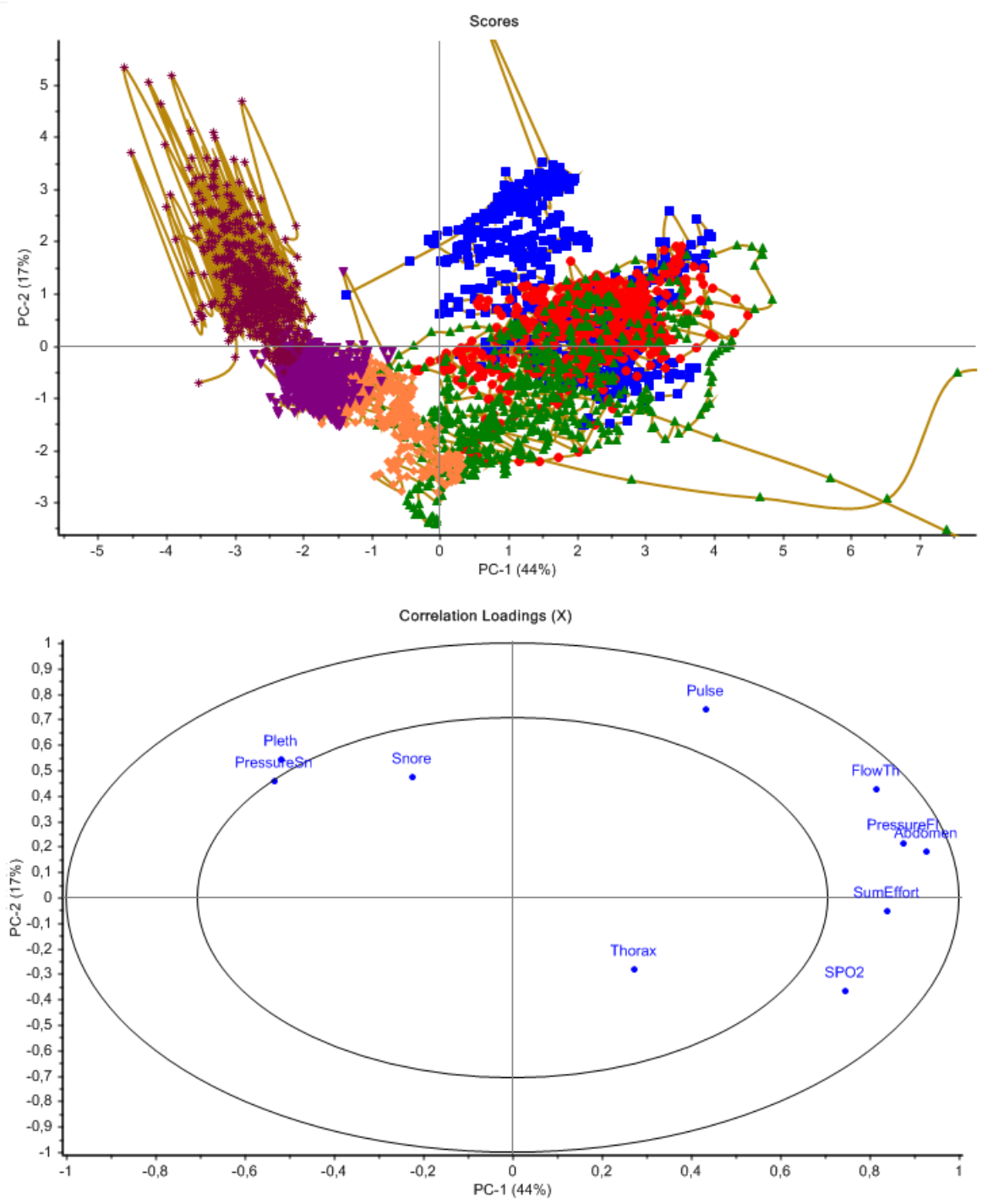


Figure 6.7: Scores and loadings for PCA of patient 2396 (AHI = 0), PCs 1 and 2. Scores are colored and connected to illustrate breathing changing characteristics over time, each group represents 10 minutes

6.2 PLSR of multiple patients: simple prediction of sleep characteristics

We perform predictive analysis of multiple patients. In this section, red color will always indicate validation values.

6.2.1 PLSR of a small subset, using histograms

In our initial analysis, we narrow the scope to working with a subset of the data. We choose 18 patients, 10 of which are 'healthy' in the sense of having zero AHI (nothing more is known about their health). The remaining 8 are the most extreme apnea cases available to us, all with AHI above 40.

Considering a 18×961 normalized histogram-formulated data matrix using all of the preprocessing steps outlined in Figure 5.1, we preform PLSR with AHI as the Y variable, 7 maximum components, and full cross validation.

The first principal component clearly separates the patient groups in the scores plot (Figure 6.8), and claims to explain more than 90% of the total variance of the data set (Figure 6.9). The algorithm does not detect any significant outliers (Figure 6.10).

With such a small set of samples we do not really expect this model to stand up under scrutiny, and indeed it does not. After validation only 50% is explained (Figure 6.9), and including the validated scores shows a much less obvious clustering (Figure 6.11). The validated explained variance is not able to climb above 50%, no matter the number of components we include.

The results overall are promising, although due to the extreme nature and small size of the data set, probably quite overoptimistic. However, the fact that the groups of patients not only separate, but separate clearly along the very first principal component, lends some credence to the histogram matrix formulation. There may indeed be characteristics of sleep-apnea record-

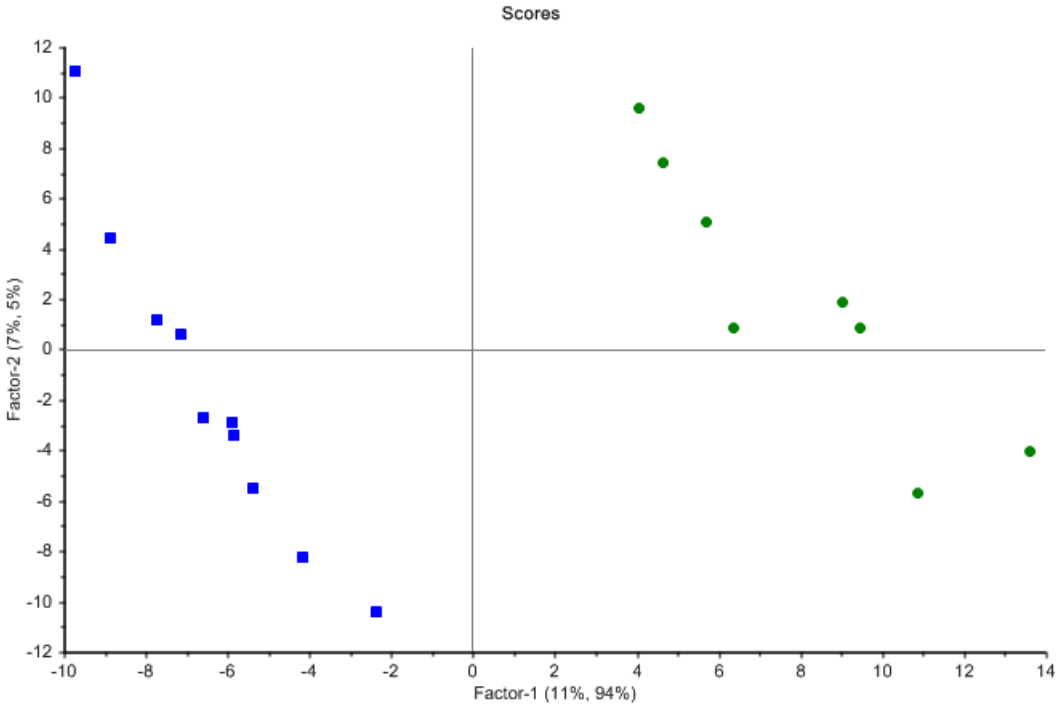


Figure 6.8: Scores for PLSR of 18 patients, from histograms of fully preprocessed data. Green dots have high AHI, blue have zero

ings that can be picked up through simply examining patient sensor amplitude distributions.

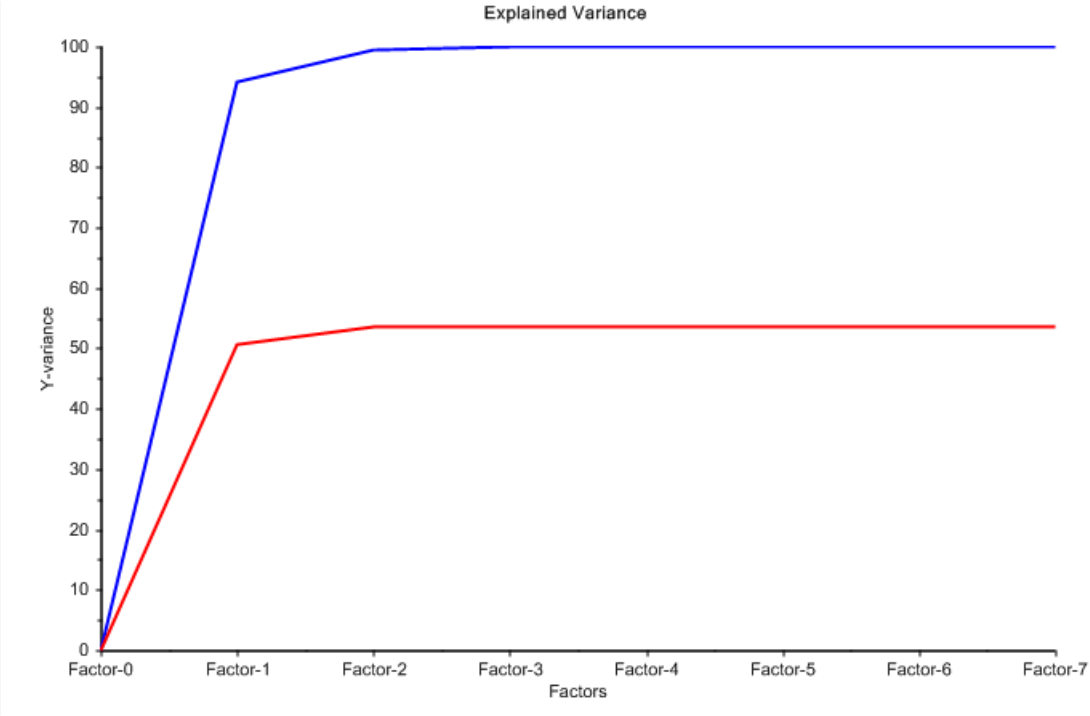


Figure 6.9: Explained variance of PLSR of 18 patients, from histograms of fully preprocessed data. Blue indicates original model, red is validation

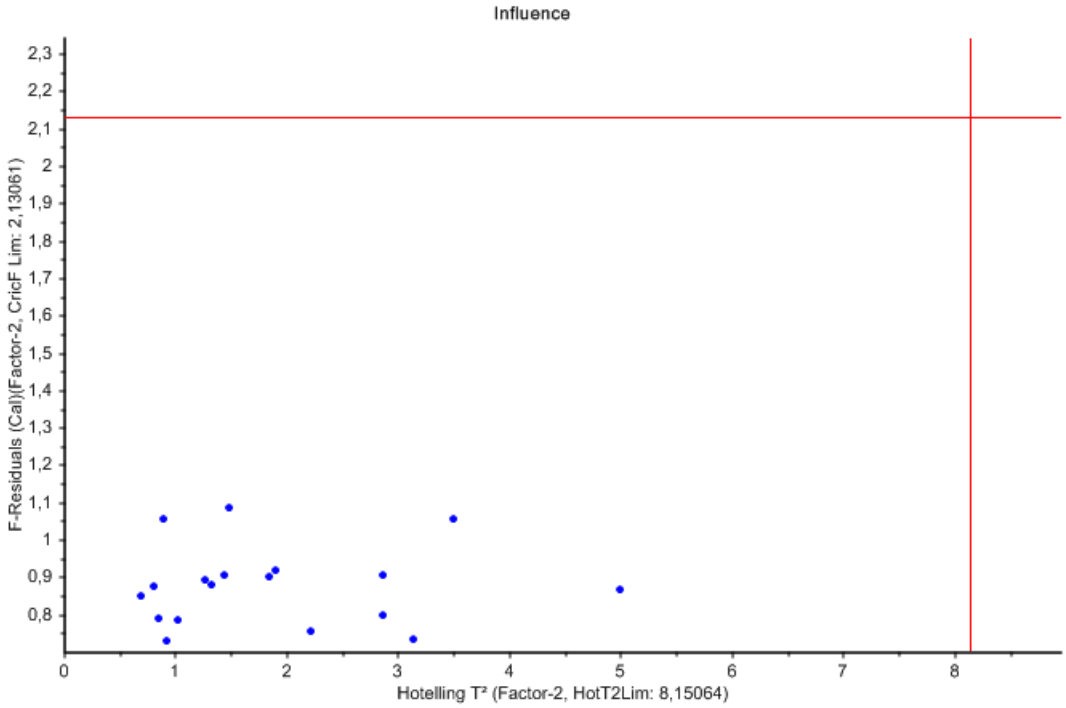


Figure 6.10: Influence plot for PLSR of 18 patients

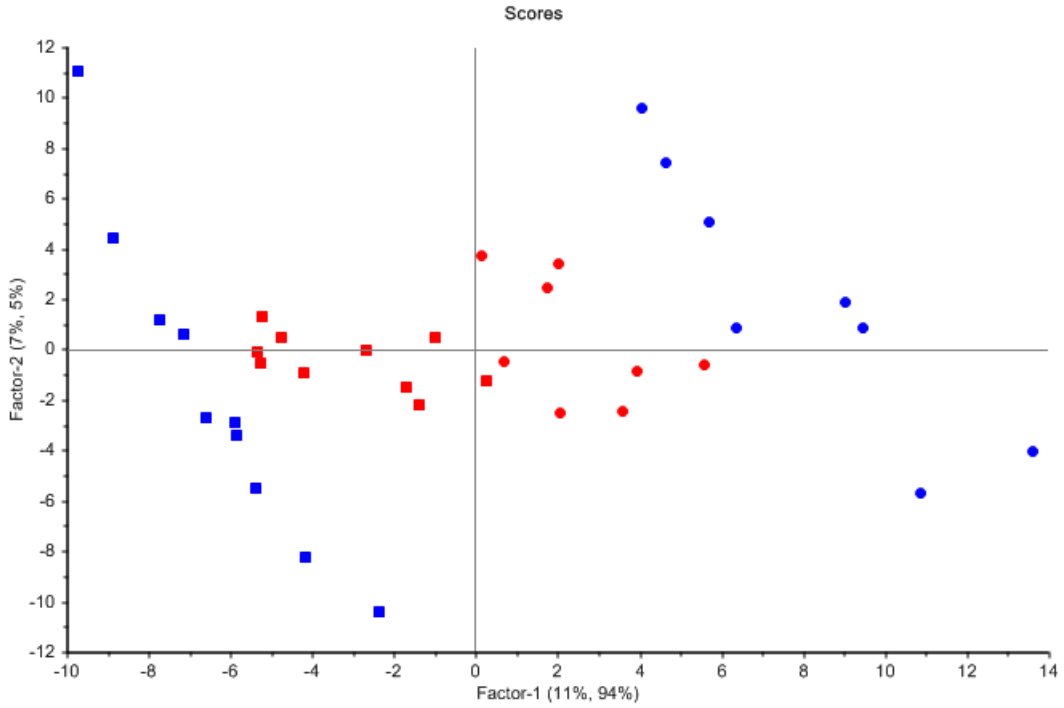


Figure 6.11: Scores for PLSR of 18 patients, from histograms of fully preprocessed data. Blue indicates original model, red is validation. Squares have high AHI, circles have zero

6.2.2 Comparing PLSR of preprocessed and raw data

In an effort to evaluate the preprocessing we employ, we construct a normalized histogram data matrix using the raw, unprocessed data, in its original resolution. The result is an 18×24310 matrix, or in other words, the preprocessing done reduces the size of the data by a factor greater than 25. Preforming analysis of this raw data is noticeably more cumbersome, and *The Unscrambler®X* is somewhat ill-suited to the task.

Proceeding exactly the same as in the previous section (AHI as the Y variable, 7 maximum components, full cross validation), we again see group separation in the first PC, although this time with a noticeable outlier (Figure 6.12). Examining the explained variance (Figure 6.13), we see that the first PC still claims to explain a majority of the variation in the data set, although it no longer jumps to 100% at the second. As with the preprocessed data, this analysis also does not take well to validation, and still, with only 18 samples, we cannot expect it to.

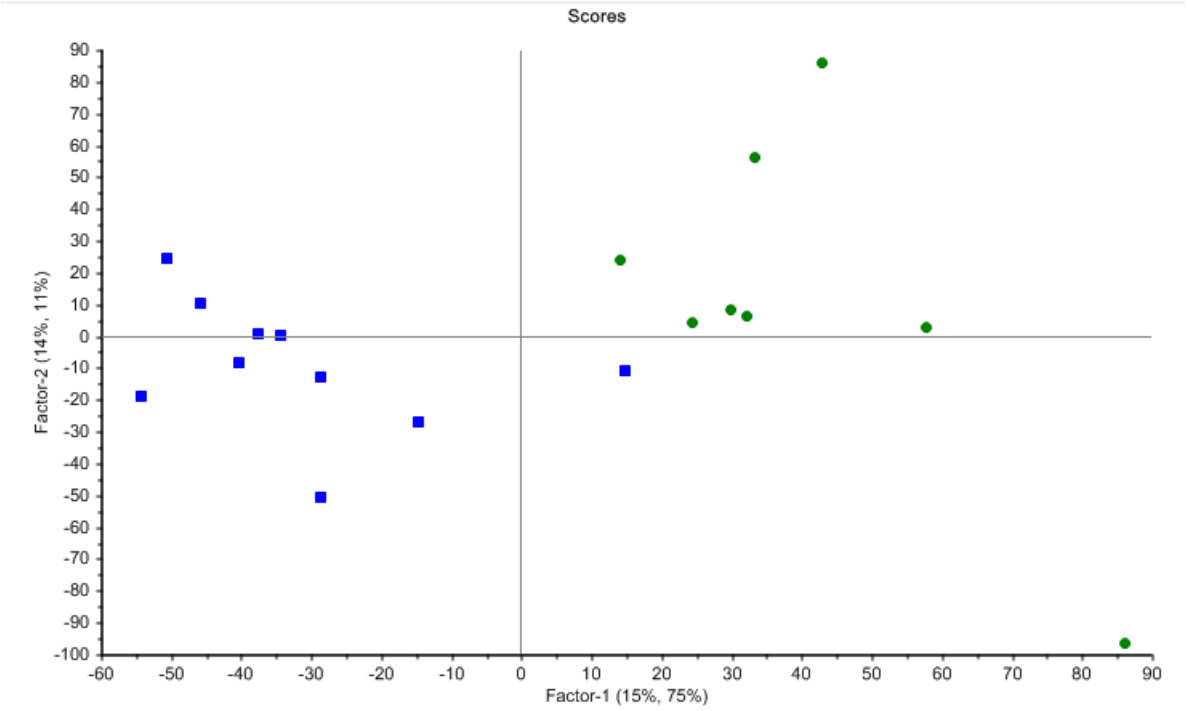


Figure 6.12: Scores for PLSR of 18 patients, from histograms of raw data. Green dots have high AHI, blue have zero

Looking still at the explained variance, we may interpret that this model picks up a larger amount of dynamic variation, indicating that some detail is lost through the preprocessing. However, the significant similarities of the scores plots of Figures 6.8 and 6.12, and the fact that the groups still tend to separate along the first principal component, leads us to continue to believe that the majority of the variance and dynamics related to apnea are preserved through the preprocessing. Therefore, at least within the scope of this thesis, the ease of computation afforded by the data reduction and preprocessing presented in Section 5 will be worth the loss of detail.

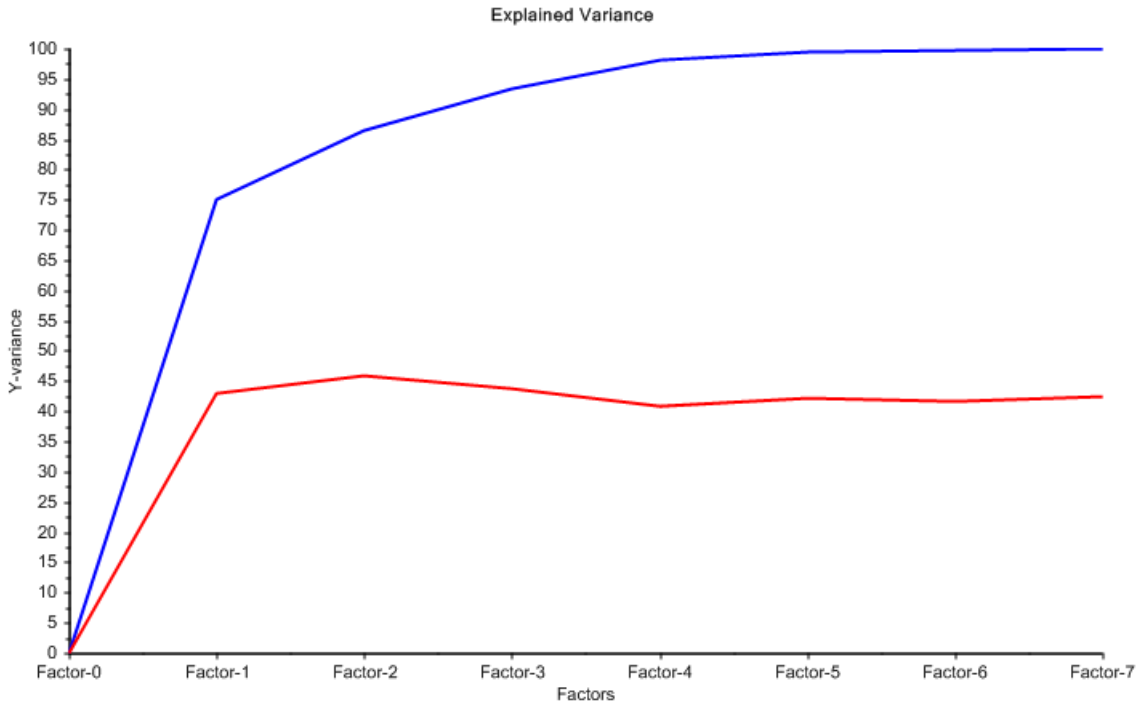


Figure 6.13: Explained variance for PLSR of raw data, 18 patients

6.2.3 PLSR of histogram-formulated data

We now wish to perform analysis on all the available data. The final data set consists of 70 patients, 44 of which have little or no sign of apnea. The AHI distribution of this sample set can be seen in Figure 6.14.

The fully preprocessed histogram-formultated data forms a sparse 70×1163 matrix. We perform PLSR as in the previous sections, with 7 maximum components, and full cross validation. However, we want to examine the possible influence of all our Y variables. Therefore, to begin our analysis, we leave nothing out, and include all the Y variables available (see Section 5.6.2).

Using all Y-variables

Figure 6.15 shows the explained variance from this analysis. We can see that the first principal component appears to pick up something significant, increasing both in the model predicted and the validated explained variance. After the first component, or perhaps the fourth is we are

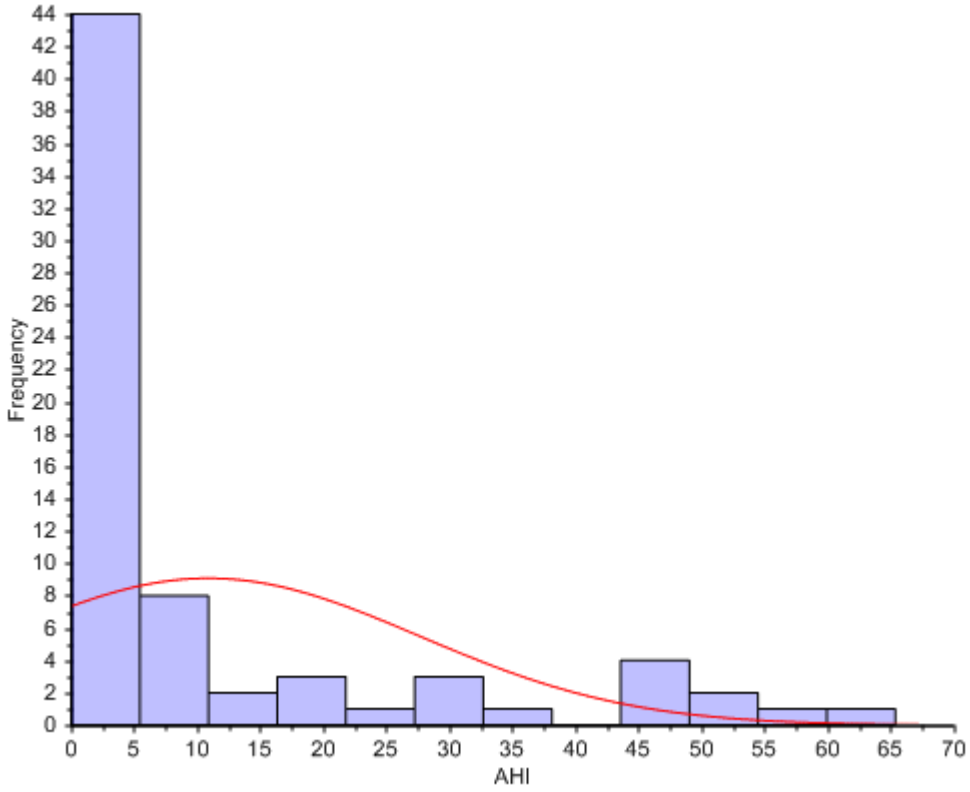


Figure 6.14: AHI distribution for the 70 test samples

being generous, the model is unable to significantly increase its validated performance. While using all the Y-variables, we are unable to achieve more than 15% explained variance.

We can verify this by repeating the calculations with an increased number of components. Setting the maximum components to the highest allowed number (*The Unscrambler®X* allows a maximum of 65 components), we see in Figure 6.16 the predicted performance converging towards 100% as we slowly force our model to perfectly fit the data, while the validated predictive ability not only fails to increase beyond 15%, but sinks slightly as we increase the number of components. The maximum validated performance occurs at approximately PC 4.

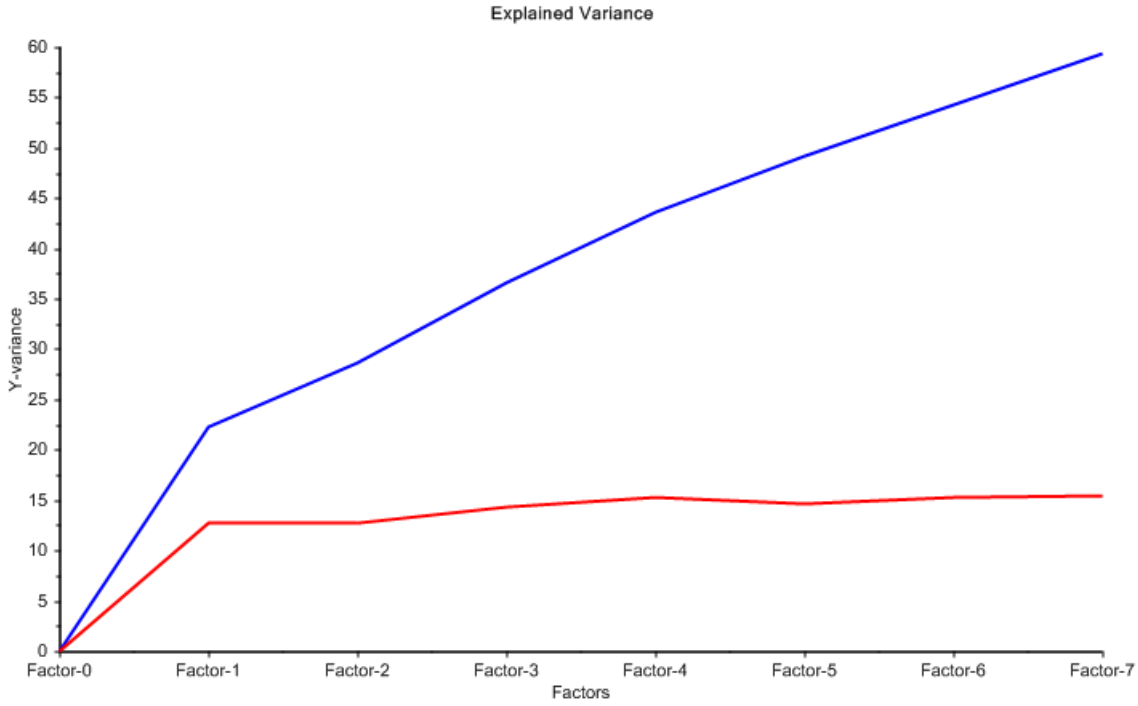


Figure 6.15: Explained variance for PLSR of 70 patients, using all available Y-variables

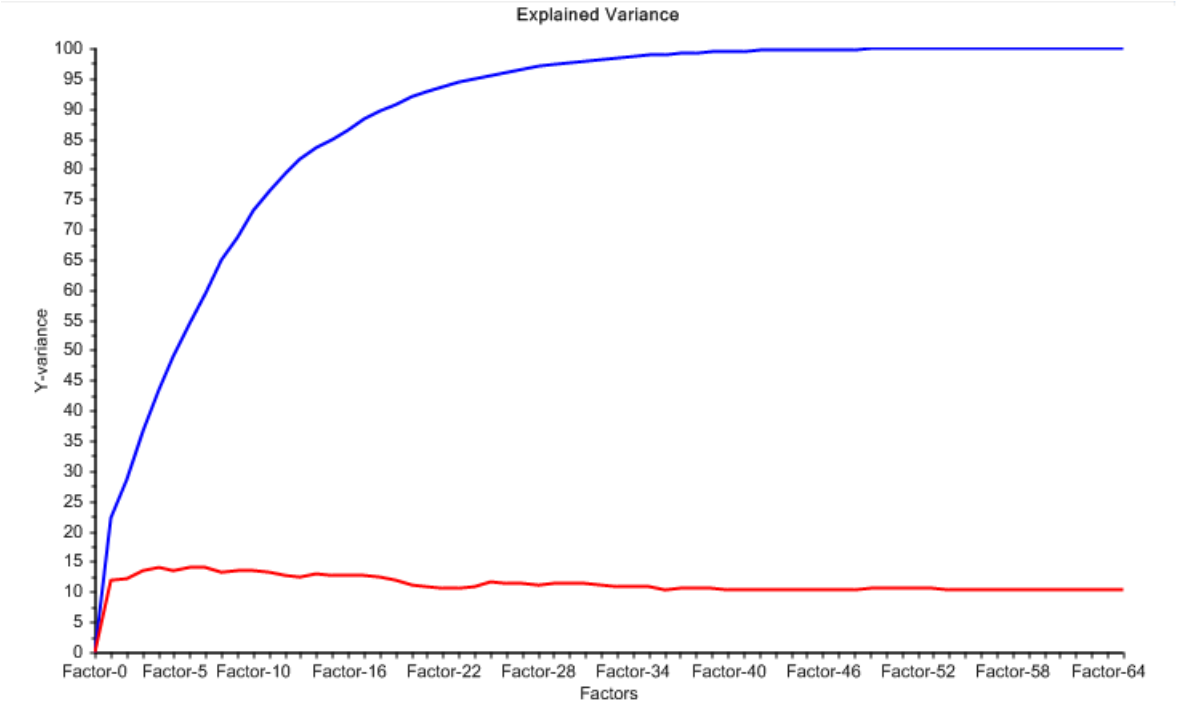


Figure 6.16: Explained variance for PLSR of 70 patients, using all available Y-variables and allowing for 65 components

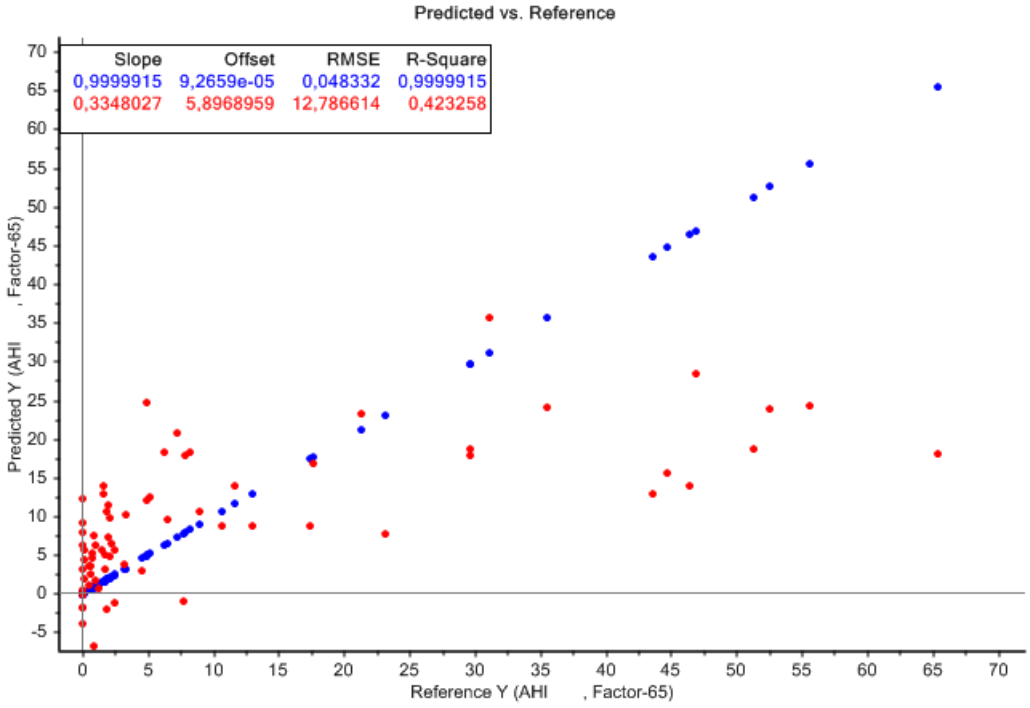


Figure 6.17: Including 65 factors gives 'perfect' prediction, but validation shows that this is far from true

We may now begin to narrow our search, looking for the Y-variables that we can best predict from our data set. Figure 6.18 shows the explained variance of each individual Y-variable in the seven-component analysis, and Table 6.1 shows the explained variance after each component for the six best explained variables.

The top two variables are ASum and FlowE. We would expect these to behave similarly, as they contain almost the same data - ASum is the sum of all apnea events, and FlowE is the same plus the variables BodE (body event - a large body movement disturbing the recordings) and FlowL (flow limitation - a reduction in airflow that has not been scored as an apnea event). Three of the other top performers are all individual apnea events - CA (Central Apnea), OSA (Obstructive Apnea), and HypA (Hypopnea), and are three of the four components of ASum (the fourth being the least common apnea type, mixed apnea (MixA)).

The only top 6 variables without a direct relation to the other four is AHI and DSI, the desaturation index. Both are averages per hour for the entire night; AHI is average number of apnea events per hour, DSI is average number of desaturation events per hour. It is not unexpected that these two behave similarly, as an apnea event will be accompanied by a desaturation event (although the reverse is not necessarily true).

That ASum is the best performing Y-variable when performing this unbiased unweighted analysis is a very promising result. Thus far, we have been using AHI as our primary predicted variable. However, since AHI is the average per hour throughout the entire night, it contains information from outside the one hour we have chosen to work with. ASum is thus a better choice, being the sum of seconds of apnea events in the data we are actually using.

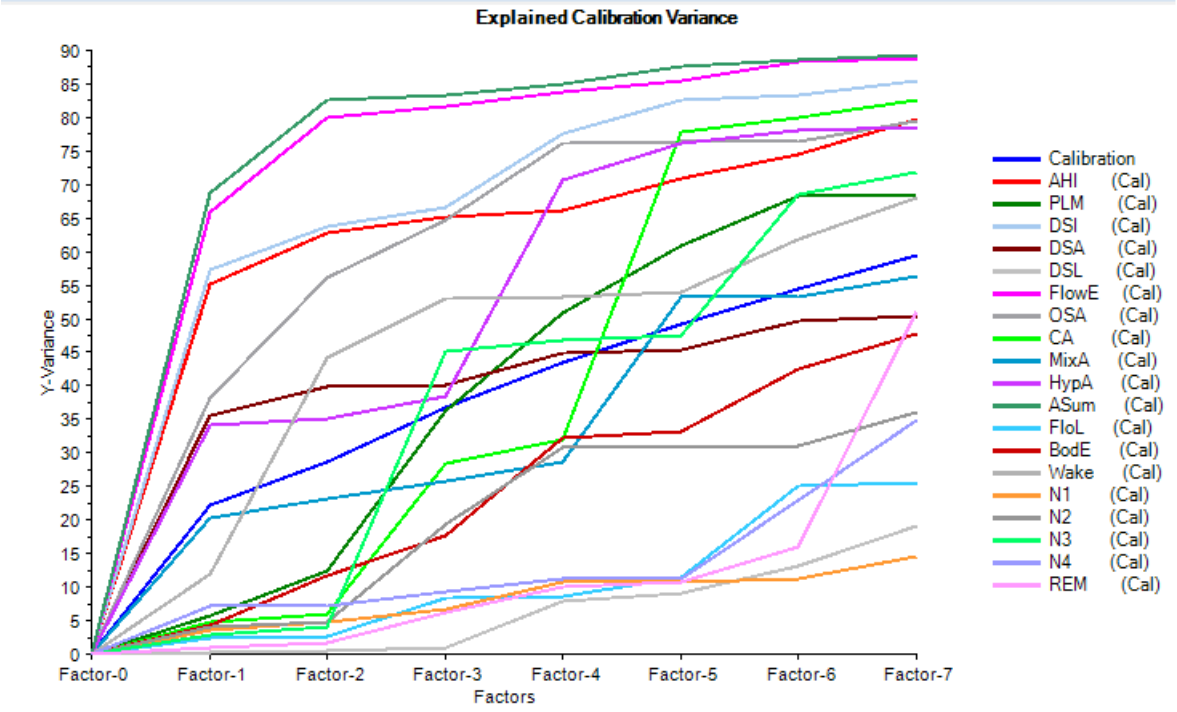


Figure 6.18: Explained variance each Y variable in PLSR of 70 patients. Calibration values only

Factor	ASum	FlowE	DSI	CA	OSA	AHI	HypA
Factor 1	68,76444	65,88893	57,20355	4,79009	38,10143	55,09525	34,0527
Factor 2	82,60652	80,04798	63,74527	5,923184	56,12182	62,86323	35,07401
Factor 3	83,33675	81,61374	66,51811	28,42834	64,69478	65,08066	38,32149
Factor 4	85,00224	83,69391	77,51285	31,95509	76,22023	66,22291	70,65981
Factor 5	87,53642	85,56057	82,54999	77,84238	76,36018	70,93343	76,13943
Factor 6	88,63451	88,29353	83,26016	79,87802	76,39275	74,5199	78,04833
Factor 7	89,25108	88,83456	85,39812	82,70749	79,45043	79,75278	78,54467

Table 6.1: The explained variance of the five best performing y variables after 7 components

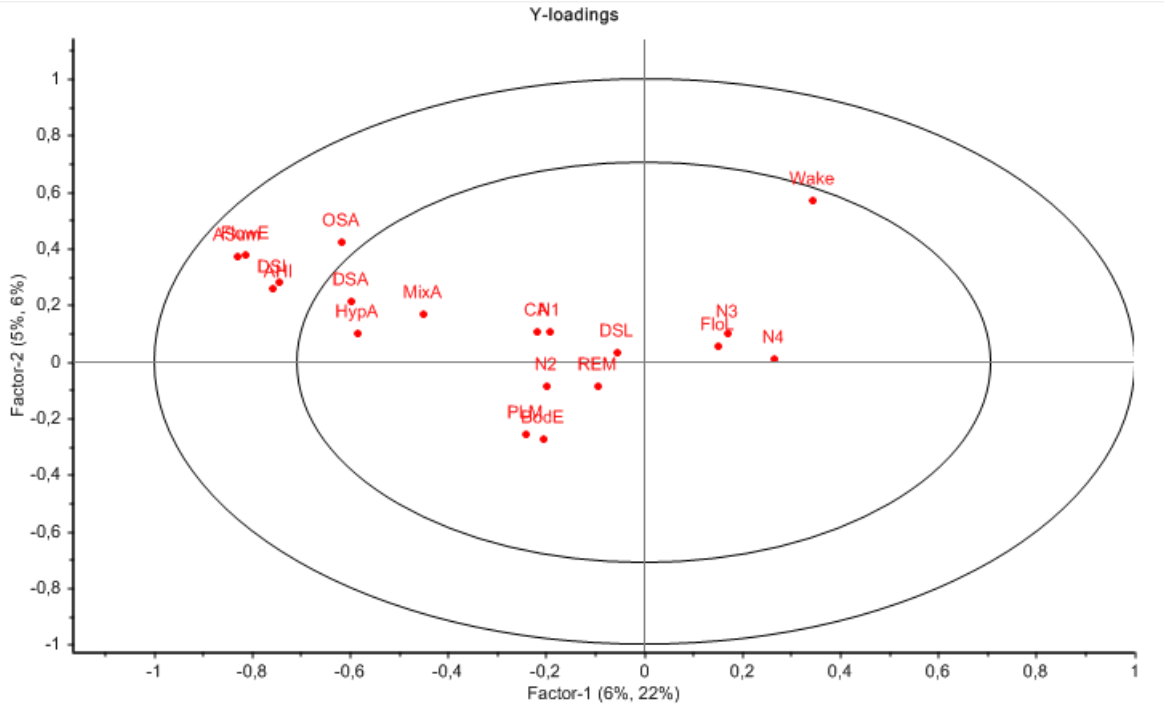


Figure 6.19: Y loadings for PLSR of 70 patients, using all available Y-variables

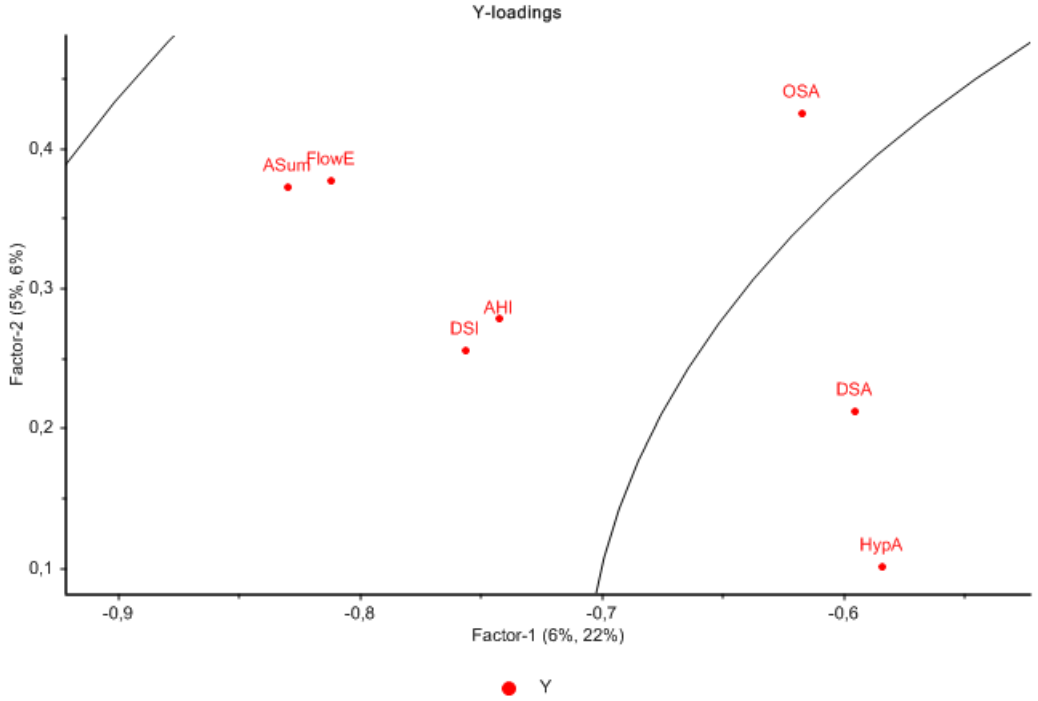


Figure 6.20: Y loadings for PLSR of 70 patients, zoomed in on the most important variables

Using ASum only

We now repeat the analysis, using only ASum as the Y-variable and using four components. We see from Table 6.2 and Figure 6.21 that the model reaches close to 100% explained variance within four components, but that the validation is not able to reach more than a maximum of 52.7%, which it does at the second component.

Factor	Calibration	Validation
Factor 1	78,64323	46,61226
Factor 2	92,85884	52,71743
Factor 3	98,54221	51,85374
Factor 4	99,51511	50,99939

Table 6.2: Explained variance for calibration and validation for four factors, PLSR with Y=ASum

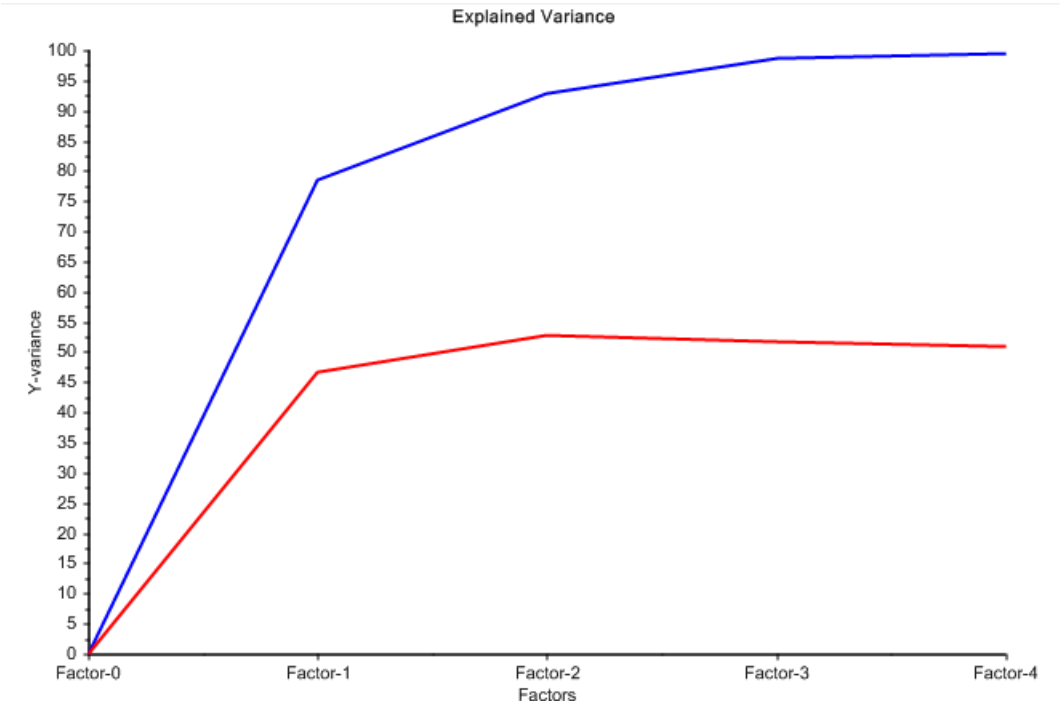


Figure 6.21: Explained variance for PLSR of 70 patients, with ASum as the Y-variable

In Figure 6.22 we see that the ten points with the greatest sum of apnea events branch out from the rest. In figure 6.23 we see the values of ASum predicted by the model plotted against the actual values, after only two components are included.

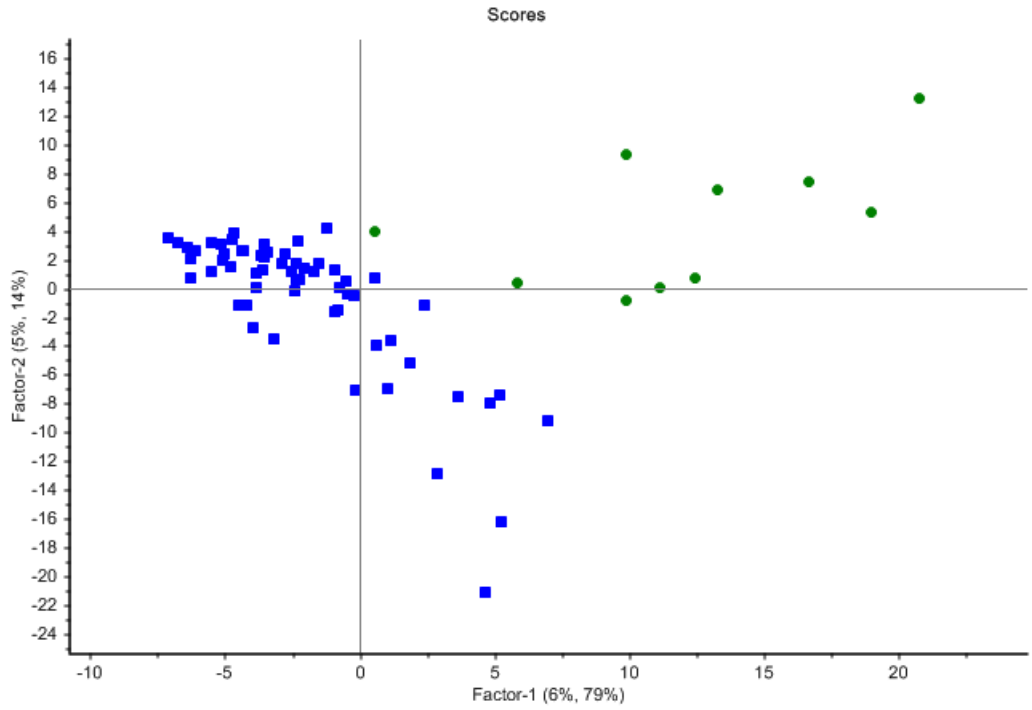


Figure 6.22: Scores plot for the first two components of PLSR of 70 patients, ASum as Y. Green dots mark the 10 patients with highest ASum

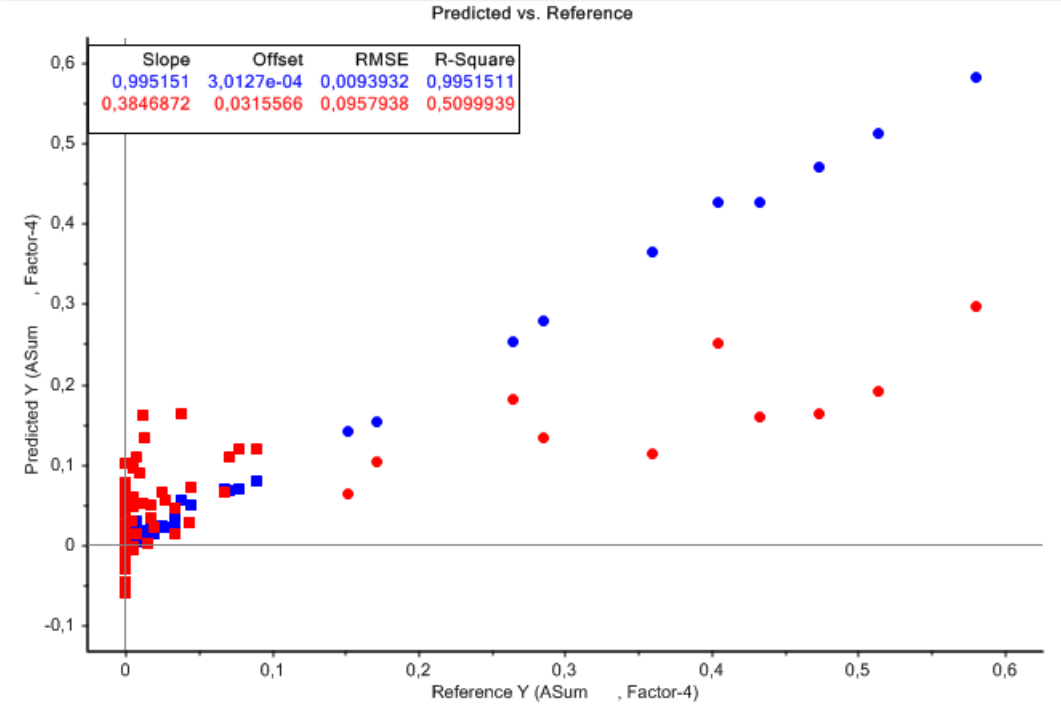


Figure 6.23: Predicted vs. reference for ASum prediction. Circles indicate the top 10 ASum patients, squares are the rest

Weighting the objects

From Figure 6.23 we can see that the model appears to be more accurate for the lower values, while not quite being able to reach the highest ones. This may be due to the imbalances in our data set. The number of healthy patients far outweigh the number of high-apnea ones, so it should come as little surprise that 'healthiness' is prioritized above 'sickness' in the resulting model.

To counteract this effect, we downweight all but the ten patients with the highest ASum. We choose to use a factor of 4 (i.e. multiply all but the top ten by 0.25), because this is the approximate ratio of no AHI to very high AHI in our data.

The result is a significant improvement on the predictive ability and validity of the model. **We have thus managed to achieve a model that has good predictive ability for the data set, and that also performs well under validation.**

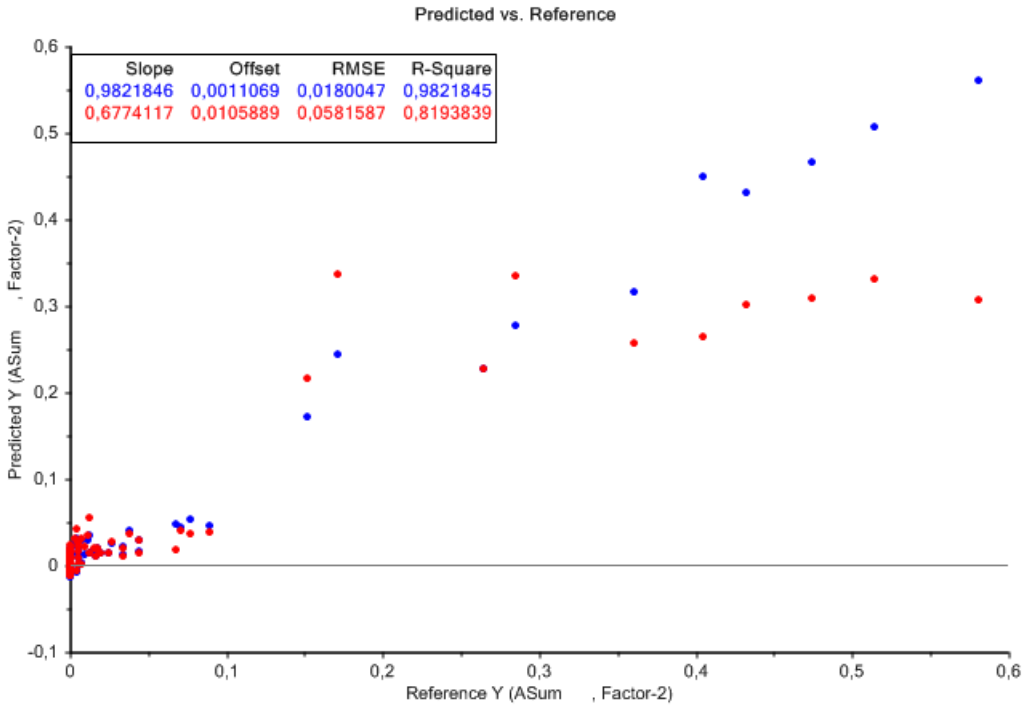


Figure 6.24: Predicted vs reference for PLSR of 70 patients with ASum as the Y-variable, after data has been weighted to remove imbalances between sick and healthy patients. The ability to recognize sick patients has improved significantly in comparison to the unweighted results in figure 6.23

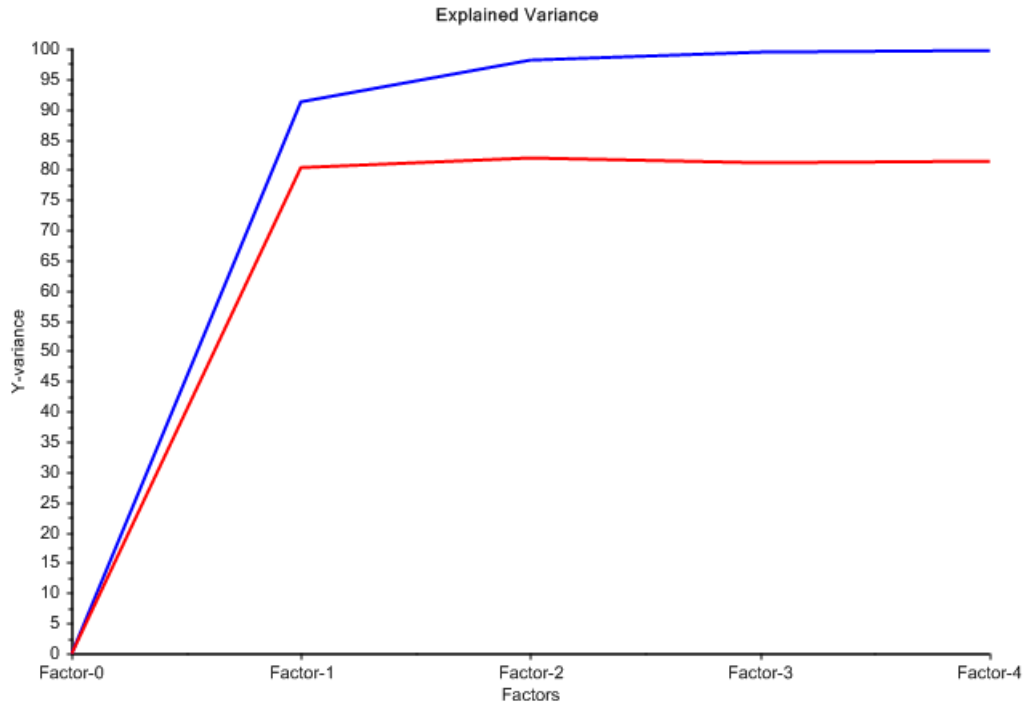


Figure 6.25: Explained variance for weighted PLSR of 70 patients with ASum as the Y-variable

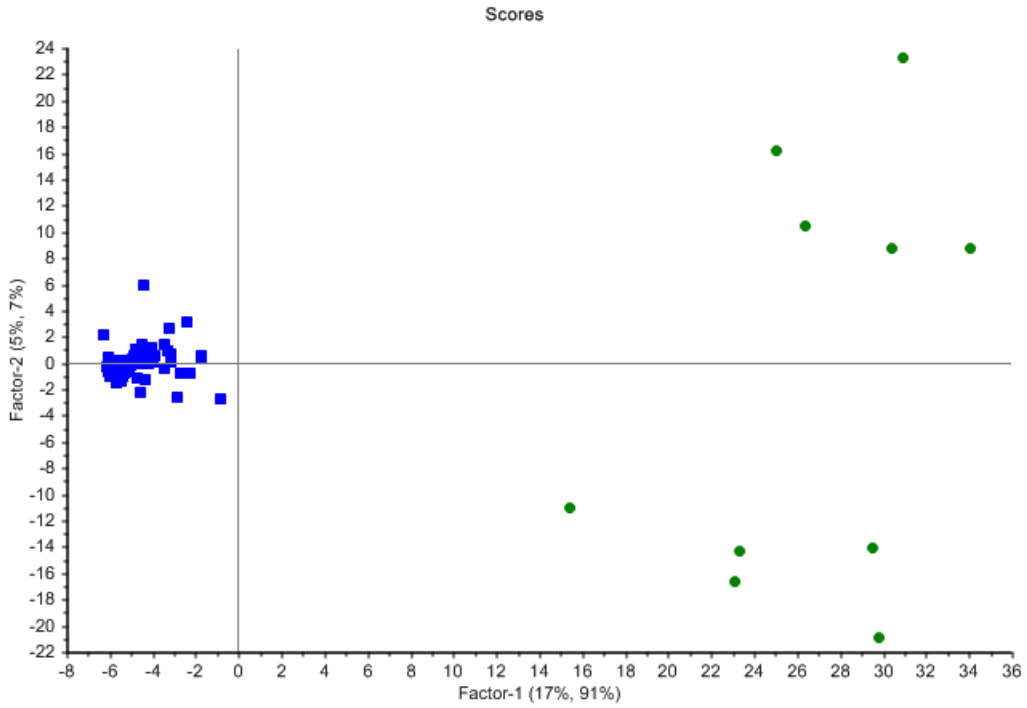


Figure 6.26: Scores for weighted PLSR of 70 patients with ASum as the Y-variable. Green dots indicate the top ten highest ASum values

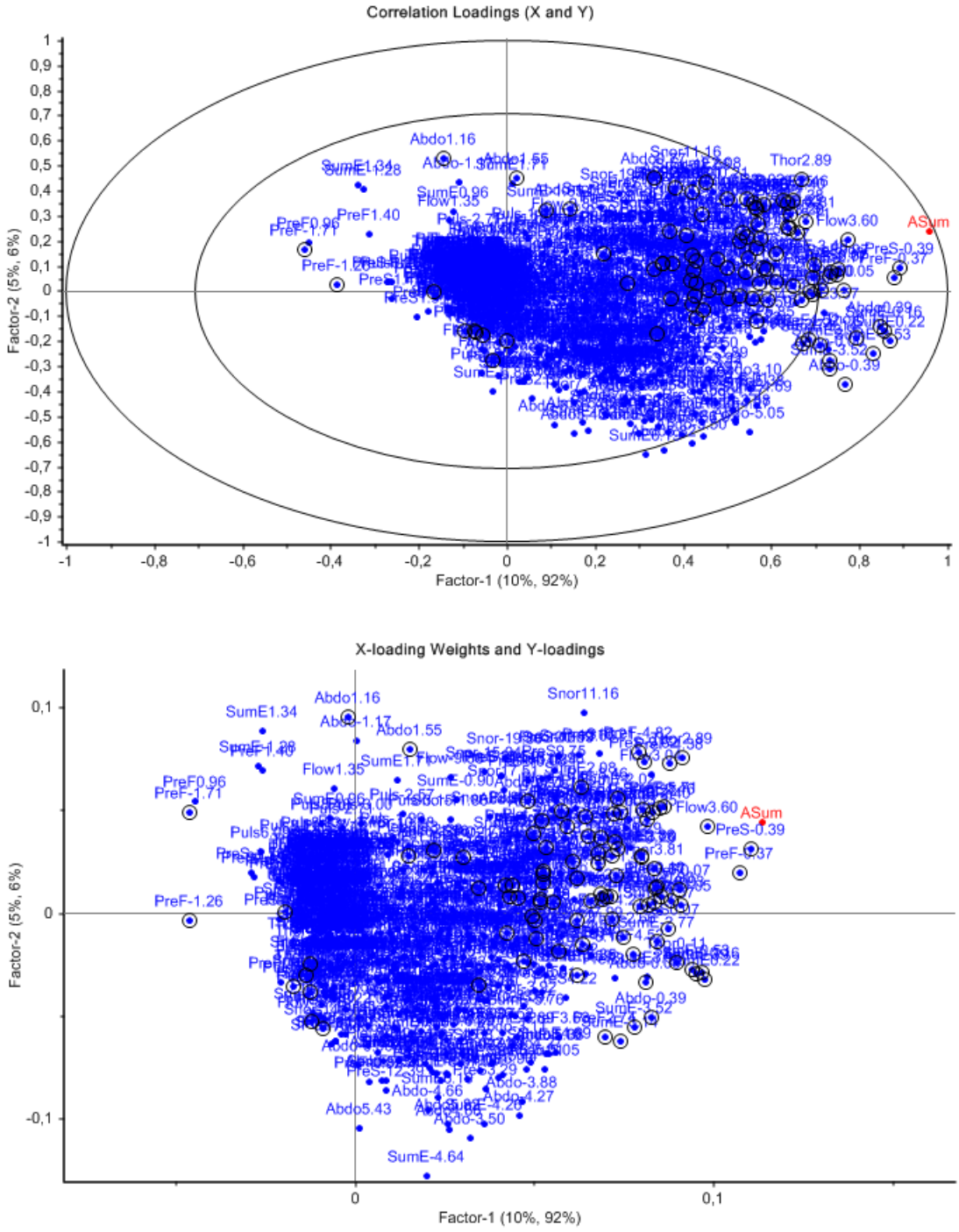


Figure 6.27: Loadings and loadings weights for weighted PLSR of 70 patients with ASum as the Y-variable

6.2.4 PLSR of frequency-formulated data

We attempt the same analysis using frequency-formulated data. We have 70×512 matrices formed by the Y-variables and binned FFT and PSD. Proceding as in the previous section, we are not able to get results of the same quality. The best explained variance achieved through this method with this data is shown in Figure 6.28, which uses a matrix formed from the power spectra of the sensors. There is still a tendency of higher apnea separating along the first principal component (this time in the positive direction), indicating that there is some connection between what structure the model is able to recognize and the apnea of the patient.

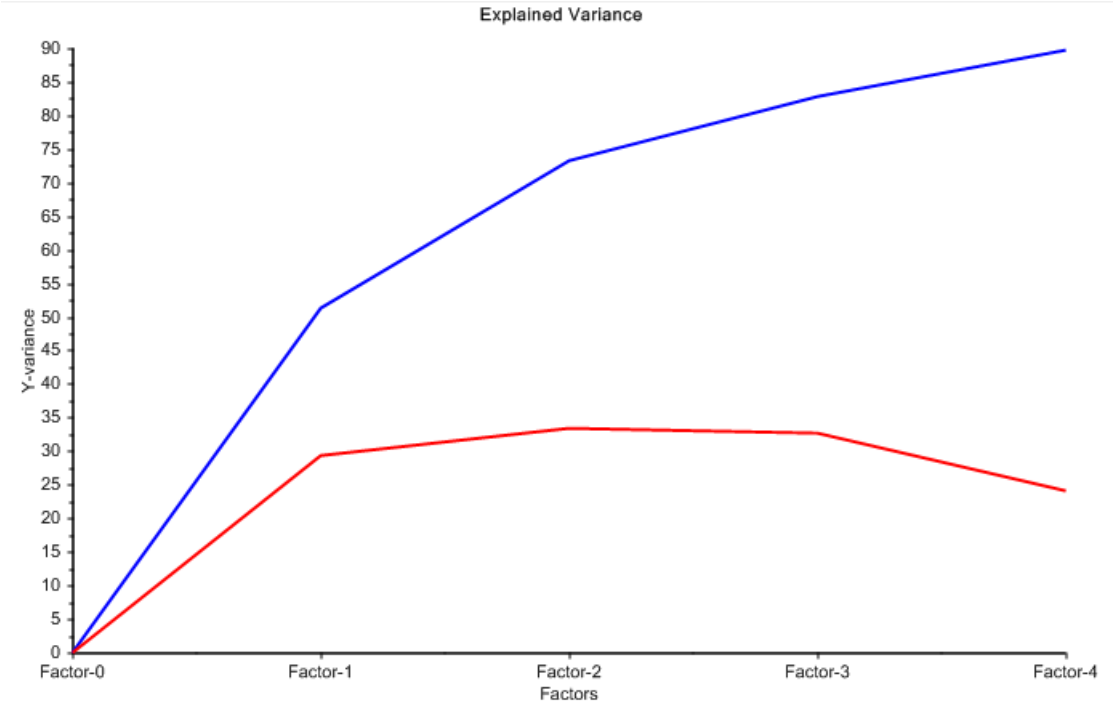


Figure 6.28: Explained variance for weighted PLSR of 70 patients with ASum as the Y-variable, X matrix composed of binned power spectra

We are, through an ardous process of trial and error, able to achieve an apparently much better result using the method of variable selection. By gradually excluding the columns of least importance to the model, we are able to get iteratively better results, and in the end, we have a model with almost the same predictive ability as in the previous section, as shown in Figure 6.30.

This result looks excellent, but it is false. In calculating this, we are using only 8 of the 512

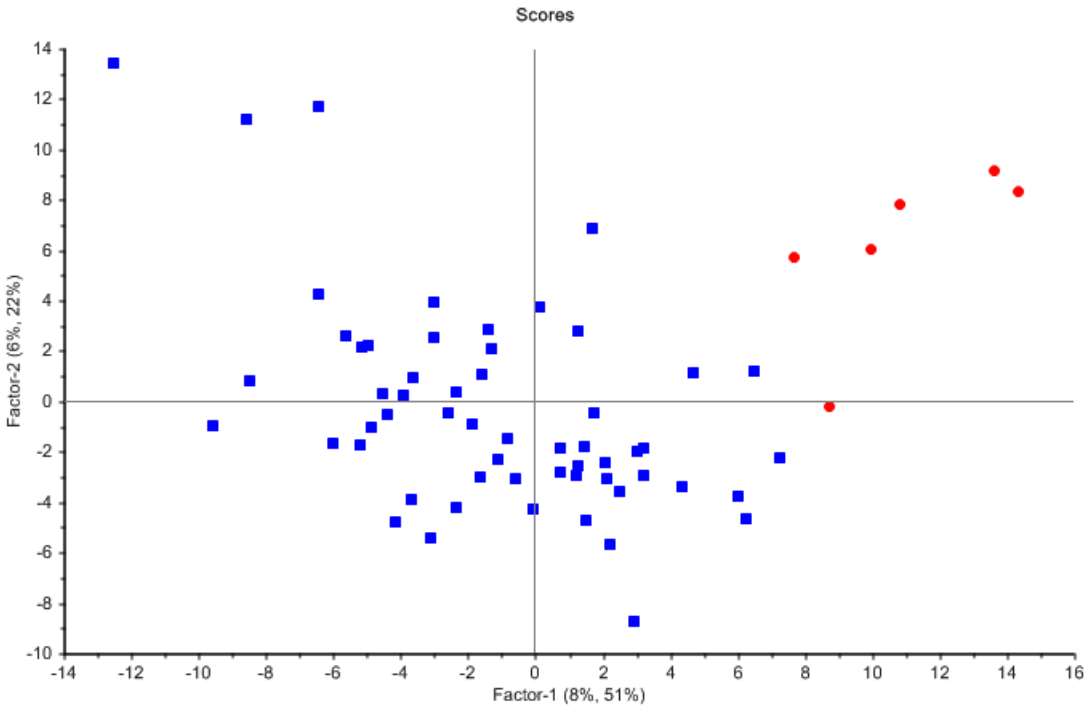


Figure 6.29: Scores for weighted PLSR of 70 patients with ASum as the Y-variable, X matrix composed of binned power spectra

columns in the original matrix. We include this here to illustrate the dangers of forcing good results to appear even when the data does not facilitate them. The only true conclusion we can draw is that **in their pure form, frequency-formulated data matrices appear less suited for apnea prediction than histogram amplitude distributions**.

There are several reasons why the frequency analysis might work less well than the simpler amplitude distribution. First of all, the data we are working with are not stationary, so one of the fundamental assumptions of the Fourier transform is not met. A different type of frequency analysis, such as wavelet analysis or the Hilbert-Huang transform, might be better suited to the task.

Furthermore, apnea might simply not be very strongly related to the spectral content of the signals. This might not appear to be the case when examining the raw signals of a patient like 2407 (see Figure 5.3), where there appears to be a strong periodic tendency to the apnea. However, as we have noted, this patient exhibits the Cheyne-Stokes breathing pattern, which is cer-

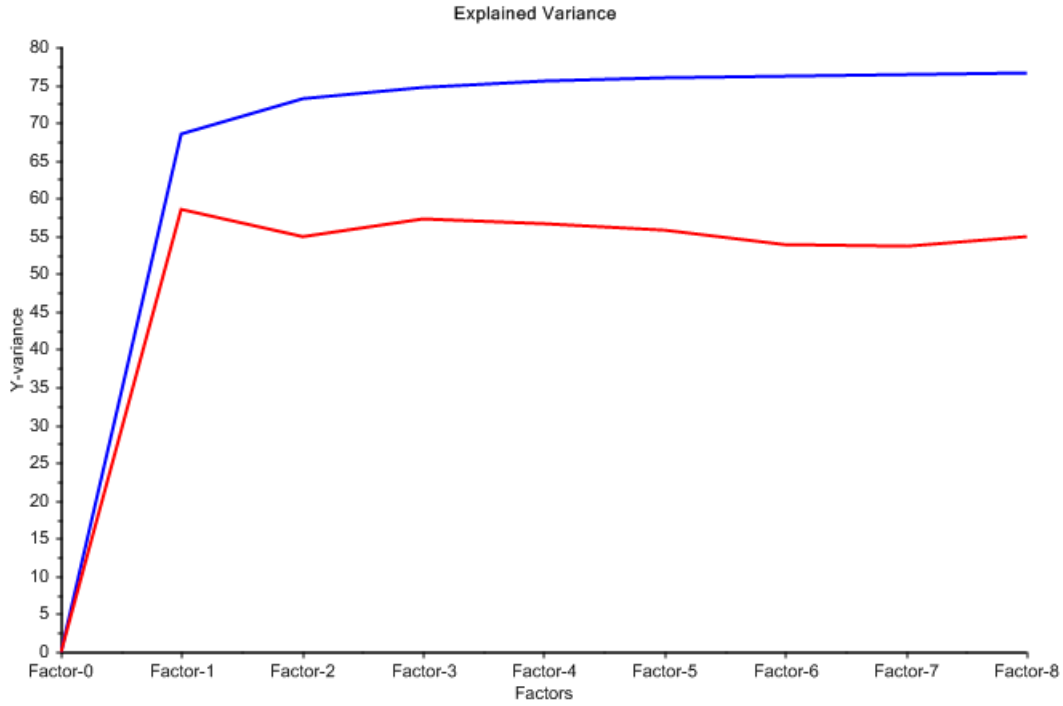


Figure 6.30: Excellent explained variance is achieved for weighted PLSR of 70 patients with OSA as the Y-variable,but only after performing vigorous variable selection

tainly not the case for all apnea patients. Perhaps, then, frequency based analysis are appropriate for recognizing Cheyne-Stokes breathing, but less suited for analyzing patients with a more sporadic apnea pattern.

7

Discussion

This section will discuss and evaluate the methods presented in the preceding chapters. Points of improvement and avenues of further work will be explored.

7.1 Discussion

The model presented in Section 6.2.3 achieves good predictive ability of ASum for the data we have available, using a subset of 10 sensors from the full PSG recordings. Although we were only able to achieve such a good result after weighting the data based on prior knowledge, reducing the weight to a factor half (as opposed to one quarter) achieves similar results. As such, we are fairly confident that the weighting achieves the desired result of balancing the data set.

The loadings and loadings weights plots in Figure 6.27 are hard to read because of the large number of variables, but show that the ASum characteristic played a large part in building the model. As such, it might be harder to build a model of sleep apnea without this clinical knowledge.

It is difficult to draw any conclusions as to which sensors are the most important for the results from these plots. The histogram formulated matrix is sparse, so the more informative variables will most likely be the centers of the histogram, rather than any given sensor subset. Nevertheless, it might be possible to map the loadings plot more carefully (coloring each sensor,

for example), to see if patterns reveal themselves.

7.2 Limitations of the Data Set

Not meant for this kind of analysis

The data used in this work was not originally meant for this kind of analysis. As such, there was a number of challenges related to its use.

The lack of ability to retroactively track the specific hardware used to perform the recordings is detrimental, rendering us unable to discover connections and errors due to equipment. The difference in the detail level of the analysis performed also led to several data sets being discarded (e.g. those lacking respiratory scoring could not be used in this work).

The original hospital database was messy and rife with spelling errors. This led to project delays and less data being available. As mentioned in Chapter 4, a shift in equipment in 2012 meant older data became unavailable as well. Furthermore, the original software did not have options for mass exporting of data to a universal format. This task, when done manually, took significant time, and caused further project delays.

Once exported, the data itself had little to no explanatory documentation, and unravelling the purposes and meanings of the files was yet another time consuming and difficult process.

Lack of a truly 'healthy' test group

All patients used had been referred to the hospital for some reason or another. While not all have respiratory disturbances, it is likely that they all nevertheless have some problem related to sleep. As such, we do not have a truly healthy test group, only patients who are 'healthy with respect to respiration'. It is unknown how other sleep disorders might affect the results of our analysis.

No compensation for the first-night effect

Only a very small set of patients had more than one night of PSG recordings. For the significant majority, data was gathered for one night only, and as such we have no correction for the first-night effect of adaption difficulties[12].

Unbalanced with respect to AHI

The final data set was not balanced with regards to symptoms of respiratory disturbances. As shown in Section 6.2.3, good predictive ability in regards to apnea was achieved only once the data was weighted to compensate for this. It would of course be preferable if this was not necessary - if the original data set had a more balanced distribution of zero to severe apnea.

7.3 Improvements

7.3.1 Include a priori knowledge of scoring rules

As mentioned in Chapter 3, the soft modelling methods explored in this thesis use a bare minimum of a priori information. The goal of this was to see if the methods could discover structures in the data without aid. Seeing now that some structure does appear, the next step could be to extend the analysis using available information, most prominently the scoring rules of the AASM Manual[4].

It is not unreasonable to believe that greater success could be achieved by combining the multivariate soft modeling with more traditional univariate methods, and perhaps apply other pattern recognition methods again to that result.

7.3.2 Points of improvement in the preprocessing

Missing data correction

The missing data compensation scheme used in this thesis is very simplistic, and was only implemented for those sensors deemed to be the most problematic after visual inspection. There

might still be junk data included in the analysis, and the effects of these are unknown. It would be beneficial to spend some time to develop a more sophisticated method for data cleaning and quality control, individually for each sensor if necessary.

Furthermore, it was realized after much of the analysis had already been done that some pulse oximeter data still appeared erroneous after the zero rectification preformed in Section 5.4. Perhaps it would be better to simply discard all data after a dislodgement, rather than keeping data that might or might not be corrupt.

Potential error in the treatment of SpO2

The SPO2 is measured in percentage. Perhaps it is unwise to simply center it. On second thought, it might be better to scale it down (e.g. by 100), rather than subtracting the mean.

7.3.3 Different approach to frequency analysis

There are multiple more advanced methods of frequency analysis that could be explored. Short-time Fourier Transform, Wavelet analysis, or the Hilbert-Huang transform may all be better suited to this kind of nonstationary data. It might be possible to use frequency analysis to search specifically for the Cheyne-Stokes breathing pattern.

7.3.4 Expanding beyond a single hour

This thesis made the simplification of reducing each data set to one hour of steady sleep. This simplification relies on access to a pre-scored hypnogram, which is unavailable in most commercial use-cases. If we were to expand beyond this single hour of deep sleep, we would have to expand the work to include, if not a full sleep staging algorithm, then at least some manner of recognizing whether or not the subject is sleeping at all. Upon this expansion, the potential difference of respiration and apnea characteristics between light and deep sleep must also be taken into account, adding an inevitable time-dimension to the work.

7.3.5 Including more sensors from the original data

The 10 sensor subset explored in this thesis was originally chosen for convenience, to aid in the initial infrastructure development. However, there was never time to go back and include a greater range of the original data. As such, there are significant unexplored and untreated data still left in our database.

7.4 An alternative idea for person-decoupled matrix formulation

The multi-person matrices formulated in this thesis all have individual patients as objects, with their data summarized as rows in the data matrix. This makes the analysis person-dependent, and perhaps harder to generalize. An alternative approach would be to attempt to decouple data and individuals, by separating epochs of data and shuffling them randomly. This would result in a data matrix with a much greater amount of objects, but significantly more sparse in terms of variables.

Provided appropriate preprocessing, this approach might result in a general respiration-apnea model independent of the person of origin. However, it might also result in some patients dominating the model. In any case, it is an avenue worth exploring.

7.5 Further Work: Outline of an extensive research project

If this work were to be continued, we would begin from scratch by gathering a new set of data. The process would be tracked, and the equipment used as well as the individual analyst scoring the results would be recorded. The test group would consist of a balanced number of zero to severe apnea cases, and we would like to gather data from at least 200 patients.

Additionally, the original idea behind this work was to find a more lightweight and easy way of recognizing apnea. As such, a data gathering project should include commercial wearable

products, performing double registrations with traditional PSG equipment and these on-market products. These market actors have typically focused only on their own product in what little validation they have performed[27]. This work would include several, and look for a combination of sensors that could be useful.

In an ideal world, we would also gather multiple nights of data for each person. This might not be possible with PSG recordings, but perhaps a PSG subset and/or the lightweight commercial wearable sensors could be sent home with the patients, and allowed to gather data from, say, two consecutive weeks of sleep.

Then, with data range increased and quality improved, we would test the same methods developed in this thesis. We would also like to expand the analysis by exploring the patient-independent 30s epoch idea, as well as configurations involving algorithms based on the scoring rules in the AASM Manual[4]. We would also include more of the original sensors from the PSG data than what this thesis has explored. We would then explore which combination of sensors give the best results, and thus gain an understanding of the necessary parameters to include in a product capable of recognizing sleep apnea.

This outlines a research project that might eventually lead to a commercial at-home sleep monitor with a validated clinical ability to recognize respiratory disturbances in its customers.

8

Conclusion

This thesis has described the process of gaining access to, exporting, and formatting polysomnography data for analysis. A selection of preprocessing and data reduction techniques are described and implemented, and multivariate analysis and modelling is performed on several different configurations of the data.

Principal Component Analysis models of single patient data sets, using time instances as objects and individual sensors as variables, revealed that the main characteristics of apnea present in a 10 sensor subset are preserved through the preprocessing and data reduction.

To build a predictive model based on multiple patients, we let the patients be the objects and explore two methods of variable definition: binned spectral components after a transformation to the frequency domain, and amplitude distributions in the form of histograms.

Partial Least Squares Analysis of frequency-formulated data appear to indicate that the time domain may be more suitable for apnea recognition than the frequency domain. However, this conclusion cannot be properly drawn before more sophisticated frequency analysis tools are attempted than just the Fourier transform.

Partial Least Squares Analysis of a subset of histogram-formulated antithetical patients (either very high apnea or none at all) has the two groups separating very clearly along the first

principal component, indicating that our methods are applicable at least in extreme cases. However, this subset is too small to stand up to validation.

Finally, PLSR of all available data in histogram formulation achieves a model with good predictive ability, that also performs well under validation. The model uses four principal components, and weights the data to compensate for the asymmetrical distribution of healthy and sick patients in the data set.

This work has thus succeeded in building a model for a sleep apnea characteristic with good predictive ability for the data set. We conclude by offering a discussion of the limitation of our work, as well as suggesting the outline of an extensive research project that might eventually lead to a commercial at-home sleep monitor with a validated clinical ability to recognize respiratory disturbances in its customers.

Appendix A

Acronyms

In alphabetical order

AASM American Academy of Sleep Medicine

AHI Apnea-Hypopnea Index

ASum Sum of apnea

BodE Body Event

CA Central Sleep Apnea

DSA Desaturation Average

DSI Desaturation Index

DSL Desaturation Length

ECG Electrocardiography

EEG Electroencephalography

EMG Electromyography

EOG Electrooculography

FloL Flow Limitation

FlowE Flow Events

HypA Hypopnea

MA Mixed Apnea

NIPALS Nonlinear Iterative Partial Least Squares

OSA Obstructive Sleep Apnea

PC Principal Component

PCA Principal Component Analysis

PLM Periodic Leg Movement

PLSR Partial Least Squares Regression

PSG Polysomnography

REK Regional Etisk Komité (Regional Ethical Comittee)

REM Rapid Eye Movement

RIP Respiratory Inductive Plethysmography

SpO2 Blood oxygen saturation

Bibliography

[1] American Sleep Apnea Association. Sleep apnea, 2015.

[2] Aristotle. On sleep and sleeplessness. *http://classics.mit.edu/Aristotle/sleep.html*, 350 B.C.E.

[3] H. Banaee, M. U. Ahmed, and A. Loutfi. Data mining for wearable sensors in health monitoring systems: A review of recent trends and challenges. *Sensors*, 13:17472–17500, 2013.

[4] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. L. Marcus, and B. V. Vaughn. *The AASM Manual for the Scoring of Sleep and Associated Events*. American Academy of Sleep Science, Darien IL, 2nd edition, 2012.

[5] S. Chokroverty. *Atlas of Sleep Medicine*. Elsevier Inc., 2014.

[6] Descartes. Descartes and the pineal gland. *http://plato.stanford.edu/entries/pineal-gland/index.html*, 2013.

[7] M. DiAz-GonzaLez, R. Hormiga, A. Mari, Rivero, D. Y., and S. M. M. Morales. Concordance among measurements obtained by three pulse oximeters currently used by health professionals. *Journal of Clinical and Diagnostic Research*, 8, 2014.

[8] F. Dieterle. Overfitting, underfitting and model complexity.

[9] M. Engrstrøm. Polysomnografi ved utredning av svnlidelser. *Tidsskrift for Den Norske Legeforening*, 1:58–62, 2013.

[10] M. Engrstrøm. Investigation of obstructive respiratory disturbance during sleep. *Tidsskrift for Den Norske Legeforening*, 21:1962–4, 2015.

[11] K. H. Esbensen and D. Guyot. *Multivariate Data Analysis - in practise*. Camo Process AS, 5 edition, 2002.

[12] A. et al. The first night effect: An eeg study of sleep. *Psychophysiology*, 2:263–266, 1966.

[13] L. B. et al. The first-night effect may last more than one night. *Journal of Psychiatric Research*, 35(3):165–172, 2001.

[14] C. Guilleminault. *Clinical Neurophysiology of Sleep Disorders*. Elsevier Health Sciences, 2005.

[15] S. T. Håkensen. Vanlig med pustestopp i svne, 2013.

[16] L. F. Haas. Hans Berger (1873-1941), Richard Caton (1842-1926), and electroencephalography. *Journal of Neurology, Neurosurgery and Psychiatry.*, 74(1):9, 2003.

[17] R. Hierons. Willis's contribution to clinical medicine and neurology. *Journal of the nurological sciences*, 4:1–13, 1967.

[18] J. E. Jackson. *A User's Guide to Principal Components*. Wiley, 1991.

[19] R. N. Kingshott and N. J. Douglas. The effect of in-laboratory polysomnogaphy on sleep and objective daytime sleepiness. *Sleep*, 23(8), 2000.

[20] D. B. Kirsch. Pathophysiology of sleep apnea. *Chest*, 139(4):939–946, 2011.

[21] H. Martens and T. Næs. *Multivariate Calibration*. Wiley, 1989.

[22] MathWorks. Matlab histogram documentation.

[23] S. C. Mukhopadhyay. Wearable sensors for human activity monitoring: A review. *IEEE Sensors Journal*, 15, 2015.

[24] U. D. of Health and H. Services. What is sleep apnea?, 2012.

[25] S. E. of Philosophy. Alcmaeon. *http://plato.stanford.edu/entries/alcmaeon/*, 2003.

[26] A. Rechtschaffen and A. E. Kales. *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. Public Health Service, U.S. Government Printing Office, 1nd edition, 1968.

[27] K. Russo, B. Goparaju, and M. T. Bianchi. Consumer sleep monitors: is there a baby in the bathwater? *Nature and Science of Sleep*, 7:147–157, 2015.

[28] T. Shochat. Impact of lifestyle and technology developments on sleep. *Nature and Science of Sleep*, 4:19–31, 2012.

[29] SOMNOMedics. Somnoscreen™plus brochure.

[30] E. Stratmoen. A brief history of dream analysis and interpretation.

[31] W. W. Tryon. Issues of validity in actigraphic sleep assessment. *Sleep*, 27(1):158–165, 2004.

[32] S. Wold, M. Sjostrom, and L. Eriksson. Pls-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58:109–130, 2001.