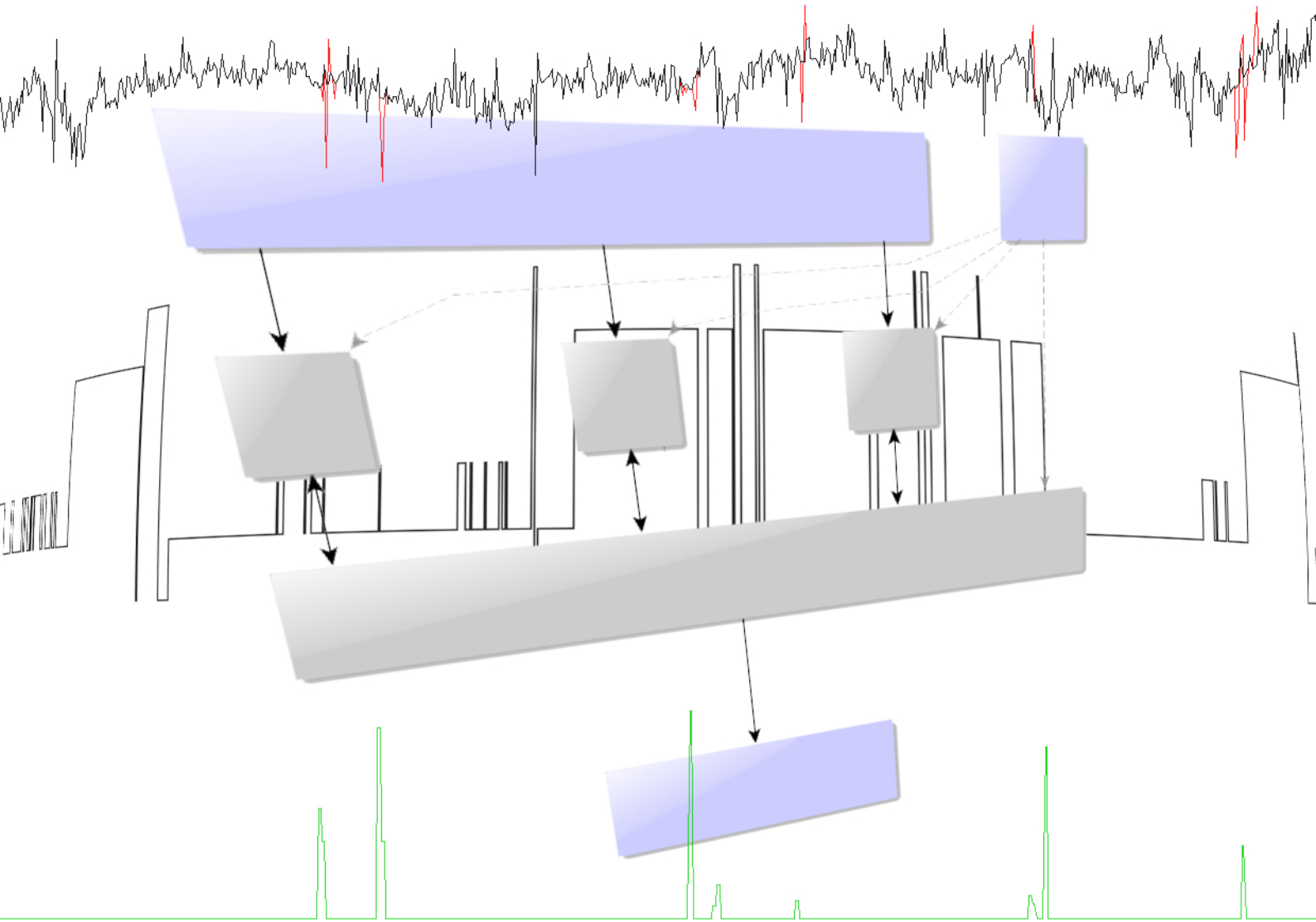


A framework for sleep staging based on unobtrusive measurements



Declaration

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions.

The work was performed under the guidance of Dr. ir. Verus Pronk and Dr. Henriette van Vugt in the capacity of supervisors of the research process at the Philips Research *High Tech Campus*, Eindhoven, The Netherlands, and Dr. ir. Robbert Jan Beun in the capacity of first supervisor at Utrecht University, the Netherlands. The aforementioned supervisors together with dr. Rogier van Eijk constitute the graduation committee.

Benny van der Vijgh

7th of November 2010,
Utrecht, The Netherlands.

Preface

This thesis is based on research conducted between February 2009 and February 2010 at the Philips Research institute *High Tech Campus* in Eindhoven and is submitted in partial fulfilment of the requirements for a Dutch Master of Science Degree in Cognitive Artificial Intelligence at Utrecht University.

As is blissfully inevitable in the field of cognitive science, the research and this thesis are both a multidisciplinary effort, containing elements from several disciplines such as physiology, information technology, digital signal processing, statistics and neurology. Therefore, this thesis is written in a manner which allows readers who do not have a firm basis in all these disciplines to still grasp the contents of this work. All basic notions from the different disciplines are explained and included in the glossary for later recall if necessary.

I would like to take this opportunity to acknowledge the invaluable effort that the three supervisors, who took guidance over the research and writing process, have taken.

Dr. ir. Verus Pronk and Dr. Henriette van Vugt provided the supervision of the research process at the High Tech Campus on a daily basis and I wish to thank them for the freedom they offered me with regard to the direction and implementation of the research and the consistently positive feedback. Dr. ir. Robbert Jan Beun embodied in his function of first supervisor the supervision of Utrecht University on the research and writing process. I wish to thank him for the freedom that was offered to me in the research process, the meetings in which the research was discussed in a broader perspective and the very constructive feedback.

With regard to the multidisciplinary aspect of the research, it took several meetings with experts in the respective fields in order to implement their hard needed knowledge in this research. I therefore wish to acknowledge the often invaluable knowledge and efforts these experts have offered.

First of all, I would like to thank Dr. Roy J.E.M. Raymann from Philips Research for providing me with the necessary prerequisites for conducting this research and for the large amount of feedback that I received in the research and writing process. Furthermore, I would like to thank Igor Berezhnoy, Rene Derkx, Aad Sempel and Ronald Aarts from Philips Research for their time given for this research introducing me in the fields of respiration and digital signal processing.

Table of Contents

Declaration.....	1
Preface.....	3
Glossary.....	6
Abstract.....	8
1 Introduction.....	10
2 Exploring the basis.....	13
sleep.....	13
sleep staging.....	13
existing models.....	17
3 Constructing the framework.....	19
4 The framework in detail: the data layer.....	24
dataset parameters.....	24
dataset pre-processing.....	24
5 The framework in detail: the module layer.....	28
module one, sleep/wake classification.....	28
module two, REM sleep classification.....	32
module three, light sleep/deep sleep classification.....	36
6 The framework in detail: the fusion layer.....	40
definitive classification selection.....	40
definitive classification confidence level.....	40
7 Results.....	43
module one, sleep/wake classification.....	43
module two, REM sleep classification.....	45
module three, light sleep/deep sleep classification.....	47
fusion process and framework as a whole.....	49
8 Discussion.....	53
data layer.....	53
module one, sleep/wake classification.....	54
module two, REM sleep classification.....	56
module three, light sleep/deep sleep classification.....	58
fusion process and framework as a whole.....	61
9 Conclusions.....	63
further research.....	64
Bibliography.....	65
Appendix A algorithm specification.....	67
module one.....	67
module two.....	67
module three.....	68
fusion module.....	69
peak-to-peak algorithm.....	71
body movement-artefact correction algorithm.....	72
Appendix B individual participant classifications.....	74
module one.....	74
module two.....	86
module three.....	98
3-step hypnogram.....	110

Index of Tables

1. Overview statistical metrics used in this work.....	9
2. Overview of physiological changes between the different sleep stages, with ↑ indicating a increase of the specific biosignal, ↓ indicating a decreasing biosignal and ↓↓ indicating an even more pronounced decrease. Light and deep sleep are compared to the wake state and REM is compared to NREM (light and deep sleep) as this is the typical sequence during a NREM-REM cycle.	17
3. Dataset containing all participants partitioned in three groups depending on difference in age and applied measurement technique. Participants forming the training set are given in the blue cells and participants from whom the thorax based respiration rate is used are printed bold.....	29
4. Optimal group parameter set module one.....	33
5. Optimal group parameter set module two.....	37
6. Search regions and step sizes used during two-step parametrizing of the parameter set of module three. Individual results were used to determine the search region for the group parameterization. Note that the step size is equal for all groups and individual parameterization for each step.....	41
7. Optimal group parameter set module three.....	41
8. Fusion table for module one and two.....	43
9. Module one results.....	45
10. Module two results.....	47
11. Module three results.....	49
12. Framework results 3-step hypnogram.....	51
13. Confusion matrices for 3-step hypnogram (in %) per set, group and overall (TR and TST stand for training and test set respectively).....	52
14. Agreement for 4-step hypnogram.....	54

Diagram Index

1. Example hypnogram.	21
2. Overview of data and module layer structure.....	26
3. Data and module layer, BM and RR stand for body movement and respiration rate respectively..	27
4. Data, module and fusion layer, BM and RR stand for body movement and respiration rate respectively.....	28
5. Complete framework, BM and RR stand for body movement and respiration rate, respectively..	29
6. Excerpt of raw abdominal respiration effort biosignal (taken from participant 102) in grey together with results of consecutive pre-processing steps. Black denotes the filtered raw signal (third order high-cut Butterworth filter, cut-off 0.42Hz), green lines mark candidate peaks for start expiration and blue marks peaks ultimately selected as start of expiration.....	33
7. Raw and body movement corrected abdominal based respiration rate biosignal in red and black respectively. Body movement biosignal used for body movement correction is depicted in green. Whole night data, taken from participant 104.....	34
8. ROC per set for module one.....	57
9. ROC per set for module two.....	59
10. ROC per set for module three.....	61

Glossary

Agreement

Real value ranging between 0 and 1 indicating the agreement with the golden standard. 0 indicates no agreement and 1 indicates complete agreement. Agreement is calculated by dividing the amount of true positives and negatives by the total classifications made. For a schematic overview see table 1.

Biosignal

All signals that can be measured from biological beings such as humans. Examples are body movement, respiration effort and heart rate.

Electrocardiogram (ECG)

Registration of the electrical activity of the heart over time, captured and externally recorded by skin electrodes. Is used as a diagnostic and research tool for examining the activity of the cardiac muscle.

Electroencephalogram (EEG)

Registration of the electrical activity of the brain produced by the firing of neurons over time, captured and externally recorded by electrodes placed on the scalp. Is used as a diagnostic and research tool for examining the amount and location of brain activity.

Electromyogram (EMG)

Registration of the electrical activity produced by skeletal muscles over time, captured and externally recorded by skin electrodes. Is used as a diagnostic and research tool for examining the activity of muscles over time.

Electrooculogram (EOG)

Registration of the resting potential of the retina over time, captured and recorded by skin recordings placed above or besides the eyes. Is used as a diagnostic and research tool for examining the movement of the eye.

Epoch

Unit of time used in hypnograms, representing 30 seconds in most hypnograms.

Hypnogram

Graph indicating the sequence of sleep stages during sleep. For every epoch during sleep a (numeric) sleep stage classification is given. The numeric classification uses the following coding of the sleep stages: 6 indicates the wake state, 5 stands for the REM sleep stage and 1 to 4 represent the NREM stages 1 to 4 respectively. For an example of a hypnogram see illustration 1, page 15.

MT epoch

Epoch in a hypnogram classified as MT, indicating that there existed too much body movement to successfully classify the epoch.

Negative predictive value

Real value ranging between 0 and 1 indicating the fraction of negative classifications made by the classifier that are correct. Here 0 indicates none of the negative classifications are actual negatives, 1 indicating all negatives classifications are actual negatives. For a schematic overview see table 1.

Polysomnogram (PSG)

Simultaneously recording of an electroencephalogram (EEG), electrooculogram (EOG) and electromyogram (EMG), if needed supplemented with additional recordings, for example an ECG.

Positive predictive value

Real value ranging between 0 and 1 indicating the fraction of positive classifications made by the classifier that are correct. Here 0 indicates none of the positive classifications are actual positives, 1 indicating all positive classifications are actual positives. For a schematic overview see table 1.

Sensitivity

Real value ranging between 0 and 1 indicating the fraction of actual positives that are correctly identified as such by the classifier. Here 0 indicates none of the actual positives are classified as positive, 1 indicates all actual positive classifications are classified as such. For a schematic overview see table 1.

Specificity

Real value ranging between 0 and 1 indicating the fraction of actual negatives that are correctly identified as such by the classifier. Here 0 indicates none of the actual negatives are classified as negative, 1 indicates all actual negative classifications are classified as such. For a schematic overview see table 1.

Respiratory effort

Biosignal indicating the amplitude of the respiration, either from abdominal or thorax based respiration. From this biosignal the respiration rate can be derived.

ROC (Receiver Operating Characteristic)

Graphical plot of the sensitivity versus 1 – specificity. Different classifiers can be plotted in one ROC to give in a graphical manner an overview of the different classifiers with respect to the respective performance in sensitivity and specificity.

		Golden standard		
		positive	negative	
Classifier	positive	true positive (tp)	false positive (fp)	→ positive predictive value $tp / (tp + fp)$
	negative	false negative (fn)	true negative (tn)	→ negative predictive value $tn / (tn + fn)$
		↓ sensitivity $tp / (tp + fn)$	↓ specificity $tn / (tn + fp)$	agreement $(tp + tn) / (tp + tn + fp + fn)$

Table 1: Overview statistical metrics used in this work.

Abstract

For diagnosis and research purposes, a person's sleep should be objectively quantified. Traditionally, sleep staging is performed by human experts via visual inspection of polysomnographic (PSG) data measured during the sleep period. There are several drawbacks to this golden standard: the data are acquired in a rather obtrusive manner and not in the natural sleep environment of the sleeper, hereby affecting the sleep. Furthermore, this manual scoring is time consuming and prone to human errors. Automatic sleep staging models exist but typically perform binary classifications, e.g. sleep/wake, REM/NREM or light/deep sleep classification and on the whole use obtrusively obtained measurements. In this research we aim to develop a framework for classifying sleep stages from unobtrusive measurements, being measurements during which the participant is not disturbed and no sensors are attached. Our current implementation is presented that constructs 3- and 4-step hypnograms (containing wake, REM and NREM in the 3-step hypnogram and wake, REM, light sleep and deep sleep in the 4-step hypnogram) utilizing body movement and respiration rate biosignals obtained during sleep. This current implementation of the framework consists of three modules that each use a unique way of interpreting one or more of these biosignals and a fusion process. The modules partly employ adapted versions of existing models for sleep stage classification and partly newly developed models, grounded in sleep physiology. The modules classify epochs of a person's sleep based on one or more biosignals and relevant subject parameters such as age and gender. The framework ensures that each epoch is classified by at least two different modules. A specialized fusion process analyses the output from the different modules, merges the output into a final hypnogram and calculates a confidence level for each epoch classification.

To compare the sleep stages as derived from the current implementation of the framework with the golden standard, a secondary analysis on an existing data set of 23 participants is performed. The results are promising and show potential: the three modules perform binary classifications with 90%, 85% and 72% agreement to the golden standard, the fusion process corrects 61% of the classifications on which the modules disagree and fuses the binary classifications in definitive 3- and 4-step hypnograms. During this process the overall agreement drops significantly because all erroneous classifications of the modules are added up, but due to the correction of erroneous classifications by the fusion process these hypnograms maintain 69% and 51% agreement with the golden standard, respectively.

This degree is expected to increase significantly when the existing modules are improved, additional modules are added to allow the fusion process to correct more erroneous classifications and the parameterization of the modules is done simultaneously.

Also, the framework contains various technical advancements not found in other contemporary sleep staging methods. Exemplary here is the great flexibility of the framework, enabling the researcher or clinician to easily add, remove or alter modules. Also complete parameterization, ensuring that at all times the clinically relevant parameters of the participant are taken into account, can be mentioned here. The current implementation of the framework is expected to be a basis for unobtrusive sleep staging that can be performed in the natural sleep environment of the sleeper and without influencing or disturbing sleep, hereby potentially lowering the threshold for (preliminary) sleep diagnostics, research and therapy.

What probing deep has ever solved the mystery of sleep?

– Thomas Bailey Aldrich

1 Introduction

Humans spend around a third of their lives being asleep. During sleep, vital processes take place that are important for many aspects of our lives such as (child) development, learning and memory, metabolism, mood and even the survival of the person.

Given this importance of sleep and the amount of time spent asleep, it is important to gain insight into the process of sleep. This is usually done by determining the different sleep stages a person goes through during a night's sleep. Traditionally, this sleep staging is performed by human experts via visual inspection of polysomnographic (PSG) data, which contains simultaneous recordings of the activity of the brain, eye movement and muscular activity, acquired during the night. But there are several drawbacks to this golden standard:

- the data are acquired in an obtrusive manner and not in the natural sleep environment, thereby affecting the sleep of the person.
- the manual scoring leaves room for human errors and is time consuming and costly because of the needed time and equipment.

Currently, in the sleep research field no alternatives for sleep staging exist that tackle all these drawbacks while providing insight in all important sleep stages. The existing models that partially tackle the drawbacks can be divided into three classes:

- models performing *automated* sleep staging utilising *obtrusively* measured data, tackling the issues attached to manual scoring by employing automated algorithms, but by utilising obtrusively measured data the sleep of the person is still affected.
- models performing *automated partial* sleep staging utilising *unobtrusively* measurable data, by using data that is acquired without disturbing the sleeper or attaching sensors and by processing this data automatically, these models tackle the issues attached to both manual scoring and the obtrusive acquisition of data. Unfortunately, these models are only capable of classifying certain sleep stages, for example sleep/wake classification or REM classification, so not a complete insight in the process of sleep is obtained.
- models performing *automated full* sleep staging utilising *unobtrusively* measurable data, tackling all the drawbacks while providing a complete insight in the process of sleep, but the models that are currently available in the field either provide no algorithm definition for performing sleep staging or the algorithm presented is unreproducible.

A more in-depth treatment of the models from the different classes is beyond the scope of this chapter and will be postponed to chapter 2, where the basis for the framework, together with current models, is explored.

Given the current state of affairs in which no model exists that tackles all the important drawbacks of the golden standard, this work aims to investigate the possibility for developing a method that does tackle all these drawbacks while providing insight in all the important sleep stages during the night.

Hence the following research question is formulated:

Is it possible, and if so, to what degree, to perform automated sleep staging using solely unobtrusive measurements ?

In order to answer this research question we aim to develop a framework that is able to tackle the aforementioned drawbacks of the golden standard and that can be used to investigate the possibilities of determining sleep stages using solely unobtrusive measurements.

To utilise the existing body of work on models performing *automated partial* sleep staging utilising *unobtrusively* measurable data it was chosen to design a framework that combines different classifiers of this type in order to construct a complete sleep stage model. This framework contains both existing partial sleep stage classifiers in adapted form as well as newly developed partial sleep stage classifiers, all able to operate on unobtrusively acquirable data. In this manner the framework can perform sleep staging automatically using solely data as input which can be acquired unobtrusively, hereby aiming to tackle the drawbacks of the golden standard while providing insight in all the important sleep stages during the night.

Given the current state of measurement techniques available, the unobtrusively acquirable data that can be used by the framework for sleep staging is limited. This is because not every biosignal that can be used for sleep staging can be measured unobtrusively. Currently, the only biosignals that are indicative of sleep stages and can be measured unobtrusively are body movement, respiration effort and heart rate. Therefore, in the framework only these biosignals will be used as input for determining sleep stages.

To further define the requirements set to the framework, from an artificial intelligence point of view, there are roughly two different approaches to the algorithms that will be contained in the framework. Either the algorithms will be a connectionist model or it will belong to the class of classical models. This partitioning pertains to the sense in which the algorithm regards and processes the input. In the connectionist model, input is processed using many interconnected computational units, representing the neural connections of the human brain. This often results in models processing input in a manner not intelligible to humans, but at the same time allowing for complex computations not always feasible in the classical approach. In a classical model the input is processed using transparent, rule-based decision algorithms, rendering a processing that is insightful to humans. Since the computations used in the framework are not beyond the scope of classical methods, we chose to use classical algorithms, utilising rules grounded in physiology. In this manner, the grounds on which the classifications are based as well as the decision making are transparent, giving more insight, which is helpful for further research and analysis purposes.

We end up with a set of requirements for the framework needed for answering the research question:

- use the unobtrusively acquirable biosignals body movement, respiration effort and heart rate as input, to not affect the sleep during the night.
- use automated algorithms that perform sleep staging without the need of human intervention, to eliminate possibilities for human errors and reduce costs and time needed.
- perform complete sleep staging, to give a complete insight in the process of sleep during the night, to be an alternative to the golden standard.
- using (adapted) existing and newly developed partial classifiers, to use the existing body of work.
- use classical algorithms, which are rule based with the rules grounded in physiology, to make the decision making transparent.

In order to answer the research question, this framework will be constructed and consecutively be tested to determine the agreement of the classifications made by the framework with the golden standard. For constructing this framework first we investigate what behaviour the required biosignals body movement, respiration effort and heart rate display during the different sleep stages. Based on this behaviour, existing as well as newly developed partial sleep classifiers will be combined in a manner that ensures all requirements set will be fulfilled. This framework will then be trained and tested using an existing dataset of 23 healthy participants with the sleep stages manually scored.

The construction and testing of the framework in order to answer the research question is described in detail in this thesis. Following this first introductory chapter is a chapter exploring the basic notion of sleep and sleep stages, with emphasis on the differences in the biosignals between the different sleep stages that are used by the framework to see in what way these biosignals can best be used by the framework. Also included in this chapter is the postponed in-depth review of existing models. In the consecutive third chapter the construction of the framework using these differences in biosignals and an overview of the resulting framework is given. Chapters 4, 5 and 6 are dedicated to describing the framework in more detail. In chapter 7 the test results of the framework are given as the classifications made by the framework on the dataset are compared with the golden standard. These results are discussed in chapter 8 after which the final conclusions are given in chapter 9. To conclude, chapter 10 contains recommendations for further research and recommendations for improvement of the framework.

2 Exploring the basis

This chapter explores the basis for the framework. First, a definition of sleep and sleep staging together with the underlying principles of these concepts are presented in this chapter. Secondly, the changes during the different sleep stages in the biosignals that will be used as input for the framework, being body movement, respiration effort and heart rate are explored. Furthermore, a more in depth overview of the different classes of existing sleep stage classifiers described in chapter 1 is given.

sleep

For the phenomenon of sleep many different definitions exist. So does the Oxford English Dictionary define sleep as *'the unconscious state or condition regularly and naturally assumed by man and animals, during which the activity of the nervous system is almost or entirely suspended, and recuperation of its powers takes place; [...]'*. Another standard work, the Macmillan Medical Dictionary states sleep to be *'a naturally recurring state of relatively suspended sensory and motor activity, characterized by total or partial unconsciousness and the inactivity of nearly all voluntary muscles.'*

Although the different definitions give a small taste of the many aspects pertaining to sleep, most of the definitions are tailored to a specific discipline, such as psychology, neurology or medicine. For our purpose a more general definition is needed because of the multidisciplinary nature of this work and will suffice without affecting the research as such by being too general. Therefore the general definition offered by the Oxford English dictionary is adhered to in this work as this work provides the sufficing general definition given above and the dictionary is considered the standard dictionary for the English language.

sleep staging

Within sleep, different stages can be discerned based on neural, tonic and ocular activity, the activity of the brain, muscles and eyes respectively. Utilizing the in the 1930's discovered EEG, a registration of the electrical activity of the heart, and based on earlier work by, amongst others, Aserinsky and Kleitman (Aserinsky, Kleitman, 1953) and Jouvet et al. (Michel Jouvet, 1959) in the 1950's, Rechtschaffen and Kales defined five different sleep stages. They introduced standardized criteria to record sleep and stage sleep in their 1968 landmark work *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects* (Rechtschaffen, Kales, 1968). These standardized criteria for recording and staging are the golden standard in the clinical and sleep research field. The five sleep stages defined by Rechtschaffen and Kales are staged based on a polysomnogram, a recording consisting of minimally a simultaneously recorded electroencephalogram (EEG), recording of the activity of the brain, an electrooculogram (EOG), recording of the eye movements and an electromyogram (EMG), recording of the muscular activity. Based on these recordings, sleep time is partitioned into epochs, 30-second time units, and each epoch is classified into rapid eye movement (REM) sleep or non-rapid eye movement (NREM) sleep. NREM sleep is further divided into four stages, NREM 1 to NREM 4, based on the pattern of the brainwaves. Together with epochs during which the participant is awake, this results in a total of five sleep stages and the wake state.

NREM stages 1 and 2 are commonly referred to as light sleep and stages 3 and 4 as deep sleep or slow-wave sleep. Our framework follows this tradition, thereby effectively differentiating four separate stages: REM sleep, light sleep (NREM stages 1 and 2 combined), deep sleep (NREM stages 3 and 4 combined) and wake. During a nights' sleep, a so-called NREM-REM cycle, in which REM sleeps follows NREM sleep, occurs approximately every 90 minutes, resulting in four to six cycles for an average night of sleep. This can be seen in the hypnogram in diagram 1. In a hypnogram the sequence of sleep stages during the night is depicted by giving the sleep stage classifications for each epoch during the night using a numeric coding. In this coding 6 indicates the wake state, 5 stands for the REM sleep stage and 1 to 4 represent the NREM stages 1 to 4 respectively. In this hypnogram three NREM-REM cycles can be discerned, where the sleep stages gradually incline from NREM 1 or NREM 2 to NREM 4 and eventually REM sleep. The different sleep stages are described briefly below together with the relevant physiological processes occurring during these stages.

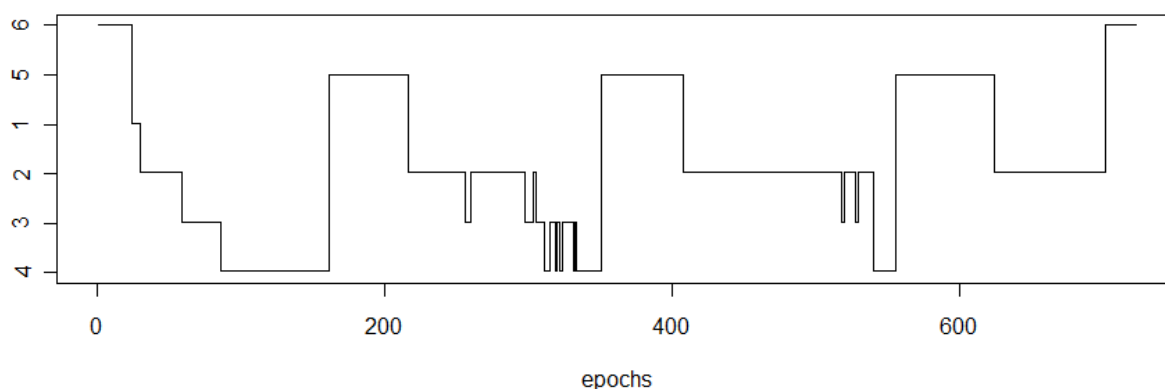


Diagram 1: Example hypnogram.

Light sleep (NREM stages 1 and 2)

Light sleep accounts for 50-65% of the total sleep time and is a transitional stage from full wakefulness to sleep arising when mechanisms maintaining wakefulness abate. It can furthermore appear during transitions from sleep to wake or after brief body movements. During light sleep there is an increase in parasympathetic activity, relating to the part of the nervous system that takes over when the body relaxes more, and a decrease in sympathetic activity, the part of the nervous system that takes over when we become alert and excited. Relatively low voltage, mixed frequency brain waves with potential slow eye movements and a decrease in muscular activity characterizes this stage.

Reflecting these changes is a shallower, more rapid and more unstable respiration pattern, quantified in a heightened respiration rate which shows a relatively high amount of variation (Douglas et al., 1982; Penzel et al., 2007 and Appelberg et al., 2007). Heart rate drops as the transition from full wakefulness to sleep develops, although during this stage the mean heart rate is still higher compared to deep sleep. Similar, heart rate variability, the variability of the heart rate over time, decreases compared to wakefulness while having a higher mean heart rate variability when compared to deep sleep (Bušek et al., 2005; Zhuang et al., 2005; Voronin et al., 2006 and Vigo et al., 2009).

Body movement decreases drastically when compared to wakefulness although occasional low amplitude body movements can occur (Tryon, 2004; Pollak, 2001; Paquet et al., 2007 and Ancoli-Israel et al., 2003).

Deep sleep (NREM stages 3 and 4)

Around 15-20% of the sleep time is spent in deep sleep, with a higher prevalence during the first NREM-REM cycles. Deep sleep constitutes so-called slow-wave sleep and is characterized by moderate amounts of high-amplitude, slow-wave brain activity for stage three and large amounts of high-amplitude, slow-wave brain activity in stage four. No eye movement is registered by the EOG and muscular activity is decreased compared to light sleep or wakefulness.

At this time, a steady, relatively fast respiration pattern can be discerned, reflected in a higher respiration rate displaying only minor variance (Douglas et al., 1982; Penzel et al., 2007; Appelberg et al., 2007 and Schäfer, Schläfke, 2009). Reflecting the autonomic stability characteristic for NREM sleep and especially for deep sleep, the heart rate drops as does the heart rate variability, resulting in a relatively slow and steady heart rate (Bušek et al., 2005; Zhuang et al., 2005; Voronin et al., 2006 and Vigo et al., 2009). Body movement is almost completely absent during deep sleep (Tryon, 2004; Pollak, 2001; Paquet et al., 2007 and Ancoli-Israel et al., 2003).

REM sleep

REM sleep accounts for 20-25% of sleep time with a higher prevalence during the second half of the night, the latter NREM-REM cycles. Rem sleep constitutes an active form of sleep during which most dreams are assumed to occur. REM sleep can be divided into tonic and phasic REM sleep, a distinction not reflected in the sleep stages. During tonic REM sleep, paralysis of most muscular groups occurs, whereas during phasic REM sleep the characteristic rapid eye movement and incidental bursts of EMG activity occur. During REM sleep, there is a further increase in parasympathetic activity coupled with a further decrease in sympathetic activity compared to NREM sleep.

This results in fast, shallow and irregular respiration behaviour, marked by higher respiration rates with high deviation compared to NREM sleep and wakefulness (Douglas et al., 1982; Penzel et al., 2007; Appelberg et al., 2007 and Schäfer, Schläfke, 2009). Heart rate becomes variable and during phasic REM sleep, significant increases of heart rate can occur as a result of bursts of parasympathetic activity. This is reflected in a higher heart rate when compared to NREM sleep and a significant higher heart rate variability (Bušek et al., 2005; Zhuang et al., 2005; Voronin et al., 2006 and Vigo et al., 2009). Body movement is absent during tonic REM sleep due to aforementioned sleep paralysis of most muscular groups whereas during phasic REM sleep little twitches can be discerned (Tryon, 2004; Pollak, 2001; Paquet et al., 2007 and Ancoli-Israel et al., 2003).

Wake

Wakefulness can occur several times during sleep and is characterized by brainwave alpha activity, brainwaves within the 8-12 Hz frequency range, or low voltage, mixed frequency activity in the EEG, usually combined with a high tonic EMG and possible rapid eye movements and blinks visible in the EOG. This is reflected by a steady, relatively deep respiration characteristic with a low respiration rate compared to NREM and REM sleep stages (Douglas et al., 1982; Penzel et al., 2007; Appelberg et al., 2007 and Schäfer, Schläfke, 2009). Heart rate is higher than any of the sleep stages during the night, with a moderately high heart rate variability (Bušek et al., 2005; Zhuang et al., 2005; Voronin et al., 2006 and Vigo et al., 2009) compared to sleep. Body movement occurs with the highest frequency and the highest power during wakefulness compared to sleep stages (Tryon, 2004; Pollak, 2001; Paquet et al., 2007 and Ancoli-Israel et al., 2003)

In table 2 a summary of the physiological changes in body movement, respiration effort and heart rate during the different sleep stages is given. It can be seen here that each sleep stage has a unique pattern of physiological changes.

sleep stage	body	respiration		heart	
	movement	rate	variability	rate	variability
awake	active				
light sleep	↓	↓↓	↓	↓	↓
deep sleep	↓↓	↓	↓↓	↓↓	↓↓
REM sleep	absent / twitches	↑	↑	↑	↑

Table 2: Overview of physiological changes between the different sleep stages, with ↑ indicating a increase of the specific biosignal, ↓ indicating a decreasing biosignal and ↓↓ indicating an even more pronounced decrease. Light and deep sleep are compared to the wake state and REM is compared to NREM (light and deep sleep) as this is the typical sequence during a NREM-REM cycle.

Important to note is that the correlation between biosignals and sleep stages is not always as straightforward as depicted here. Firstly, although there is a large body of literature consistently describing the physiological changes during the different sleep stages, these changes are usually found using obtrusive measurement techniques such as PSG. With the use of unobtrusive measurement techniques the signals recorded typically contain more noise due to the unobtrusive nature of the technique employed, causing the correlation to be more elusive. Examples of unobtrusive measurement techniques contain, amongst others, the use of cameras and microphones to determine the respiration rate and the use of bed-actigraphy (BACT), where load-sensing cells support the bed and determine the amount of body movement of the participant during the night. A complete coverage of unobtrusive measurement techniques is beyond the scope and aim of this work, for our purposes it suffices to be aware of the existence of these techniques and the fact that the signals recorded typically contain more noise compared to the signals recorded by the golden standard.

A second cause impairing the correlation between biosignals and sleep stages when using unobtrusive measurement techniques concerns the biosignals recorded. The golden standard employs primarily the activity of the brain for classifying sleep stages whereas unobtrusive measurements record other biosignals since it is currently not possible to measure brain activity unobtrusively. This entails that we have to determine the state of the brain from these signals, introducing noise since there is no one-to-one mapping between brain activity and the physiological changes in the recorded biosignals.

Thirdly, the correlation between sleep stages and biosignals can be modulated by demographics, such as, for example, the age and gender of the person. For example, it can be the case that between an elderly woman and a young male there are differences to be found in the changes in respiration rate for certain sleep stages.

These three reasons are the main cause of a lower information value of the unobtrusively acquired biosignals compared to the golden standard, which is the major issue for existing (partial) sleep-stage classifiers using these kinds of signals. The framework aims to overcome this issue by combining several classifiers using different biosignals in order to obtain as much information as possible and by taking important demographics such as age and gender into account in determining the sleep stages.

existing models

Here a more in depth overview of existing classifiers of the different classes described in the previous chapter is given. The classes described in the previous chapter are:

- models performing *automated* sleep staging utilising *obtrusively* measured data, tackling the issues attached to manual scoring by employing automated algorithms, but by utilising obtrusively measured data the sleep of the person is still affected.
- models performing *automated partial* sleep staging utilising *unobtrusively* measurable data, tackling the issues attached to both manual scoring and the obtrusive acquiring of data, but these models are only capable of classifying certain sleep stages, for example sleep/wake classification or REM classification, so not a complete insight in the process of sleep is gained.
- models performing *automated full* sleep staging utilising *unobtrusively* measurable data, tackling all the drawbacks while providing a complete insight in the process of sleep, but the models that are currently available in the field either provide no algorithm definition for performing the sleep staging or the algorithm presented is unreproducible.

Within the first class of models, the models performing *automated* sleep staging utilising *obtrusively* measured data, an interesting example is recent work by Virkkala, (Virkkala, 2009) utilising the eye movements during the night for determining the moments that the person is in deep sleep, the stage during sleep when the person sleeps the deepest. In their approach, machine learning algorithms are trained on a large dataset to recognize the pattern made by the eyes during deep sleep. In the validation, the algorithm achieved a 93% agreement with the golden standard.

This is an impressive agreement using an automated algorithm, but unfortunately the eye movement during the night can currently only be measured obtrusively, using electro-oculography, (EOG), which requires electrodes to be attached to the sides of the eyes. Other examples in this class of models include work by Shinar et al., (Shinar et al., 2001), Sako et al., (Sako et al., 2001), and Piryatinskaa et al. (Piryatinskaa et al., 2009).

Of the second class of models, performing *automated partial* sleep staging utilising *unobtrusively* measurable data, not many models exist. Out of these models the most classic example is the algorithm by presented by Cole et al. (Cole et al., 1992) in their 1992 work *Automatic Sleep/Wake identification from wrist activity* in which an algorithm is presented that is able to differentiate between wake and sleep using the body movement biosignal, being any signal measured from biological beings such as humans, which can be measured unobtrusively. The algorithm employs a thresholding technique that is able to classify wake and sleep moments during the night with an agreement of 90% with the golden standard. Another, more recent work on automated classifiers performing *automated partial* sleep staging on *unobtrusively* acquirable biosignals constitutes the work by Chung et al. (Chung et al., 2009) presented in their article *REM sleep estimation only using respiratory dynamics*, published at the end of 2009. Here an elaboration of their earlier algorithm described in their 2007 article (Chung et al., 2007) is given. This algorithm employs adaptive thresholds on respiration rates to determine the location of REM sleep during the night, which it is able to perform with a agreement of 91% with the golden standard. More examples of models in this class include work by Karlen et al. (Karlen et al., 2008) and work by Redmond and Heneghan (Redmond, Heneghan, 2006).

Finally, the models performing *automated full sleep staging* utilising *unobtrusively* measurable data, the third class of models, currently either provide no algorithm definition for performing sleep staging or the algorithm presented is not reproducible. The most pronounced example hereof is the model presented by Watanabe et al., in *Noncontact method for sleep stage estimation*, published in 2004. In this publication an algorithm is presented for complete sleep staging based on heart rate. Unfortunately the algorithm was incomprehensible to us and it proved irreproducible due to unclear mathematical notations.

3 Constructing the framework

Given the differences between the sleep stages in the unobtrusively measurable biosignals body movement, respiration effort and heart rate as explored in the previous chapter and the requirements set to the framework in the first chapter, we can start the construction of the framework needed for answering the research question. This construction is given in this chapter after which in the consecutive chapters the constructed framework is treated in detail and is formally specified.

We start by combining two of the requirements, the use of the biosignals body movement, respiration effort and heart rate as input and the use of (adapted) existing and newly developed partial classifiers. Therefore we create two layers, one containing the biosignals discretized per epoch, which we call the data layer, and one containing the partial classifiers, the module layer. We furthermore let the biosignals in the data layer be used as input to the modules in the module layer, which in principle can contain any number of modules. The two layers are depicted in diagram 2.

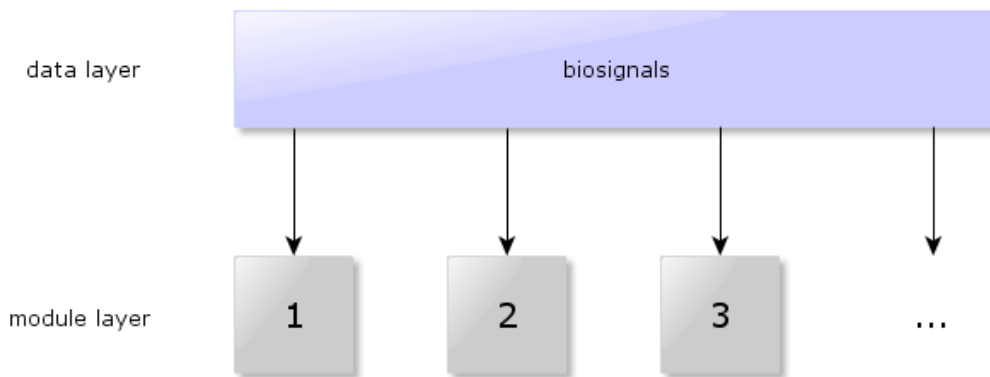


Diagram 2: Overview of data and module layer structure.

In the current implementation of the framework, the module layer contains three modules, all utilising the physiological changes described and summarized in table 2 in the previous chapter. The first two are adapted existing modules, the first an adapted version of the Cole et al. (Cole et al.,1992) sleep/wake classifier taking the body movement biosignal as input and the second module an adapted version of the Chung et al. (Chung et al., 2007 and Chung et al., 2009) REM classifier using respiration rate as input. These classifiers were chosen because of their compliance with two of the other requirements, firstly the use of algorithms performing sleep staging without the need of human intervention and secondly the requirement that the algorithms are transparent, classical rule based algorithms with the rules grounded within physiology. The third, newly developed module classifies light and deep sleep using both respiration rate and body movement, naturally also using the changes in biosignals given in table 2. Because the modules only use the biosignals body movement and respiration rate as input for their classification algorithms, without the need for heart rate, only these biosignals are contained in the data layer. The output of every module is a classification for every epoch of the night. The first module classifies all the epochs of the night as either sleep or wake, the second module aims to detect all REM epochs and the third module aims to partition NREM into light and deep sleep. This module takes therefore, by exception, only biosignals of NREM epochs as input instead of all the epochs of the night.

The current implementation of the data and module layer is given in diagram 3.

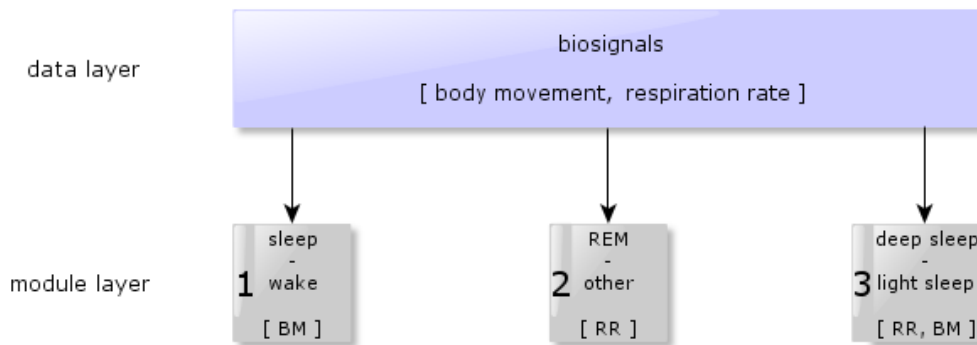


Diagram 3: Data and module layer, BM and RR stand for body movement and respiration rate respectively.

To fulfil the remaining requirement set for this framework, performing complete sleep staging, the output of the different modules needs to be fused into one final complete hypnogram. To this end a new layer is constructed where all the output of the different modules is gathered, called the the fusion layer. This output of the different modules is examined per epoch in the fusion layer. When the classifications of the different modules are in agreement, the most exact agreement is chosen. For example. When module one classifies a certain epoch as sleep and the second module classifies this epoch of REM, the modules are in agreement and the final classification for this epoch will be REM as this is the most exact classification. When the classifications are in disagreement, the most reliable classification is chosen. The reliability of the different classifications for every module is determined during the parameterization of the framework on the training set, which is specified in chapter 5 where the module layer is discussed in detail. Secondly, besides determining the final classification for each epoch, the confidence level for each of these final classifications is determined. This level is given by a value between 0 and 1, indicating the confidence of the framework that this classification is correct (0 means completely no confidence and 1 means complete confidence in the classification). The confidence level for each epoch classification in the definitive hypnogram depends on the reliability of all the classifications out of which the definitive classification is selected. This process is specified in detail in chapter 6 where the fusion layer is described in detail. Finally, it is possible to communicate from the fusion layer to the module layer, which can be needed in exceptional cases. Currently this is only needed for module three, where there is communication back from the fusion layer to the module concerning which epochs are classified as NREM by the first two modules, which is necessary since the third module currently only processes NREM epochs. The data layer and module layer, together with the fusion layer are given in diagram 4.

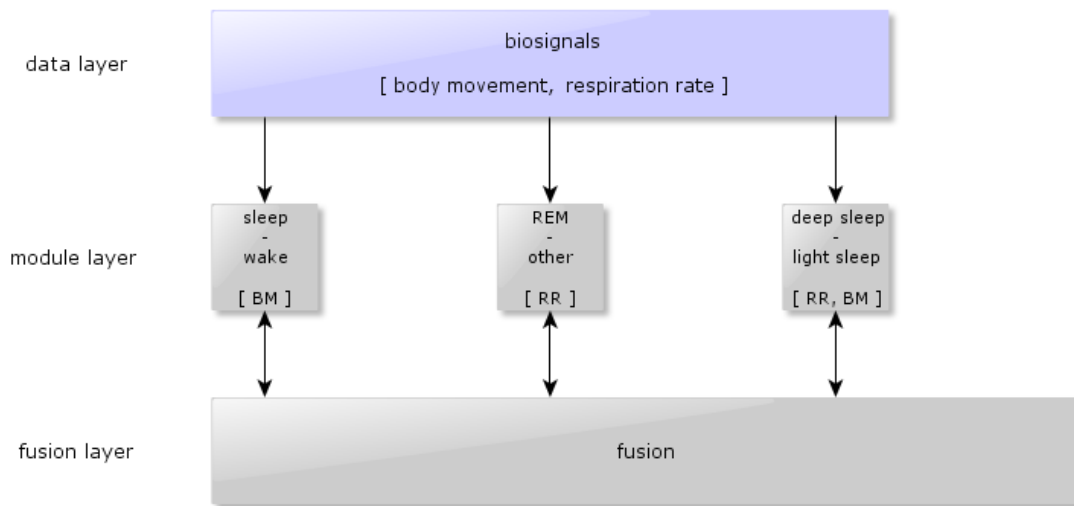


Diagram 4: Data, module and fusion layer, BM and RR stand for body movement and respiration rate respectively.

Finally, one extra component is added which allows the framework to take into account relevant parameters of the person whose night's sleep is being classified. As described in the previous chapter, this is important to account for the differences in physiological changes during the sleep stages for different demographic groups.

The component added is called the parameter component, also located in the data layer, containing meta data about the person sleeping such as for example age and gender. In the current dataset, the participants are grouped into three groups based on their age and gender, so the parameter component uses the group number of a participant to determine the meta data of this participant. The framework uses this information to select the optimal parameter set for each module and to determine the most reliable classification as determined during the parameterization on the training set when classifications are in disagreement.

To conclude, the output of the framework is outputted as a 4-stage hypnogram containing the stages wake, REM, light and deep sleep per 30-second epoch with a confidence interval for each classification. The complete framework is given in diagram 5.

In order to allow the framework to select the optimal parameter set for each module, given the relevant meta data contained in the parameter component, the dataset of 23 participants was divided into a training set of 15 participants and a test set of 8 participants. The different modules were parameterized on this training set to determine the parameter set at which the module performs optimally within the framework given the relevant meta data of the participant. For example, it can be the case that the optimal parameter set for an elderly woman will be different than for a young male for a certain module. Furthermore, the reliability of the classifications for every module working with the optimal parameter set was determined, so it is known how reliable a classification of a given module will be for a participant knowing its relevant parameters.

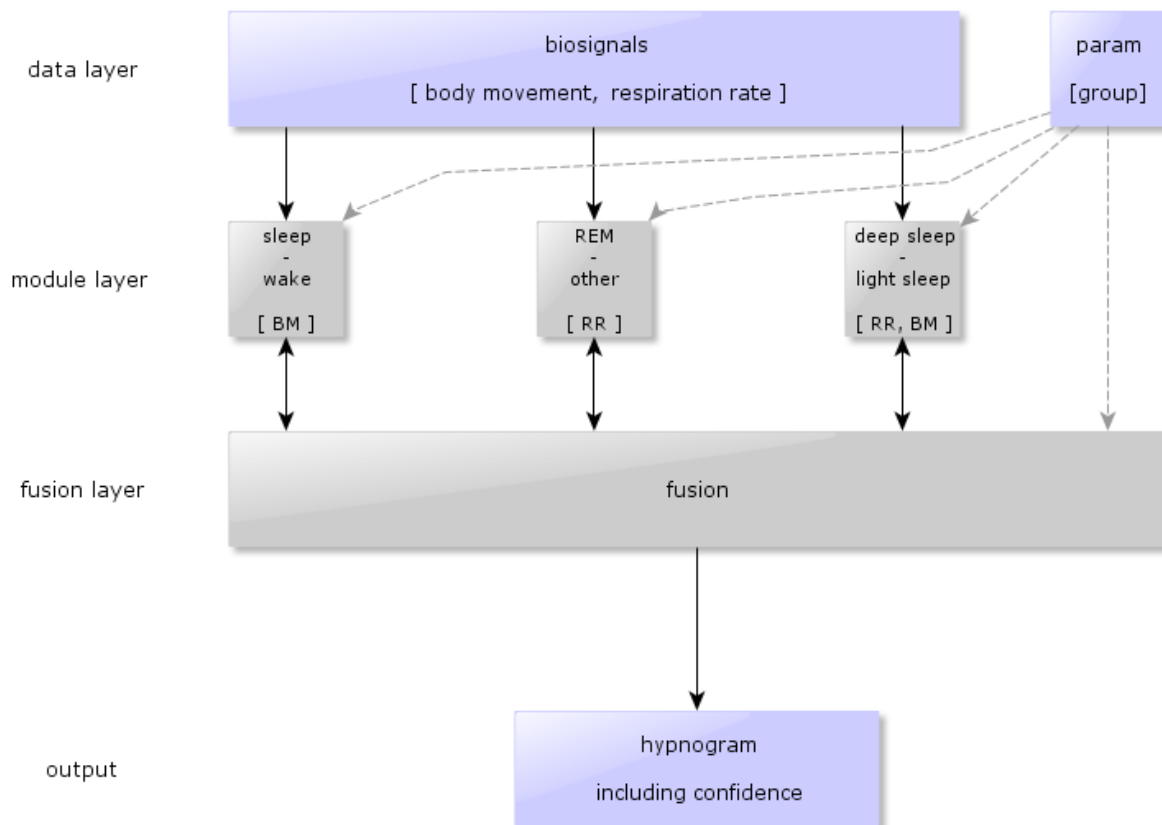


Diagram 5: Complete framework, BM and RR stand for body movement and respiration rate, respectively.

By constructing the framework in this manner, the framework fulfils the requirements set in the first chapter in order to answer the research question at hand and, even more, has several additional technical advantages not available in contemporary sleep staging methods. These advantages are briefly described here.

- Using different modules renders the framework very flexible.
Every module can be turned on or off at any given time, new modules can be attached or current ones removed without impairing the framework. This can be beneficial when certain classifications are not of interest or certain biosignals are compromised. In these cases the modules outputting these classifications or using these signals can be turned off. In this implementation it is, for example, possible to switch off module three rendering a definitive 3-staged hypnogram containing the stages wake, REM and NREM.
- The modularity in the framework allows for increasing confidence.
The more modules that are present in the module layer outputting their classifications to the fusion layer, the more information the fusion process has to determine the definitive hypnogram. This allows the agreement with the golden standard and the confidence levels of the definitive hypnogram to be increased with every module.

- The parameterization of the framework ensures that no individual fitting per participant is needed.

Due to the complete parameterization of the framework no per-participant fitting is required. Furthermore this parameterization enables the framework to constantly take the relevant participant parameters such as gender and age into account, during the classification in the module layer as well as during the fusion process.

- The framework is compliant with many current existing unobtrusive measurement set-ups.

The biosignals currently used by the framework, body movement and respiration rate, are derived biosignals, meaning that these are not the raw biosignals measured on a person, but are calculated on the basis of the raw biosignals as will be described in chapter 4. Through the use of these biosignals, the current implementation of the framework is compliant with many existing unobtrusive measurement set-ups. This is because although the raw biosignals potentially differ between the set-ups, these derived biosignals can almost always be derived from the different raw biosignals. This reduces the costs for equipment and makes it possible to perform sleep staging in the natural sleep environment of the person.

- The fusion process enables the framework to utilise the strengths of different biosignals and classification methods.

The rationale behind this approach is that specific biosignals and classification methods are better suited for certain classifications or participant groups than other biosignals or methods. For example, the biosignal body movement performs well on the detection of sleep epochs and less well on the detection of wake epochs. This means the reliability for wake classifications of modules employing body movement as input will drop accordingly and therefore will be less likely to be chosen if there exists disagreement. This enables the framework to allow some amount of over- or underestimation of a module in order to classify those epochs that normally wouldn't be picked up since these errors may be eliminated by other modules. In this way, more epochs can be found that normally would remain unclassified using the same model.

The same holds for certain classification methods that perform clearly better for specific participant groups. Classifications from modules employing these methods will be less reliable for these kinds of participants and therefore will be less likely to be selected if there exists disagreement. So in this manner the strengths of different biosignals and classification methods are exploited because it enables the framework to correct erroneous classifications made by modules and to correct for pitfalls attached to certain methods or biosignals by selecting at all times the most reliable classification out of an as wide as possible spectrum of available biosignals and approaches to these signals. This makes it possible that for every epoch always the best method and biosignal available in the framework is used.

The current framework is implemented in R (version 2.7.2 on Windows Vista) because this statistical programming language contains very little overhead and allows for quick statistical computations needed by the framework. In the next chapters the current implementation is specified in detail.

4 The framework in detail: the data layer

In this chapter the content of the first layer of the current implementation of the framework, the data layer, is described. The data contained in the data layer originates from an existing dataset and contains the abdominal and thorax respiration effort and body movement biosignal of 23 participants.

First, the relevant parameters of the dataset are given. Next, the pre-processing performed on this data to render it suitable for input to the module layer is described.

dataset parameters

The 23 participants in the dataset are partitioned into three groups based on the difference in age and the applied measurement technique, being either clinical or ambulatory. The group a participant is placed in is stored as meta data in the parameter component to allow the framework to take these parameters into account as described in the previous chapter. The first group contains 6 healthy young participants measured in a clinical set-up, the second group comprises 6 healthy elderly participants measured in a clinical set-up and the third group is composed of 11 elderly participants measured in an ambulatory setting. All participants are free of sleep disorders and are not under medication known to influence sleep or any of the relevant physiological processes. All sleep stages are manually scored, based on PSG data. From the participants the abdominal and thorax respiration effort, recorded using a band around the abdomen or thorax measuring the rising and falling caused by the respiration, and the body movement biosignal, measured using a motion sensitive sensor attached to the wrist, called an actigraph, are recorded during the first night and present in the data layer.

Furthermore is the dataset partitioned into a training set on which the different modules will be parameterized and a test set on which the performance of the framework is determined. The ratio between the training and test set with regard to size, for the whole dataset as well as for the different subgroups, is kept as close as possible to 2:1 (training set size : test set size) to allow for sufficient physiological data to determine the physiology-based parameters while maintaining enough data for testing the framework based on these parameters. The gender distribution in the test set is kept as close as possible to being sex-matched in order to notice and account for possible differences in physiological parameters between gender. The distribution of participants between the training and test set was performed randomly. This dataset contained in the data layer is schematically given in table 3 with the participants forming the training set printed in bold.

dataset pre-processing

In order to render the data usable as input for the module layer, the biosignals included therein, being respiration effort and body movement, need to be pre-processed. This pre-processing is described below, with all the algorithms referred to included in appendix A.

abdominal and thorax respiration effort signal pre-processing

Using Somnologica software (Flaga hf, Reykjavik, Iceland) the raw abdominal and thorax respiration effort signal, representing the rising and falling of the thorax or abdomen caused by respiration, was exported to plain text with the original sampling rate of 10 Hz.

An example of such a signal, an excerpt of the abdominal respiration effort biosignal of participant 102, is given in grey in diagram 6.

Using Matlab, a third order high-cut Butterworth filter with a cut-off frequency of 0.42 Hz is applied to the raw respiration data to attenuate high frequency noise. The cut-off frequency corresponds to a respiration rate of over 25 times per minute, which is far too high to be found during sleep in healthy participants, hereby effectively attenuating high frequency noise since these frequencies can not be ascribed to respiration. An example of applying this filter on raw respiration data is given in black in diagram 6 where it is applied to the excerpt of the respiration effort signal depicted in grey in this diagram.

Secondly, to calculate the respiration rate from this filtered signal, the time between every two consecutive maxima, indicative for the start of an expiration, was determined. To this end a custom R-code peak-to-peak algorithm was developed and applied to the filtered data. This peak-to-peak algorithm marks local maxima as candidate peaks for expiration when they consist of at least 0.3s increasing values followed by at least 0.3 decreasing values in order to prevent marking small local maxima not indicative for expiration as candidate peaks. The location of these candidate peaks are indicated in diagram 6 with green lines. The value of 0.3s for the increasing and decreasing signal values are based on visual inspection of the raw respiration effort signal and the output of the algorithm, where the value of 0.3 proved to enable the algorithm to eliminate as many as possible local maxima not indicative for expiration as candidate peaks.

Thirdly, within a window of 2.2s (corresponding to the 0.42Hz filter applied) centred at the candidate peaks, the highest peak is taken to be the start of the expiration. This to prevent small local maxima often following or preceding the actual start of expiration to be falsely classified as the start of an expiration. These highest peaks taken as the start of expiration are depicted in diagram 6 by blue lines. When all expiration starting points are determined, the respiration rate per 30-second epoch is determined by counting the amount of expirations within each epoch.

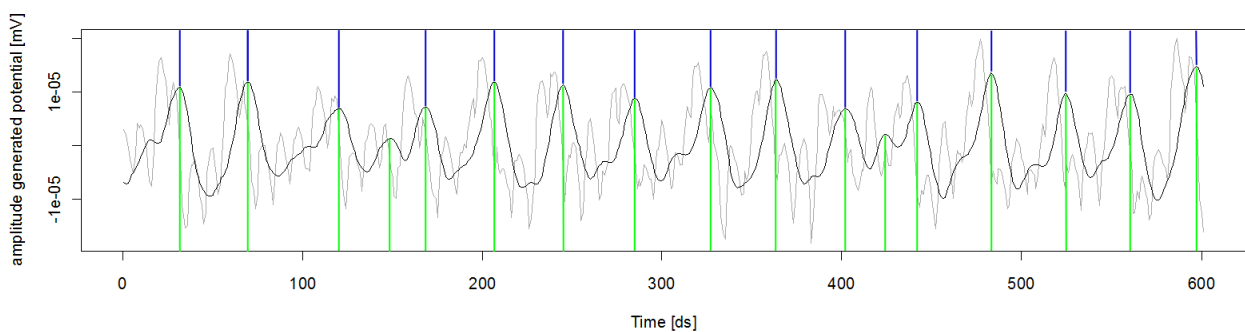


Diagram 6: Excerpt of raw abdominal respiration effort biosignal (taken from participant 102) in grey together with results of consecutive pre-processing steps. Black denotes the filtered raw signal (third order high-cut Butterworth filter, cut-off 0.42Hz), green lines mark candidate peaks for start expiration and blue marks peaks ultimately selected as start of expiration.

An example of such a respiration rate signal is given in diagram 7, where in red this respiration rate signal is given, for the most part of the diagram coinciding with the black signal, which will be described consecutively.

One final operation is applied to render the signal suitable as input to the module layer. To remove ectopic local maxima from the signal caused by body movement, a custom R-code body movement-artefact correction algorithm was developed and applied to this signal. This algorithm uses the body movement

signal to search for possible ectopic peaks. It marks peaks in the body movement signal and constructs a window in the corresponding respiration rate signal in which ectopic peaks can occur due to this body movement. This window is constructed around 5 epochs preceding and 10 epochs following the body movement peak to account for potential phase shifts between the body movement and respiration rate signals. To determine if ectopic peaks indeed occur in this window, the overall mean and standard deviation of the respiration rate inside the window is determined after which this is compared to the mean and standard deviation of 5 epochs preceding and following the window. If there exist peaks in the window that are more than 2 times the standard deviation from the mean, they are marked as ectopic peaks. These peaks are then attenuated by averaging over the mean of the 5 epochs preceding the window and the 5 epochs following the window as this proved to render a usable attenuation of the ectopic peaks. An example of the result of this correction is given in diagram 7 where in black the result of applying the algorithm to the original signal, given in red, is depicted together with the body movement signal used for the correction, given in green. This resulting signal is the respiration rate signal, containing a positive value for each epoch, indicating the respiration rate in times per minute and will be used as input for the module layer.

Given this final respiration rate signal, per participant the most suitable respiration rate signal, thorax or abdominal based, is chosen. To this end, first the amplitude of both signals is compared and the signal with the highest amplitude is chosen since the framework uses the difference in respiration rate rendering the signal with the highest amplitude the most usable for the framework. Secondly, the signal is inspected visually for potential abnormalities that were not picked up by the automatic pre-processing, such as compromised parts of signals due to measurement equipment failure. If these exist in either the abdominal or thorax based signal, but not in the other signal, the latter signal is used, otherwise the originally chosen signal will be used. In table 3 it is indicated which respiration rate signal, thorax or abdominal based, is used for each participant; participants of whom the thorax based respiration rate is used are printed italic.

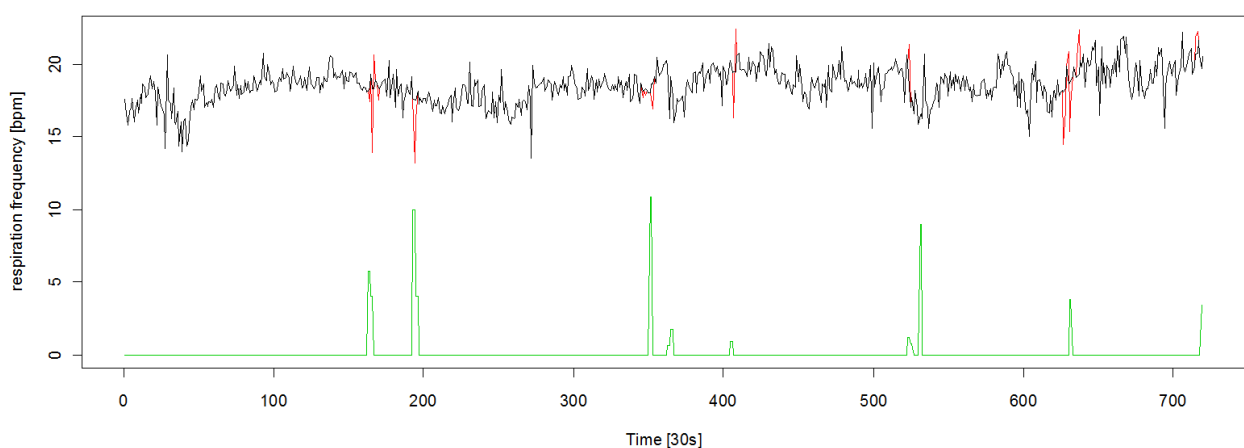


Diagram 7: Raw and body movement corrected abdominal based respiration rate biosignal in red and black respectively. Body movement biosignal used for body movement correction is depicted in green. Whole night data, taken from participant 104.

body movement signal pre-processing

With a few participants the body movement recordings contained no zero values, indicating that during the whole night movement was registered, although the scoring of the PSG data of these participants showed they were asleep. Instead, very low values were consistently recorded at epochs during sleep at which zero values, indicating no body movement, should be present. Because of the impossibility to produce such low values consistently during the night, especially during deep sleep and tonic REM stages as the scored sleep stages revealed, these low values were replaced by zero values, assuming a misconfigured measurement device. No further pre-processing was performed, rendering the final signal that will be used as input for the module layer to contain a positive value for each 30-second epoch, without a unit of measurement, where higher values indicate more body movement. An example of this final body movement signal is given in green in diagram 7.

	male	female
Group 1 young clinical	101	102
	103	104
		106
		108
Group 2 elderly clinical	201	202
	203	204
	205	
	207	
Group 3 elderly ambulatory	301	302
	303	304
	305	306
	307	308
	309	310
	311	

Table 3: Dataset containing all participants partitioned in three groups depending on difference in age and applied measurement technique. Participants forming the training set are given in the blue cells and participants from whom the thorax based respiration rate is used are printed bold.

5 The framework in detail: the module layer

In this chapter the second layer of the current implementation of the framework, the layer containing the different modules processing the biosignals contained in the data layer, is described. Each module is described separately including a description of the model underlying the module and the corresponding physiological basis for this model. Furthermore, the algorithm implementing this model and the parametrization of the module on the training set are described. In the algorithm description references to variables and constants in formulas are printed in italics to ensure readability. The optimized group parameter sets are given in this chapter. Furthermore, all module algorithms are given in appendix A.

module one, sleep/wake classification

The first module is designed to distinguish wake from sleep epochs using solely the body movement biosignal from the data layer. The discrimination between sleep and wake is based on the observation that during wake epochs body movement occurs more frequently and with a higher amplitude compared to sleep epochs (Tryon, 2004; Pollak, 2001; Ancoli-Israel et al., 2003 and Paquet et al., 2007) as described in chapter 2 and summarized in table 2 in the same chapter.

This observation was modelled by, amongst others, (for examples see Johnson et al., 2007; Hedner et al., 2004; Sadeh et al., 1994 and Ancoli-Israel et al., 2003), Cole et al. (Cole et al., 1992) on whose model this module is based. The model used here is an abstracted version of the model used by Cole et al., meaning that all variables can be parameterized to improve performance of the module within the framework including parameters that were considered constants in the original Cole et al. model.

In our model, every epoch is classified as either wake or sleep, based on the amount of body movement during that epoch and the surrounding epochs. This is done by taking the weighted summation of the body movement signal (the body movement signal is discussed in the fourth chapter and an example of this signal is given in diagram 7 of the same chapter) within a window of epochs surrounding the epoch of interest. If this summation exceeds a particular threshold, the epoch of interest is classified as wake, otherwise as sleep.

By using this summation window the model aims to attenuate potentially present recordings of twitches and small movements not indicative for wake epochs, to allow for robust sleep/wake classification using the body movement signal. Furthermore, a weight distribution within the summation window enables the model to determine the influence that individual epochs have on the classification, which allows for parameterization on the different participant groups.

algorithm

The algorithm sequentially inspects every epoch of the body movement biosignal, which, as described in chapter 4, contains a positive value for each epoch during the night, indicating the amount of body movement for that epoch. Starting at the first epoch of the night and ending with the last epoch, a discriminant value is computed for each epoch that is indicative for the participant being awake or asleep. The algorithm is specified stepwise in the following description.

1. The sequence B denotes the body movement signal. The length of B is equal to T , where T denotes the total number of epochs during the night, and in index t of B the body movement value for epoch t during the night is given, with t in $\{1, 2, 3, \dots, T\}$.

For every epoch being inspected, a window of interest E surrounding this epoch of interest B_t is constructed. The size and shape of the window of interest are determined by the pre and $post$ variables, denoting the number of epochs in the window before and after the epoch of interest respectively.

$$E = \langle B_{t-pre}, B_{(t-pre)+1}, \dots, B_t, B_{t+1}, \dots, B_{t+post} \rangle$$

2. As not all values within this window of interest are equally important in determining whether the participant was awake or asleep during the epoch of interest, the values are weighted using a sequence w with length equal to E , representing the weight assigned to the each epoch in the window of interest. The element-wise multiplication of w and E renders the weighted window of interest E' with length equal to $pre + post + 1$, the same length as w and E .

$$E' = w \cdot E$$

3. Finally, in order to obtain the discriminant value, indicative for the participant being awake or asleep, of the epoch of interest B_t , the summation is taken of the elements of the weighted window of interest E' and scaled using a scaling factor α with α between $[0, 1]$.

$$\text{discriminant}(B_t) = \alpha \cdot \sum_{i=1}^{i=(pre+post+1)} E'_i$$

4. If the discriminant value of a given epoch exceeds the threshold of 1, the epoch is classified as wake, otherwise as sleep^I. The classification is then stored in sequence H^{M1} , containing the classifications made by module one with in index t the classification for epoch t and length equal to T . The classifications are stored numerically, with classification 6 denoting wake and 1 denoting sleep.

$$H_t^{M1} = \begin{cases} 6, & \text{discriminant}(B_t) > 1 \\ 1, & \text{discriminant}(B_t) \leq 1 \end{cases}$$

parameterization

The parameters in the algorithm described above need to be optimized to enable the module to select the parameter set at which it contributes maximally to the agreement of the framework as a whole with the golden standard. This parameter set is optimized per participant group using the data from the training set and consists of the pre and $post$ parameters determining the size and shape of the summation window, the weight sequence w and the scaling factor α .

I In principle the threshold can be set to any positive value as long as α is changed accordingly.

In order to determine this optimal parameter set, first the optimum for this parameter set needs to be determined. Intuitively, this optimum would be chosen as the maximum agreement with the golden standard, meaning as many correct wake and sleep classifications as possible. However, since module one is the only module in the framework that can classify wake stages, this module should classify as many wake stages as possible because any wake stage not found by this module will not be recognized by the framework at all. It is at the same time important to ensure that, by pursuing this goal, not more actual sleep epochs are wrongly classified as wake than can be corrected for by the fusion layer. Therefore the optimum for this parameter set will be chosen to allow the module to maximize both overall agreement as well as the fraction of wake epochs found. We will refer to this fraction of wake epochs found as the *sensitivity*, which, in statistics, refers to the fraction of actual positives that are found by a classifier in a binary classification task. As module one indeed performs a binary classification task by classifying all epochs of the night as either wake or sleep, we will mark wake as the positive class and sleep as the negative class in module one, in order to use the concept of sensitivity to make the determination of the optimum for the parameter set easier and the description of this process more clear. A search is conducted for the optimal parameter set that achieves the highest value for a convex combination of overall agreement and the fraction of wake epochs found, the sensitivity. This combination is set at the point where the agreement with the golden standard and the sensitivity are the highest while making not more errors than can be corrected in the fusion layer, which is defined as making not more than 10% additional errors compared to the agreement achieved by module one when using the parameter set for maximum agreement with the golden standard. To determine this convex combination different combinations were systematically pursued. This resulted in an optimum location for participant groups 1 and 3 of $8/9$ * agreement and $1/9$ * sensitivity and an optimum of $7/9$ * agreement and $2/9$ * sensitivity for participant group 2. This means that using these convex combinations as the respective optima, the optimal parameter sets for the different groups achieve the highest possible agreement and sensitivity while not making more than 10% additional errors compared using a parameter set that solely achieves the highest possible agreement with the golden standard. In this manner we aim to optimize the parameters in module one to find as many wake epochs as possible while not making more errors than can be corrected for in the fusion process.

Now the optimum for the parameter set is determined, the parameter set that is located at this optimum can be determined. This parameterization is done in three steps due to the high time complexity of the parameterization algorithm caused by the large amount of parameter values possible. During each step one (set) of the parameters is optimized reducing the dimensionality of the search space drastically. First the *pre* and *post* variables are optimized, next the optimal weight sequence w is determined and finally the scaling factor α is optimized. In every step first the parameter set per participant is individually optimized followed by the optimization per group based on these individually optimized sets.

In the first step the search region for optimising the *pre* and *post* parameters is constructed around the optimal values found by Cole et al. (Cole et al., 1992) for these two parameters. This means that both parameters are varied between 0 and 4 (optimal values for *pre* and *post* were found by Cole et al. to be at 4 and 2, respectively). Furthermore, the scaling factor α is varied between 0.01 and 0.0001 to ensure enough expressive power in the parameter set to successfully optimize the *pre* and *post* parameters. The optimized group *pre* and *post* values are given in table 6.

In the second step the weight sequence w is optimized where the optimal weight values found by Cole et al were used as a starting point for the optimization. The search region was constructed around these

values with step sizes of 10% of the starting value of the respective elements of the weight sequence with the boundaries at five steps (so 50%) above and below the starting values. The *pre* and *post* variables were kept constant to the optimal group values determined in step I and the scaling factor α was varied between 0.01 and 0.0001 to ensure enough expressive power in the parameter set to successfully optimize the weight parameters. optimized group weight sequence w values are given in table 6, here *weight i* denotes the i^{th} element in w and if the respective element does not exist in the sequence w , it is set to 0.

During the final optimization step the *pre*, *post* and weight sequence w parameters are set to the optimal group values found in step I and II and only the value of the scaling factor α is varied within a more specific search region with values between 0.005 and 0.00001 and a step size of 0.00001. optimized group scaling factor α values are given in table 6.

This table will be used by module one to select the optimal settings for its parameter set of *pre*, *post*, w and α , given the relevant participant parameter, being group 1, group 2 or group 3, which is stored in the parameter component.

Also the predictive value, the value indicating the chance that a given classification is correct, is determined for each optimal parameter setting, so that is known how reliable the classifications made by module one are for the different parameter settings. These values will be used during the fusion process in the case of disagreement between the modules, this will be discussed in detail in chapter 6.

	pre	post	weight 1	weight 2	weight 3	weight 4	weight 5	weight 6	weight 7	α
Group 1	3	1	0	30	16	25	121	8	0	0.00005
Group 2	4	2	45	27	12	31	108	8	50	0.00021
Group 3	4	1	55	33	16	31	121	8	0	0.00006

Table 4: Optimal group parameter set module one.

module two, REM sleep classification

The second module utilizes the respiration rate biosignal contained in the data layer and the therefrom derived respiration rate variability to detect REM sleep stages. The detection is based on the existence of heightened respiration rates and increased respiration rate variability during REM sleep compared to other sleep stages in a nights' sleep (Douglas et al., 1982) as was described in chapter 2 and summarized in table 2 in the same chapter. The model underlying this module originates from earlier work by Chung et al. (Chung, et al., 2009 and Chung et al., 2007) and employs the complete Chung et al. model and two improvements reducing errors at the beginning and end of the night that were frequent with the original model. These improvements will be covered in step 4 and 6 of the algorithm specification below.

In the model two adaptive thresholds based on the respiration rate and respiration rate variability signals are employed, classifying any epoch as REM at which both the respiration rate and the respiration rate variability exceed their respective thresholds.

algorithm

1. The respiration rate biosignal, which, as described in chapter 4, contains a positive value for each epoch indicating the respiration rate in times per minute, is smoothed by taking for each epoch the mean within a window of 30 epochs (900s) around this epoch of interest. This is done to obtain a more robust signal for classification, with the window size of 30 epochs being the window size found by Chung et al. to render a signal containing the important features needed for REM classification while eliminating most of the noise present in the original signal. The sequence R denotes the respiration rate and has length equal to T , where T denotes the total number of epochs during the night, and in index t of R the respiration rate for epoch t during the night is given, with t in $\{ 1,2,3,\dots,T \}$. Furthermore \bar{R} denotes the smoothed respiration rate signal, which has the same length and index structure as sequence R .

$$\bar{R}_t = \frac{1}{30} \cdot \sum_{i=t-15}^{i=t+14} R_i$$

For the first 15 epochs and the last 14 epochs it is not possible to take the mean of 30 epochs around the epoch of interest. This is because values of epochs R_t with $t < 1$ or $t > T$ are not available since these epochs would denote epochs before and after the measurements. Therefore, for these epochs the mean is taken of the amount of epochs available until the start or end of the measurements, so until R_1 and R_T .

2. Given the smoothed respiration rate signal \bar{R} from step 1, the respiration rate variability is computed. The absolute value of the difference between the respiration rate and the smoothed respiration rate is taken as the respiration rate variability signal ΔR . Here ΔR denotes a sequence containing the respiration rate variability of the whole night in beats per minute with in index t the respiration rate for epoch t with t in $\{ 1,2,3,\dots,T \}$ where T is the total number of epochs.

$$\Delta R_t = | R_t - \bar{R}_t |$$

3. The respiration rate variability signal is also smoothed to retrieve a more robust signal. This is done in the same manner as with the respiration rate in step 1, taking for each epoch the mean within a window of 30 epochs (900s) around this epoch of interest and treating the beginning and end of the night in the same manner.

$$\Delta \bar{R}_t = \frac{1}{30} \cdot \sum_{i=t-15}^{i=t+14} \Delta R_i$$

Here the window size of 30 epochs is again the window size used in the original model, rendering a signal containing the important features whilst suppressing noise.

4. For both the respiration rate and the respiration rate variability, adaptive thresholds are constructed by taking for each epoch of the respective signals the mean within a window of 300 epochs (9000s) around these epochs of interest. The window size of 300 epochs was found by Chung et al. to render thresholds that simultaneously follow the long term behaviour of the original signals while being rigid enough to act as thresholds. Furthermore, by the adaptive threshold obtained in this manner for respiration rate and respiration rate variability a static offset is added that can be varied to account for the different baseline respiration rates found between participant and participant groups.

In the original model of Chung et al. the thresholds for the beginning and end of the night were computed in the same manner as in step 1 and 3, using the available epochs when it is not possible to use a window of 300 epochs because this would entail using epochs before R_I or after R_T . This resulted in the first 150 and last 149 epochs of the thresholds to be computed of less epochs than the remaining epochs, resulting in less of the desired rigidity and therefore more errors at the beginning and end of the night.

Therefore, two adjustments are incorporated to prevent these errors. The first adjustment regards the errors made at the end of the night and is incorporated in this step. The second adjustment, reducing the errors at the beginning of the night, is covered in step 6.

The adjustment covered in this step, regarding errors made at the end of the night, comprises the fact that for the last 149 epochs not only the remaining available epochs until R_T are used, but also the average of the last 300 epochs before the end of the night, which we will refer to as the *trend*. In this manner we determine the threshold using more epochs, ensuring enough rigidity.

Below are the formulas for the construction of the respiration rate and respiration rate variability thresholds in the first part of the night, followed by the formulas for the construction of the respiration rate and respiration rate variability thresholds for the last part of the night, incorporating the adjustment described.

$$\text{threshold}(R_t) = \left(\frac{1}{300} \cdot \sum_{i=t-150}^{i=t+149} R_i \right) + \text{offset} \quad [\text{with } t \leq T-149]$$

$$\text{threshold}(\Delta R_t) = \left(\frac{1}{300} \cdot \sum_{i=t-150}^{i=t+149} \Delta R_i \right) + \text{offset} \quad [\text{with } t \leq T-149]$$

$$\text{threshold}(R_t) = \left(\frac{1}{(T-i)+150} \cdot \sum_{i=t-150}^{i=T} R_i + \text{trend} \right) + \text{offset} \quad [\text{with } t > T-149]$$

$$\text{threshold}(\Delta R_t) = \left(\frac{1}{(T-i)+150} \cdot \sum_{i=t-150}^{i=T} \Delta R_i + \text{trend} \right) + \text{offset} \quad [\text{with } t > T-149]$$

5. Each epoch at which the smoothed respiration rate and smoothed respiration rate variability exceed their respective thresholds and the smoothed respiration rate variability is above the offset, to ensure the variability is high enough to be indicative for REM sleep, is classified as REM sleep. All other epochs remain unclassified, belonging to either the NREM or wake stage. The classifications are stored in sequence H^{M2} , containing the classifications made by module two with in index t the classification for epoch t and length equal to T . The classifications are stored numerically, with classification 5 denoting REM and 0 denoting not REM.

$$H_t^{M2} = \begin{cases} 5, & \text{smooth}(R_t) > \text{threshold}(R_t) \wedge \text{smooth}(\Delta R_t) > \text{threshold}(\Delta R_t) \wedge \text{smooth}(\Delta R_t) > \text{offset} \\ 0, & \text{otherwise} \end{cases}$$

6. Finally, to eliminate errors at the beginning of the night made by the algorithm, any epochs classified as REM epochs in the first 100 epochs, 50 minutes, are reclassified as not REM, since it is physiologically virtually impossible to be in REM sleep this early after sleep onset.

$$H_t^{M2} = 0 \quad \text{for } t \leq 100$$

parameterization

For this module only the *offset* parameter in the algorithm as described above needs to be optimized in order for the model to select the optimal parameter set for each participant group. While there is at the moment no other module to classify REM sleep, this would normally call for optimization towards sensitivity to ensure classification of REM stages, the positive class in module two, comparable with the method employed with module one for wake stages. But in this current implementation the epochs that are classified as NREM by module one and two are given to module three for classification instead of module three inspecting the whole night as is the standard in the framework structure. This calls for optimization towards *specificity*, to ensure the least amount of errors propagating towards module three. In statistics, specificity refers to the fraction of actual negatives that are found by a classifier in a binary classification task.

As module two indeed performs a binary classification task by classifying all epochs of the night as either REM or not REM, and REM is marked as the positive class and not REM as the negative class in module two, we can use the concept of specificity to make the determination of the optimum for the parameter set easier and the description of this process more clear. In order to find the middle ground between optimising towards sensitivity and optimising towards specificity, this module is currently optimized towards agreement. Because there is only one parameter to be optimized this is done per group without optimising individually first. The search region for the *offset* parameter optimization was between 0 and 1 using step size 0.001. Within none of the groups the parameter optimization reached the boundaries of the search region, ratifying the chosen search region. The results of the group parameterization are given in table 5.

This table will be used by module two to select the optimal setting for it's *offset* parameter, given the relevant participant parameter, being group 1, group 2 or group 3, which is stored in the parameter component.

Also the predictive value, the value indicating the chance that a given classification is correct, is determined for each optimal parameter setting, so that is known how reliable the classifications made by module two are for the different parameter settings. These values will be used during the fusion process in the case of disagreement between the modules, this will be discussed in detail in chapter 6.

	offset
Group 1	0.151
Group 2	0.986
Group 3	0.683

Table 5: Optimal group parameter set module two

module three, light sleep/deep sleep classification

The third module employs all the biosignals contained in the data layer. Using the physiological differences in body movement, respiration rate and respiration rate variability during deep and light sleep, the module discriminates between these sleep stages.

As described in chapter 2 and summarized in table 2 in the same chapter, during deep sleep there is no body movement and a steady, high respiration rate pattern reflected by a relatively high valued respiration rate signal coupled with a low respiration rate variability signal. In light sleep, body movement can occur and a more unstable respiration rate pattern can be discerned, visible in a lower respiration rate compared to deep sleep and a heightened respiration rate variability (Douglas et al., 1982).

Based on these differences, a model is developed that starts out with all epochs classified as deep sleep after which in two steps light sleep epochs are searched for. The rationale behind starting out with all epochs set to deep sleep is that these epochs are harder to find given the two used biosignals and because the proportion of deep sleep is much smaller compared to light sleep, making the detection of these stages even further challenging.

First this detection of light sleep epochs is performed by utilizing a summation window on the body movement signal which adds up all the body movement values within the window. If the summation exceeds zero, so when body movement is detected within the window, the epoch of interest is classified as a light sleep epoch. By taking into account multiple epochs by means of this summation window, the relatively sparse positive values in the body movement signal can be used to detect a relatively large number of light sleep epochs while maintaining the number of false positives, erroneous light sleep classifications, to a minimum.

Secondly, an epoch is classified as light sleep when the respiration rate is below an adaptive threshold based on the respiration rate and the respiration rate variability is above an adaptive threshold based on the respiration variability signal, indicating a relatively low and unstable respiration, indicative for light sleep. In this manner, by utilising adaptive thresholds that follow their original signals, the model is able to detect epochs with lowered respiration rates and heightened variability indicating light sleep stages whilst taking the long term rhythm of these signals during the night into account.

algorithm

1. Here H^{M3} denotes the sequence containing the classifications made by module three with in index t the classification for epoch t with t in $\{ 1,2,3,\dots,T \}$ where T is the total number of epochs during the night. The classifications are stored numerically, with classification 4 denoting deep sleep. The algorithm starts out with all epochs set to deep sleep as denoted in undermentioned formula.

$$H^{M3} = 4^T$$

2. The sequence B denotes the body movement signal, which, as described in chapter 4, contains a positive value for each epoch during the night, indicating the amount of body movement for that epoch. Similar to the approach in the other modules, the length of B is equal to T , where T denotes the total number of epochs during the night, and in index t of B the body movement value for epoch t during the night is given, with t in $\{ 1,2,3,\dots,T \}$.

The algorithm sequentially inspects every element of B , the epoch of interest, and for each epoch a discriminant value is computed that is indicative for the participant being in the light sleep stage. This discriminant value is equal to the summation of the body movement values within a window of epochs surrounding the epoch of interest. The size and shape of the summation window is determined by the pre and $post$ variables, where pre and $post$ denote the amount of epochs in the summation window before and after the epoch of interest respectively.

$$\text{discriminant}_{M3}(B_t) = \sum_{i=t-pre}^{i=t+post} B_i$$

- When the discriminant value of a given epoch exceeds zero, indicating that body movement occurred in or around this epoch, the epoch is classified as light sleep, otherwise the epoch remains classified as deep sleep as is denoted in the following formula.

$$H_t^{M3} = \begin{cases} 2, & \text{discriminant}_{M3}(B_t) > 0 \\ H_t^{M3}, & \text{discriminant}_{M3}(B_t) \leq 0 \end{cases}$$

Here 2 denotes light sleep in the sequence H^{M3} .

- Consecutively, the algorithm utilizes the respiration rate biosignal, which, as described in chapter 4, contains a positive value for each epoch indicating the respiration rate in times per minute. First the respiration rate variability signal is computed which is equal to the difference signal of the respiration rate signal. This is given in the undermentioned formula.

$$\Delta R_t = R_t - R_{(t-1)} \quad [\text{with } R_1 = 0]$$

As described earlier in module two, R and ΔR both denote sequences, containing respectively the respiration rate and the respiration rate variability of the whole night in beats per minute, with index t the respiration rate for epoch t with t in $\{ 1, 2, 3, \dots, T \}$ where T is the total number of epochs during the night.

- To obtain the adaptive thresholds needed for classification, both the respiration rate signal and the respiration rate variability signal are smoothed by taking for each epoch the mean of the epochs within a window centred around the epoch of interest with the size equal to the smoothing factor β as denoted respectively in the following two formulas.

$$(\bar{R}_t) = \left(\sum_{i=t-0.5 \cdot \beta}^{i=t+0.5 \cdot \beta} R_i \right) / \beta + 1$$

$$(\Delta \bar{R}_t) = \left(\sum_{i=t-0.5 \cdot \beta}^{i=t+0.5 \cdot \beta} \Delta R_i \right) / \beta + 1$$

Furthermore a static offset is added to both smoothed signals that can be varied to account for the different baseline respiration rates found between participant and participant groups. This is denoted in the following two formulas.

$$\text{threshold}(R_t) = (\bar{R}_t) + \text{offset}_R$$

$$\text{threshold}(\Delta R_t) = (\Delta \bar{R}_t) + \text{offset}_{\Delta R}$$

6. When in any epoch the respiration rate is below and the respiration rate variability above their respective thresholds, the epoch is classified as light sleep, otherwise the epoch retains the classification obtained in step one as given in the following formula.

$$H_t^{M3} = \begin{cases} 2, & R_t < \text{threshold}(R_t) \wedge \Delta R_t > \text{threshold}(\Delta R_t) \\ H_t^{M3}, & R_t > \text{threshold}(R_t) \vee \Delta R_t < \text{threshold}(\Delta R_t) \end{cases}$$

parameterization

In order for this module to select the optimal parameter set for each participant group, the parameters in the algorithm as described above need to be optimized using the data from the training set.

There are five parameters that need to be optimized. These are the *pre* and *post* parameters determining the size of the summation window in step one, the smoothing factor β determining the smoothing of the respiration rate and respiration variability and the corresponding offset_R and $\text{offset}_{\Delta R}$ determining the respective thresholds.

These parameters are optimized towards agreement because, at the moment, this is the only module for both light sleep and deep sleep that can detect these stages. Therefore, there is no additional gain in optimising the parameters towards overestimation of a certain sleep stage for better results in the fusion layer where these results will be challenged by other modules.

The optimising of the parameters towards agreement is done simultaneously for all parameters. This is done in two steps to keep the optimization within reasonable time boundaries, which is due to the high time complexity caused by the high dimensionality of the search region.

In the first step the parameters are simultaneously optimized within a physiologically plausible search region with a relative large step size. If in this step certain parameters are optimized at the boundaries of the search region this step is repeated using a larger search region for the given parameter. In this manner an overview of the locations of the optima of the different parameters is constructed.

In the second step the parameters are further optimized by setting the search region around the found optima and using a smaller step size for each parameter if possible.

These two-steps were first executed for each participant individually to identify possible outliers and to determine the search region for the group parameter optimization, which is set around the lowest and highest values found for all participants per group. The different search regions and step sizes used in the individual and group optimization for both steps are given in table 6.

		step I			step II		
		lower bound	upper bound	step size	lower bound	upper bound	step size
individual	pre	1	10	1	Dependent on individual results step I .		1
	post	1	10	1		1	
	β	130	150	5		1	
	offset _R	0	1	0.1		0.01	
	offset _{ΔR}	0	0.5	0.1		0.01	
Group 1	pre	11	19	1	17	19	1
	post	6	11	1	9	11	1
	β	125	160	5	150	160	1
	offset _R	0	1.5	0.1	0	0.2	0.01
	offset _{ΔR}	0	0.21	0.1	0	0.1	0.01
Group 2	pre	15	17	1	15	17	1
	post	6	11	1	9	11	1
	β	120	150	5	140	150	1
	offset _R	0	1.52	0.1	1.4	1.5	0.01
	offset _{ΔR}	0	0.15	0.1	0	0.1	0.01
Group 3	pre	10	18	1	16	18	1
	post	8	11	1	9	11	1
	β	127	172	5	140	150	1
	offset _R	0	2.5	0.1	0	0.1	0.01
	offset _{ΔR}	0	0.21	0.1	0	0.1	0.01

Table 6: Search regions and step sizes used during two-step parametrizing of the parameter set of module three. Individual results were used to determine the search region for the group parameterization. Note that the step size is equal for all groups and individual parameterization for each step.

The results for the group parameterization are given in table 7.

This table will be used by module three to select the optimal settings for its *pre*, *post*, β , *offset_R* and *offset _{Δ R}* parameters, given the relevant participant parameter, being group 1, group 2 or group 3, which is stored in the parameter component.

Also the predictive value, the value indicating the chance that a given classification is correct, is determined for each optimal parameter setting, so that is known how reliable the classifications made by module three are for the different parameter settings. These values will be used during the fusion process in the case of disagreement between the modules, this will be discussed in detail in chapter 6.

	pre	post	β	offset _R	offset _{ΔR}
Group 1	18	10	158	0.1	0
Group 2	16	10	149	1.5	0
Group 3	17	10	148	0	0

Table 7: Optimal group parameter set module three.

6 The framework in detail: the fusion layer

In this chapter the current implementation of the fusion layer, the third and final layer of the framework, is described. In the fusion layer the output of the different modules is fused into one definitive hypnogram. First the selection of the definitive classifications for this hypnogram is described after which the method for determining the confidence of these classifications is covered, including the formal description of the algorithm used. For completeness, the complete algorithm is given in appendix A.

definitive classification selection

Per epoch the classifications made by the different modules for this epoch are compared.

In case these classifications are in agreement, the most specific classification is selected as the classification for this epoch in the final hypnogram. For example, if module one classifies a certain epoch as 'sleep' and module two classifies this same epoch as 'REM', the corresponding epoch in the final hypnogram will be classified as 'REM', this being the more specific classification.

In case the classifications disagree, the classification with the highest predictive value given the relevant participant parameters contained in the parameter component is selected for the definitive hypnogram. If after this selection a more specific classification exists in the module classifications for this epoch, such as light and deep sleep are more specific than NREM sleep, this classification is selected for the definitive hypnogram. This event can not occur in the current implementation of the framework because of the low number of modules in the module layer.

definitive classification confidence level

The confidence level for each epoch classification in the definitive hypnogram depends on the predictive value of all the classifications out of which the definitive classification was selected. The confidence level is equal to the predictive value of the definitive classification, determined during the parameterization of the respective module, on which an alteration is applied based on the (dis)agreement with the other module classifications for this epoch. For example, module one classifies an epoch as 'wake' and module two classifies the same epoch as 'REM'. Assuming that for this specific participant the classification of module two has a higher predictive value, the classification 'REM' will be used in the definitive hypnogram. The confidence level will be equal to the predictive value for REM classifications in module two with a negative alteration applied because there existed disagreement between the different modules.

The alteration on the confidence level of the definitive classification is determined by the summation of the predictive values of the classifications given by the other modules. In this summation for predictive values of disagreeing modules the negative is taken and the summation is scaled, as described in the algorithm section beneath, to ensure the resulting final confidence level to remain valued between 0 and 1. In this manner the confidence given to a definitive epoch classification is based on the amount of (dis)agreement between all the different modules with respect to this epoch.

algorithm

1. As at the moment module three is only capable of performing classification on NREM epochs and not of performing classification on whole night epochs, the fusion process first fuses the classifications of module one and two in order to determine the NREM epochs, which will subsequently be fused with the output of module three.

For the fusion process of module one and two the classifications given by module one and two are compared for every epoch of the night. In case the classifications are not conflicting, the most specific classification is selected. If the classifications are conflicting, the classification with the highest predictive value, as determined during the parameterization of the specific module, is selected. The fusion of the classifications of module one and two is schematically given in table 8 where the possible combinations of classifications and the resulting final classifications are given.

		module one classification	
		sleep [1]	wake [6]
module two classification	REM [5]	REM	conflict, select classification with highest predictive value.
	not REM [0]	NREM	Wake

Table 8: Fusion table for module one and two.

2. Consecutively, the epochs classified as NREM are partitioned into light sleep and deep sleep by module three.
3. For the determination of the confidence of the definitive classification as derived in the first two steps, first the base confidence is determined, which is equal to the predictive value of the selected classification as determined during the parameterization, denoted below with $pv(H_t^{Mm})$.

$$c_{\text{base}}(t) = pv(H_t^{Mm})$$

here H_t^{Mm} corresponds to the selected, final classification made by module m for epoch t .

4. The alteration on the base confidence level is equal to the summation of the predictive values of the classifications given by the other modules for the same epoch, where for the disagreeing modules the negative of the predictive values are taken. In this manner agreement of other modules with the selected classification contributes to a higher confidence in this final classification while disagreement lowers this confidence. This denoted in the following two formulas.

$$c_t^{Mm'} = \begin{cases} pv(H_t^{Mm'}), & H_t^{Mm'} \text{ agrees with } H_t^{Mm} \\ -pv(H_t^{Mm'}), & H_t^{Mm'} \text{ disagrees with } H_t^{Mm} \end{cases}$$

Here $c_t^{Mm'}$ denotes the confidence of modules not delivering the final classification in their respective classifications, which is negated if these are in disagreement with the final classification.

$$c_{alt}(t) = \sum_{m'=1}^M c_t^{Mm'} \quad \text{with } m' \neq m$$

Here M is the total number of modules and m the module delivering the final classification.

5. This alteration is scaled to ensure that, when applied to the base confidence, the final confidence will remain normalised between 0 and 1. This is done by scaling the alteration by the maximal alteration possible divided by the amount of modules contributing to this alteration. So, in case the alteration is positive, meaning that overall the other modules agree with the selected classification and therefore the base classification will be increased, the maximum alteration that can be applied to the base confidence with the final confidence level remaining under 1 is $1 - c_{base}(t)$. In order to ensure the alteration is not larger than this number, the alteration is scaled by this term, divide by the amount of modules which have contributed to this alteration, equal to $M-1$, needed due the summation. If the alteration is negative the maximum alteration that can be applied is equal to $c_{base}(t)$ divide by the amount of modules which have contributed to this alteration. This is denoted in the following formula.

$$c_{alt}(t) = \begin{cases} \frac{1 - c_{base}(t)}{m-1} \cdot c_{alt}(t), & c_{alt} \geq 0 \\ \frac{c_{base}(t)}{m-1} \cdot c_{alt}(t), & c_{alt} < 0 \end{cases}$$

6. To determine the final confidence level for the classification in the definitive hypnogram the alteration is applied to the base confidence level. In this manner the confidence of the definitive classification is taken as the basis, and, depending on whether the other modules agree or disagree, this confidence is respectively heightened or lowered, where the confidence of the (dis)agreeing classifications determine the amount of this alteration.

$$c(t) = c_{base}(t) + c_{alt}(t)$$

7 Results

This chapter describes the classifications made by the framework and compares these with the golden standard classifications. First the test results of the three modules working independently are described after which the statistics of the fusion process and the results of the framework as a whole for the test set are treated. The results are specified in terms of the following statistical metrics: agreement, sensitivity, specificity, and negative- and positive predictive value. Of these metrics, primarily the first three metrics are used in the descriptions because of their high explanatory value and if needed the negative- and positive predictive value are taken into account. In the calculation of the results epochs scored as MT by the golden standard, indicating that there existed too much body movement for successfully classifying the epoch, are excluded from the calculation. In this chapter solely a description of the results is given, discussion of these results will be treated in the next chapter. All classifications, made by the framework as a whole and by all modules separately, are given per participant in appendix B.

module one, sleep/wake classification

Module one achieves an overall agreement to the golden standard of around 90% with few deviation between the groups. Relatively low recognition of wake epochs, the positive class, and a near flawless classification of sleep epochs, the negative class, can be seen. The precise metrics per participant, set and group can be found in table 9 followed by an exposition of the most significant of these metric results.

		sensitivity	ppv	specificity	npv	agreement						
		sensitivity	ppv	specificity	npv	agreement	sensitivity	ppv	specificity	npv	agreement	
Group 1 young clinical	101	0.07	1	1	0.96	0.96						
	103	0.3	0.57	0.99	0.9	0.96						
	102	0.11	0.75	1	0.97	0.97	0.09	0.04	0.93	0.97	0.91	106
	104	0.03	1	1	0.9	0.9	0.12	1	1	0.75	0.76	108
	mean	0.13	0.83	1	0.93	0.95	0.11	0.52	0.97	0.86	0.84	mean
	sd	0.12	0.21	0.01	0.04	0.03	0.02	0.68	0.05	0.16	0.11	sd
Group 2 elderly clinical	201	0.19	0.29	0.92	0.86	0.8						
	203	0.37	0.54	0.93	0.87	0.83						
	202	0.19	0.8	0.97	0.68	0.69	0.51	0.61	0.96	0.95	0.92	205
	204	0.48	0.26	0.6	0.79	0.56	0.44	0.07	0.93	0.99	0.92	207
	sd	0.31	0.47	0.86	0.8	0.72	0.48	0.34	0.95	0.97	0.92	sd
	mean	0.14	0.25	0.17	0.09	0.12	0.05	0.38	0.02	0.03	0	mean
Group 3 elderly ambulatory	301	0.07	0.22	0.93	0.78	0.74						
	303	0.03	0.07	0.97	0.96	0.94						
	305	0.37	0.46	0.97	0.96	0.94						
	307	0.2	1	1	0.76	0.78	0.31	0.71	0.99	0.92	0.91	309
	302	0.21	0.22	0.94	0.94	0.89	0.14	0.12	0.96	0.96	0.92	311
	304	0.3	0.9	1	0.95	0.95	0.05	0.28	0.97	0.8	0.78	306
	308	0.59	0.67	0.99	0.98	0.97	0.32	0.82	0.99	0.9	0.89	310
	mean	0.25	0.51	0.97	0.9	0.89	0.21	0.48	0.98	0.9	0.88	mean
	sd	0.19	0.36	0.03	0.09	0.09	0.13	0.34	0.02	0.07	0.06	sd
	TRAINING						TEST					

Table 9: Module one results.

Between the different groups the amount of underestimation of wake epochs (sensitivity) differs. Group 1 achieves the lowest sensitivity of 0.11, compared to group 3 with sensitivity values around 0.21 and group 2 with a sensitivity value of 0.48 .

This higher sensitivity in group 2 is reflected in the agreement, which is also higher compared to the other two groups, around 0.92 versus 0.84 and 0.88 for group 1 and 3 respectively. Also noticeable with respect to the sensitivity is participant 306 exhibiting the low value of 0.05 for this metric compared to the group mean of 0.21. With respect to the specificity metric no noteworthy differences can be discerned.

Between the test and training set no noticeable differences can be discerned for the different groups, the results generally are better in the test set, which is not necessarily to be expected but indicates that no overfitting occurred.

The metrics per set are visualized in the Receiver Operating Characteristic, (ROC) in diagram 8.

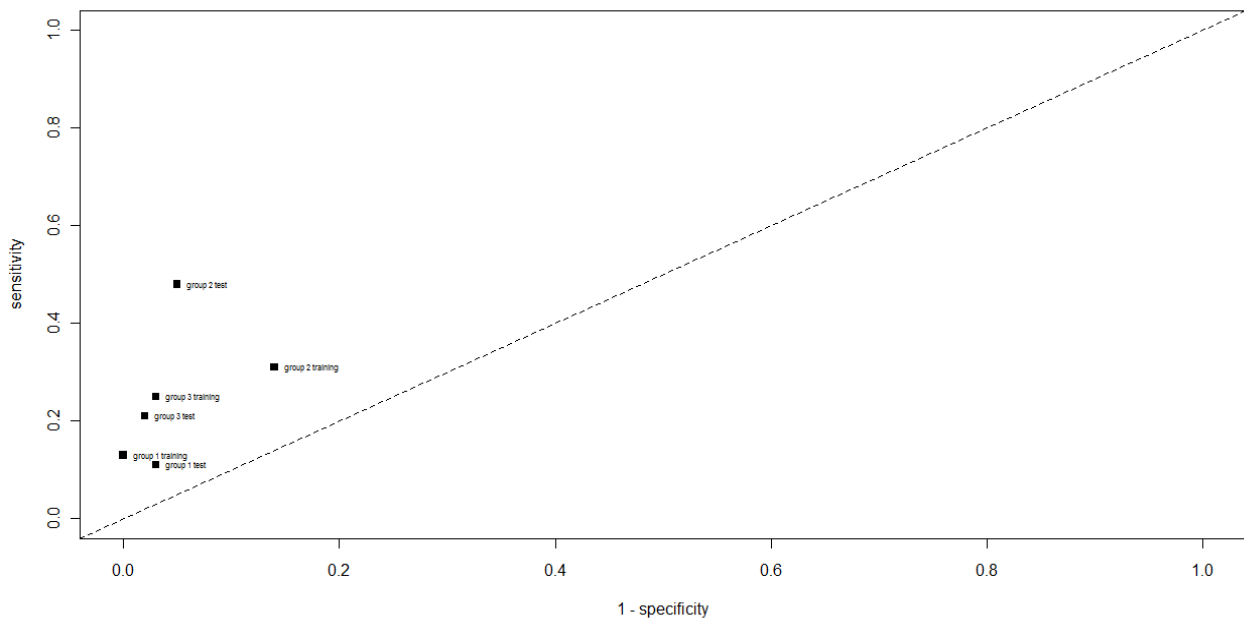


Diagram 8: ROC per set for module one

In a ROC the sensitivity is plotted against 1 minus the specificity. In this manner the classifier can be visually inspected and different sets can be easily compared. The more the classifier is located in the top left corner, the higher the agreement is with the golden standard.

In this ROC the aforementioned absence of overfitting can be seen in the fact that the training and test sets of the groups are either located close together or, as is the case with group 2, the test set is located closer to the top left corner, indicating higher agreement.

module two, REM sleep classification

Module two, performing REM classification, achieves an agreement between 77% and 88% with the golden standard, depending on the group. This agreement is brought about by relatively few recognized REM epochs, the positive class, so a low sensitivity, and a near perfect recognition of epochs belonging to other sleep stages, the negative class, so a high specificity. The precise metrics per participant, set and group can be found in table 10 followed by an exposition of the most significant of these metric results.

		sensitivity	ppv	specificity	npv	agreement	sensitivity	ppv	specificity	npv	agreement	
Group 1 young clinical	101	0.16	0.55	0.97	0.83	0.82						
	103	0.69	0.8	0.95	0.92	0.92						
	102	0.7	0.85	0.96	0.91	0.91	0.82	0.68	0.88	0.93	0.86	106
	104	0.19	0.41	0.94	0.85	0.82	0.71	0.56	0.92	0.96	0.89	108
	mean sd	0.44 0.3	0.65 0.21	0.96 0.01	0.88 0.04	0.87 0.05	0.77 0.08	0.62 0.08	0.9 0.03	0.95 0.02	0.88 0.02	mean sd
Group 2 elderly clinical	201	0.56	0.8	0.97	0.9	0.9						
	203	0	0	0.93	0.89	0.84						
	202	0	0	0.9	0.92	0.84	0.06	0.07	0.87	0.84	0.75	205
	204	0	0	0.91	0.89	0.82	0.09	1	1	0.78	0.79	207
	sd mean	0.14 0.28	0.2 0.4	0.93 0.03	0.9 0.01	0.85 0.03	0.08 0.02	0.54 0.66	0.94 0.09	0.81 0.04	0.77 0.03	sd mean
Group 3 elderly ambulatory	301	0.04	0.04	0.84	0.86	0.74						
	303	0.44	0.83	0.96	0.81	0.82						
	305	0.41	0.9	0.98	0.81	0.83						
	307	0.11	0.16	0.91	0.86	0.8	0	0	0.98	0.75	0.74	309
	302	0.41	0.96	1	0.85	0.86	0.25	0.66	0.94	0.72	0.72	311
	304	0.34	0.77	0.96	0.8	0.8	0.63	0.79	0.94	0.89	0.87	306
	308	0.1	1	1	0.8	0.81	0.21	0.42	0.92	0.81	0.77	310
	mean sd	0.26 0.17	0.67 0.4	0.95 0.06	0.83 0.03	0.81 0.04	0.27 0.26	0.47 0.35	0.95 0.03	0.79 0.08	0.78 0.07	mean sd
TRAINING						TEST						

Table 10: Module two results.

Between the different groups the amount of recognized REM epochs (sensitivity) differs greatly. The sensitivity in group 1, containing young clinically measured participants, is considerably higher when compared to the other two groups of which the third group (elderly ambulatory measured participants) is again substantially higher than group 2 containing elderly clinically measured participants. This influences the agreement as well, the agreement in group 1 is higher than the agreement in the other groups, which are close to one another. With respect to the sensitivity also another interesting result can be discerned, participant 309, located in group 3, displays an extremely low sensitivity value of 0. With regard to specificity, the recognition of the negative class, there is no meaningful difference between the different groups.

Between the test and training set no noticeable differences can be discerned for the different groups, the results are generally better in the test set, which is not necessarily to be expected but indicates that no overfitting occurred. The metrics per set are visualized in the Receiver Operating Characteristic, (ROC) in diagram 9.

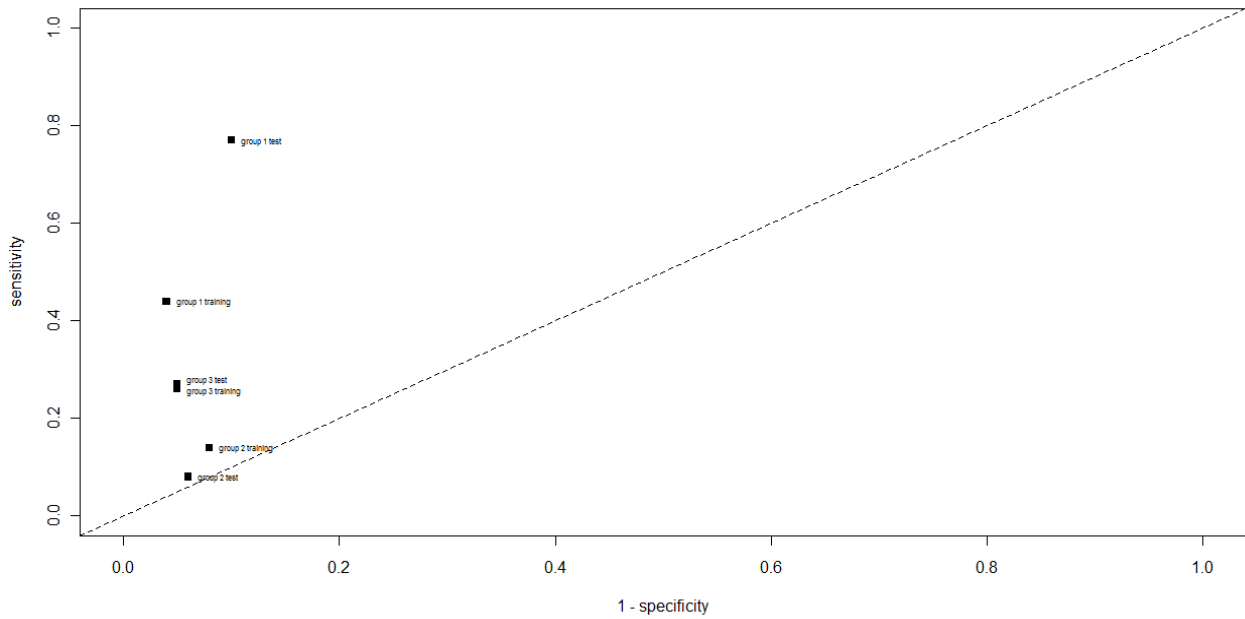


Diagram 9: ROC per set for module two

In this ROC the aforementioned absence of overfitting can be seen in the fact that the training and test sets of the groups are either located close together or, as is the case with group 1, the test set is located closer to the top left corner, indicating higher agreement.

module three, light sleep/deep sleep classification

Module three achieves an overall agreement of around 72% based on high performance in recognition of the positive class, light sleep epochs, evidenced in the high values for the sensitivity metric. At the same time this is coupled with a low performance in classification of the negative class, deep sleep epochs, as is seen in the low values for the specificity and negative predictive value (npv), indicating the fraction of negative classifications made that are correct. The precise metrics per participant, set and group can be found in table 11 followed by an exposition of the most significant of these metric results.

		TRAINING					TEST				
		sensitivity	ppv	specificity	npv	agreement	sensitivity	ppv	specificity	npv	agreement
Group 1 young clinical	101	0.69	0.94	0.79	0.33	0.7					
	103	0.92	0.76	0.33	0.36	0.74					
	102	0.73	0.84	0.53	0.36	0.68	0.89	0.67	0.05	0.18	0.63
	104	0.55	0.66	0.57	0.45	0.56	0.87	0.84	0.44	0.51	0.77
	mean	0.72	0.8	0.56	0.38	0.67	0.88	0.76	0.25	0.35	0.7
	sd	0.15	0.12	0.19	0.05	0.08	0.01	0.12	0.28	0.23	0.1
Group 2 elderly clinical	201	0.84	0.83	0.1	0.1	0.72					
	203	0.76	0.93	0.32	0.1	0.72					
	202	0.72	0.74	0.44	0.42	0.63	0.83	0.98	0.54	0.08	0.82
	204	0.98	0.79	0.15	0.67	0.78	0.84	0.9	0.29	0.19	0.77
	sd	0.83	0.82	0.25	0.32	0.71	0.84	0.94	0.42	0.14	0.8
	mean	0.11	0.08	0.16	0.28	0.06	0.01	0.06	0.18	0.08	0.04
Group 3 elderly ambulatory	301	0.89	0.93	0.07	0.05	0.84					
	303	0.7	0.86	0.1	0.04	0.64					
	305	0.64	0.9	0.44	0.13	0.62					
	307	0.68	0.95	0.45	0.09	0.67	0.74	0.95	0.22	0.04	0.72
	302	0.88	0.8	0.44	0.59	0.75	0.9	0.6	0.08	0.33	0.57
	304	0.72	0.81	0.51	0.38	0.67	0.8	0.91	0.06	0.03	0.74
	308	0.78	0.92	0.45	0.21	0.74	0.83	0.9	0.54	0.38	0.79
	mean	0.76	0.88	0.35	0.21	0.7	0.82	0.84	0.23	0.2	0.71
	sd	0.1	0.06	0.18	0.2	0.08	0.07	0.16	0.22	0.19	0.09

Table 11: Module three results.

On the group level the high performance in the recognition of light sleep (sensitivity) differs slightly with group 1 outperforming group 2 and group 3. With respect to the low performance on deep sleep classification group 2 achieves higher values for the relevant metric specificity while the negative predictive value is the highest in group 1. The agreement with the golden standard is higher in group 2 compared to group 1 and group 3 with practically identical values for this metric.

Between the test and training set no noticeable differences can be discerned for the different groups, the results generally are better in the test set, which is not necessarily to be expected but indicates that no overfitting occurred. Exception here on is group 1 and to a lesser degree group 3, where the specificity metric is much lower in the test set compared to the training set, the results are not too low to expect overfitting. The metrics per set are visualized in the Receiver Operating Characteristic, (ROC) in diagram 10.

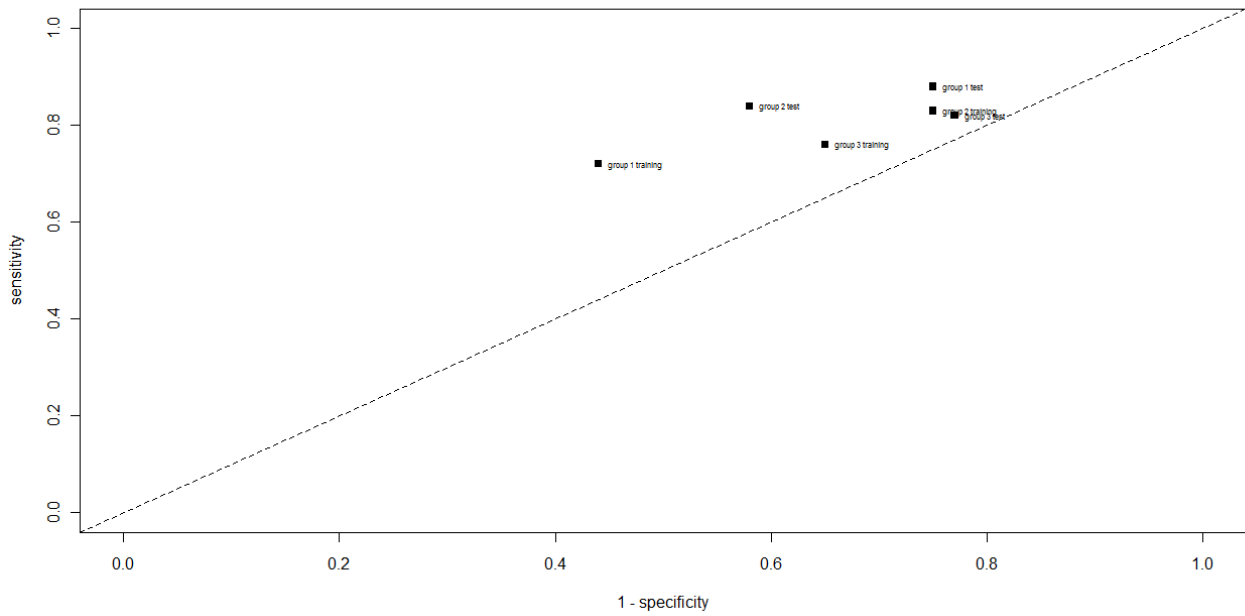


Diagram 10: ROC per set for module three

In this ROC the aforementioned absence of overfitting can be seen in the fact that the training and test sets of the groups are located close together. Also, with group 1 and to a lesser degree group 3, the training set is located closer more to the left, indicating higher specificity as described earlier.

fusion process and framework as a whole

The description of results, achieved by the current implementation of framework working as a whole, so including the fusion of the hypnograms outputted by the different modules, is divided into two sections. First the results of the framework working with only module one and two, resulting in a 3-step hypnogram consisting of the classifications NREM, REM and wake, are discussed. Secondly the results obtained by the framework working with all modules, so including module three, resulting in 4-step hypnogram, are given.

3-step hypnogram

The framework with only module one and two working, returning a 3-step hypnogram, achieves an overall agreement between 67% and 72%. This hypnogram is based on the classifications made by module one and two and the fusion process which corrects between 50% to 67% of the erroneous classifications on which disagreement existed between the modules.

First the results regarding the fusion, such as the amount of conflicts between the modules, the amount of these conflicts that are corrected and the resulting agreement with the golden standard, are given per participant, set and group.

Secondly the agreement with the golden standard is further specified in the confusion matrix of the 3-step hypnogram per set and group.

The fusion statistics per participant, set and group can be found in table 12.

		# conflicts	# corrected	% corrected	from module	agreement	# conflicts	# corrected	% corrected	from module	agreement		
Group 1 young clinical	101	0	0	NA	2	0.77							
	103	0	0	NA	2	0.85							
	102	0	0	NA	2	0.87	2	2	100	2	0.76	106	
	104	0	0	NA	2	0.72	2	0	0	2	0.68	108	
	mean	0	0	NA	2	0.8	2	1	50	2	0.72	mean	
	sd	0	0	NA	0	0.07	0	1.41	70.71	0	0.06	sd	
Group 2 elderly clinical	201	11	2	18.2	1	0.74							
	203	3	3	100	1	0.71							
	202	7	7	100	1	0.62	27	18	66.7	1	0.72	205	
	204	27	1	3.7	1	0.47	0	0	NA	1	0.72	207	
	sd	12	3.25	27.08	1	0.64	13.5	9	66.67	1	0.72	sd	
	mean	10.52	2.63	51.75	0	0.12	19.09	12.73	NA	0	0	mean	
Group 3 elderly ambulatory	301	6	3	50	2	0.54							
	303	0	0	NA	2	0.77							
	305	6	1	16.7	2	0.77							
	307	10	0	0	2	0.6	0	0	NA	2	0.67	309	
	302	15	11	73.3	2	0.77	4	3	75	2	0.67	311	
	304	0	0	NA	2	0.75	0	0	NA	2	0.66	306	
	308	0	0	NA	2	0.78	6	2	33.3	2	0.66	310	
	mean	5.29	2.14	40.54	2	0.71	2.5	1.25	50	2	0.67	mean	
	sd	5.79	4.06	32.92	0	0.1	3	1.5	29.49	0	0.01	sd	
TRAINING						TEST							

Table 12: Framework results 3-step hypnogram.

Due to the fact that the current implementation of the framework contains only three modules, not so many conflicts arise during the fusion process, per participant this amounts to an average of 5 or 6 conflicts but this amount varies greatly. Of the conflicts arising in the fusion process, around 61% is corrected.

On the group level there are large differences to be discerned with respect to the amount of conflicts, the percentage of these conflicts that are corrected and the resulting agreement with the golden standard. In group 1 virtually no conflicts between module one and two exist, in contrast to group 2 with an average of around 13 conflicts per participant and to a lesser degree group 3 with around 3 conflicts per participant. Also the percentage of these conflicts which are corrected by the fusion process differs. In group 1 this is equal to 50%, while in group 2 this is around 67% and in group 3 equal to 50%. The agreement with the golden standard of the resulting 3-step hypnogram also differs between the groups, group 1 and 2 display an agreement of 0.72, compared to 0.67 for group 3.

The confusion matrix per set, group and overall of the 3-step hypnogram is given in table 13. Here one confusion matrix is given for every set and group and one matrix for the overall performance of the current implementation of the framework when outputting a 3-step hypnogram. In any of these confusion matrices, the rows display the classifications made by the framework and the columns show the classifications according to the golden standard. This results in the correct classifications to appear in the primary diagonal. The cell values in the matrices are in percentages of the total classifications within the respective matrices.

group 1 TR					group 1 TST				
class \ stand	NREM	REM	wake	total	class \ stand	NREM	REM	wake	total
NREM	70.2	10.79	4.67	85.66	NREM	54.7	4.11	11.97	70.77
REM	3.2	9.35	0.28	12.83	REM	7.31	14.75	0.84	22.89
wake	0.74	0.28	0.49	1.51	wake	4.04	0	2.3	6.33
total	74.14	20.41	5.45	100	total	66.04	18.86	15.1	100
group 2 TR					group 2 TST				
class \ stand	NREM	REM	wake	total	class \ stand	NREM	REM	wake	total
NREM	53.49	7.46	12.36	73.31	NREM	67.76	16.75	2.09	86.6
REM	1.3	2.64	4.26	8.2	REM	3.14	1.47	0.7	5.3
wake	9.4	2.11	6.97	18.49	wake	4.12	1.12	2.86	8.09
total	64.19	12.22	23.59	100	total	75.02	19.33	5.65	100
group 3 TR					group 3 TST				
class \ stand	NREM	REM	wake	total	class \ stand	NREM	REM	wake	total
NREM	62.39	14.83	8.09	85.31	NREM	56.96	18.12	9.57	84.66
REM	3	6.33	1.08	10.41	REM	3.13	7.16	0.73	11.02
wake	1.56	0.48	2.23	4.27	wake	1.39	0.73	2.2	4.32
total	66.95	21.64	11.4	100	total	61.48	26.01	12.51	100
group 1 ALL					group 2 ALL				
class \ stand	NREM	REM	wake	total	class \ stand	NREM	REM	wake	total
NREM	65	8.55	7.12	80.67	NREM	58.27	10.58	8.92	77.77
REM	4.58	11.16	0.47	16.2	REM	1.92	2.25	3.07	7.23
wake	1.84	0.19	1.1	3.13	wake	7.63	1.78	5.59	15
total	71.42	19.89	8.69	100	total	67.82	14.6	17.58	100
group 3 ALL					OVERALL TST				
class \ stand	NREM	REM	wake	total	class \ stand	NREM	REM	wake	total
NREM	60.36	16.06	8.65	85.07	NREM	60.92	13.21	8.37	82.5
REM	3.05	6.64	0.95	10.64	REM	3.14	6.67	1.31	11.11
wake	1.5	0.58	2.22	4.29	wake	2.92	0.75	2.71	6.38
total	64.91	23.28	11.82	100	total	66.98	20.63	12.39	100

Table 13: Confusion matrices for 3-step hypnogram (in %) per set, group and overall (TR and TST stand for training and test set respectively).

On the overall level for the test set (matrix in right lower corner) it can be seen that the lion's share of the erroneous classifications is made with NREM classifications. Around 21% of the total classifications consist of erroneously classified NREM epochs, 13% wrongly classified REM epochs and 8% wrongly classified wake epochs. The other erroneous classifications account for only small percentages of the total classifications.

On the group level, the overall behaviour, where by far the largest portion of faulty classifications are made when epochs are erroneously classified as NREM, is maintained.

In group 1 the percentage of these erroneous classifications is 15.67, distributed over 8.55% wrongly classified REM epochs and 7.12% wrongly classified wake epochs. Furthermore displays group 1 a relatively high percentage of NREM epochs which are falsely classified as REM epochs. The other erroneous classifications constitute small percentages of the total classifications. These percentages are depicted in table 13, in the *group 1 ALL* confusion matrix.

For group 2 19.5% of the total classifications consist of faulty NREM classifications, partitioned in 10.58% falsely classified REM epochs and 8.92% wake epochs. Also a relatively high percentage of NREM epochs are falsely classified as wake epochs in this group. The other erroneous classifications constitute small percentages of the total classifications. These numbers can be found in table 13, in the *group 2 ALL* confusion matrix.

In group 3 24.71% of the total classifications consist of faulty NREM classifications, partitioned in 16.06% falsely classified REM epochs and 8.65% wake epochs. The other erroneous classifications constitute small percentages of the total classifications, also to be found in table 13, in the *group 3 ALL* confusion matrix.

4-step hypnogram

In the current implementation, the framework with all three modules working renders a 4-step hypnogram. As earlier described, the third module in this implementation is not able to process whole night epochs and therefore only takes epochs as input that are classified by module one and two as NREM after the fusion. The output of the third module, containing light and deep sleep epochs, is then inserted into the 3-step hypnogram. This entails that there is no additional fusion involved other than the fusion described in the 3-step hypnogram in the making of the 4-step hypnogram. Therefore no fusion statistics are described here because these are non-existent.

Furthermore no confusion matrix is given for the 4-step hypnogram while there is no additional information to be gained compared to the information available in the 3-step hypnogram. This is because the REM and wake classifications made don't change between the 3- and 4-step hypnograms, and the partitioning of NREM in light and deep sleep and the confusion between these stages are described in the results section of module three.

It is also entailed that the agreement with the golden standard for every participant, set and group is completely determined by the agreement of the 3-step hypnogram and the agreement of the third module for each of these levels.

For matters of completeness, the agreement of the 4-step hypnogram with the golden standard is given in table 14.

8 Discussion

In this chapter the methods used and results obtained by the framework are discussed. First the data set and the pre-processing of the data set in the data layer are discussed to give the correct context for the consecutive discussion of the module and fusion layer acting on this data. Next for every module in the module layer the parameterization and obtained results are discussed and followed by an overview of the most important points found during this discussion. Thirdly the fusion statistics of the fusion layer and the definitive hypnograms outputted by the framework are discussed.

data layer

The data pre-processing employed in the data layer and the resulting data set as described in chapter 4 will be discussed here. First the pre-processing will be discussed after which some focus points regarding the data set are discussed.

dataset pre-processing

There are two methods employed during the pre-processing in the data layer that need to be discussed here. The first method in need of discussion is the body movement-artefact correction algorithm. As described in chapter 4, this algorithm is employed on the respiration rate signal to remove ectopic local maxima caused by body movement. With all participants this algorithm acted correctly and removed body movement peaks in the respiration rate signal, except for two participants. With participants 308 and 106 the algorithm didn't perform optimally, leaving small body movement peaks present in the respiration rate signal.

The second method employed that needs to be discussed here concerns the replacement with zero values of the physiologically impossible low values occurring in the body movement signal, as described in chapter 4. This is a plausible approach to correct the supposed faulty body movement signal, however, at the same time, it must be taken into consideration here that this method is based on the unverifiable assumption that the actigraph was misconfigured in certain cases.

data set

Two focus points regarding the data set are in need of discussion here.

First, the structure of the dataset needs to be discussed. The size of the dataset as a whole is expected to be large enough to support any conclusions drawn concerning the whole dataset. But the group sizes in the test set can potentially be too small to provide the conclusions of the needed confidence, this holds especially for group 1 and group 2, as can be seen in table 3 of chapter 4.

The second point concerns an overall focus point. The data set used for parameterization and testing of the framework contains *obtrusive* measurements gathered using PSG, the golden standard. This means that the framework, which is constructed to use *unobtrusive* signals, is based on *obtrusive* measurements. This position is unavoidable as it is necessary to validate the framework against the golden standard, but needs to be mentioned here.

module one, sleep/wake classification

parameterization

The parameterization of module one, as described in chapter 5, was done in three steps to reduce the time needed to reasonable levels. The method employed has the potential disadvantage that not all the possible combinations within the search region will be analysed. Furthermore the search region for the pre and post parameters was forcedly set-up between the values found by Cole et al. because for values outside these ranges no corresponding physiological plausible values for the weight sequence were available. Searching for these weight sequence values would have resulted in an exhaustive search, which with the current high time complexity of the parameterization was not an option.

So due to this construction of the search region and the stepwise parameterization within this search region, not all possible combinations of the parameters are analysed. But owing to the fact that the third parameter, the scaling factor, was varied during all the steps, only certain tuples of the first three parameters, the pre and post and the weight sequence parameters were not examined. And because the results of the stepwise parameterization are very close to the results found by Cole et al. and other authors (Tryon, 2004; Pollak, 2001; Paquet et al., 2007; Ancoli-Israel et al., 2003; Cole et al., 1992 and Johnson et al., 2007) the expectation is that in those not examined parameter tuples no optimum for this module will be found.

Naturally, to completely discard the option of higher optima at other points in the search space a simultaneous parameterization of all parameters needs to be executed, which calls for using another method with a lower time complexity.

results

As described in chapter 7, module one achieves an overall agreement with the golden standard of around 90%, combined with low values for the sensitivity metric. This is a high agreement on it's own, but is important to realise that the prevalence of sleep epochs during a night is around 85%, rendering a naïve classifier, which classifies all epochs during the night as sleep, performing only slightly worse.

The low values for the sensitivity metric are the main cause of the remaining 10% erroneous classifications made. These low values, indicating underestimation of wake epochs, differ on the group level. In group 1 the lowest sensitivity of around 0.11 is found, compared to group 3 a with sensitivity value of 0.21 and group 2 which has values of 0.48 for this metric.

This lowest sensitivity of group 1 is most likely due to the lesser amount of arousals (wake epoch sequences) during the night in this group compared to group 2 and group 3 (mean wake epochs during the night in group 1, group 2 and group 3 respectively: 62, 125 and 118), which make these epochs harder to detect. This is in par with results found in the literature indicating more arousals with the elderly compared to the young. (Padova et al., 2009)

Furthermore, for participant 108, in group 1, the scaling factor parameter alpha was set too low, causing body movement summation peaks indicative for wake epoch (sequences) to remain under the threshold, resulting in unclassified wake epochs, ergo low sensitivity. Another participant from group 1, participant 106, shows a relatively low correlation between wake and sleep epochs and the respective body movement signal as presupposed and modelled in module one; being that during wake epochs body

movement occurs more frequently and with a higher amplitude compared to sleep epochs. This is caused by frequent peaks in the body movement signal during sleep epochs.

The difference in sensitivity metric values between group 2 and group 3, 0.48 and 0.21 respectively, containing both elderly participants, is caused by the adjustment made to the metric combination used for the optimization during the parameterization where the ratio between agreement and sensitivity was shifted for group 2 as described in chapter 5.

Moreover, the individual reasons for the low sensitivity found with group 1 also hold for participants in group 3, adding to the lower sensitivity of the group. First the low correlation between wake and sleep epochs and the respective body movement signal can be found as a reason for the low values of the sensitivity metric for participants 309 and 311. With participant 309 the absence of body movement during wake epochs causes the low correlation and with participant 311 this is caused by peaks in the body movement signal occurring recurrently during sleep epochs. Secondly the too low value of the scaling factor group parameter alpha can be identified for being the cause of low sensitivity metric values for participants 306 and 310 where with participant also a possible phase shift between the body movement and the scored sleep stages was discerned.

Based on the discussion above three causes for the low sensitivity values, the main cause of the erroneous classifications, can be identified:

- Lower sensitivity is more prominent with young subjects due to lesser amount of arousals during the night compared to the elderly, making these few epochs harder to detect.
- The correlation between wake and sleep epochs and body movement is too low with certain participants. This occurred in 3 of the individually discussed participants.
- The scaling factor alpha parameter is set too low. This occurred in 3 of the individually discussed participants.

No solutions can be given for the more prominent low sensitivity with young subjects as this is due to unalterable physiological phenomena. Furthermore, no direct solution to eliminate the low correlation between wake and sleep epochs and body movement exists given the current model. This low correlation also pertains to the noise introduced by using unobtrusive sleep staging as described in chapter 2. Solutions for addressing the issue concerning the low value of the alpha parameter can include giving a greater share to sensitivity during parameterization. To determine the size of this share it is important that all modules are parametrized at the same time.

With regard to the literature, the results found are in par with the results found by Cole et al. and subsequent research results found in the literature (Tryon, 2004; Pollak, 2001; Paquet et al., 2007; Ancoli-Israel et al., 2003; Cole et al., 1992 and Johnson et al., 2007).

module two, REM sleep classification

parameterization

As described in chapter 5, only one parameter in the algorithm needed to be optimized. Therefore an exhaustive brute force search for this parameter within a reasonable search region and step size was executed where the optima were not found at the edges of the search region. Therefore and because of the lack of physiologically plausibility for values outside this search region these values are taken to be the optimal parameter set for this module.

results

The results presented in chapter 7 indicate that, on an overall level, module two achieves an agreement between the 77% and the 88% with the golden standard, together with low sensitivity metric values. This is a reasonable achievement, especially when the prevalence of REM sleep, constituting 20-25% of the sleep time during the night, is taken into account. However, these results also leave room for improvement, especially with respect to the low sensitivity metric, which is the main cause of the erroneous classifications. Apart from a few individualized cases, this low sensitivity metric is consistently discerned on both the group and individual level. The cause lies with local peaks in the respiration rates of participants forcing a vertical lift on the threshold used by the model, which in turn causes the heightened respiration rates during REM epochs to remain under this threshold and therefore stay unclassified. Within the elderly (group 2 and 3) this is caused by large wake epoch sequences, results which are in par with the physiological differences between these two age groups found in the literature. (Padova et al., 2009)

In group 1, containing young participants measured clinically, the sensitivity is considerably higher compared to groups 2 and 3 with values of 0.77, 0.08 and 0.27 respectively.

The low sensitivity values for group 2, populated by elderly participants measured clinically, are caused by two reasons. For participant 205 these values are due to large wake epoch sequences during the night, visible in epoch sequence 125 through 150 in appendix B. These wake sequences, because of the overall heightened respiration rate during wakefulness compared to sleep, force a vertical shift on the respiration rate threshold. Because of this overall higher value of the threshold the heightened respiration rates during REM sleep don't exceed this threshold and are therefore not classified as REM epochs resulting in a low sensitivity metric for these participants. This is in accordance with the general consensus stating that elderly people have more arousals during the night.

For participant 207, the low values are caused by the group offset value determined during the parameterization which is too high for this specific participant. The application of group parameter values for all participants throughout the framework enables the framework to operate without the need for individual per-participant training or fitting but also has as a consequence that these group parameter values potentially are not the best fit for some individual participants, as is the case with the group offset parameter value for this specific participant 207 in which case it is too high.

This heightened amount of arousals during the night also holds for the elderly participants measured ambulatory populating group 3. But here the effect on the sensitivity is less extreme compared to group 2, this is because there are less participants in group 3 compared to group 2 where this heightened

amount of arousals caused large problems for the algorithm. Only with participants 310 and 311 this effect can be discerned, visible around epoch 200 for participant 310 and around epoch 220 for participant 311 in appendix B. Furthermore, for participant 309 the group offset value determined during the parameterization was too high resulting in the extremely low sensitivity of 0. The reason why the heightened amount of arousals has more effect in group 2 compared to group 3 is not clear, perhaps the reason can be found in the difference between protocol and measurement set-up.

Given the discussion above, a potential solution for this problem of the low sensitivity, the main cause of erroneous classifications, could be to:

- constructing adaptive thresholds for the respiration rate and respiration rate variability using smoothing over smaller epoch sequences. In this manner the threshold will correlate more with the original respiration rate (variability) signal causing the threshold to more quickly descend after high respiration rate (variability) non-REM epoch sequences enabling the REM epochs to again exceed the threshold. This approach has as a further advantage that the framework will be able to perform sleep staging with a smaller lag between the actual measurement and the staging.

With regard to the literature on this subject, the most relevant literature is given by Chung et al. (Chung, et al., 2009 and Chung et al., 2007). Because in these studies only young participants were included, only the results from group 1 can be compared. The results of module two are in par with the results found by Chung et al. in these studies, on some metrics a few percent points lower, but the overall picture is the same.

module three, light sleep/deep sleep classification

parameterization

The parameterization was executed for all parameters simultaneously in two steps as described in detail in chapter 5. In the first step the search region was exhaustively searched with relatively large step sizes after which, in the second step, around the optima found in the first step a renewed search was conducted using smaller step sizes to find the optimum more accurately. The potential disadvantage of this method is that during the first the step sizes are too large so the actual maximum in the search space is 'stepped over' and instead another local maximum is found and explored in step two. Given the individual parameterization conducted before this group parameterization and the not too large step sizes in the first step together with the expansion of the search space whenever an optimum was found at the edge of the space it is expected this to not be the case.

results

Before the results described in chapter 7 are discussed, first two general discussion points will be treated. First, as a result of the algorithm utilising the original respiration rate (variability) signal instead of a smoothed version of this signal (as the second module does, described in detail in chapter 5), the classification often displays fickle patterns with classifications containing quickly alternating light and deep sleep classifications caused by the original signal consecutively diving under and emerging above the threshold which is, naturally, not in par with the actual sleep stage of the participant and the golden standard. Even more, it frequently occurs that although on a overall basis within a certain epoch sequence the respiration rate and respiration rate variability satisfy the conditions of the module for the actual sleep stage, many epochs will be wrongly classified because in these wrongly classified epochs not both the signals fulfilled the conditions caused by the quickly alternating states of being above or below the respective thresholds.

Secondly, as a general discussion point, not the absolute value of the respiration rate variability is used but the original signal, that is, the differential signal on the respiration rate. This signal includes the information needed, the variation of the respiration rate, but as it is not of value for the model whether or not this variation is positive or negative with respect to an earlier point in the signal, this signal carries unused and unwanted data, noise. Therefore the absolute version of this signal, which only carries the variation without indicating the direction of the variation is preferred.

Given these general discussion points, which will be followed up later in the discussion, the results of the module as given in chapter 7 are discussed. It can be seen in this chapter that, on an overall basis, module three achieves a rather low agreement with the golden standard of around 72%. This is due to high values for the sensitivity metric coupled with low values for the specificity and negative predictive value metrics.

On the group level the sensitivity values differ slightly with group 1 outperforming group 2 and group 3. With respect to the low performance on deep sleep classification group 2 achieves higher values for the relevant metric specificity while the negative predictive value is the highest in group 1. The agreement with the golden standard is higher in group 2 compared to group 1 and group 3 with practically identical values for this metric. These differences can be traced back for the most part to the two general

discussion points mentioned above, but also to individual differences, which are treated here per group. With participant 106, exhibiting low specificity and npv metrics, the low specificity metric is foremost caused by the occurrence of body movement during deep sleep, which was also encountered in the discussion of the results of module one for this participant. This body movement results in wrongly classified light sleep epochs causing a drop in the sensitivity metric value. The amount of body movement is not very high, but the effect of these activity epochs is enlarged by the large pre and post parameters. The low npv metric values with this participants is caused by the above mentioned methodological general discussion point regarding the fickleness of the respiration rate variability signal. These causes are illustrated in epoch sequences 24 through 159, 219 through 339 and 471 through 553 for participant 106 in appendix B. Furthermore, with respect to the latter participant, the body movement-artefact correction algorithm did not correct all ectopic maxima for this participant as described earlier in the dataset pre-processing section of this chapter.

For group 2, the same factors causing the low specificity and npv metric values can clearly be seen in epoch sequences 17 through 122, 153 through 256 and 539 through 626 for participant 205 and epoch sequences 150 through 304, 334 through 434 and 480 through 615 with participant 207 in appendix B. The participants exhibiting low valued specificity and npv metrics in group 3 have the same underlying issues as the participants in group 1 and 2. The low specificity metric values witnessed with participants 306 and 311 is caused by body movement occurring during deep sleep epoch sequences and the low npv metric values seen with participants 306 and 309 is brought about by on the one hand the absence of body movement during light sleep and on the other hand the aforementioned methodological factor of using the original respiration rate variability signal.

Based on the discussion above, it becomes clear that the main causes of the erroneous classifications are the two general discussion points discussed at the beginning of this section and which are repeated here:

- the use of the original respiration rate signal, not the smoothed signal, rendering it so fickle to be less applicable as a stable indicator for light and deep sleep
- the use of the original respiration rate variability signal instead of the absolute respiration rate variability signal, which would indicate correctly the information sought, being the variation of the respiration rate signal. Furthermore, this absolute signal would need to be smoothed to reduce the fickleness and render it applicable as a robust indicator for light and deep sleep.

With regard to these issues the possible solutions are fairly obvious,

- the use of the smoothed respiration rate signal instead of the original respiration rate signal
- the use of the smoothed absolute respiration rate variability signal instead of the original respiration rate variability signal.

There were also other points of discussion on the individual level,

- quite often body movement during deep sleep epochs seems to appear albeit very low valued. Perhaps the alteration of the body movement threshold value of zero to a smaller value could prevent these epochs to be wrongly classified as light sleep while maintaining a high sensitivity.

- to lift the low negative predictive value and specificity metrics, perhaps the algorithm should not start out with all epochs set to deep sleep which now results in epochs remaining in this state throughout the module which results in low npv and specificity metrics.
- Currently the module only takes NREM epoch sequences as input, this should be enlarged to whole night epoch sequences, in par with the other modules.

All these latter points are at the moment less relevant and should be reviewed again after the first two major points are treated, because these two points are expected to have great impact on the above lesser important points and their potential solutions.

fusion process and framework as a whole

3-step hypnogram

As described in the respective section of chapter 7, amount of conflicts occurring in the fusion process differs greatly between the different participants of group 2 and 3. This deviation is caused by the difference in agreement with the golden standard these participants have in module one and two. Participants with high agreement values in module one and two have none to very few conflicts during the fusion process because the amount of erroneous classifications is by definition low. As the agreement drops, the amount of conflicts rises, which is precisely the function of the fusion process, to correct erroneous classifications when these arise.

Also is described how the percentage of these conflicts that is corrected differs significantly. This is caused by the fact that at the moment between only two modules conflicts can arise. This means, because of the low amount of conflicts which can be rendered by two modules, most conflicts of a given participant originate from one or two epoch sequences during the sleep of the participant, so all the conflicts are located closely near each other. Therefore most of these conflicting epochs are classified the same stage by the golden standard and will also all be corrected with a same classification by the fusion process. This renders that all the epochs in such an epoch sequence are either correctly or faulty fused, thereby creating a large deviation in the percentage of corrected conflicts.

These two discussion points, the deviation in the amount of conflicts and the percentage of these conflicts that is corrected, explain the large deviation and the low amount of 5 to 6 conflicts per participant given in chapter 7. Also, the overall amount of 50% to 67% corrected erroneous classifications is based on the above described workings.

When we look into the erroneous classifications made by the fusion process in more detail, overall the largest proportion of these classifications, 21%, is made with NREM classifications, partitioned into 13% wrongly classified REM epochs and 8% wrongly classified wake epochs. Because of the relatively low influence of the fusion process due to the low amount of modules these numbers are almost completely determined by the overall behaviours of module one and two. Module one has overall a relatively low recognition of wake epochs and module two a low recognition of REM epochs. This results in epochs being classified as NREM during the fusion process which are actually wake or REM epochs.

4-step hypnogram

As described in the results chapter, only the agreement rates of the 4-step hypnogram are discussed, not the fusion statistics or the confusion matrix as these are either non-existent or already described in earlier results and therefore contain no new information.

The agreement rate of the 4-step hypnogram with the golden standard is around 0.53, based on the not very divergent agreement rates of 0.59, 0.57 and 0.49 for group 1, group 2 and group 3. These agreement values are fully determined by the agreement rates of the 3-step hypnogram and the agreement rates for module 3 as the classifications of module 3 are simply pasted into the 3-step hypnogram.

overall characteristics

In this section general characteristics of the framework that need discussion are treated.

The first characteristic in need of discussion is the currently used parameterization optimum of the different modules. As described in the respective parameterization paragraphs of chapter 5, the current optimum location of a module is selected by determining the function of the module in the framework. For example, module one is parametrized towards an optimum that takes the overall agreement of the module with the golden standard into account, as well as the sensitivity, the ability to detect wake epochs in this context, since module one is the only module within the current implementation of the framework that can detect wake epochs.

The problem with this approach is that the outcome of the framework, due to the fusion process, is not simply a linear summation of the outcomes of the different modules. This makes it extremely difficult, if not impossible, to determine the best optimum towards which a module needs to be parameterized to achieve the optimal performance *of the whole framework*.

The solution to this issue is to optimize all the modules simultaneously, using different parameterization optima sets in order to determine the best optimum towards which each module needs to be parameterized to achieve the best performance of the whole framework.

This operation will have a high time complexity, being the reason why this has not been performed for the current implementation, but also a rise in performance is to be expected, since it will be possible then to make sure that each module performs optimally in achieving the best performance of the whole framework.

The second characteristic that needs to be discussed here has already been addressed several times during the discussion, being the low amount of modules. One of the most important strengths of the framework is to correct erroneous classifications by means of the fusion process, which, as described, currently corrects 40 to 45% of these faulty classifications. In order for the fusion process to perform optimally, more modules are needed for inputting different classifications to correct more erroneous classifications.

9 Conclusions

In this chapter the conclusions of our research, based on the results and the discussion described in the previous chapters, is presented.

At the beginning of this work we presented the research question at hand,

Is it possible to, and if so, to what degree, perform automated sleep staging using solely unobtrusive measurements ?

In order to answer this question, we have developed a framework to perform automated sleep staging using solely unobtrusive measurements, being measurements during which the participant is not disturbed and no sensors are attached. This framework aims to overcome the most important drawbacks pertaining to the golden standard currently used. These are, firstly, the acquisition of sleep data in an obtrusive manner, and not in the natural sleep environment, hereby affecting the sleep of the person. And secondly, the fact that the manual scoring used in the golden standard leaves room for errors while also being time consuming and costly.

To overcome these drawbacks, the framework uses solely unobtrusively acquired biosignals, hereby overcoming the drawback of influencing the sleep during the measurement and making it possible to perform the measurements in the natural sleep environment. Furthermore, the framework performs sleep staging without the need of human intervention. This is accomplished by using automated classifiers, which are called modules in the framework. These modules perform a binary classification such as, for example, sleep/wake classification. With this approach the framework eliminates the possibility of human errors during the sleep staging and reduces the time and cost needed compared to the golden standard. Moreover, the modules used employ classical algorithms using rules grounded in physiology, which provide insight in the sleep stage classification process. The partial classifications of the different modules are fused by a fusion process we designed to obtain a full sleep stage classification.

The current implementation of this framework is comprised of three modules, two adapted existing classifiers and one newly developed classifier, the classifications of which are fused into a 3-step and 4-step hypnogram, indicating the sequence of sleep stages during the night. In the case of the 3-step hypnogram, wake, REM and NREM are given in the hypnogram and in the case of the 4-step hypnogram wake, REM, light sleep and deep sleep are given. This implementation was optimized and tested using a secondary analysis of an existing dataset.

The results of these tests indicate that the current implementation of the framework does not yet provides the desired results. The agreement with the golden standard of the framework is between 67% and 72% for the 3-step and 53% for the 4-step hypnogram, rendering the framework in need of improvement. This improvement is needed both on the level of the different modules, as well as on the level of the framework as a whole. Therefore the improvement must be sought both in the enhancement of the existing modules as well as in the addition of new modules, since more modules provide more information for the fusion process, making it possible to achieve a higher agreement with the golden standard. Also, it is important that the parameters used by the different modules will all be optimized simultaneously, to ensure that all modules contribute maximally to the agreement of the framework as a whole.

Nonetheless, the results of the current implementation of the framework are promising since these indicate that the approach taken with the framework, using partial classifiers in a framework to obtain full classification through a fusion process, has the potential to overcome drawbacks attached to the golden standard. Furthermore, besides the overcoming of these drawbacks, has the framework additional interesting properties, such as being very flexible through the use of modules and being compliant with many existing measurement set-ups.

Given these properties and approach of the framework, the framework is a tool with a lot of potential that can be used in the natural sleep environment of the sleeper, using a wide range of different measurement set-ups and without influencing or disturbing the sleep, hereby potentially lowering the threshold for (preliminary) sleep diagnostics, research and therapy. However, more research is needed to bring the agreement of the framework with the golden standard on a level to make the framework a reliable tool in these fields.

further research

As has become clear in the discussion above, further research needs to focus on getting the agreement of the framework with the golden standard on a higher level. There are three main recommendations to this further research.

First, to incorporate more modules in the framework, in order to gain more partial classifications that can be used by the fusion process to make a more robust full sleep stage classification. It is recommended that for these new modules models are selected which take different (combinations of) biosignals as input and utilise methods which have complementary metric results compared to the already existing modules within the current implementation of the framework. With complementary metric results here are meant results that strengthen each other in the fusion process. For example, two modules discriminating between sleep and wake, with the first module displaying an overestimation of sleep and the second module showing an overestimation of wake, strengthen each other in the fusion process because the erroneous classifications of both modules can be optimally corrected in the fusion process.

The second recommendation is to improve on the existing modules, in order to achieve higher agreement. To this end, the respective summaries at the end of the discussions of the different modules in chapter 8 can be used as a starting point for further research.

The final recommendation is that all modules need to be parametrized simultaneously, to ensure that all the modules contribute maximally to the agreement of the framework as whole with the golden standard. To this end most likely a heuristic optimization algorithm must be used to keep the time complexity of the optimization algorithm within reasonable boundaries, but it is part of the further research to determine the type of optimization algorithm.

Bibliography

- A. Sadeh, K. Sharkey, M. Carskadon. (1994). Activity-based sleep-wake identification: an empirical test of methodological issues. *Sleep*, 17(3), 201-207.
- Alexandra Piryatinskaa, Gyorgy Terdikb, Wojbor A. Woyczynskic, Kenneth A. Loparod, Mark S. Schere, Anatoly Zlotnikf. (2009). Automated detection of neonate EEG sleep stages. *Computer methods and programs in biomedicine*, 95.
- Appelberg, Pavlenko, Bergman, Rothen, Hedenstierna. (2007). Lung aeration during sleep. *Chest*, 131, 122-129.
- Aserinsky, Kleitman. (1953). Regularly occurring periods of eye mobility and concomitant phenomena during sleep.
- C.P. Pollak. (2001). How accurately does wrist body movement identify the states of sleep and wakefulness? *Sleep*, 24(8), 975-965.
- D. Vigo, J. Dominguez, S. Guinjoan, M. Scaramal, E. Ruffa, J. Solernó, L.Siri, D. Cardinali. (2009). Nonlinear analysis of heart rate variability within independent frequency components during the sleep-wake cycle. *Autonomic Neuroscience: Basic and Clinical*.
- Douglas, White, Pickett, Weil, Zwillich. (1982). Respiration during sleep in normal man. *Thorax*, 37, 840-844.
- G. Chung, B. Choi, J. Lee, J. Lee, D. Jeong, K. Park. (2009). REM sleep estimation using only respiratory dynamics. *Physiological Measurement*, 30, 1327-1340.
- G. Chung, B. Choi, K. Kim, Y. Lim, J. Choi, D. Jeong, K. Park. (2007). REM sleep classification with respiration rates. *6th International Special Topic Conference on ITAB, Tokyo*.
- I.M. Voronin, E.V. Biryukova. (2006). Heart rate variability in healthy humans during night sleep. *Human Physiology*, 3(32), 258-263.
- J. Hedner, G. Pillar, S. Pittman, D. Zou, L. Grote, D. White. (2004). A novel adaptive wrist body movement algorithm for sleep-wake assessment in sleep apnea patients. *Sleep*, 27(8), 1560-1566.
- J. Paquet, A. Kawinska, J. Carrier. (2007). Wake detection capacity of body movement during sleep. *Sleep*, 30(1), 1362-1369.
- L. Johnson, H. Kirchner, L. Rosen, A. Storfer-Isser, L. Cartar, S. Ancoli-Israel, J. Emancipator, A. Kibler, S. Redline. (2007). Sleep estimation using wrist body movement in adolescents with and without sleep-disordered respiration: a comparison of three data modes. *Sleep*, 30(7), 899-905.

- Michel Jouvet. (1959). The effect of electroconvulsive shocks in cats deprived of REM sleep. *Sleep*.
- P. Bušek, J. Vaňková, J. Opavský, J. Salinger, S. Nevšimalová. (2005). Spectral analysis of heart rate variability in sleep. *Physiological Research*, 54, 369-376.
- Padova, Barbato, Conte, Ficca. (2009). Diurnal variation of spontaneous eye blink rate in the elderly and its relationships with sleepiness and arousal. *Neuroscience letters*, 463(1), 40-43.
- Penzel, Riedl, Kantelhardt, Rostig, Glos, Suhrbier, Malberg, Fietze. (2007). Cardiovascular and respiratory dynamics during normal and pathological sleep. *Chaos*, 17.
- R. Cole, D. Kripke, W. Gruen, D. Mullaney, J. Gillin. (1992). Automatic Sleep/Wake identification from wrist activity. *Sleep*, 15(5), 461-469.
- Rechtschaffen, Kales. (1968). *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*.
- S. Ancoli-Israel, R. Cole, C. Alessi, M. Chambers, W. Moorcroft, P. Pollak. (2003). The role of body movement in the study of sleep and circadian rhythms. *Sleep*, 26(3), 342-392.
- Stephen J. Redmond, Conor Heneghan. (2006). Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea. *IEEE transactions on biomedical engineering*, 53(3).
- Takanori Sako, Naoto Burioka, Hisashi Suyama, Takashi Nomura, Takao Takeshima, Eiji Shimizu. (2001). Nonlinear behaviour of human respiration in sleep. *Chronobiology International*, 18(1), 71/83.
- Thorsten Schäfer, Marianne Schläfke. (2009). Respiratory changes associated with rapid eye movements in normo- and hypercapnia during sleep. *Journal of applied Physiology*, 85, 2213-2219.
- Virkkala. (2009). Automatic Sleep Stage Classification Using Electro-oculography. *Tampere University of Technology*.
- W. Tryon. (2004). Issues of validity in actigraphic sleep assessment. *Sleep*, 27(1), 158-165.
- Walter Karlen, Claudio Mattiussi, Dario Floreano. (2008). Improving actigraph sleep/wake classification with cardio-respiratory signals. *IEEE transactions*.
- Z. Shinar, A. Baharav, Y. Dagan, S. Akselrod. (2001). Automatic detection of slow-wave sleep using heart rate variability. *Computers in Cardiology*, (28), 593-596.
- Z. Zhuang, X. Gao, S. Gao. (2005). The relationship of HRV to sleep EEG and sleep rhythm. *International Journal of Neuroscience*, 115, 315-327.

Appendix A algorithm specification

In this appendix all the algorithms used by the framework are specified. This includes the algorithms underlying the modules and the fusion process as described in chapter 5 and 6 as well as the data pre-processing algorithms described in chapter 4. The algorithms are given in the R statistical language.

module one

```
function (bodymovement, pre, post, weightVector, scaling)
{
  # init vars
  totalTime <- length(bodymovement)
  classified <- discriminant <- rep(2, totalTime)

  # acigraphy sequence is enlarged with values equal to start and end to let algorithm use
  these values
  bodymovement <- c( rep(bodymovement[1],4), bodymovement, rep(bodymovement[totalTime],4) )

  # Sleep/Wake staging
  for(timepoint in 1:totalTime)
  {
    # determine discriminant sequence
    discriminant[timepoint] <- scaling * sum(( bodymovement[((timepoint-pre)+4):
      ((timepoint+post)+4)] * weightVector ))

    # fill classified sequence, 0 = sleep, 1 = wake
    if(discriminant[timepoint] < 1 ){ classified[timepoint] <- 0 }
    else{ classified[timepoint] <- 1 }
  }
  return(classified)
}
```

module two

```
function (respRate, offset)
{
  # init vars
  rrvA <- abs(respRate - rrSmoothed)
  rrvAsmoothed <- smoothing(rrvA, 30, 1)
  rrSmoothed <- smoothing(respRate, 30, 1)
  thresholdLin <- offset
  thresholdAdapInew <- smoothing(respRate, 300, 3) + offset
  thresholdAdapIIA <- smoothing(rrvA, 300, 1)
  totalEpochs <- length(rrSmoothed)
  predictedREMA <- rep(0, totalEpochs)

  # predict REM (in manner A), using variance as defined by Chung
  for(i in 1:totalEpochs)
  {
    if( (rrSmoothed[i] > thresholdAdapInew[i]) && (rrvAsmoothed[i] > thresholdLin) &&
      (rrvAsmoothed[i] > thresholdAdapIIA[i]) )
    {
      predictedREMA[i] <- 1
    }
  }

  # discard any REM epochs in first 90 epochs, physiologically impossible
  predictedREMA[1:90] <- 0
  # return classification
  return(predictedREMA)
}
```

module three

```
function ( bodymovement, respirationRate, startEpoch, endEpoch, pre, post, threshRR, threshRRV,
smoothing )
{
  # init vars
  totalEpochs <- length(hypnogram)
  zoneActi <- rep(0, ((endEpoch-startEpoch)+1) )
  step1Class <- rep(3, ((endEpoch-startEpoch)+1) ) # start out with all DS epochs
  zoneThresh <- 0 # threshold of activity permitted in zone around epoch
  totalLS <- totalClass <- correctClass <- 0

  # step I
  for(i in startEpoch:endEpoch)
  {
    # set window according to pre and post params, ensuring it doesn't exceed the given
    NREM epoch sequence
    start <- (i-pre)
    if(start < 1){ start <- 1 }
    stop <- (i+post)
    if(stop > totalEpochs){ stop <- totalEpochs }

    # determine zone activity
    zoneActi[(i-startEpoch)+1] <- sum( bodymovement[start:stop] )

    # based on zone activity and activity zone threshold eliminate deep sleep epoch
    candidates
    if(zoneActi[(i-startEpoch)+1] > zoneThresh){ step1Class[(i-startEpoch)+1] <- 1 }
  }

  # step II
  # init step vars
  step2Class <- step1Class
  respirationRateV <- computeVariability(respirationRate)
  smoothRR <- smoothing(respirationRate, smoothing, 3)
  smoothRRV <- smoothing(respirationRateV, smoothing, 3)

  # determining deep sleep based on RR and RRV ( conjunction )
  for(i in startEpoch:endEpoch)
  {
    if( ( respirationRate[i] < (smoothRR[i] + threshRR) ) & ( respirationRateV[i] >
(smoothRRV[i] + threshRRV) ) )
    {
      # if RR is high enough and RRV low enough, classify as LS
      step2Class[(i-startEpoch)+1] <- 1
    }
  }

  return(step2Class)
}
```

fusion module

```
function (h1, h2, group, hypnogram, subject)
{
  # given the classification hypnograms of ( in the current implementation ) module one, two
  # and three, and their respective metrics, return one final hypnogram

  # sanity check
  if( length(h1)!=length(h2) ){ return("input hypnograms not equal in lenght") }

  # initialise parameters
  metric_m1 <- read.csv("Location path\\moduleOne_metric.csv", header=TRUE)
  metric_m2 <- read.csv("Location path\\moduleTwo_metric.csv", header=TRUE)
  metric_m3 <- read.csv("Location path\\moduleThree_metric.csv", header=TRUE)

  T <- length(h1)
  hF <- c <- rep(-1, T)
  falseClass <- trueCorrected <- totalCorrected <- 0

  # currently, because module three can't process whole night epoch sequence, first M1 and
  # M2 fuse and the resulting NREM is given to M3
  for(i in 1:T)
  {
    if(h1[i]==1 & h2[i]==0)          # sleep and not REM -> OK -> NREM
    {
      hF[i] <- 3
      c[i] <- metric_m2[group,3]
    }
    else if(h1[i]==1 & h2[i]==5)     # sleep and REM -> OK -> REM
    {
      hF[i] <- 5
      c[i] <- metric_m2[group,2]
    }
    else if(h1[i]==6 & h2[i]==0)     # wake and not REM -> OK -> wake
    {
      hF[i] <- 6
      c[i] <- metric_m1[group,2]
    }
    else if(h1[i]==6 & h2[i]==5)     # wake and REM -> CONFLICT -> see metrics
    {
      totalCorrected <- totalCorrected + 1
      print(paste("found conflict at epoch", i), quote=FALSE)

      if( metric_m1[group,2] >= metric_m2[group,2] )
      {
        hF[i] <- 6
        c[i] <- metric_m1[group,2]
        if(hypnogram[i]==6){trueCorrected <- trueCorrected + 1}
        print(paste("module One has higher (positive) predicitive value for
        this participant (", metric_m1[group,2], "vs", metric_m2[group,2],
        ")"), quote=FALSE)
      }
      else
      {
        hF[i] <- 5
        c[i] <- metric_m2[group,2]
        if(hypnogram[i]==5){trueCorrected <- trueCorrected + 1}
        print(paste("module Two has higher (positive) predicitive value for
        this participant (", metric_m2[group,2], "vs", metric_m1[group,2],
        ")"), quote=FALSE)
      }
    }
    else{ print(paste("unrecognised epoch at index",i)) }
  }

  # at this point a three step hypnogram is constructed, of which the stats will be
  # determined
}
```

```

# first transform hypnogram to three step hypnogram
threeStepHypno <- hypnogram
threeStepHypno[hypnogram==1 | hypnogram==2 | hypnogram ==4] <- 3

# then determine the stats
# first, create confusionMatrix
confM <- matrix(rep(0,16),4,4)
for(i in 1:T)
{
  if(hF[i]!=threeStepHypno[i] & threeStepHypno[i]!=0){ falseClass <- falseClass + 1 }

  if(hF[i]==3)
  {
    if(threeStepHypno[i]==3){ confM[1,1] <- confM[1,1] + 1 }
    if(threeStepHypno[i]==5){ confM[1,2] <- confM[1,2] + 1 }
    if(threeStepHypno[i]==6){ confM[1,3] <- confM[1,3] + 1 }
  }
  if(hF[i]==5)
  {
    if(threeStepHypno[i]==3){ confM[2,1] <- confM[2,1] + 1 }
    if(threeStepHypno[i]==5){ confM[2,2] <- confM[2,2] + 1 }
    if(threeStepHypno[i]==6){ confM[2,3] <- confM[2,3] + 1 }
  }
  if(hF[i]==6)
  {
    if(threeStepHypno[i]==3){ confM[3,1] <- confM[3,1] + 1 }
    if(threeStepHypno[i]==5){ confM[3,2] <- confM[3,2] + 1 }
    if(threeStepHypno[i]==6){ confM[3,3] <- confM[3,3] + 1 }
  }
}

# sumation
confM[4,1] <- sum(confM[,1])
confM[4,2] <- sum(confM[,2])
confM[4,3] <- sum(confM[,3])
confM[1,4] <- sum(confM[1,])
confM[2,4] <- sum(confM[2,])
confM[3,4] <- sum(confM[3,])
confM[4,4] <- sum(confM[,4])

rownames(confM) <- c("NREM","REM","wake","total")
colnames(confM) <- c("NREM","REM","wake","total")
write.csv(confM, paste(subject,".csv", sep="" )

print(subject)
print(confM)
return(confM)

print("3-STEP HYPNOGRAM,", quote=FALSE)
print(paste("true corrections made / total number of corrections made / percentage true
corrected", trueCorrected, "/", totalCorrected, "/", (trueCorrected/totalCorrected)*100 ),
quote=FALSE)
print(paste("final three step hypnogram has agreement of", 1-(falseClass/T), "with the
golden standard"))

# print hypnogram
#ts.plot( ts(hF), ts(c), ts(threeStepHypno)+5, col=c("blue","red","green"),
gpars=list(t="s", xlab=paste("
",subject), ylab="confidence
golden standard" )
classified
time [30s]

# combine 3 step hypnogram with confidence
return( cbind(hF,c) )
}

```


peak-to-peak algorithm

```
function (dataVector, sampleRate, halfWindowSize)
{
  # using different method compared to determineRR, classifies a timepoint as a local
  # maximum when it's value is no longer ascending compared to it's neighbours

  # init vars
  size <- length(dataVector)
  ascending <- 0
  ascendingVector <- c(1, rep(0,(size-1)) )
  subString <- c(rep(1,halfWindowSize),rep(0,(halfWindowSize/2)))
  maximaVector <- rep(0,size)
  epochVector <- rep(0, (size%/(30*sampleRate)) )

  # iterate through dataVector and populate ascendingVector
  for(i in 2:size)
  {
    # when ascending, populate with 1, otherwise leave 0
    if(dataVector[i] > dataVector[i-1]){ascendingVector[i] <- 1}
  }

  # iterate through ascendingVector and determine substrings of halfWindowSize 1's followed
  # by halfWindowSize 0's(maxima)
  for(i in 1:(size-(2*halfWindowSize)) )
  {
    if(sum(ascendingVector[i:(i+length(subString)-1)]==subString)==length(subString))
    {
      maximaVector[(i+halfWindowSize)-1] <- dataVector[(i+halfWindowSize)-1]
    }
  }

  # populate epochVector by determining mean RR for each 30s epoch
  for(i in 1:length(epochVector) )
  {
    tmpSum <- 0
    tmpVector <- c(((i*30*sampleRate)-((30*sampleRate)-1)):(i*30*sampleRate))
      [maximaVector[((i*30*sampleRate)-((30*sampleRate)-1)):
        (i*30*sampleRate)]>0]

    # in case max one maxima was present in maximaVector,
    if(length(tmpVector)<=1)
    {
      epochVector[i] <- 0
    }
    else
    {
      for(j in 1:(length(tmpVector)-1))
      {
        tmpSum <- tmpSum + (tmpVector[j+1] - tmpVector[j])
      }

      # save mean RR in bpm
      epochVector[i] <- (60 / ((tmpSum / (length(tmpVector)-1)) / sampleRate))
    }
  }

  return(epochVector)
}
```

body movement-artefact correction algorithm

```
function (signalRR, signalActi, width)
{
  # init vars
  ectopicPeaks <- actiEpochs <- 0
  count <- index <- 1
  signalRR_rev <- signalRR

  # first locate the body movement peaks
  peaksActi <- c(1:length(signalActi))[signalActi>0]

  # and cluster them if occurring in subsequent epochs
  while(index <= length(peaksActi) )
  {
    # as long as subsequent epochs have peaks, keep counting
    while(peaksActi[index+count] == peaksActi[index]+count && (index+count) <=
length(peaksActi) )
    { count <- count + 1 }

    # and store the start and length of these subsequent peak epochs
    actiEpochs <- c(actiEpochs, peaksActi[index], count)

    # reset vars
    index <- index + (count)
    count <- 1
  }
  # cut bogus start
  actiEpochs <- actiEpochs[2:length(actiEpochs)]

  # secondly check if there are ectopic peaks or falls in the signalRR trace due to these
body movement peaks
  for(i in seq(1, length(actiEpochs), 2) )
  {
    # construct window in which RR ectopic peaks or falls can occur, set at window of 5
epochs preceding and 10 epochs following the body movement epoch window because of
possible phase shift between the signals and lasting effects of an body movement
peak
    start <- actiEpochs[i]-5
    end <- actiEpochs[i] + actiEpochs[i+1] + 5
    if(start < 1) { start <- 1}
    if(end > length(signalRR) ){ end <- length(signalRR) }

    ectopicWindow <- seq( start, end, 1)

    # to search for possible ectopic peaks in the ectopicWindow, the overall mean and
sd has to be determined around the ectopicWindow
    startMean <- start - 5
    endMean <- end + 5
    if(startMean < 1) { startMean <- 1}
    if(endMean > length(signalRR) ){ endMean <- length(signalRR) }

    preMean <- mean( signalRR[startMean:start] )
    postMean <- mean( signalRR[end:endMean] )
    preSd <- sd( signalRR[startMean:start] )
    postSd <- sd( signalRR[end:endMean] )

    # after which the mean and sd are determined by averaging for epochs preceding and
following the ectopic Window
    epochMean <- (preMean + postMean) / 2
    if(length(signalRR[end:endMean])==1){ epochSd <- preSd }
    else if(length(signalRR[startMean:start])==1){ epochSd <- postSd }
    else{ epochSd <- ( (preSd + postSd) / 2 ) }
  }
}
```

```

# if there exist peaks or falls in the ectopic window which are more than 2 times
the sd away from the mean, they are marked as ectopic peaks
if( sum(signalRR[ectopicWindow] > (epochMean + (2.5 * epochSd))) >= 1 )
{
  peaks <- ectopicWindow[signalRR[ectopicWindow] > (epochMean + (2 *
epochSd))]
  ectopicPeaks <- c(ectopicPeaks, peaks)
}

if( sum(signalRR[ectopicWindow] < (epochMean - (2.5 * epochSd))) >= 1 )
{
  peaks <- ectopicWindow[signalRR[ectopicWindow] < (epochMean - (2 *
epochSd))]
  ectopicPeaks <- c(ectopicPeaks, peaks)
}
}
tmp <- ectopicPeaks[2:length(ectopicPeaks)]
ectopicPeaks <- rep(10, length(signalRR) )
ectopicPeaks[tmp] <- signalRR[tmp]

# attenuate the ectopic peaks in signalRR at epoch for the neighbouring epochs equal to
width
signalRR_smooth <- smoothing(signalRR, 20)
signalRR_rev <- signalRR
signalRR_rev[tmp] <- signalRR_smooth[tmp]

return(signalRR_rev)
}

```

Appendix B individual participant classifications

In this appendix all the classifications made by the modules and the framework as a whole for the different participants are given. The legend for the diagrams depicting the classifications are given in the respective legends on the bottom of the pages.

module one

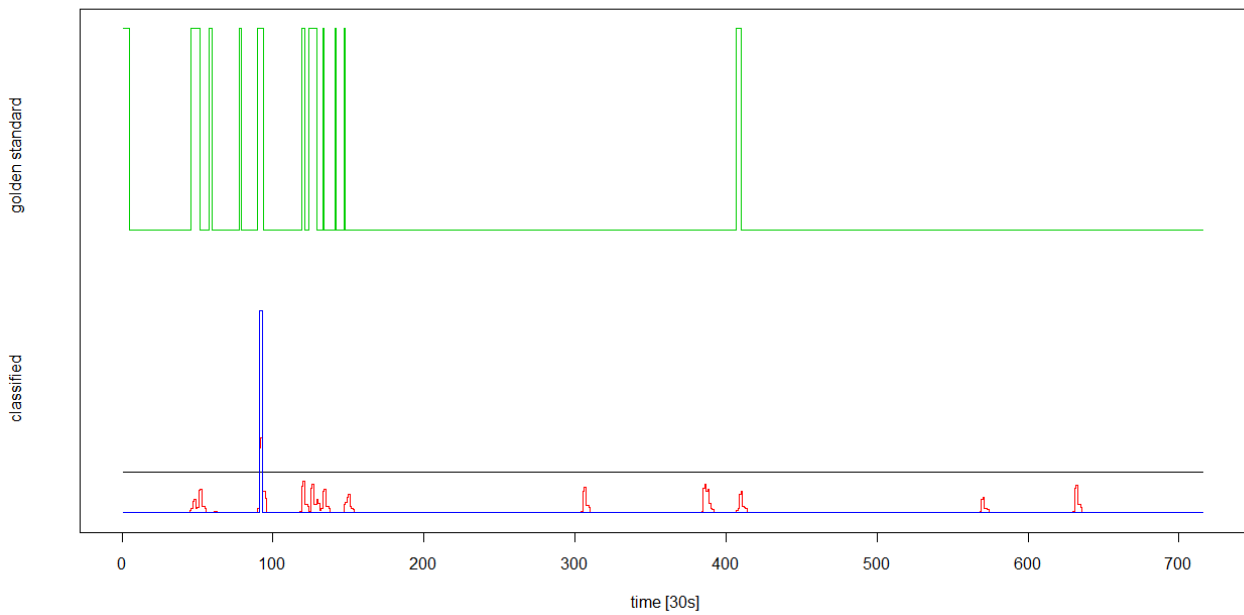


Diagram 11: Sleep/wake classification by module one for participant 101. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.



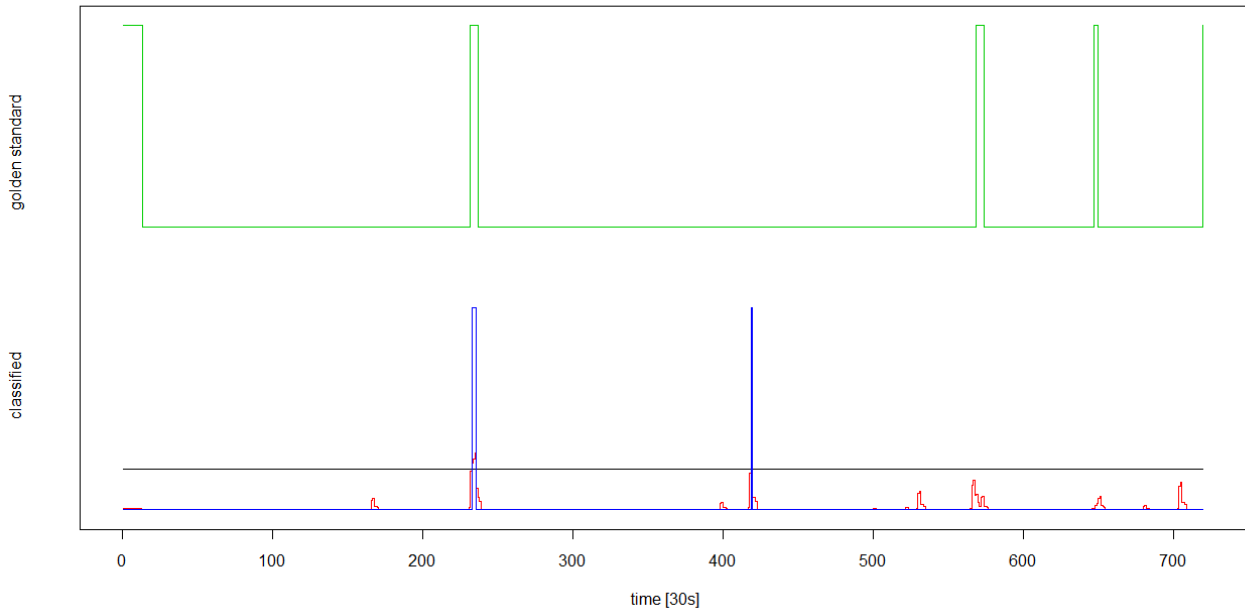


Diagram 12: Sleep/wake classification by module one for participant 102. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.

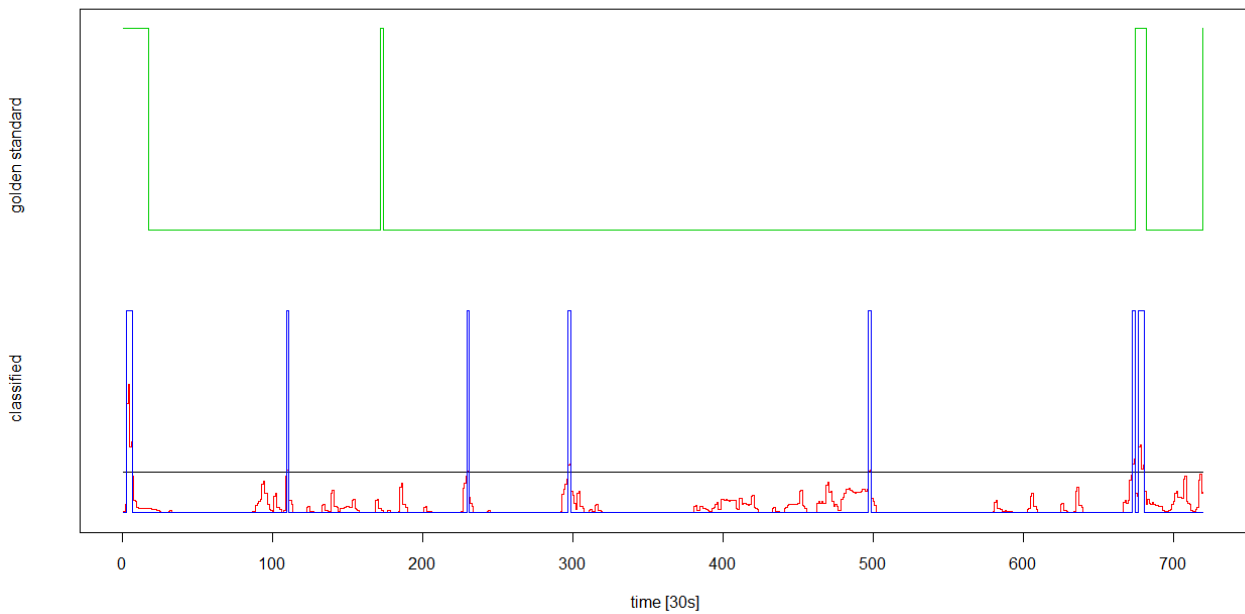


Diagram 13: Sleep/wake classification by module one for participant 103. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.



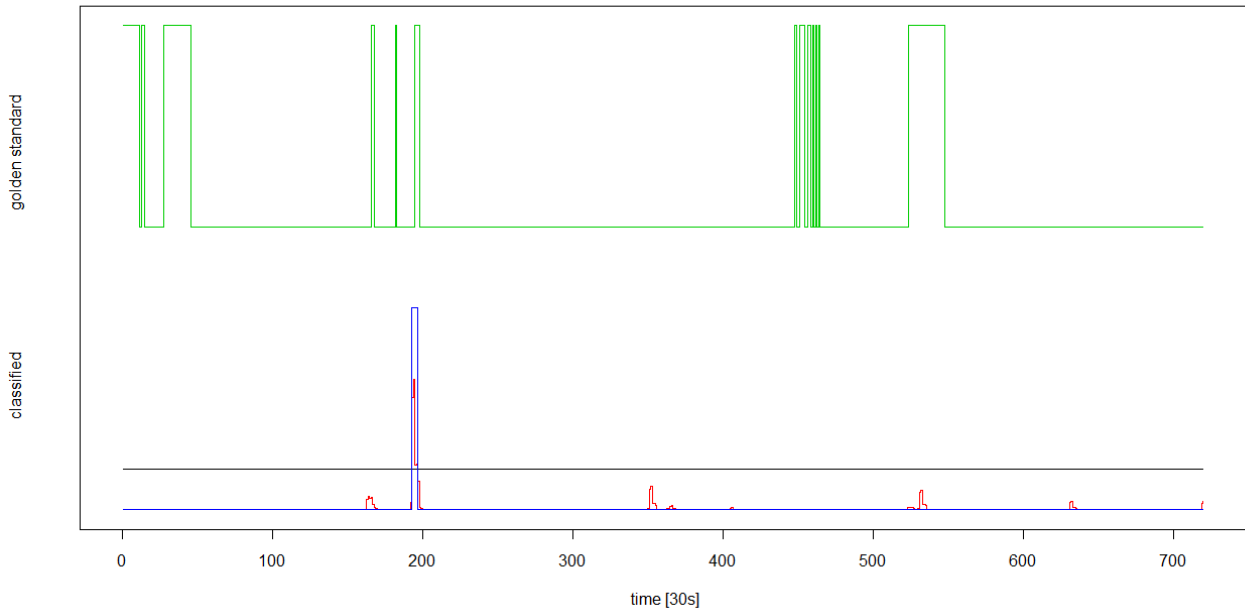


Diagram 14: Sleep/wake classification by module one for participant 104. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.

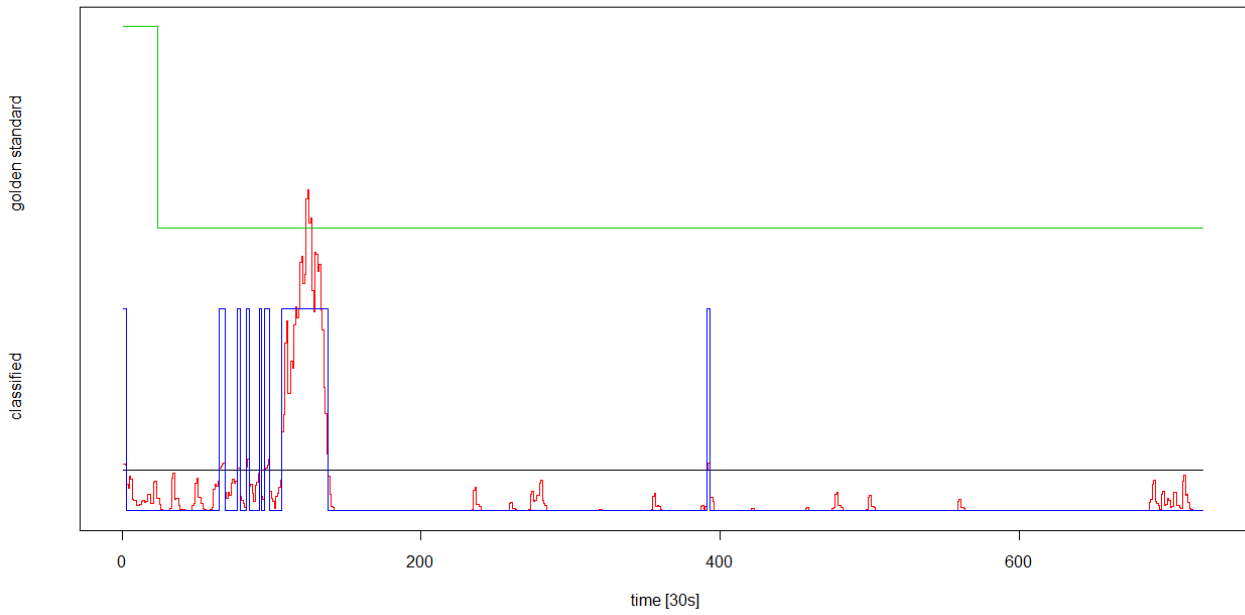
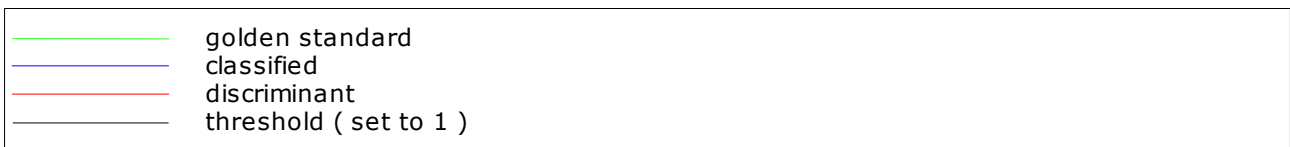


Diagram 15: Sleep/wake classification by module one for participant 106. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.



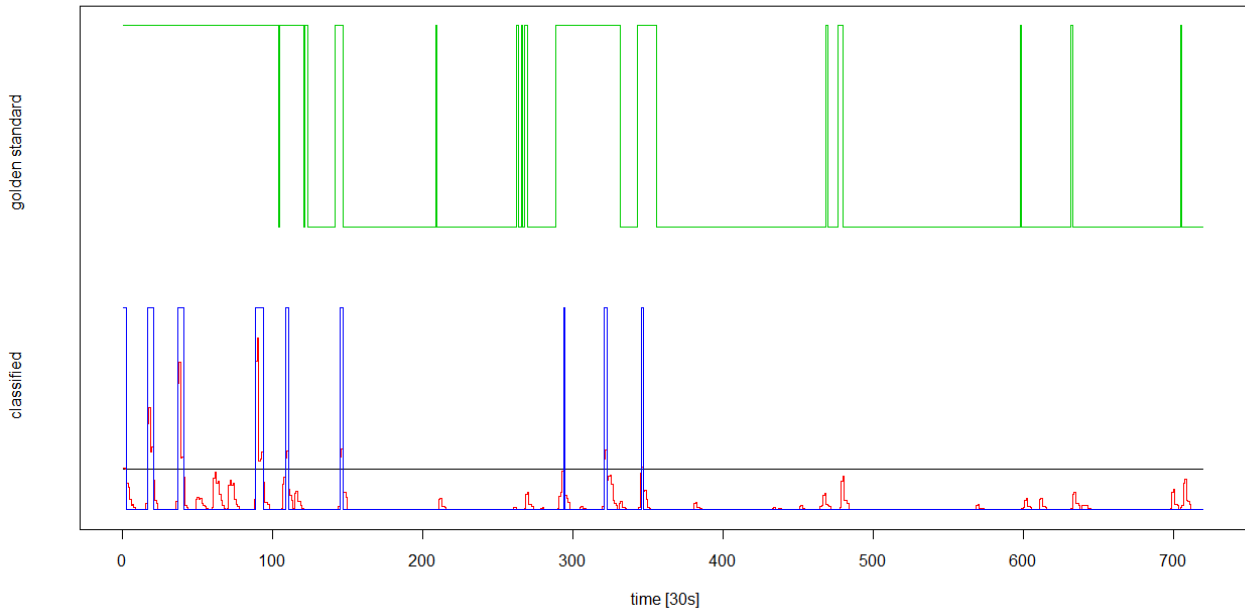


Diagram 16: Sleep/wake classification by module one for participant 108. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.

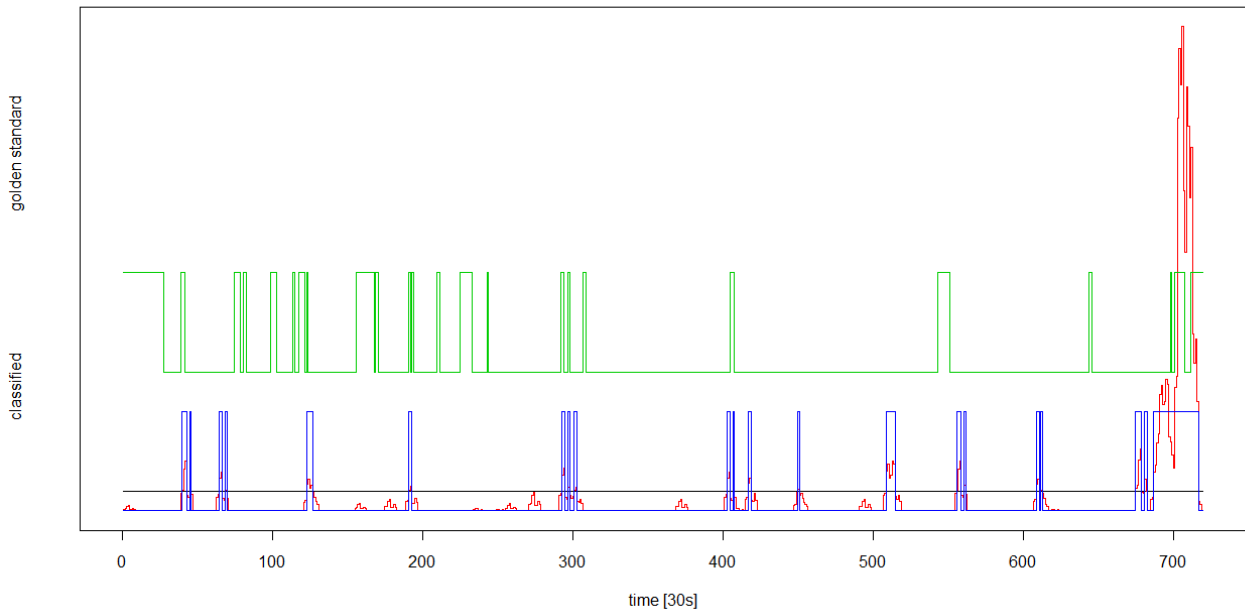


Diagram 17: Sleep/wake classification by module one for participant 201. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.



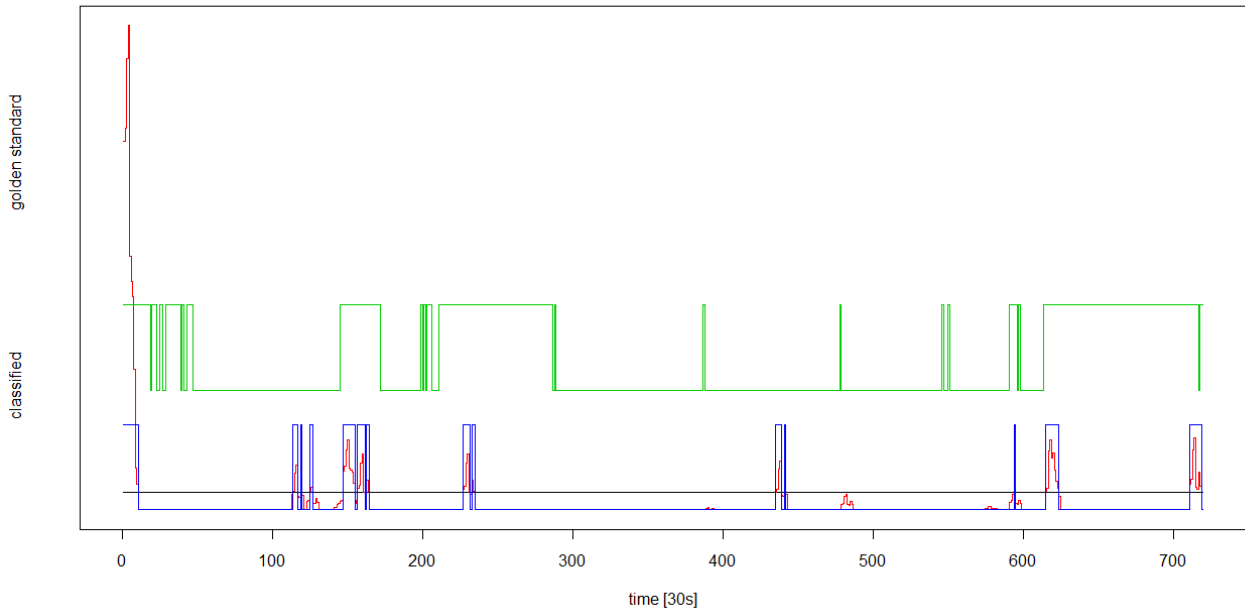


Diagram 18: Sleep/wake classification by module one for participant 202. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.

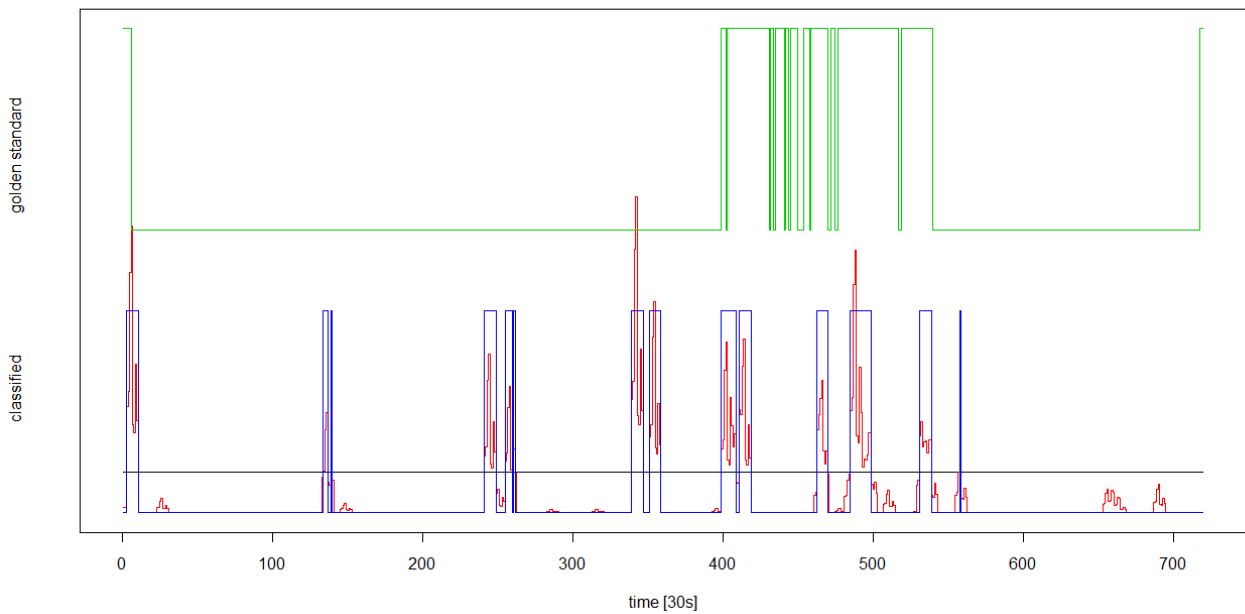


Diagram 19: Sleep/wake classification by module one for participant 203. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.



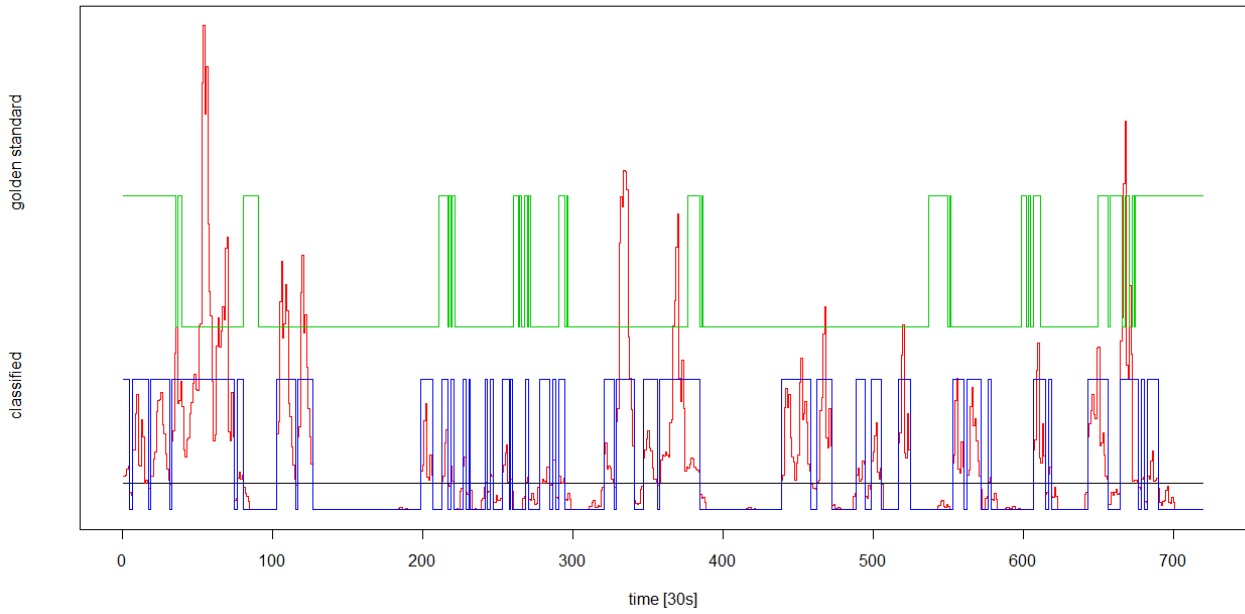


Diagram 20: Sleep/wake classification by module one for participant 204. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.

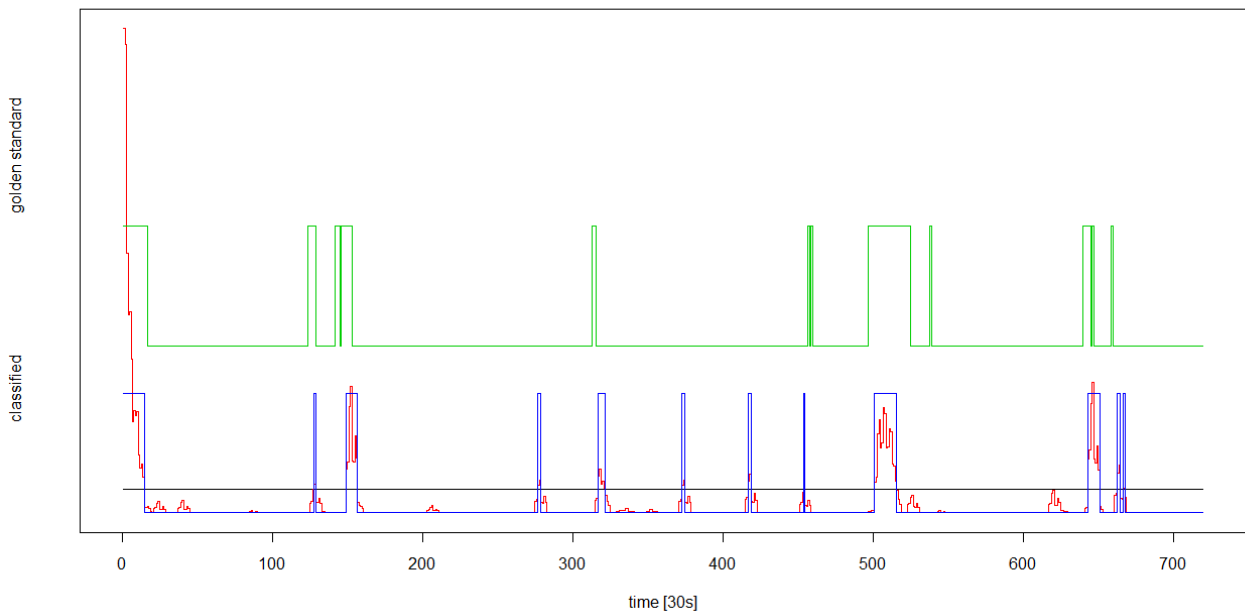


Diagram 21: Sleep/wake classification by module one for participant 205. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.



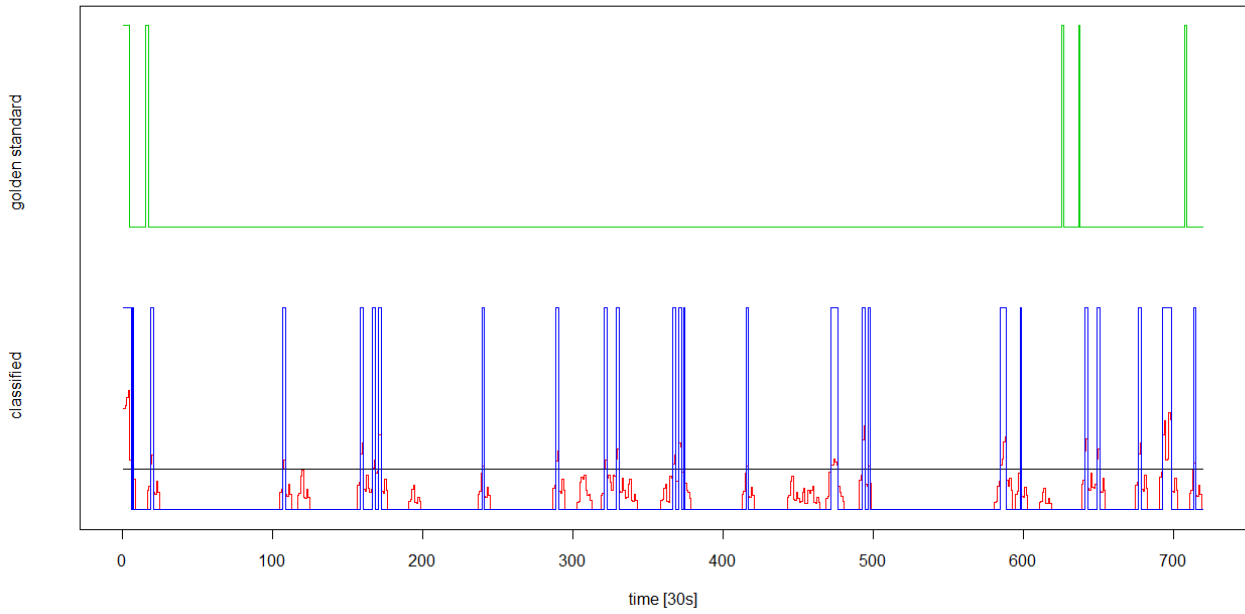


Diagram 22: Sleep/wake classification by module one for participant 207. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.

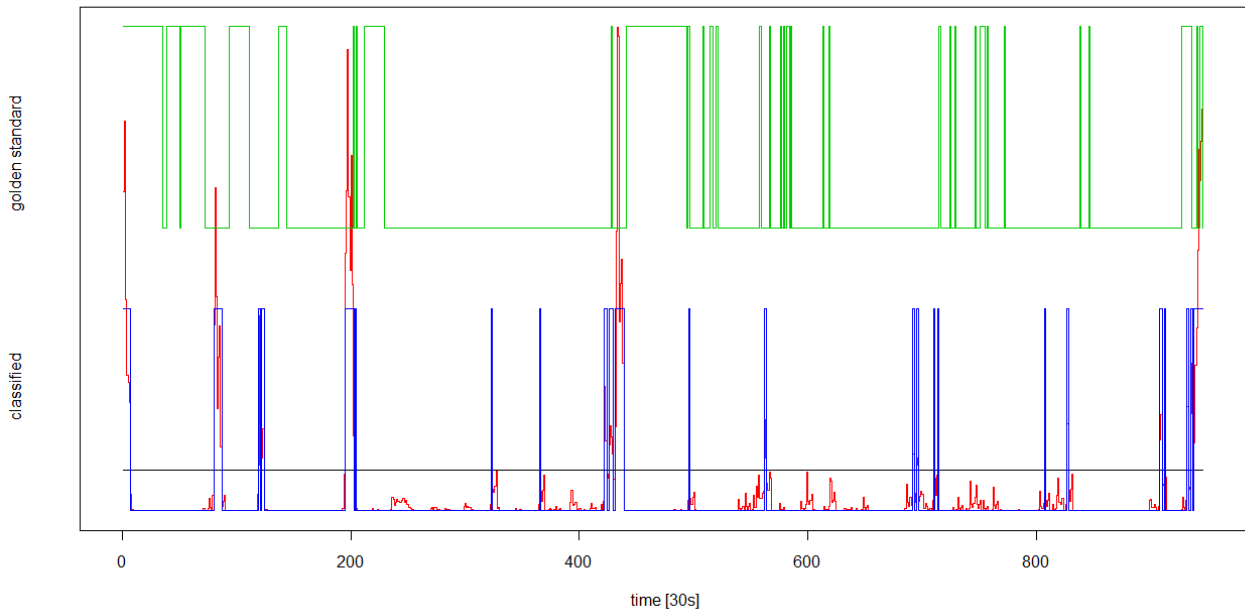


Diagram 23: Sleep/wake classification by module one for participant 301. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.



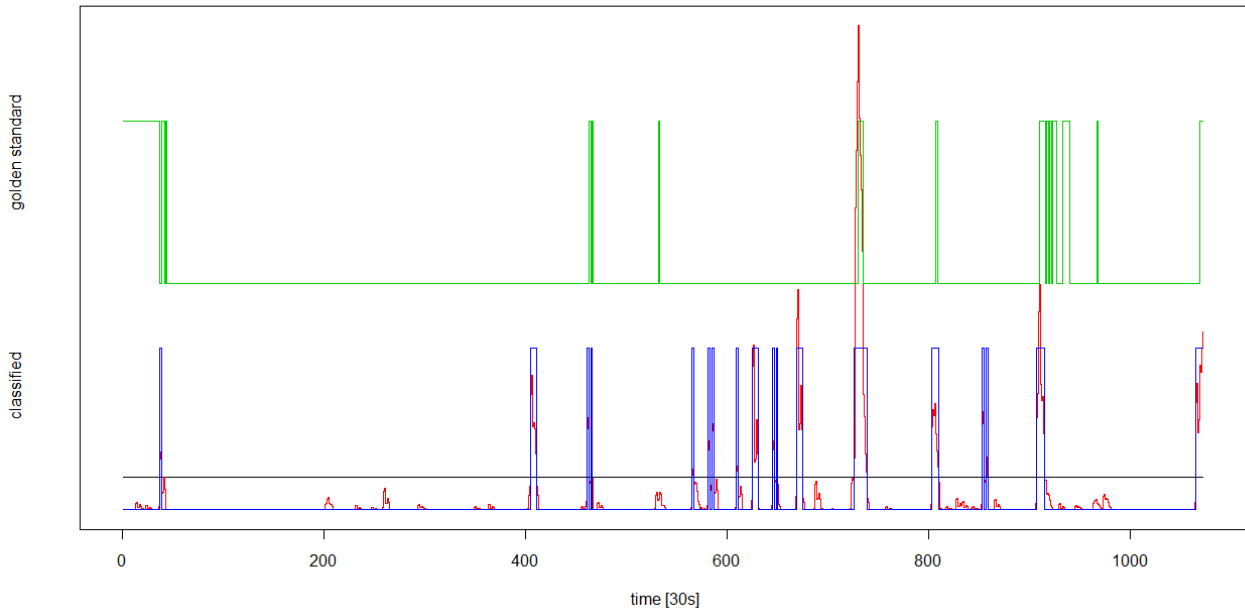


Diagram 24: Sleep/wake classification by module one for participant 302. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.

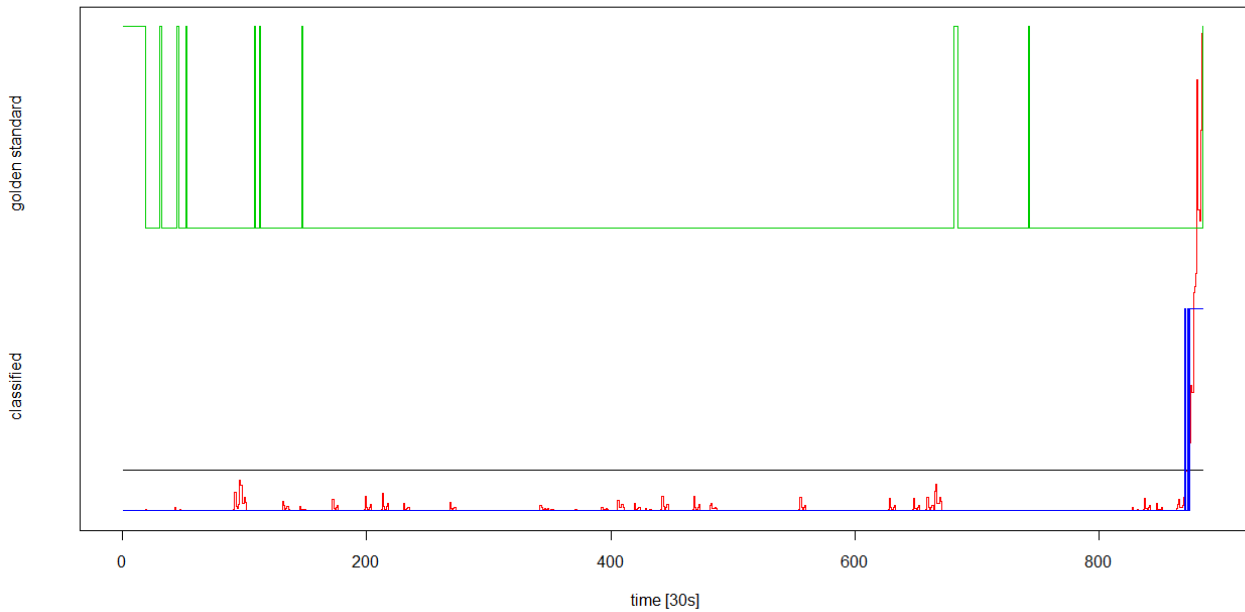


Diagram 25: Sleep/wake classification by module one for participant 303. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.



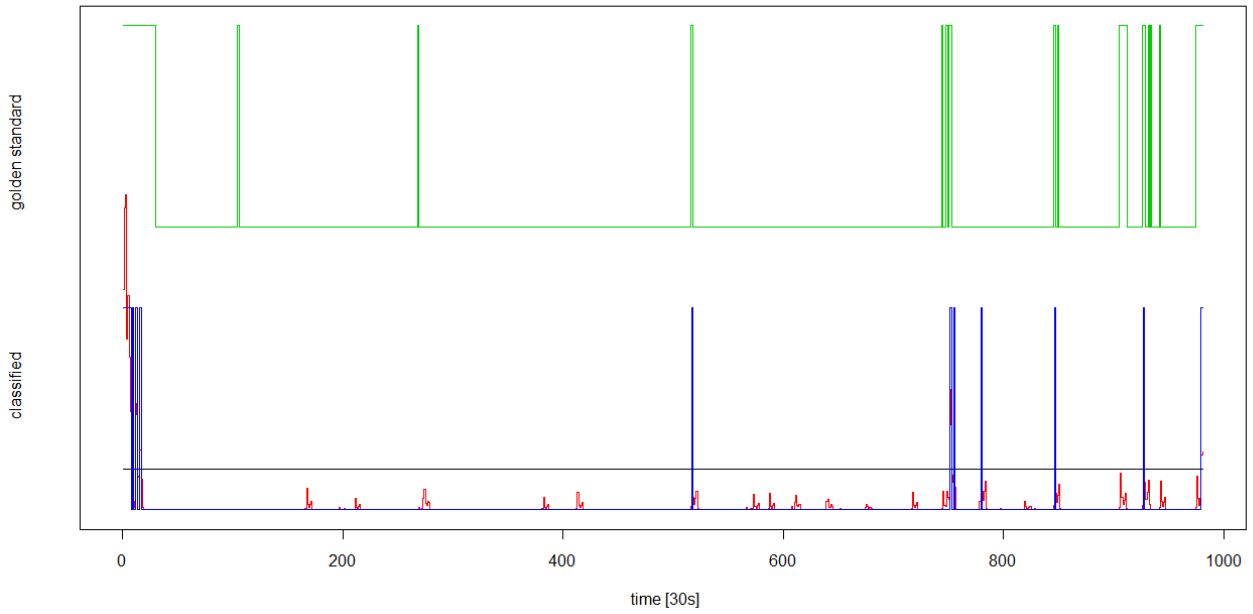


Diagram 26: Sleep/wake classification by module one for participant 304. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.

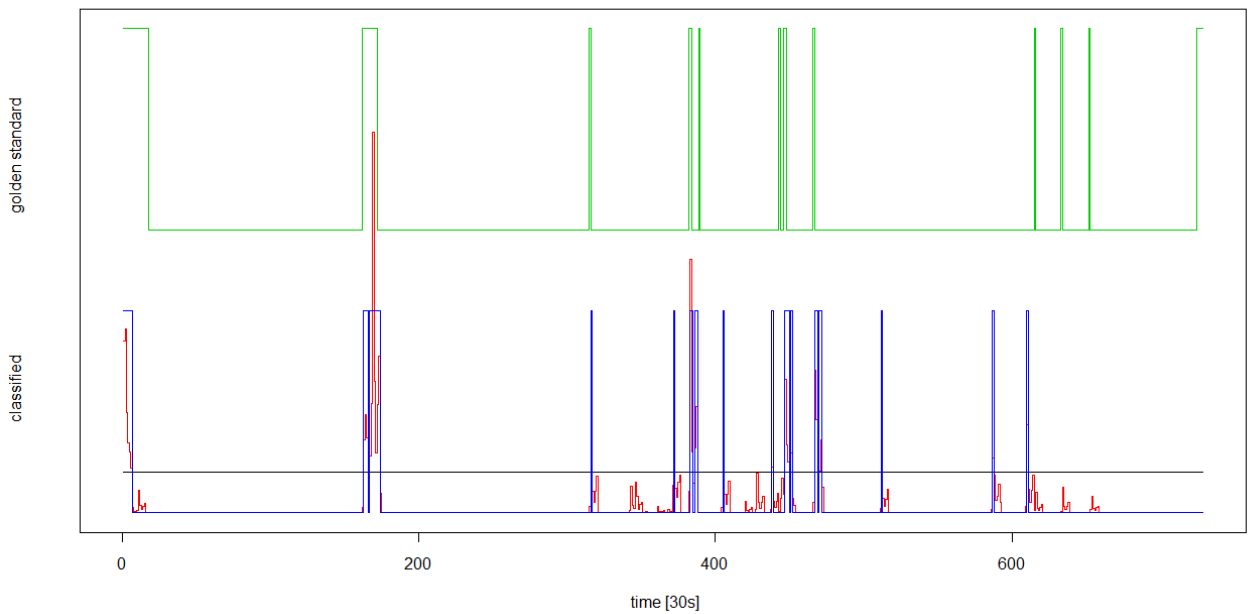


Diagram 27: Sleep/wake classification by module one for participant 305. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.



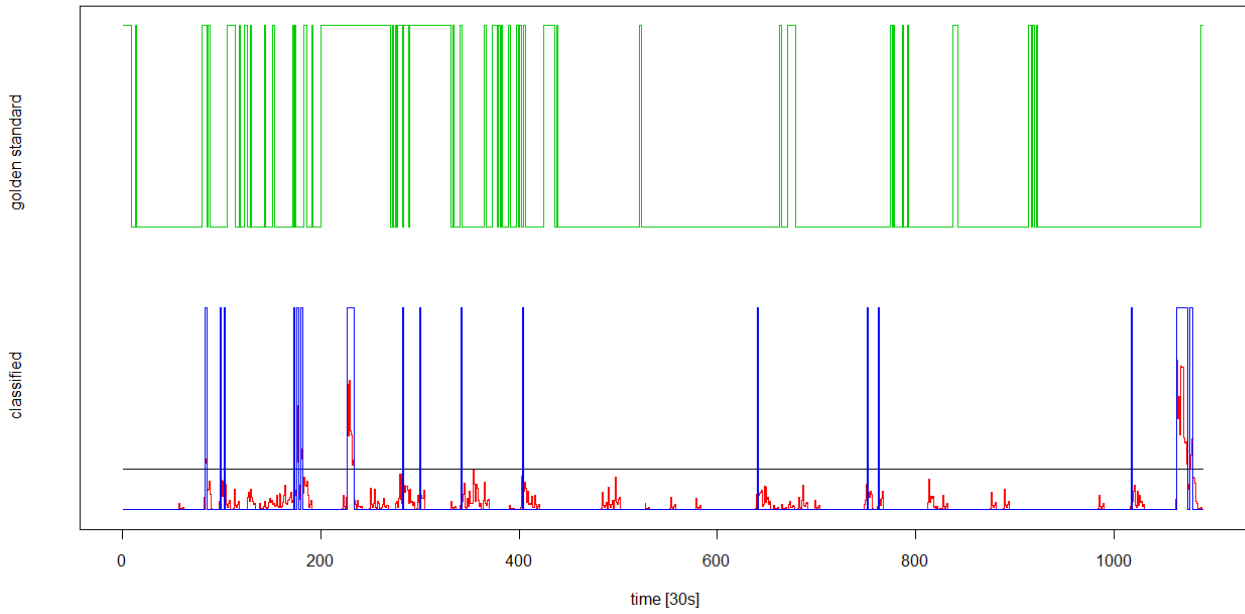


Diagram 28: Sleep/wake classification by module one for participant 306. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.

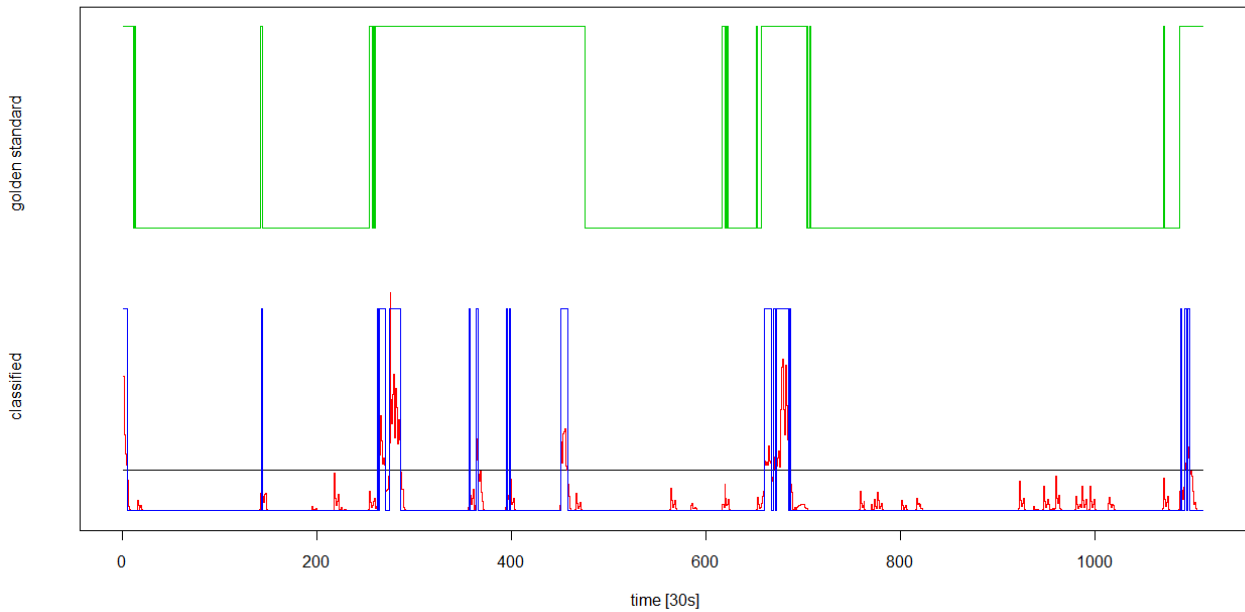


Diagram 29: Sleep/wake classification by module one for participant 307. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.



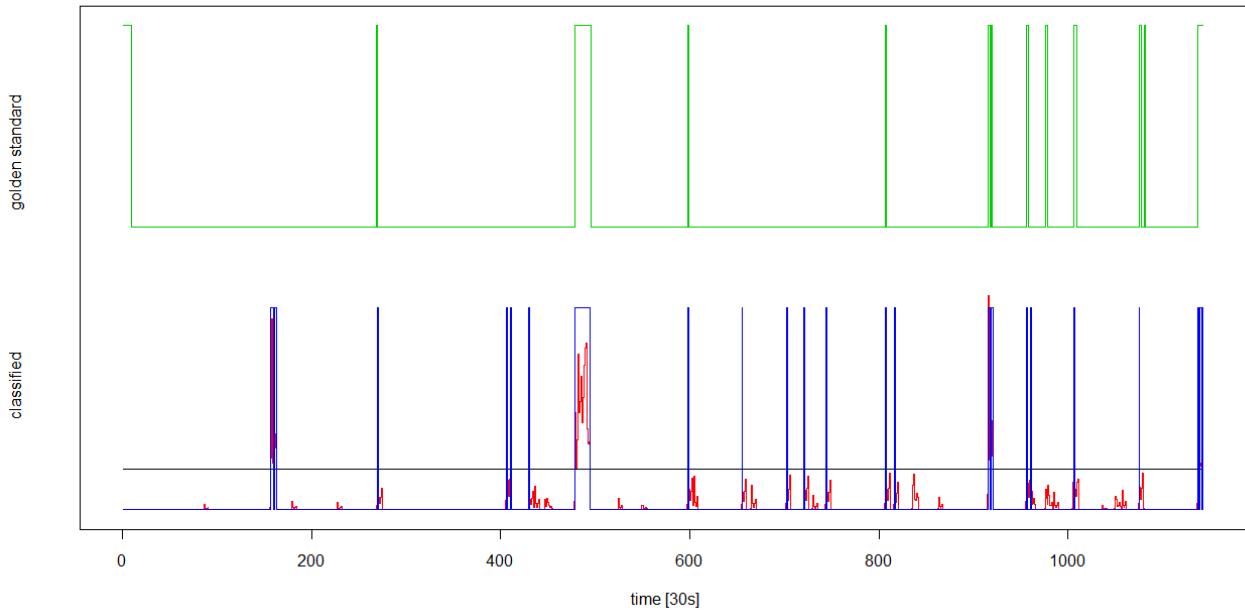


Diagram 30: Sleep/wake classification by module one for participant 308. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.

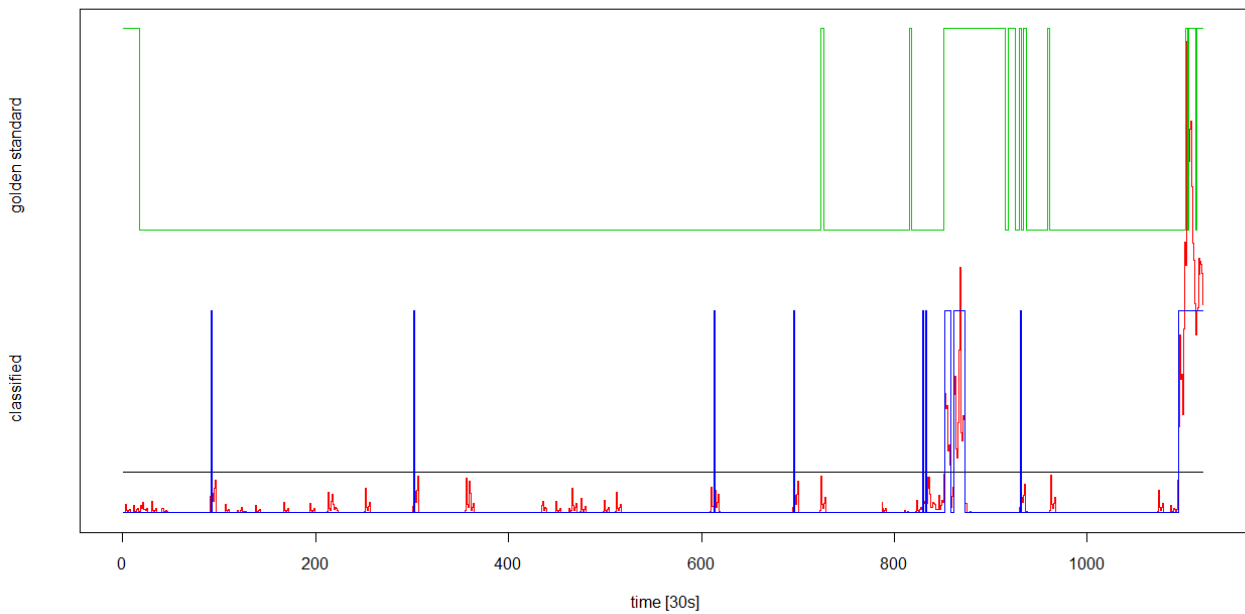


Diagram 31: Sleep/wake classification by module one for participant 309. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.



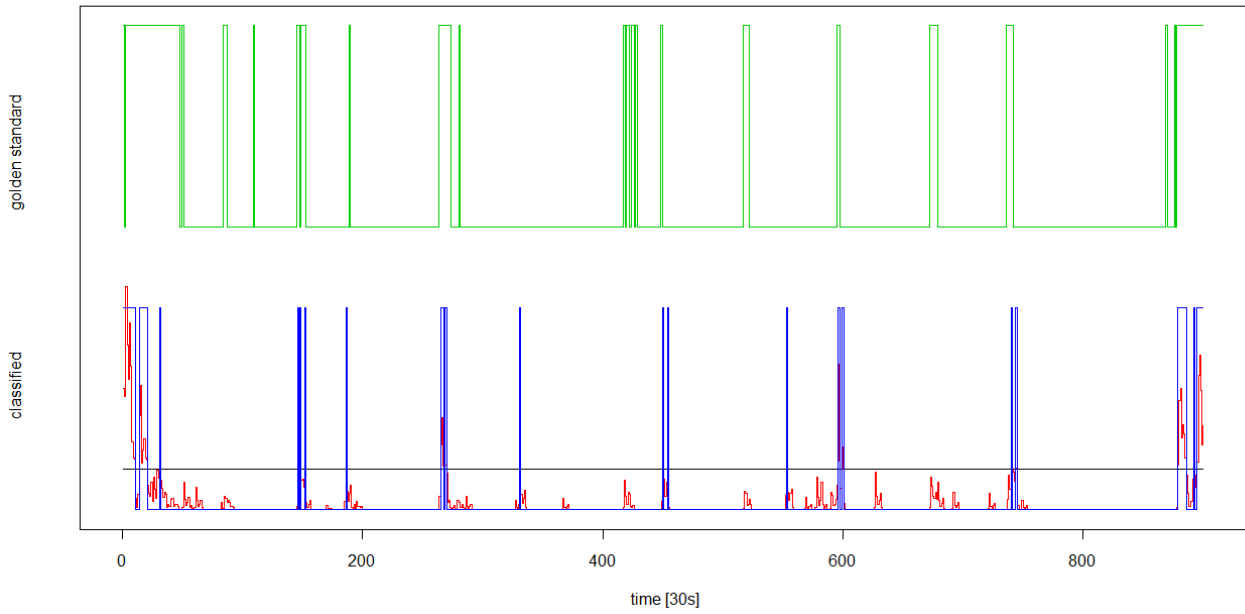


Diagram 32: Sleep/wake classification by module one for participant 310. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.

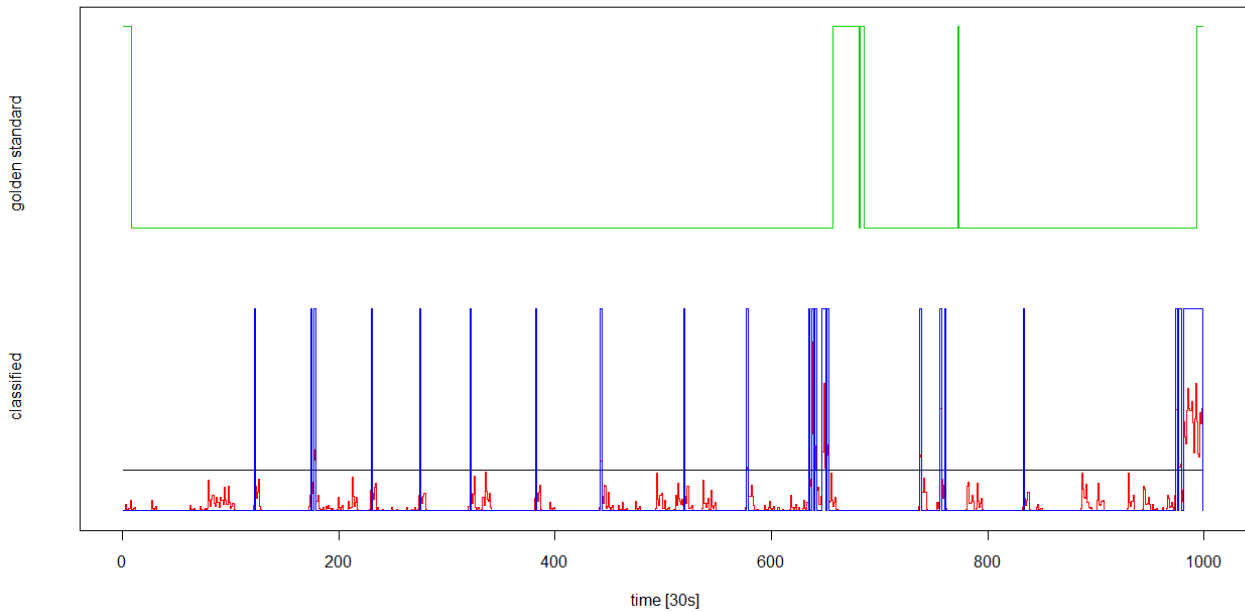


Diagram 33: Sleep/wake classification by module one for participant 311. Bars indicate wake epochs, no bar indicates sleep epochs. Discriminant is computed as described in chapter 5, if this discriminant exceeds the threshold, the epoch is classified as wake.



module two

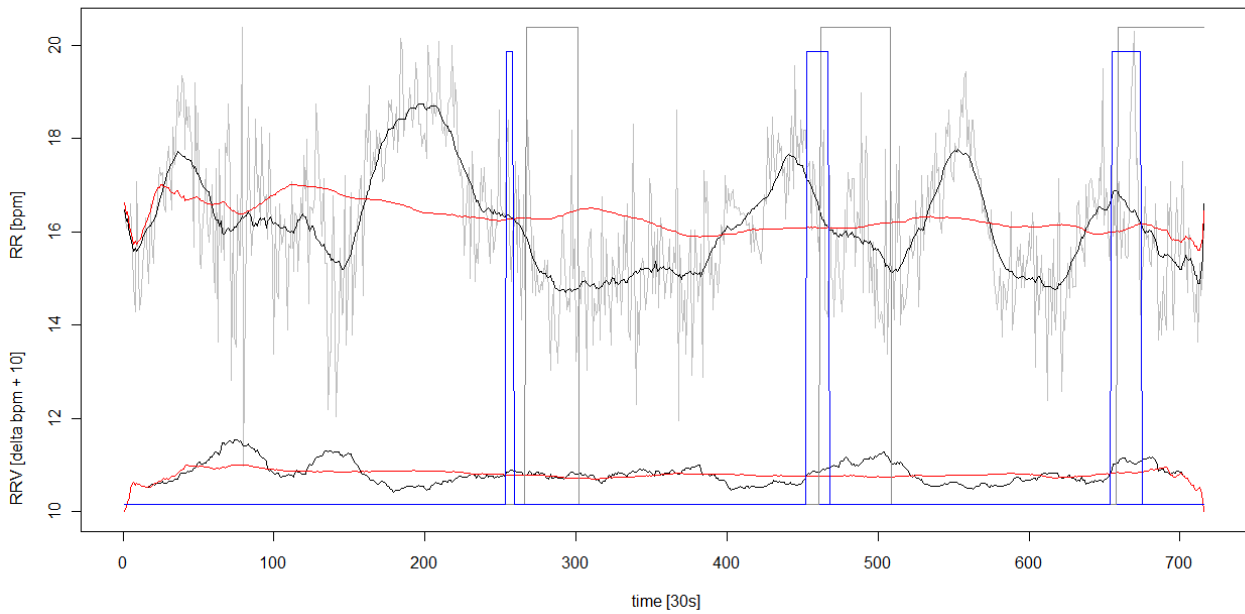
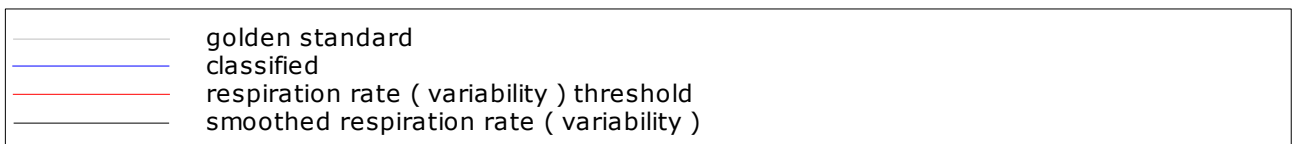


Diagram 34: REM sleep classification by module two for participant 101. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.



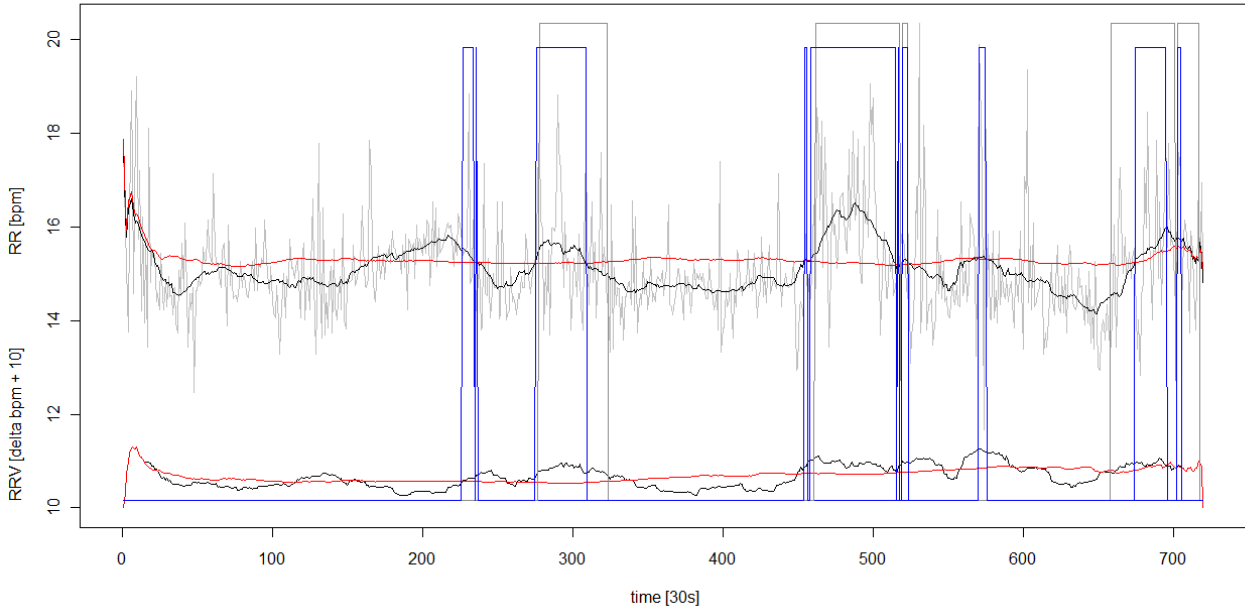


Diagram 35: REM sleep classification by module two for participant 102. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.

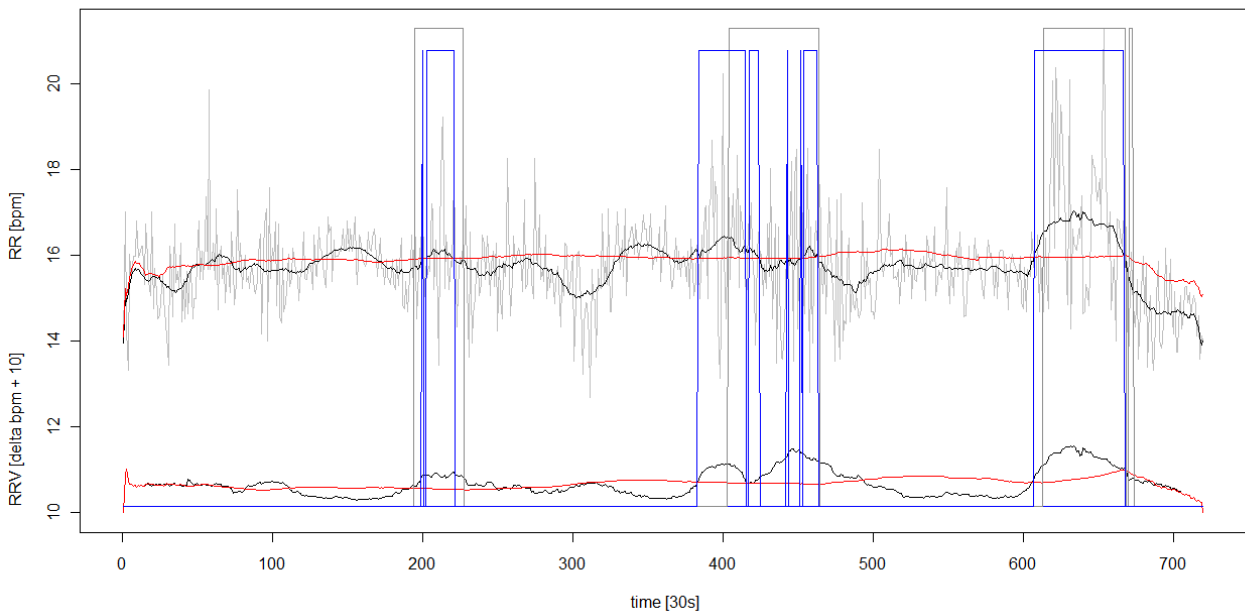
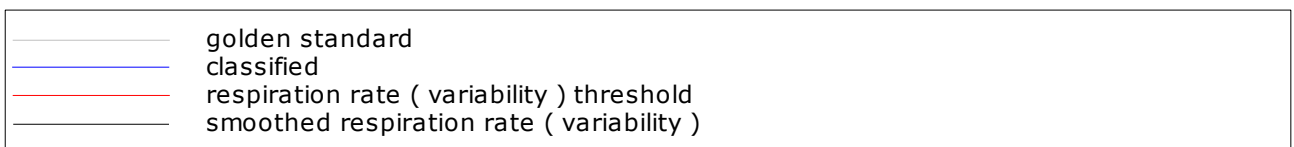


Diagram 36: REM sleep classification by module two for participant 103. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.



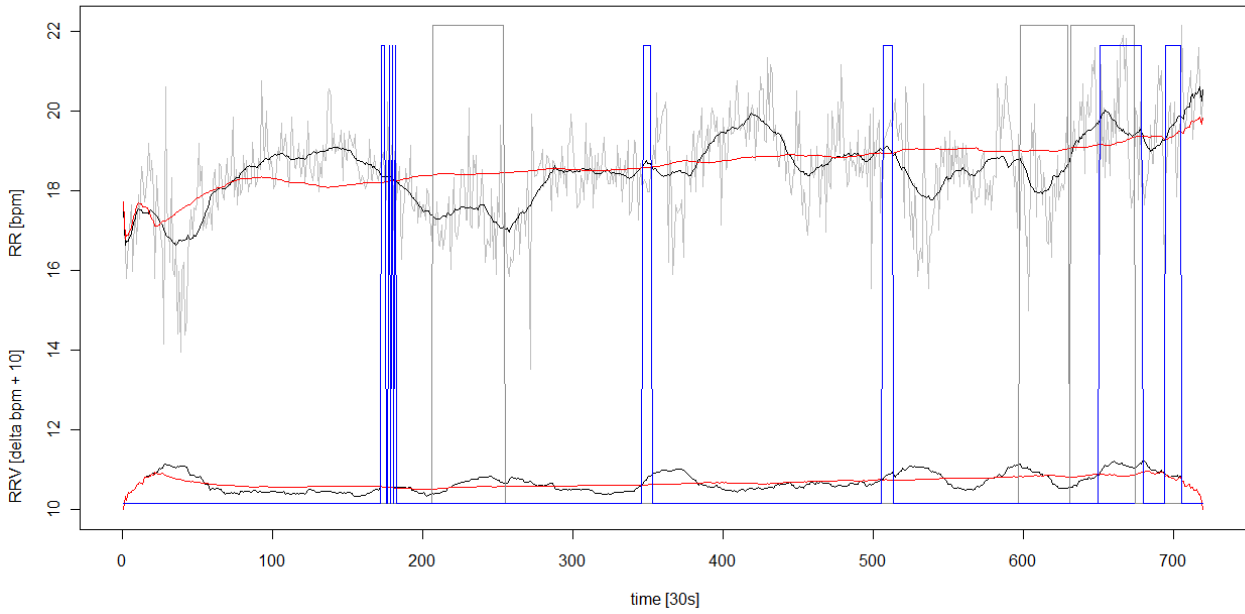


Diagram 37: REM sleep classification by module two for participant 104. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.

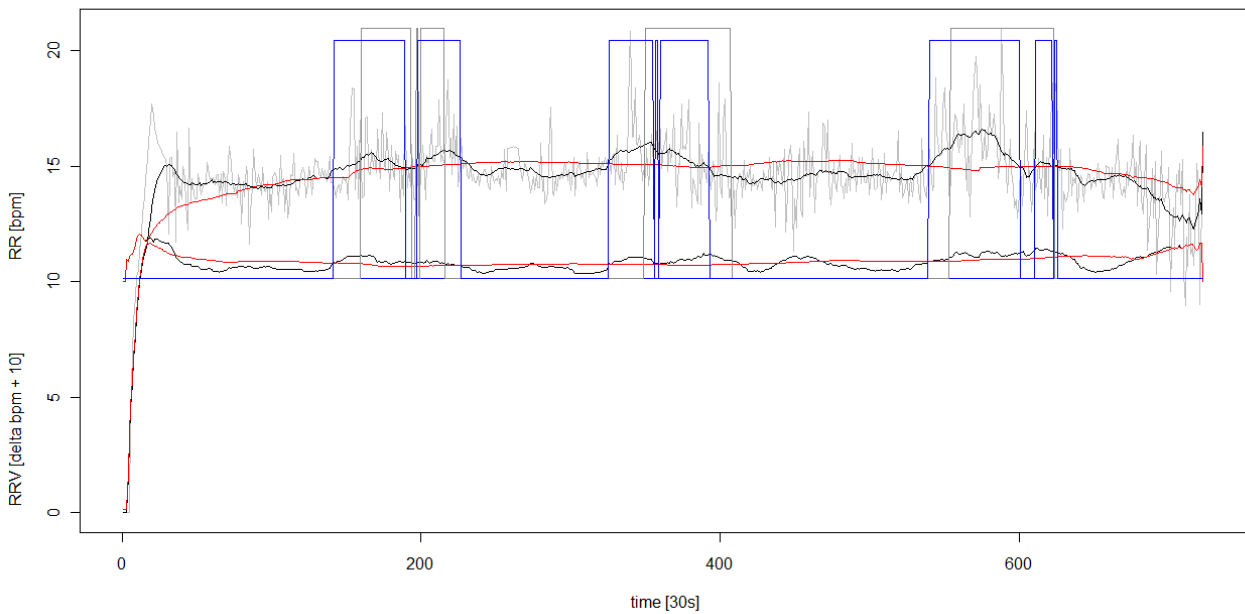
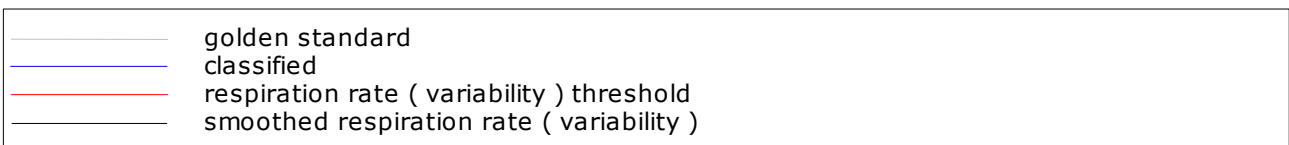


Diagram 38: REM sleep classification by module two for participant 106. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.



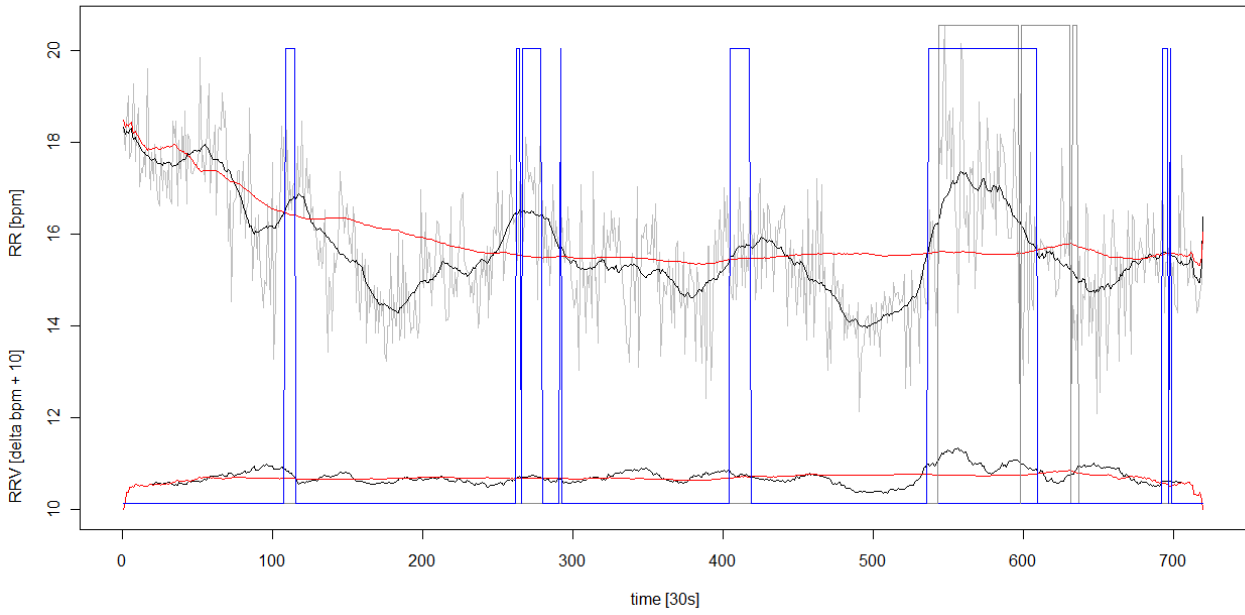


Diagram 39: REM sleep classification by module two for participant 108. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.

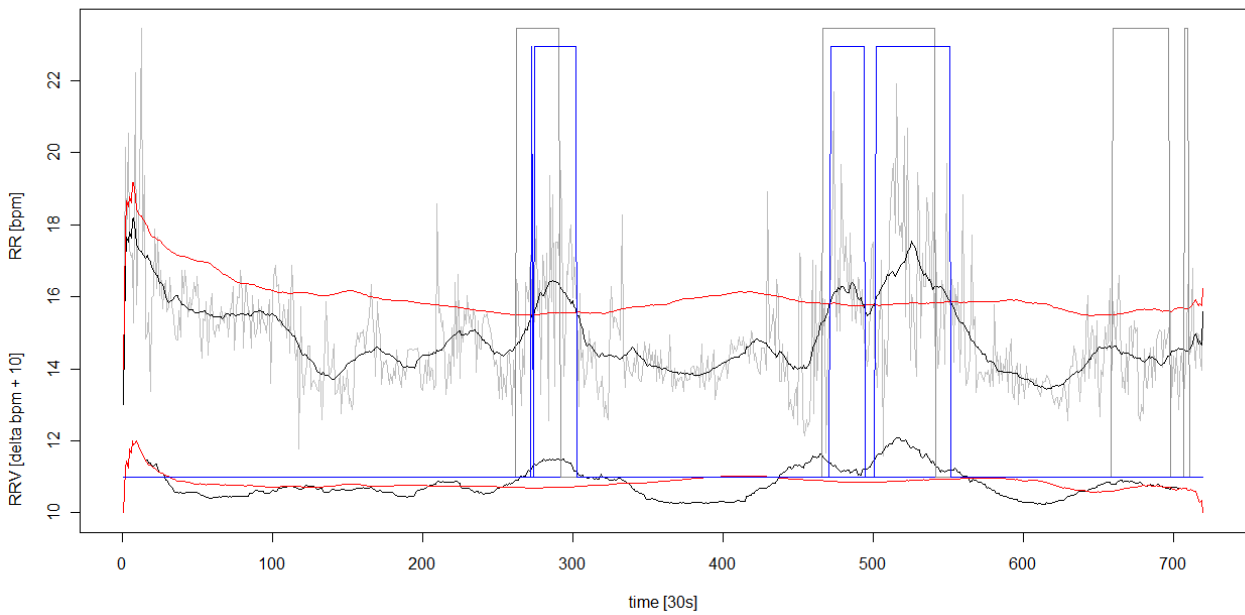
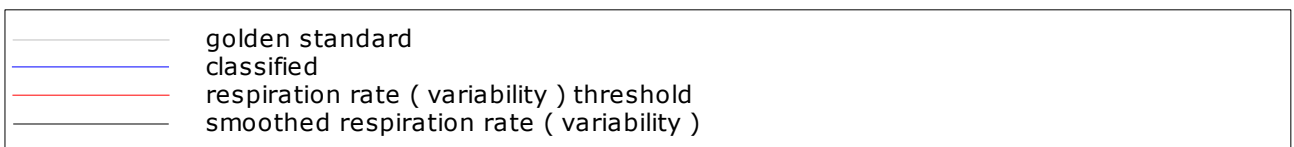


Diagram 40: REM sleep classification by module two for participant 201. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.



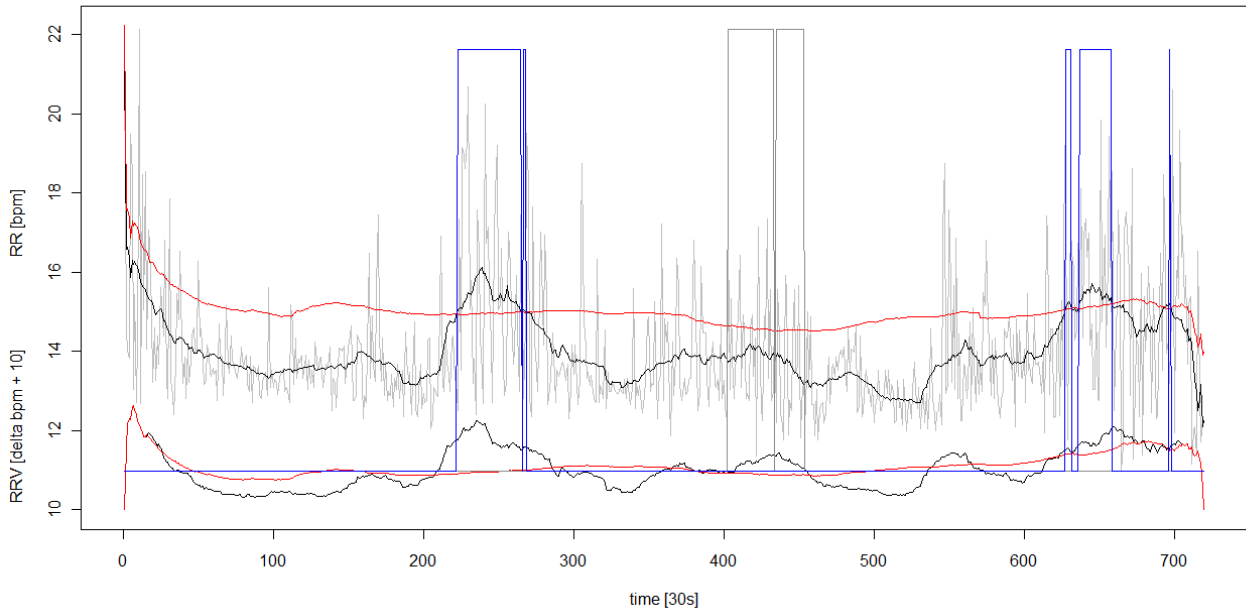


Diagram 41: REM sleep classification by module two for participant 202. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.

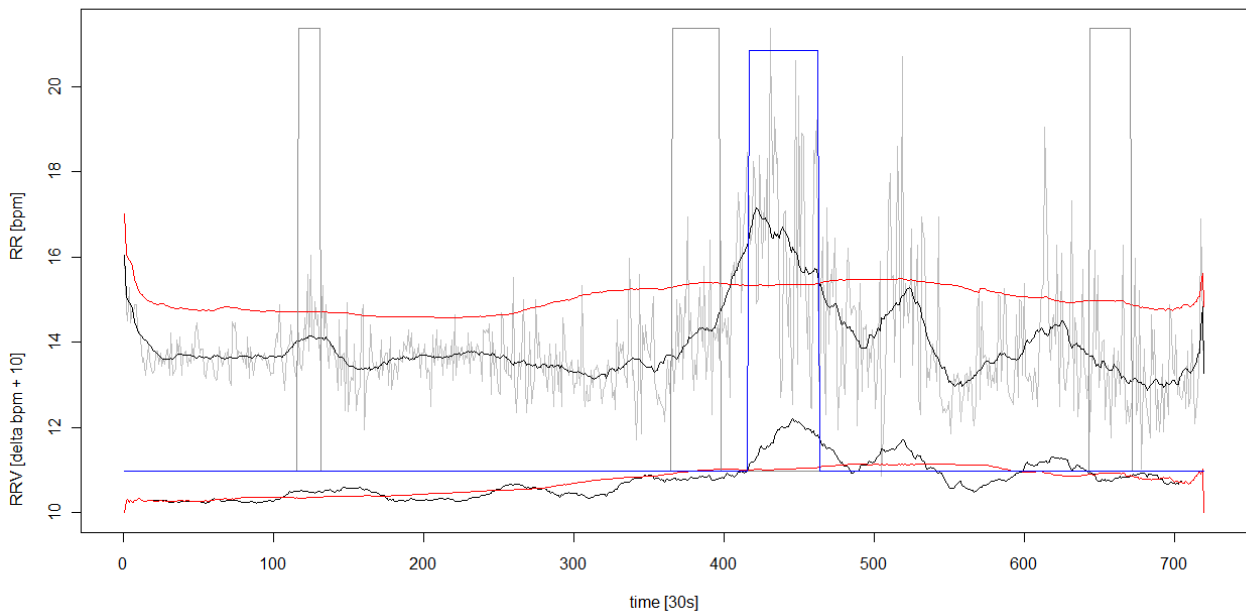
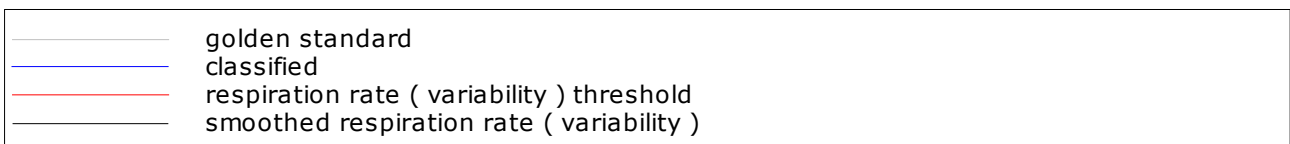


Diagram 42: REM sleep classification by module two for participant 203. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.



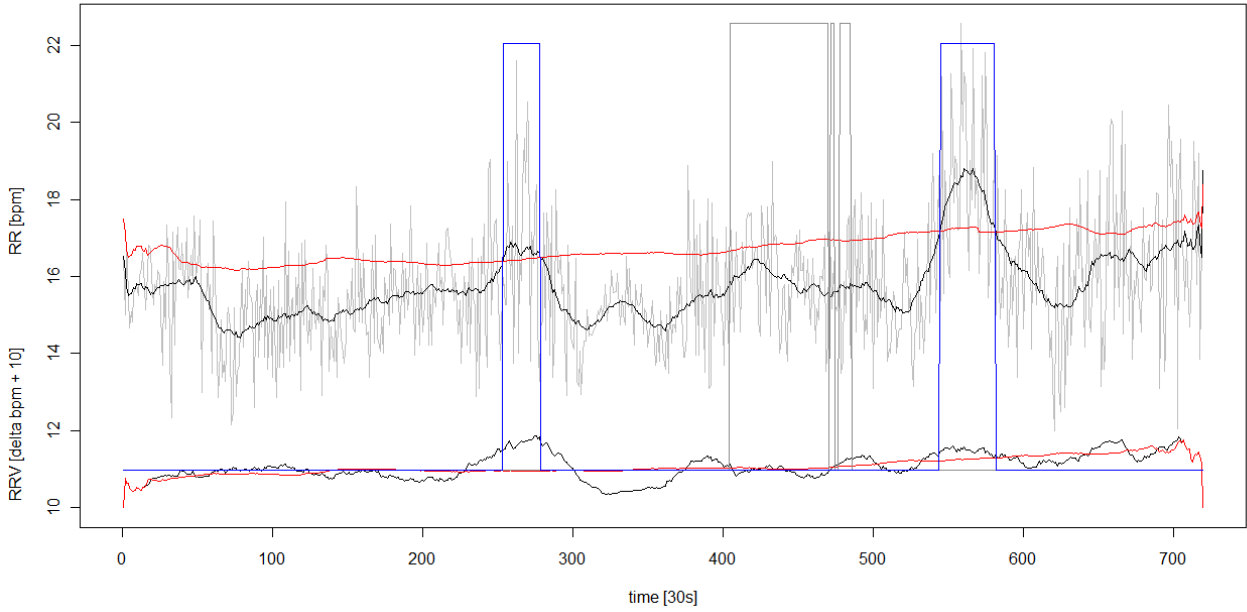


Diagram 43: REM sleep classification by module two for participant 204. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.

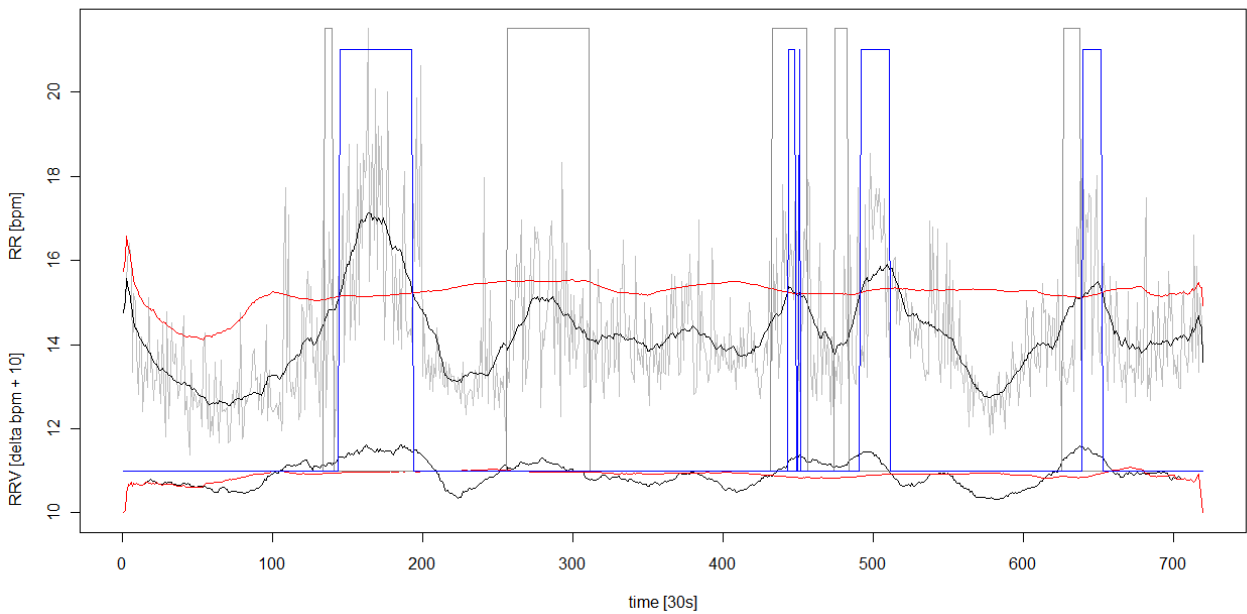
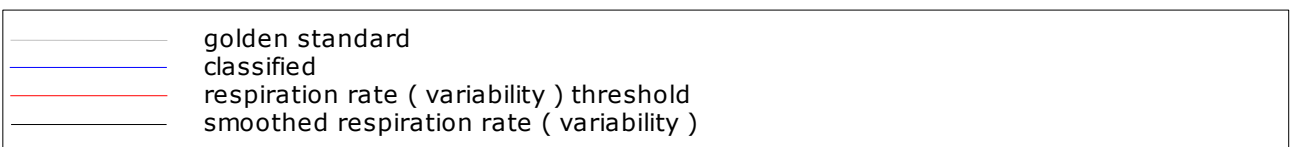


Diagram 44: REM sleep classification by module two for participant 205. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.



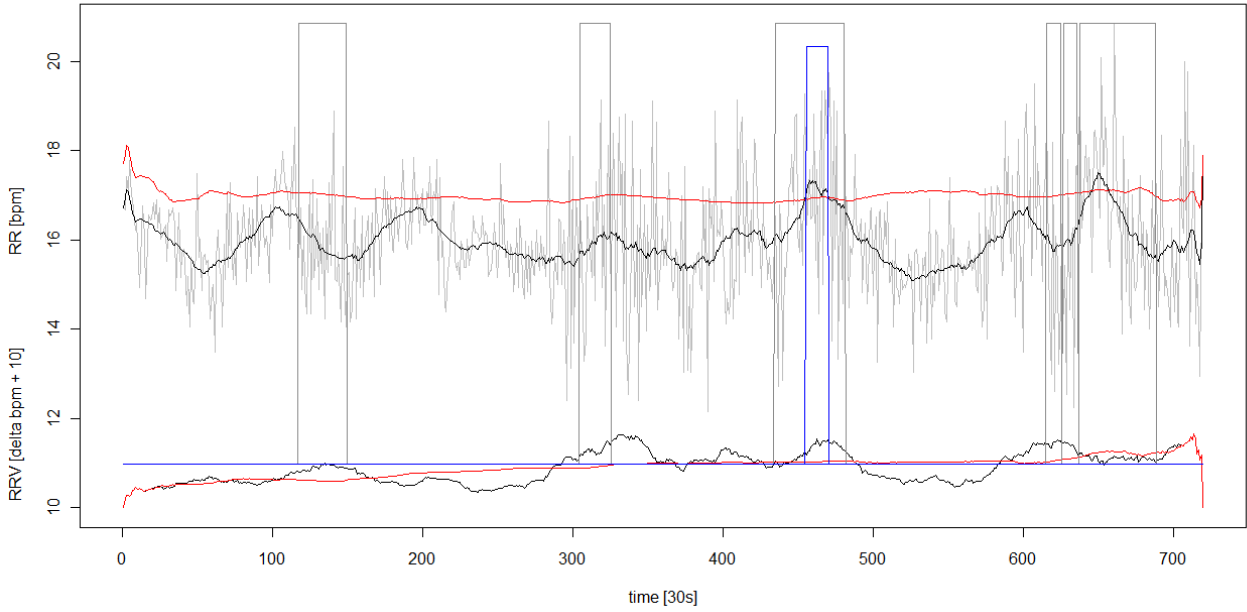


Diagram 45: REM sleep classification by module two for participant 207. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.

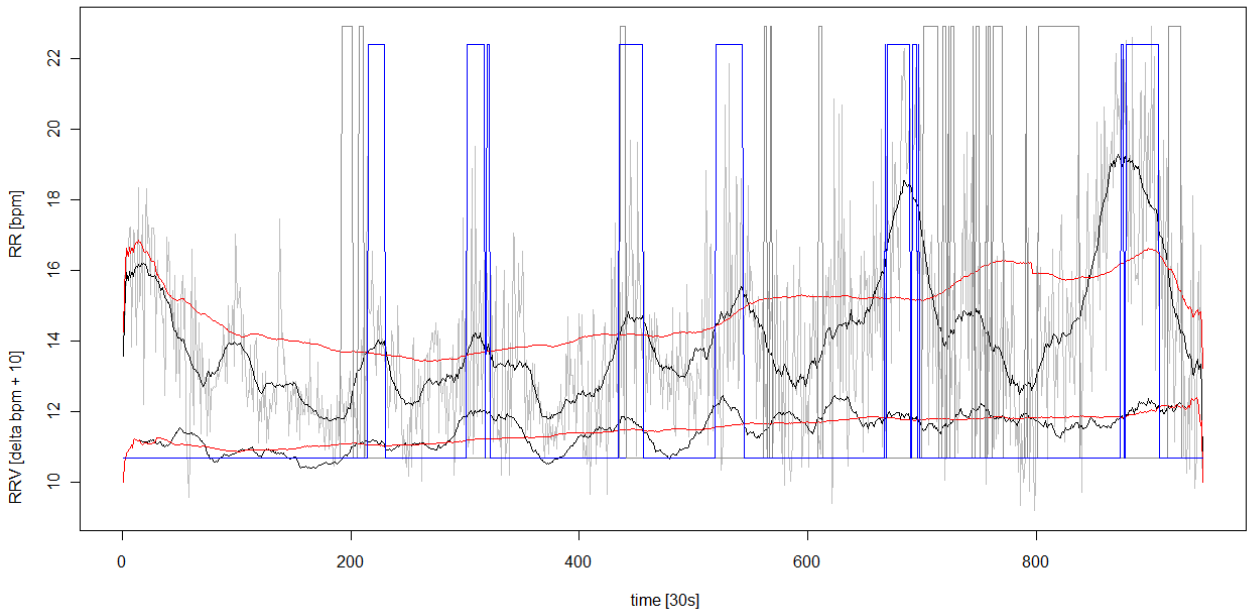
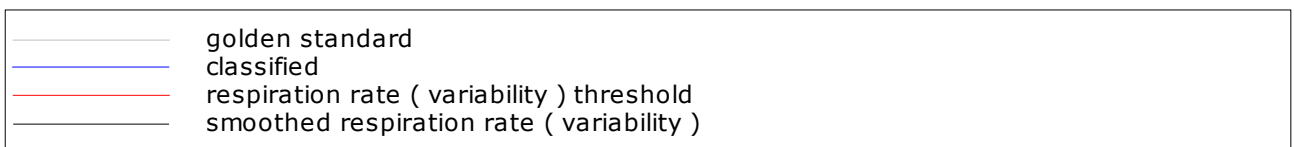


Diagram 46: REM sleep classification by module two for participant 301. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.



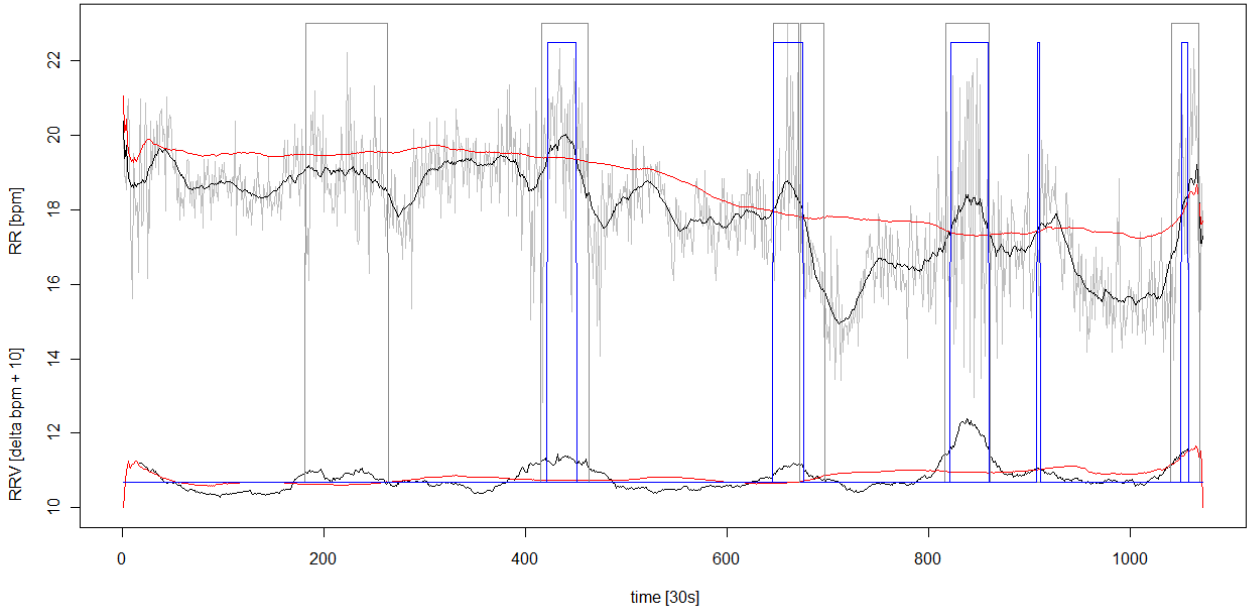


Diagram 47: REM sleep classification by module two for participant 302. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.

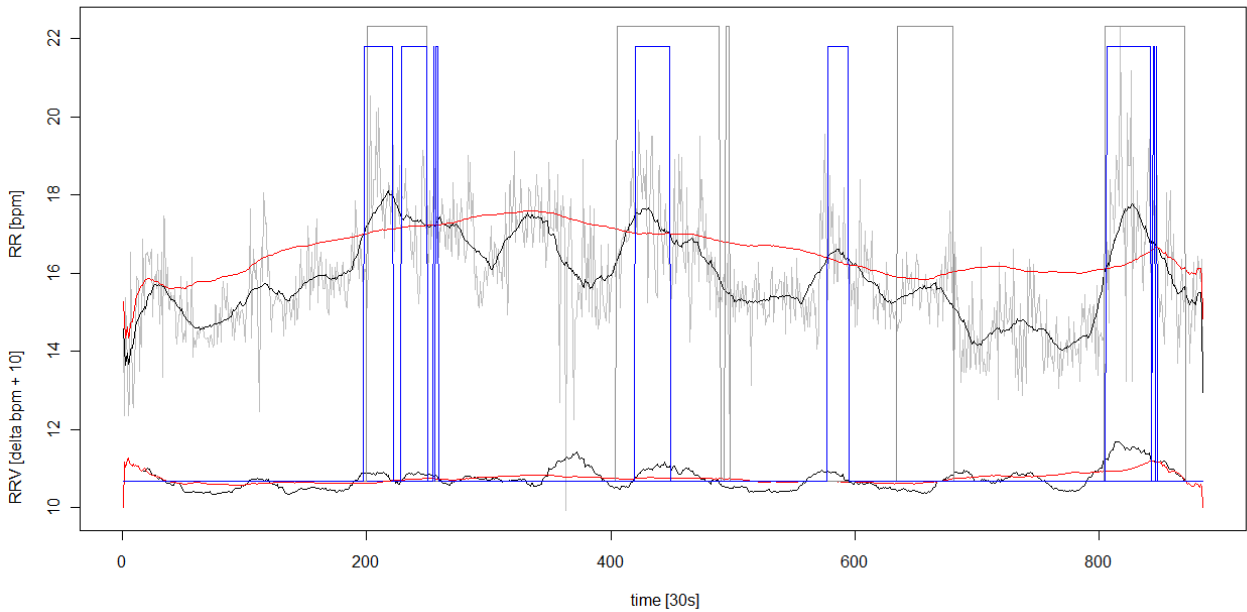
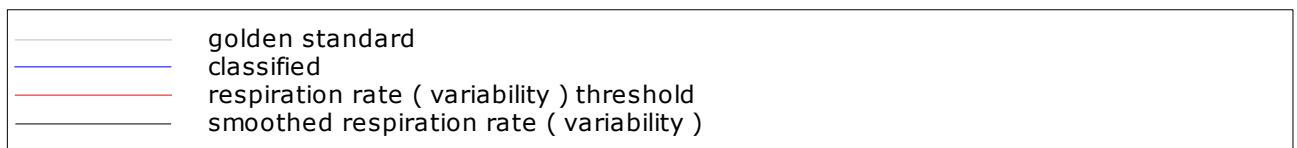


Diagram 48: REM sleep classification by module two for participant 303. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.



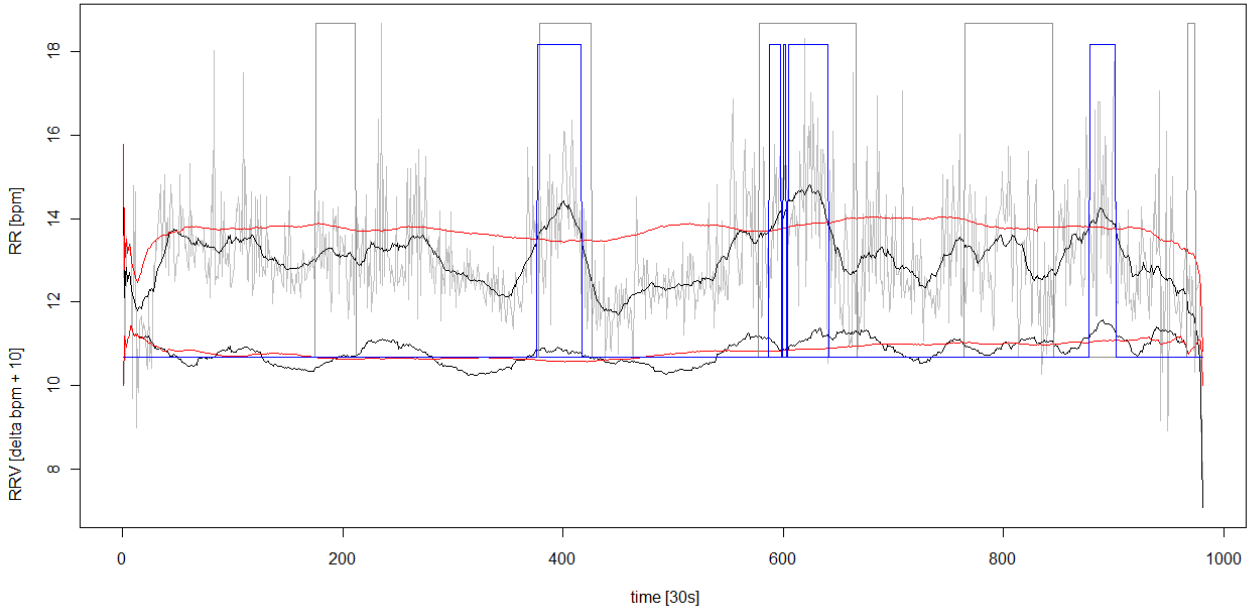


Diagram 49: REM sleep classification by module two for participant 304. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.

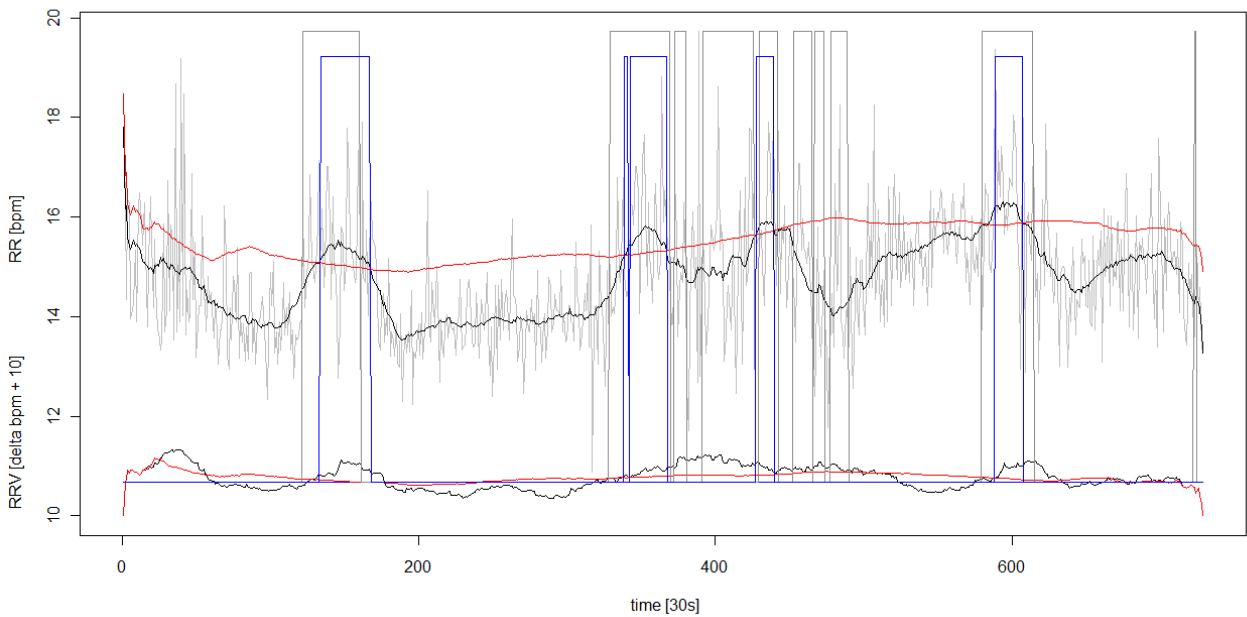
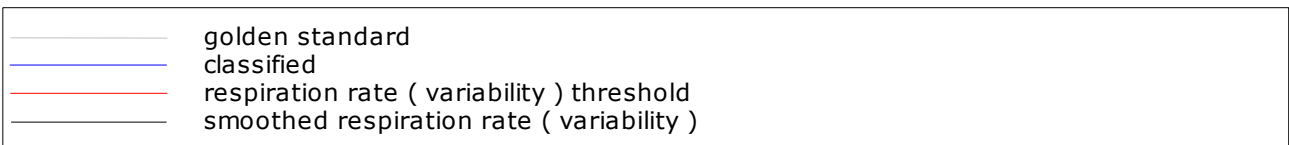


Diagram 50: REM sleep classification by module two for participant 305. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.



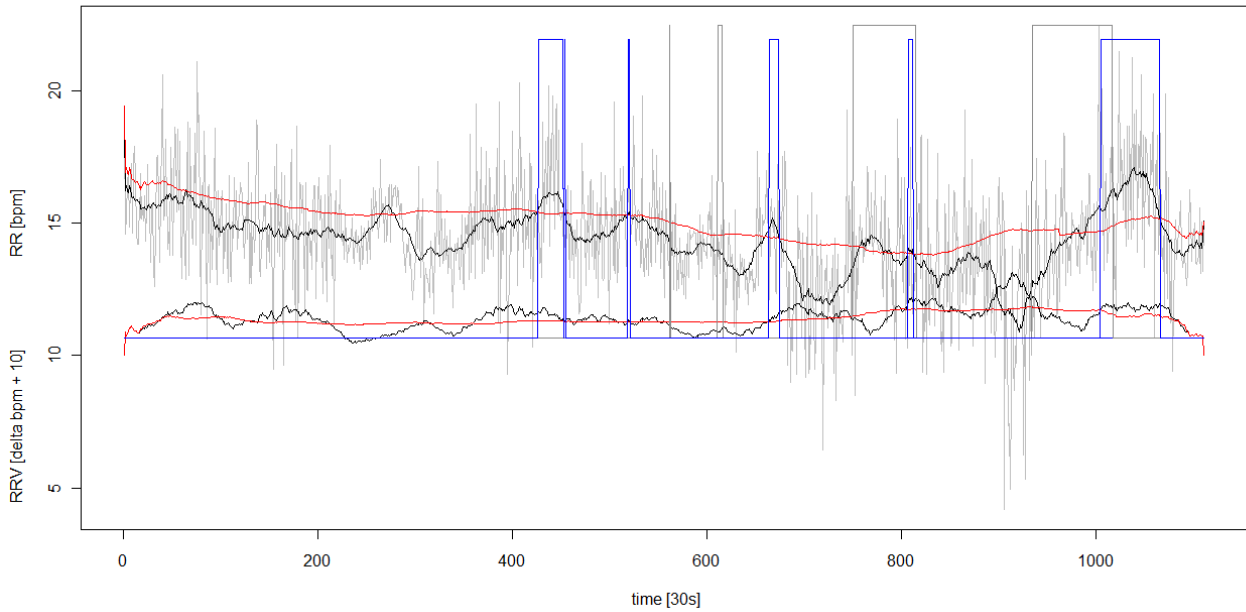


Diagram 51: REM sleep classification by module two for participant 307. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.

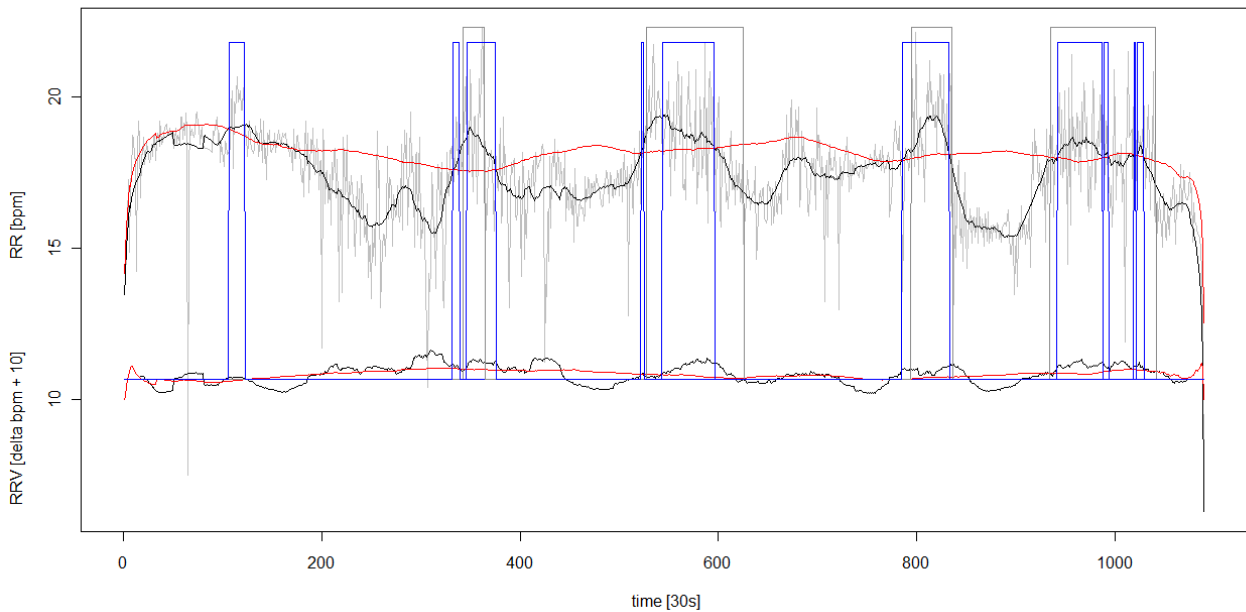
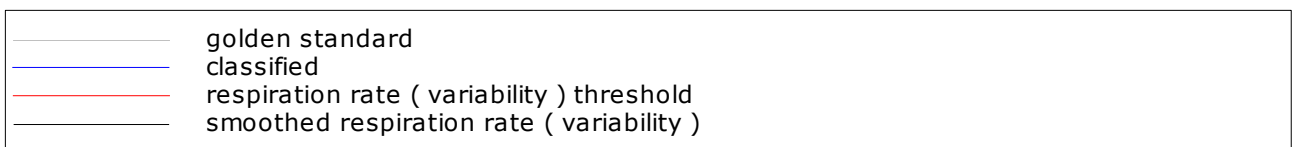


Diagram 52: REM sleep classification by module two for participant 306. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.



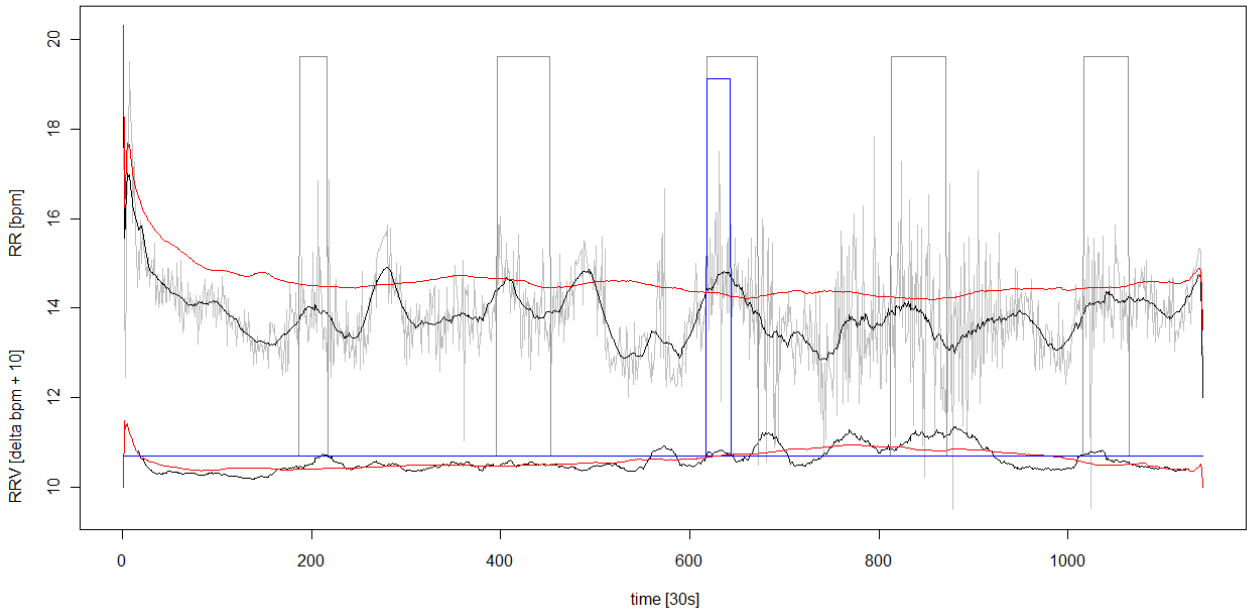


Diagram 53: REM sleep classification by module two for participant 308. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.

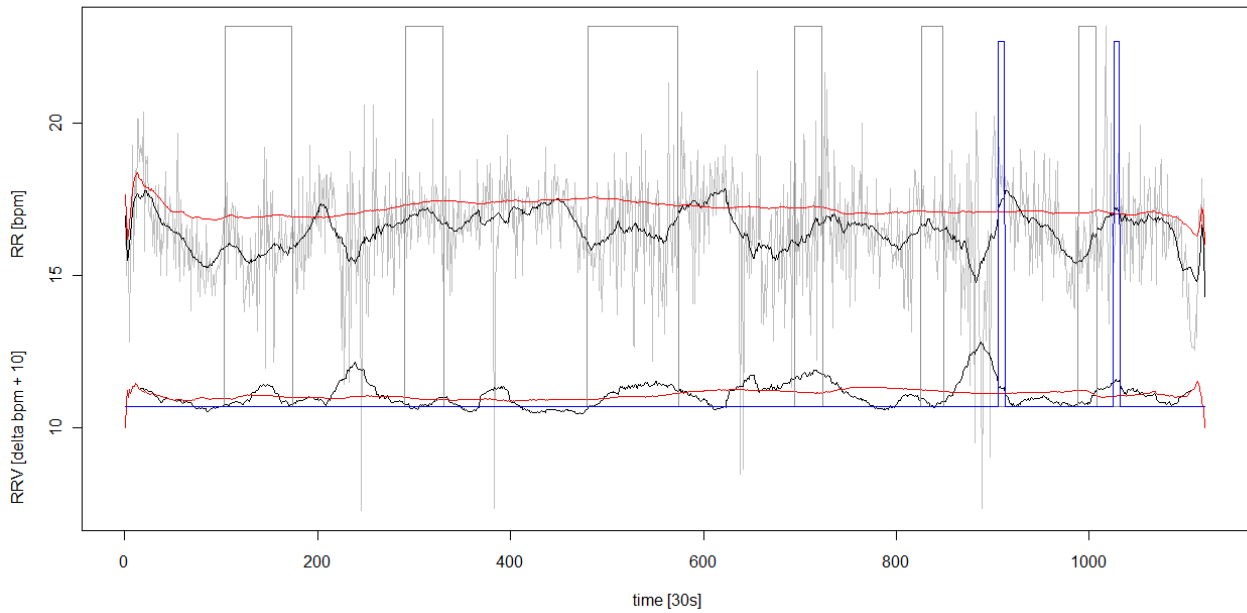
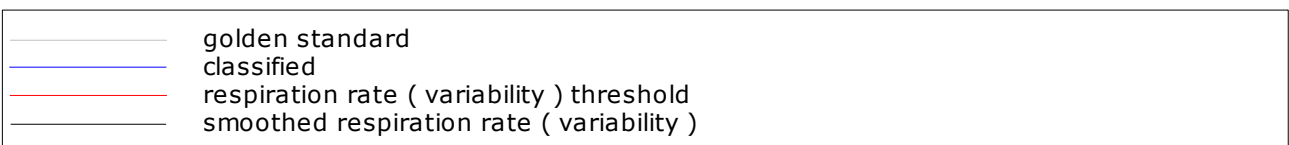


Diagram 54: REM sleep classification by module two for participant 309. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.



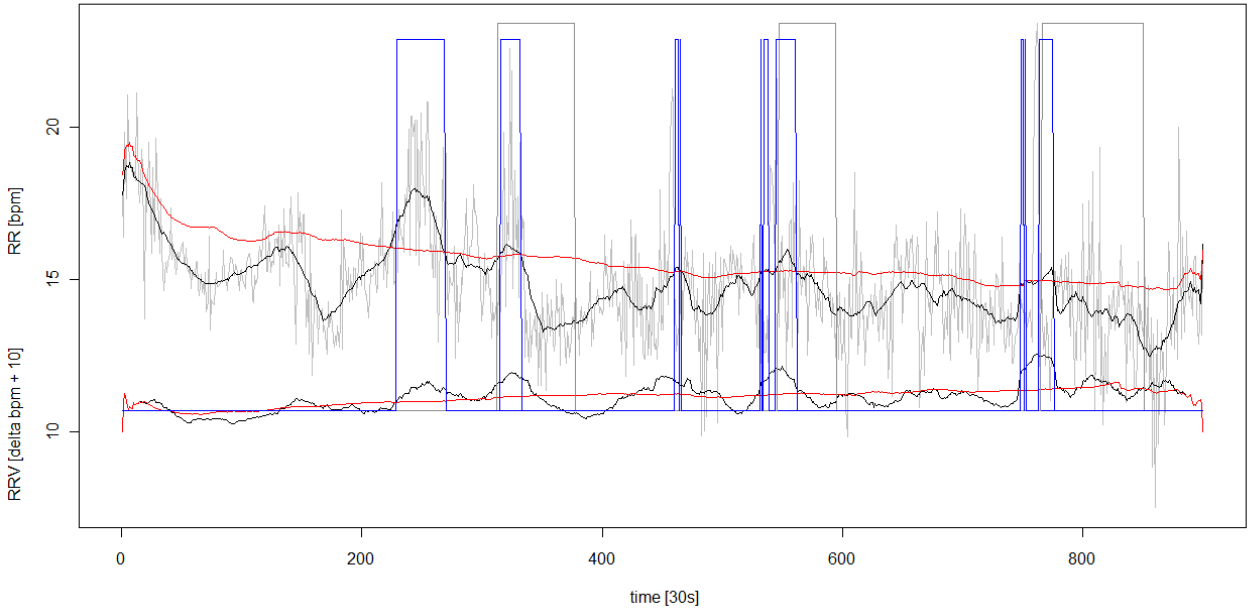


Diagram 55: REM sleep classification by module two for participant 310. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.

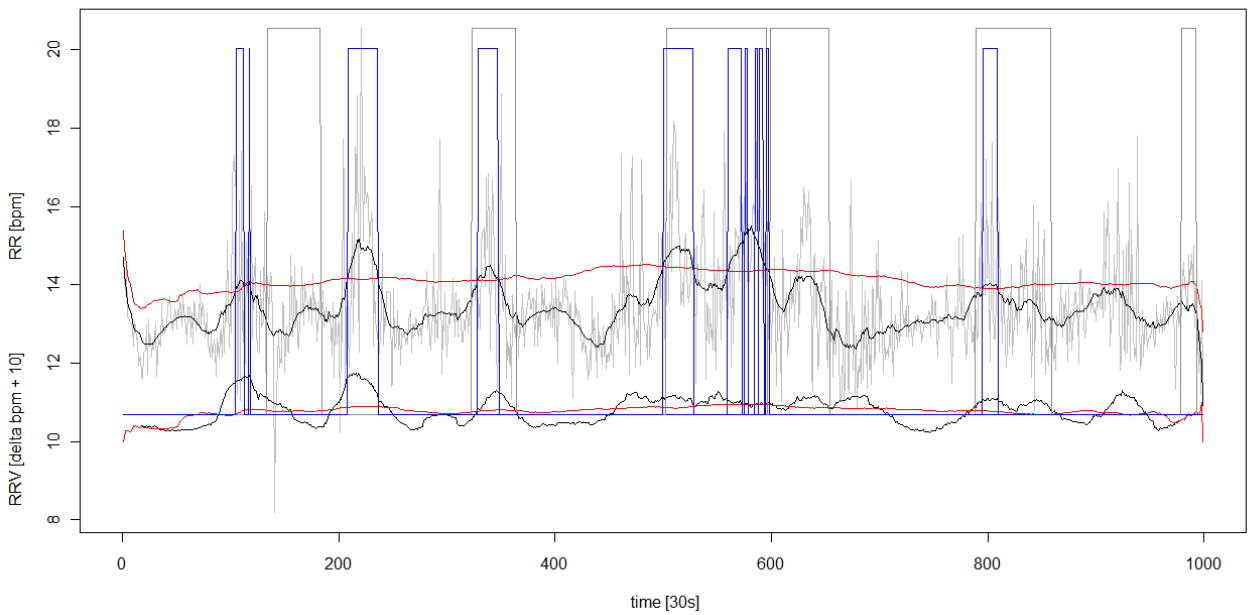
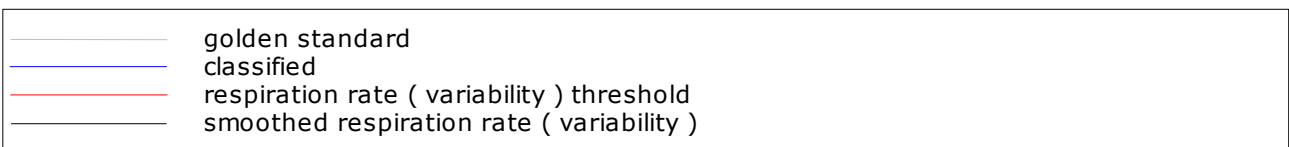


Diagram 56: REM sleep classification by module two for participant 311. Blue bars indicate classified REM epochs, grey bars indicate actual REM epochs determined by golden standard. Respiration rate (variability) thresholds are computed as described in chapter 5 and given in red, if both the respiration rate and respiration rate variability, given in black, exceed their respective thresholds, the epoch is classified as REM.



module three

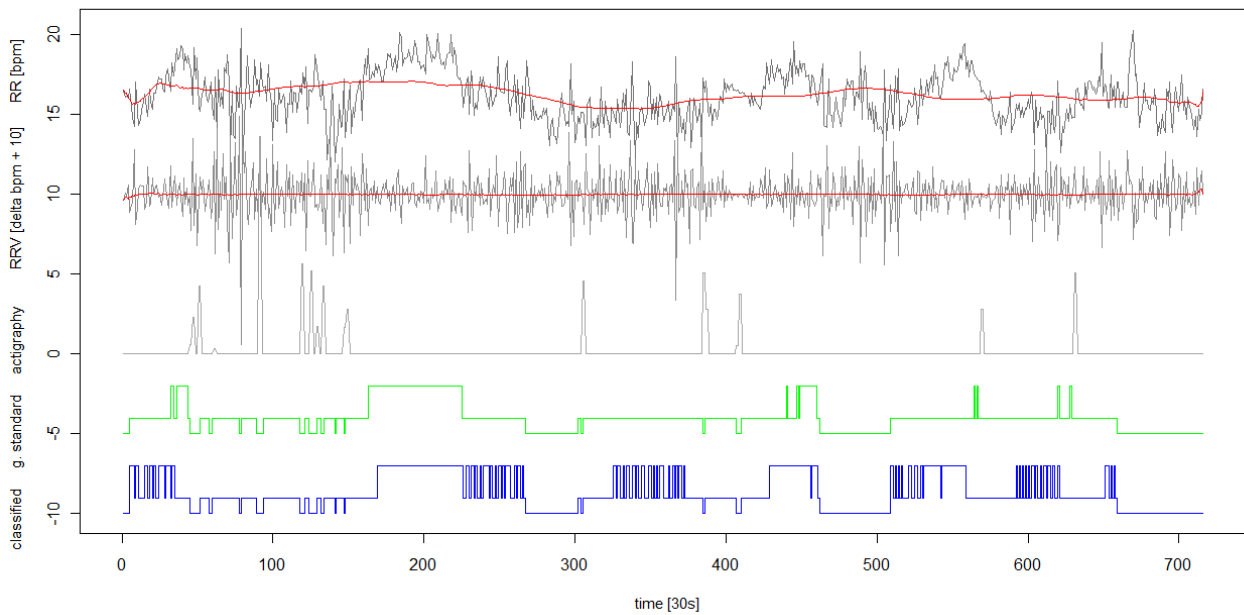
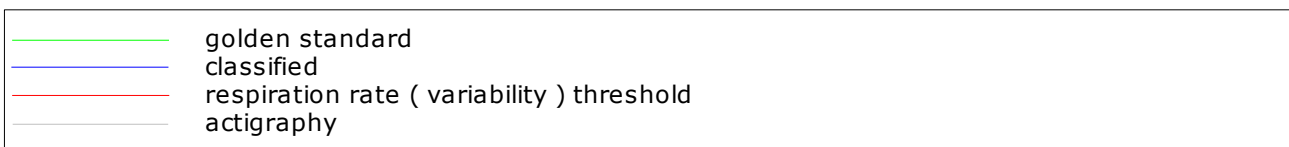


Diagram 57: Light sleep/deep sleep classification by module three for participant 101. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.



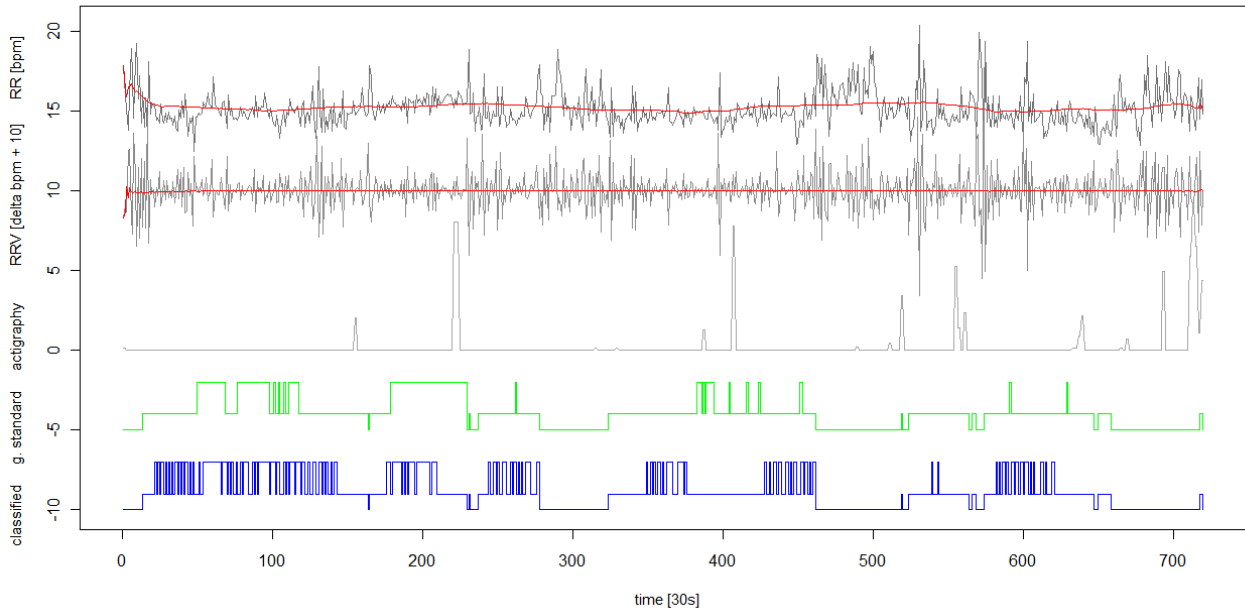


Diagram 58: Light sleep/deep sleep classification by module three for participant 102. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.

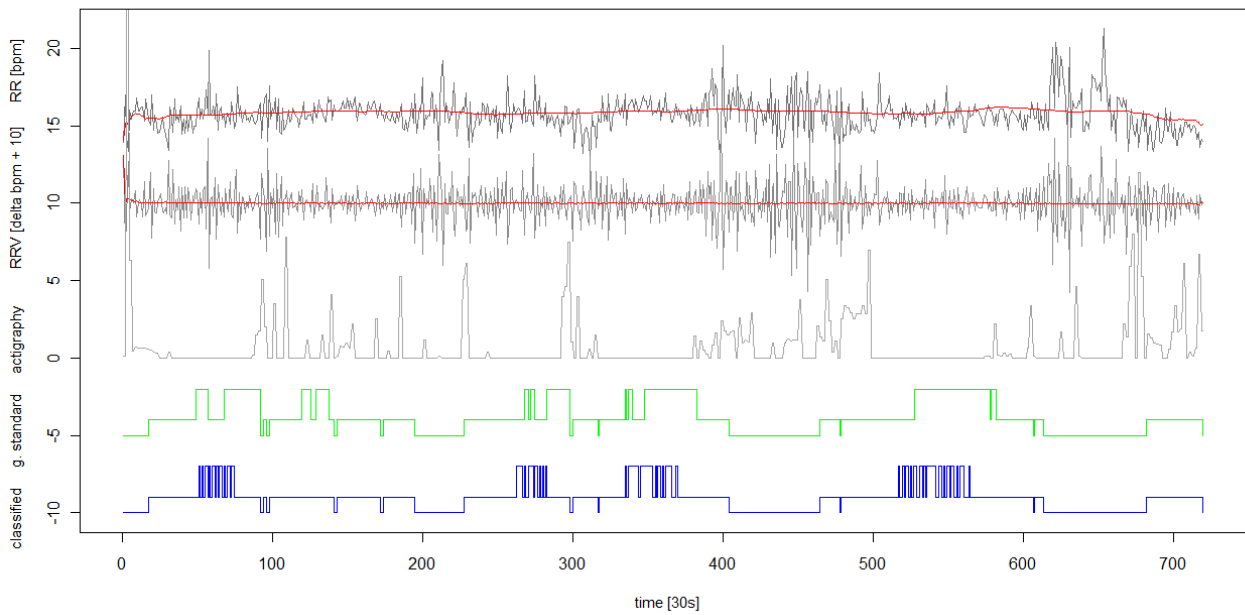
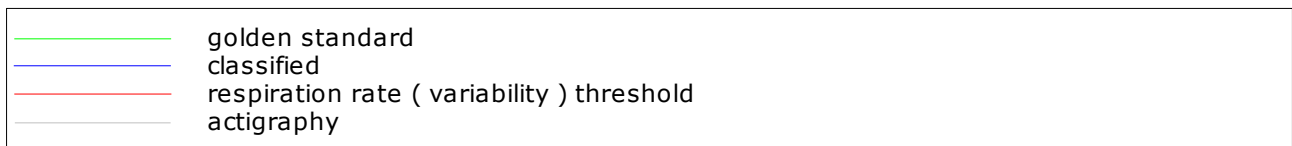


Diagram 59: Light sleep/deep sleep classification by module three for participant 103. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.



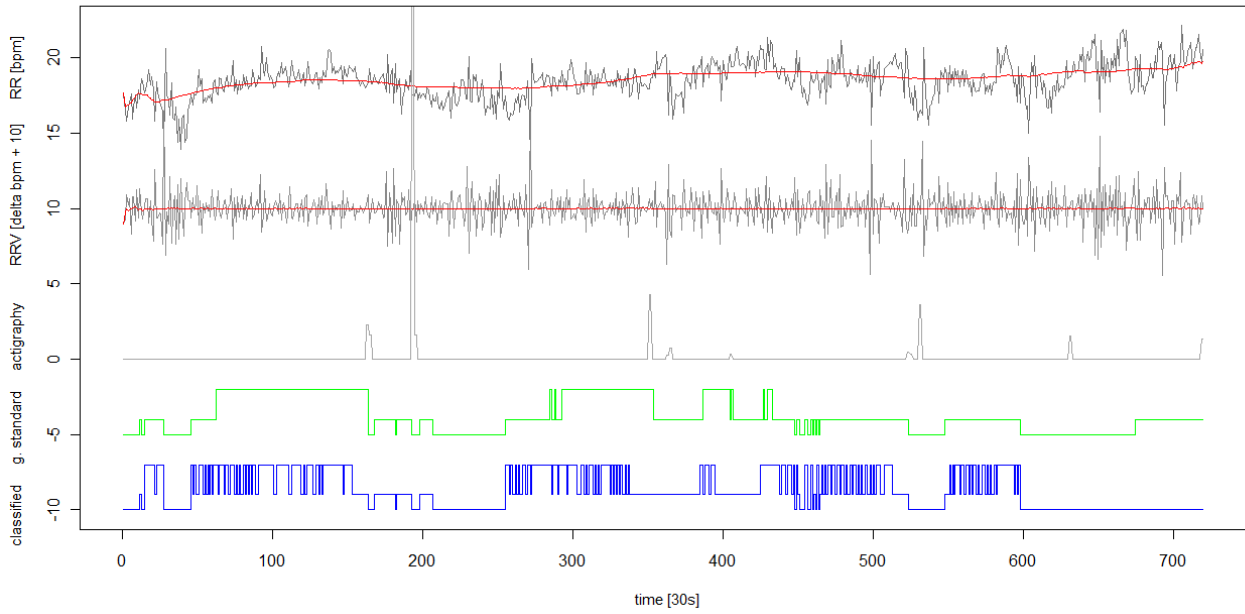


Diagram 60: Light sleep/deep sleep classification by module three for participant 104. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.

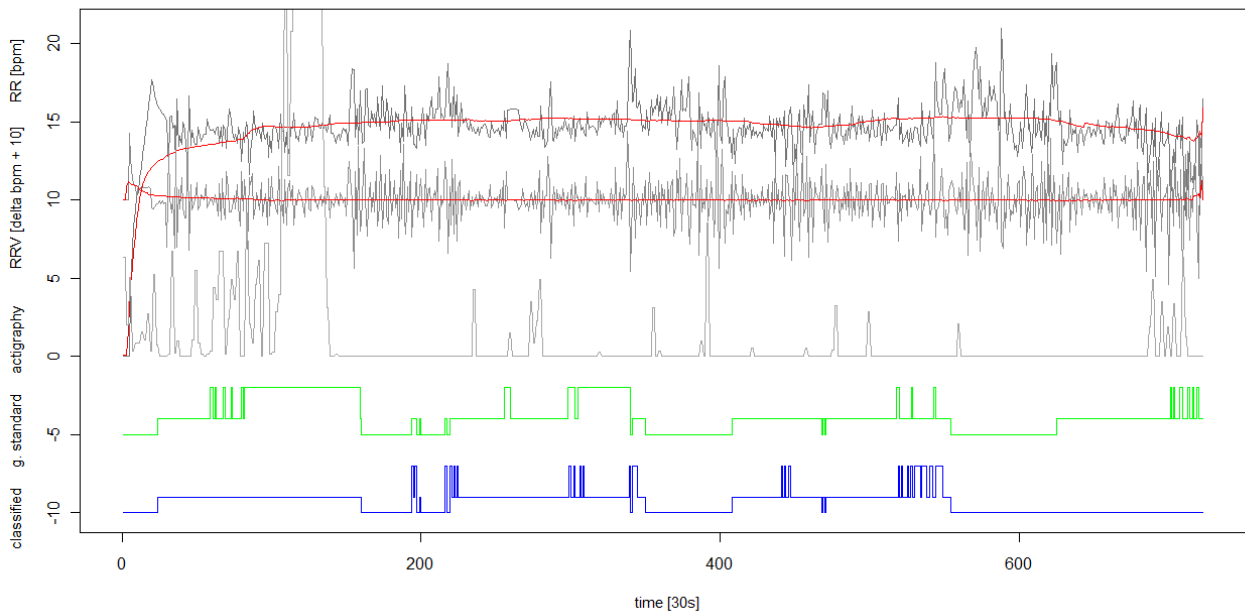
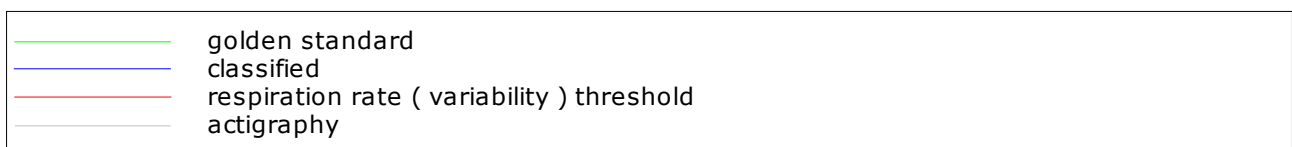


Diagram 61: Light sleep/deep sleep classification by module three for participant 106. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.



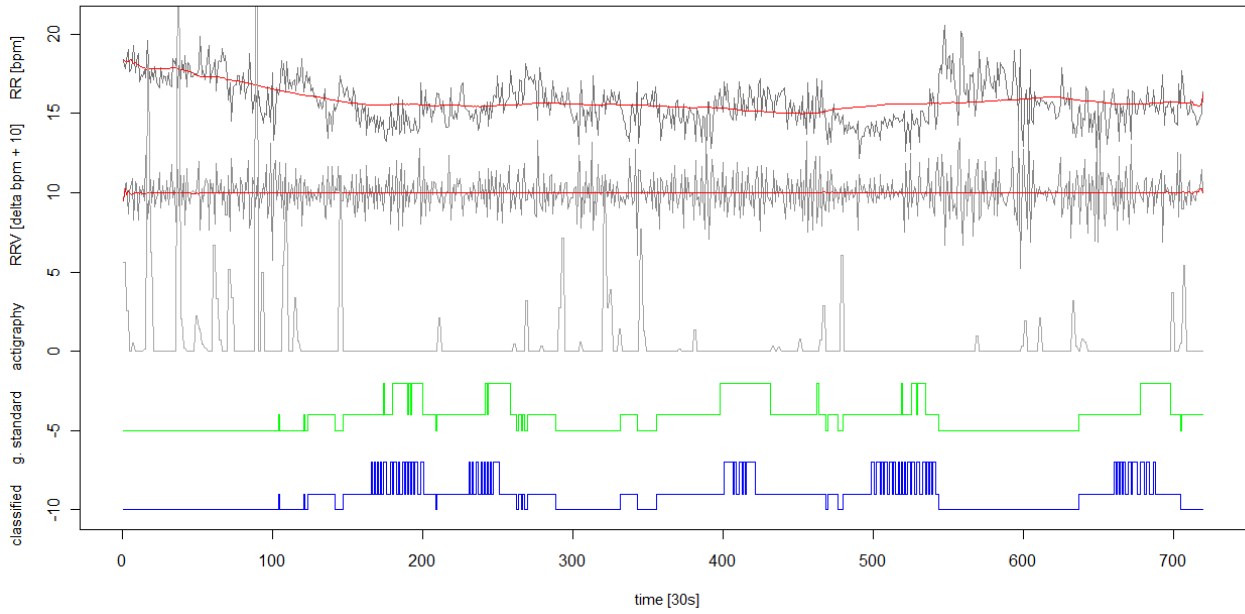


Diagram 62: Light sleep/deep sleep classification by module three for participant 108. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.

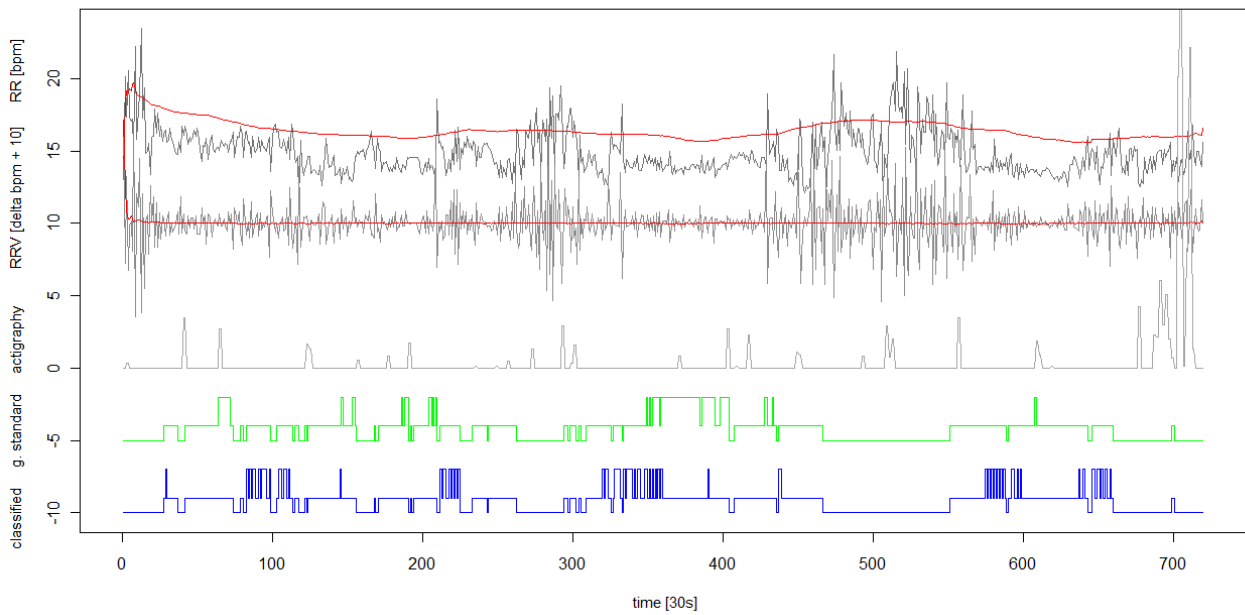
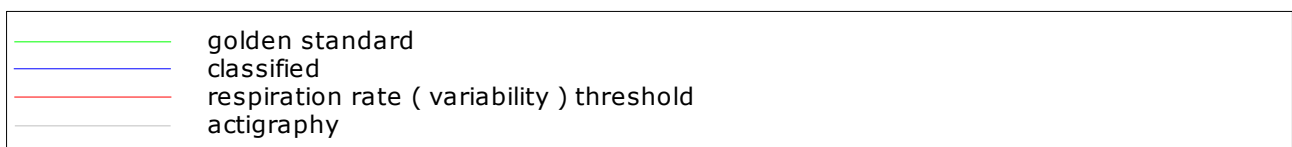


Diagram 63: Light sleep/deep sleep classification by module three for participant 201. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.



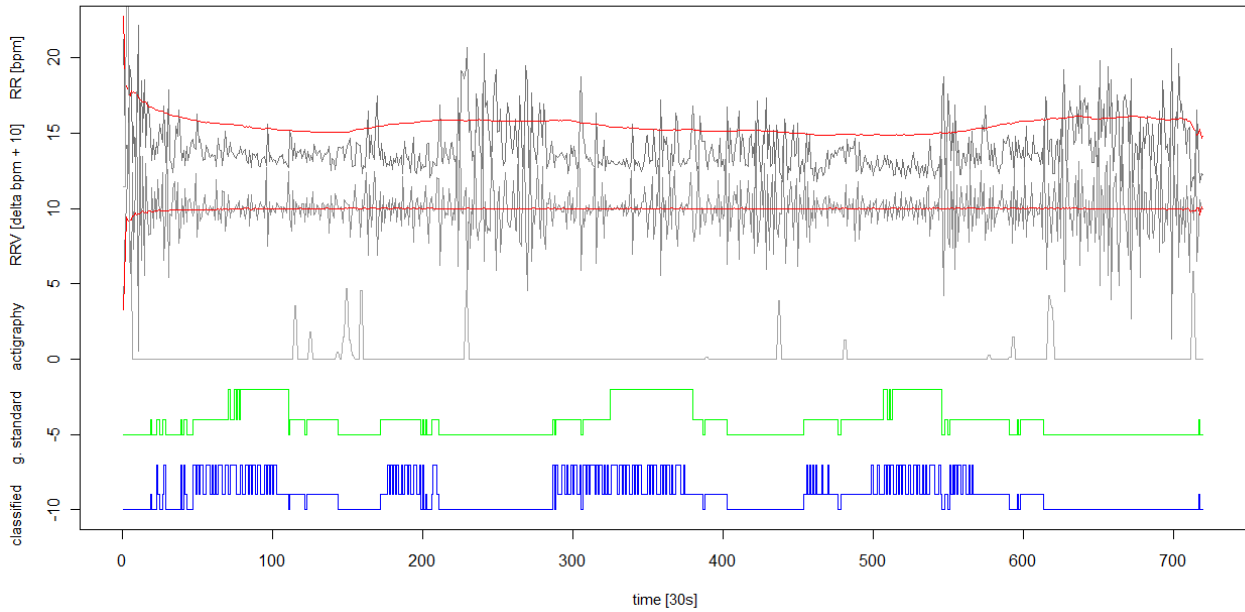


Diagram 64: Light sleep/deep sleep classification by module three for participant 203. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.

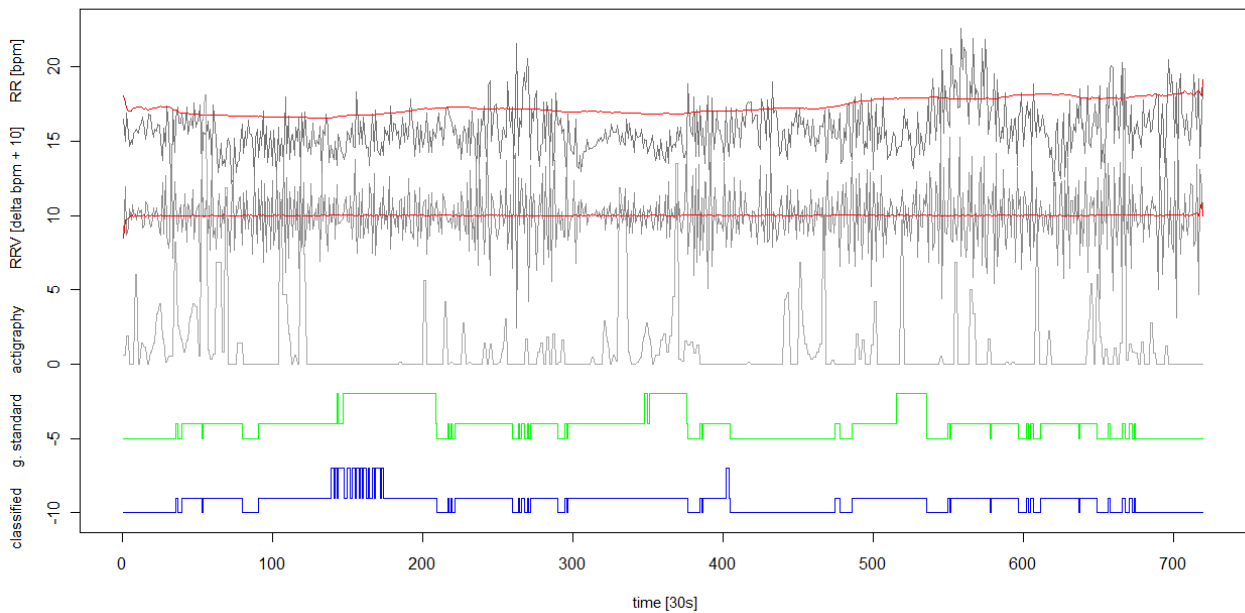
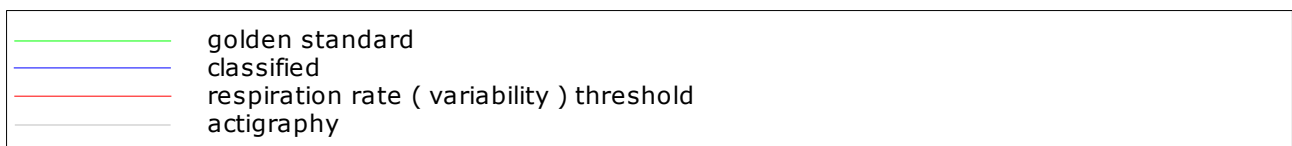


Diagram 65: Light sleep/deep sleep classification by module three for participant 204. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.



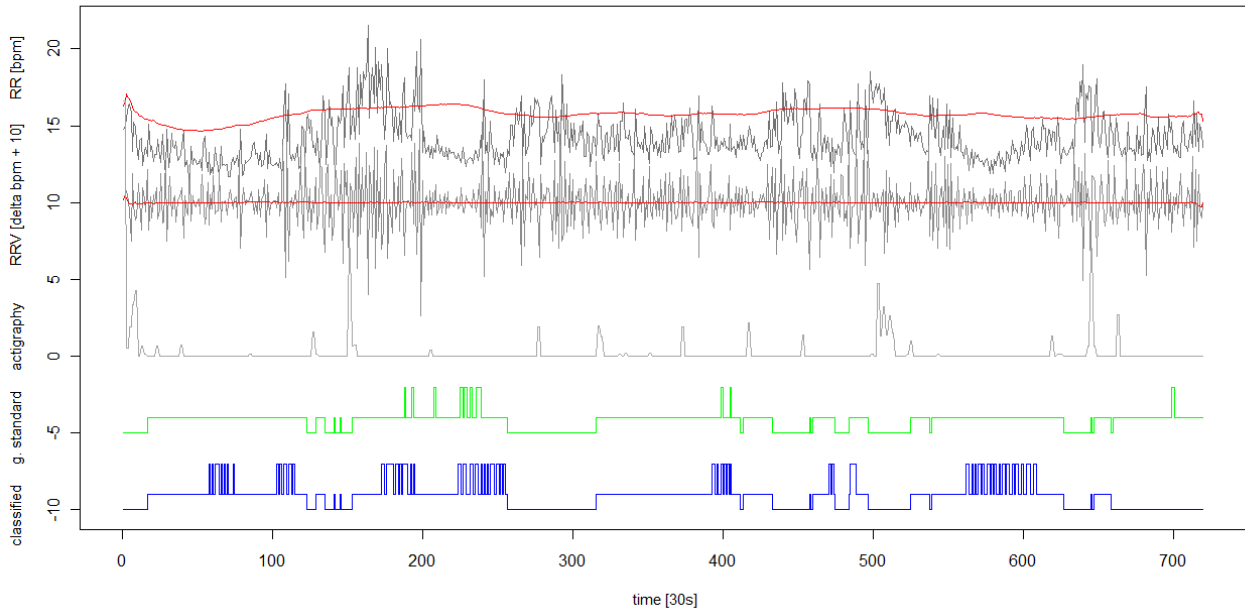


Diagram 66: Light sleep/deep sleep classification by module three for participant 205. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.

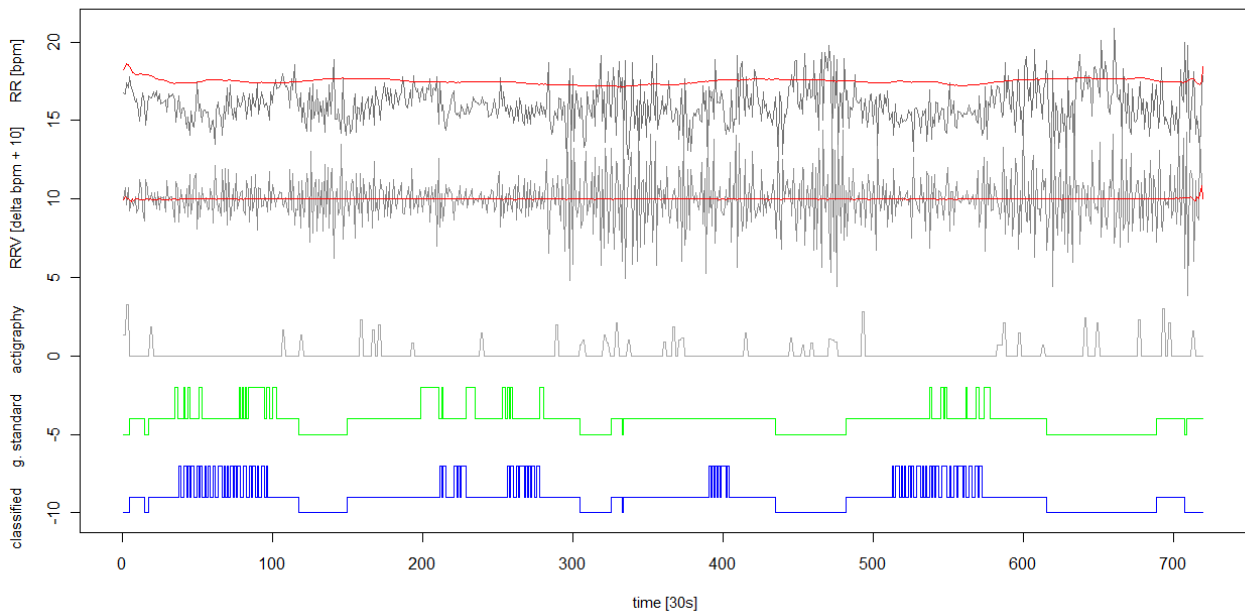
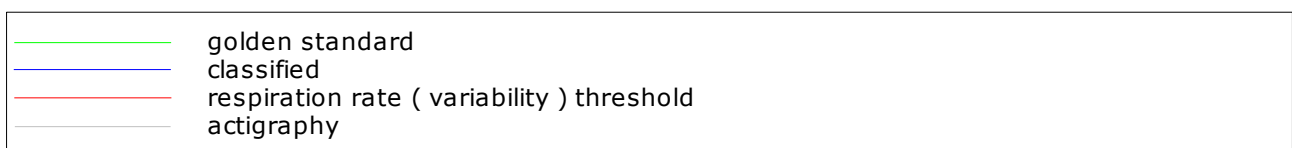


Diagram 67: Light sleep/deep sleep classification by module three for participant 207. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.



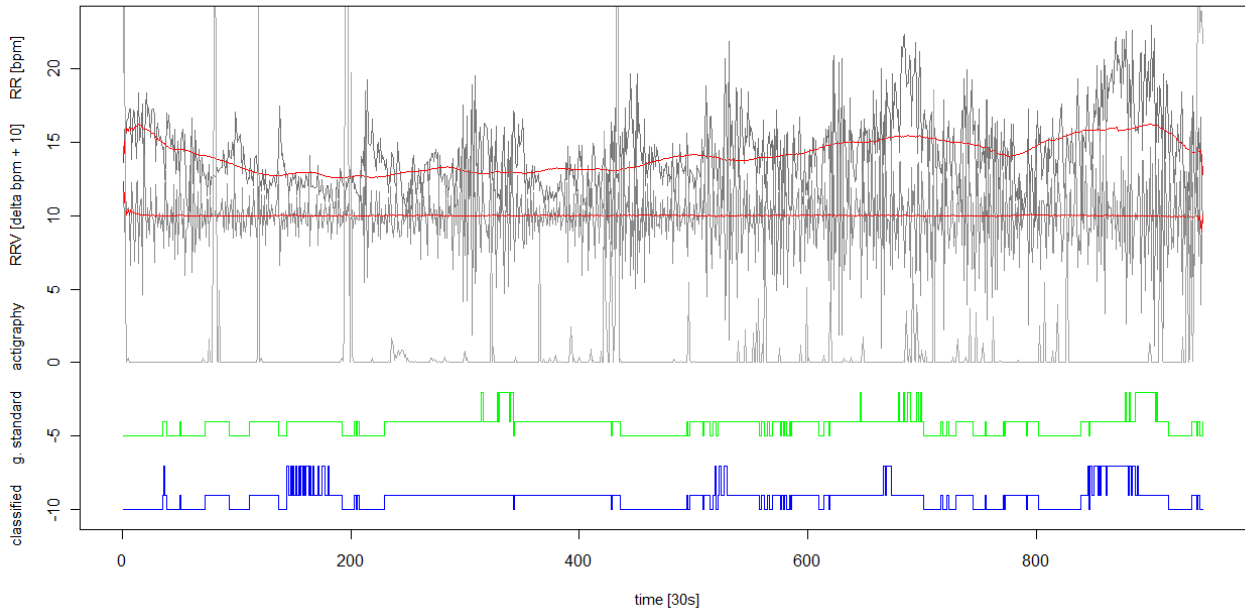


Diagram 68: Light sleep/deep sleep classification by module three for participant 301. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.

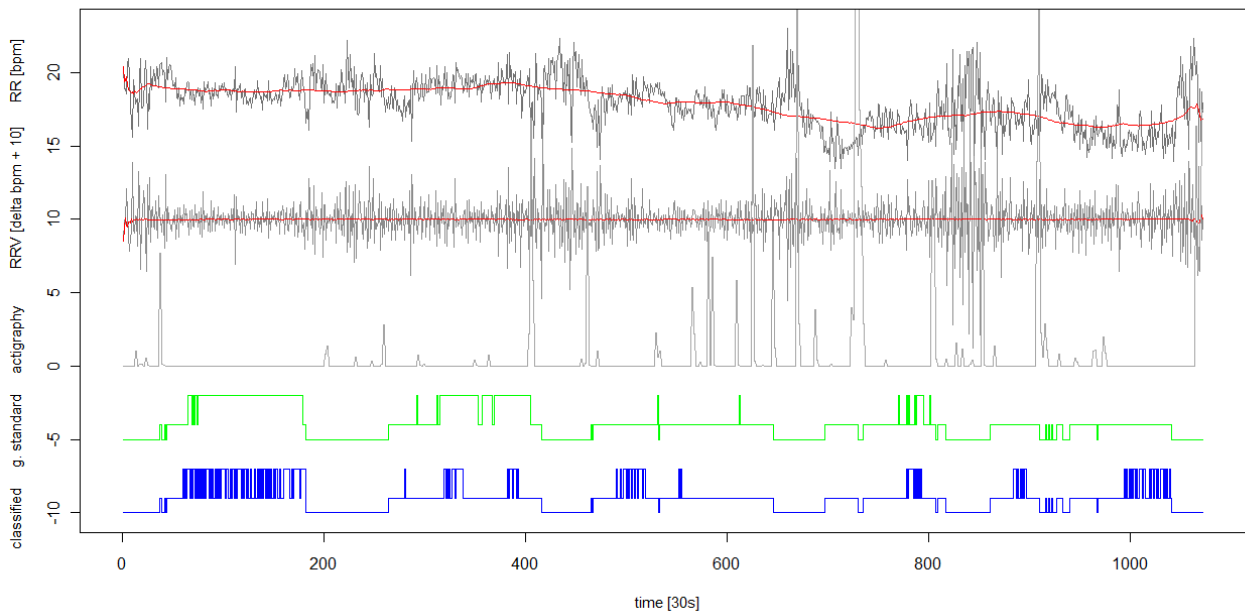
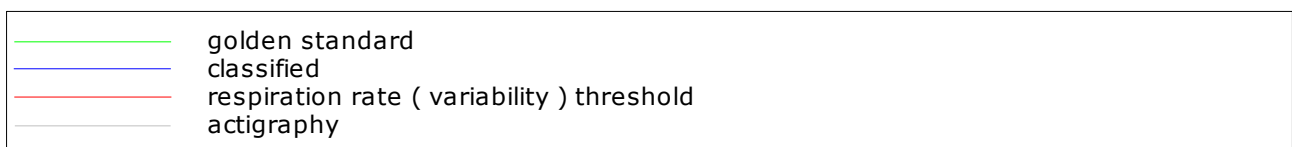


Diagram 69: Light sleep/deep sleep classification by module three for participant 302. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.



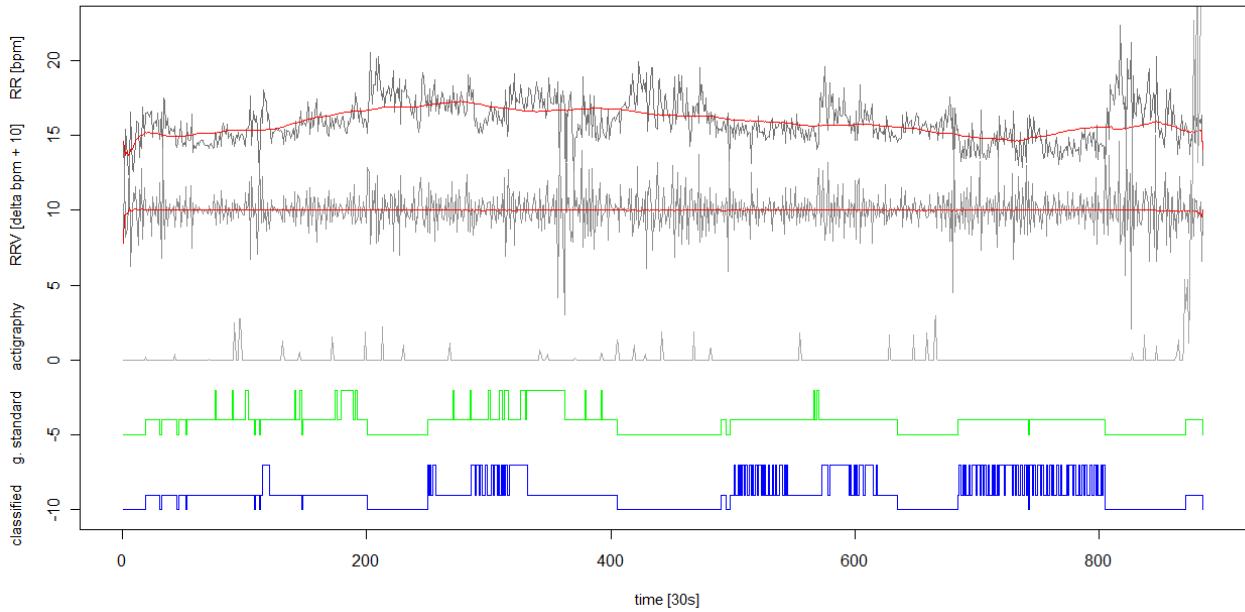


Diagram 70: Light sleep/deep sleep classification by module three for participant 303. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.

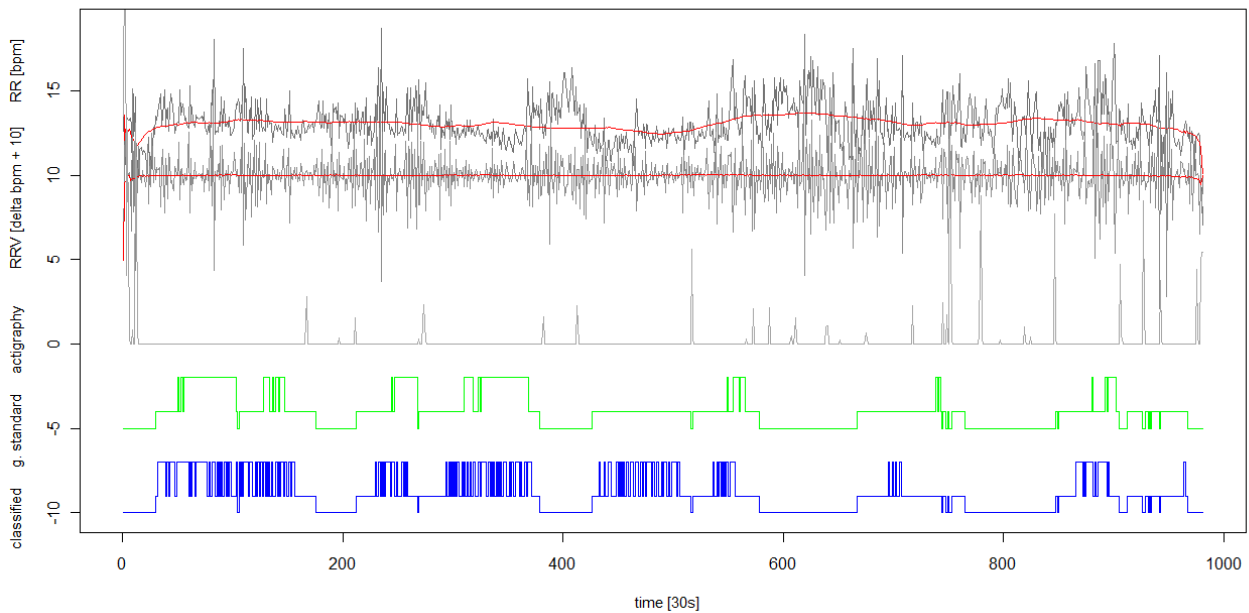
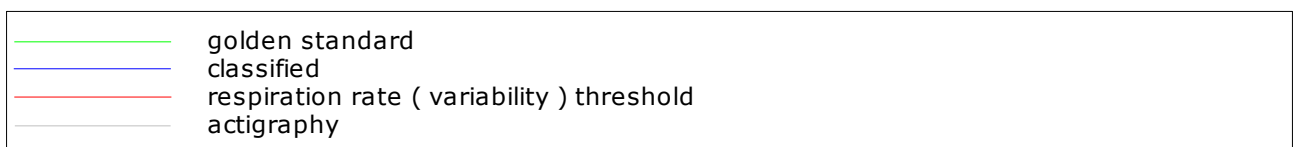


Diagram 71: Light sleep/deep sleep classification by module three for participant 304. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.



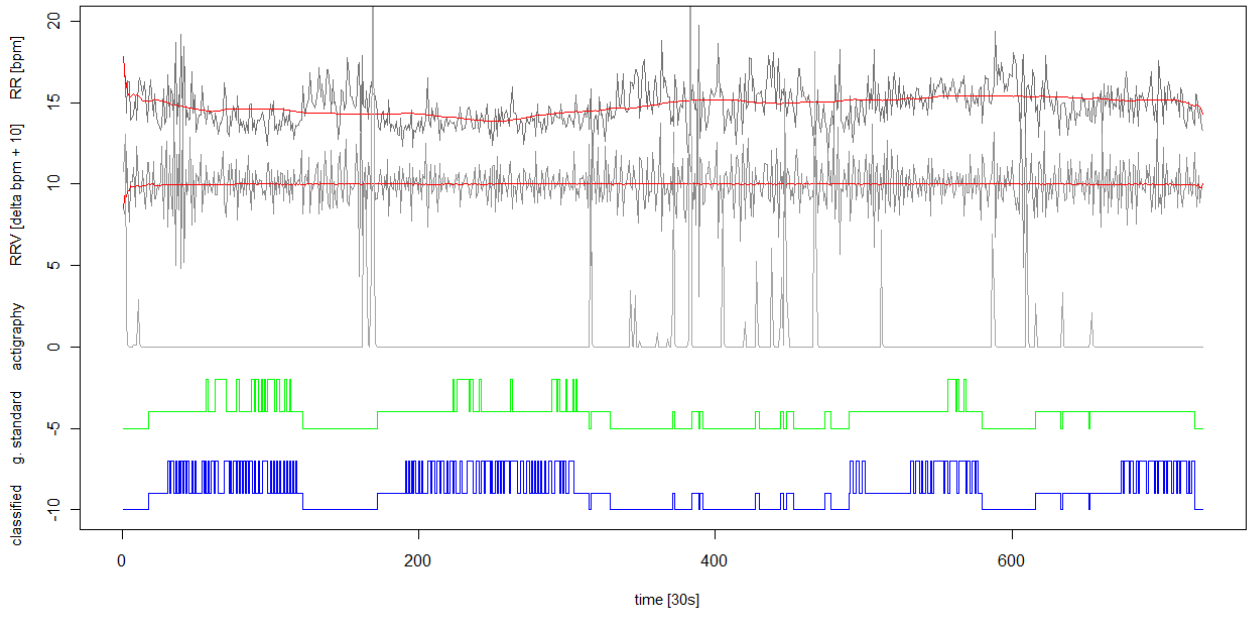


Diagram 72: Light sleep/deep sleep classification by module three for participant 305. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.

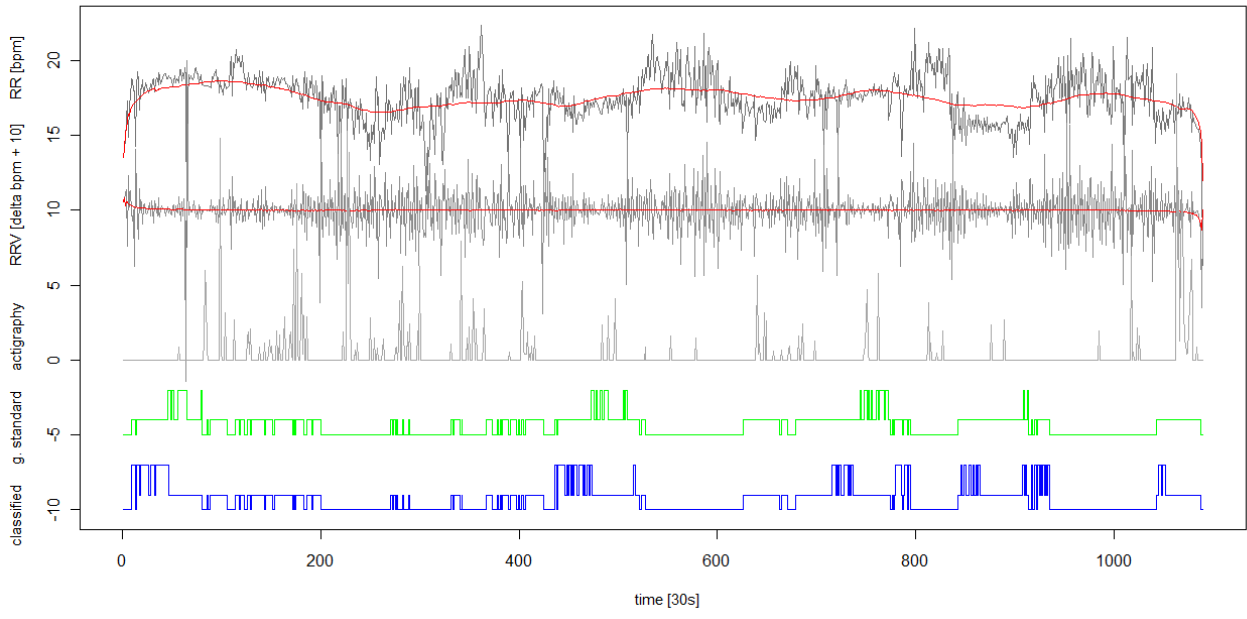
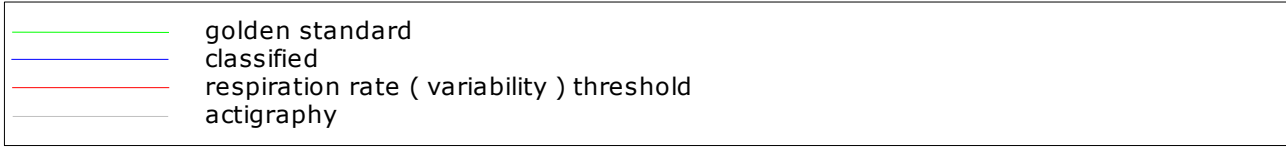


Diagram 73: Light sleep/deep sleep classification by module three for participant 306. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.



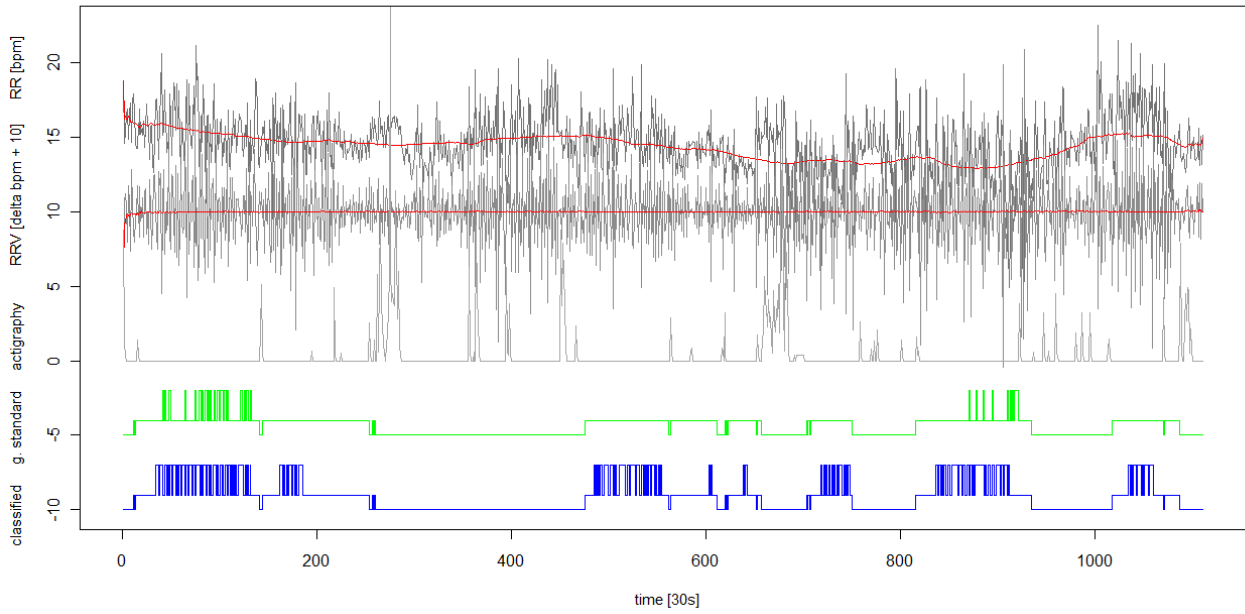


Diagram 74: Light sleep/deep sleep classification by module three for participant 307. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.

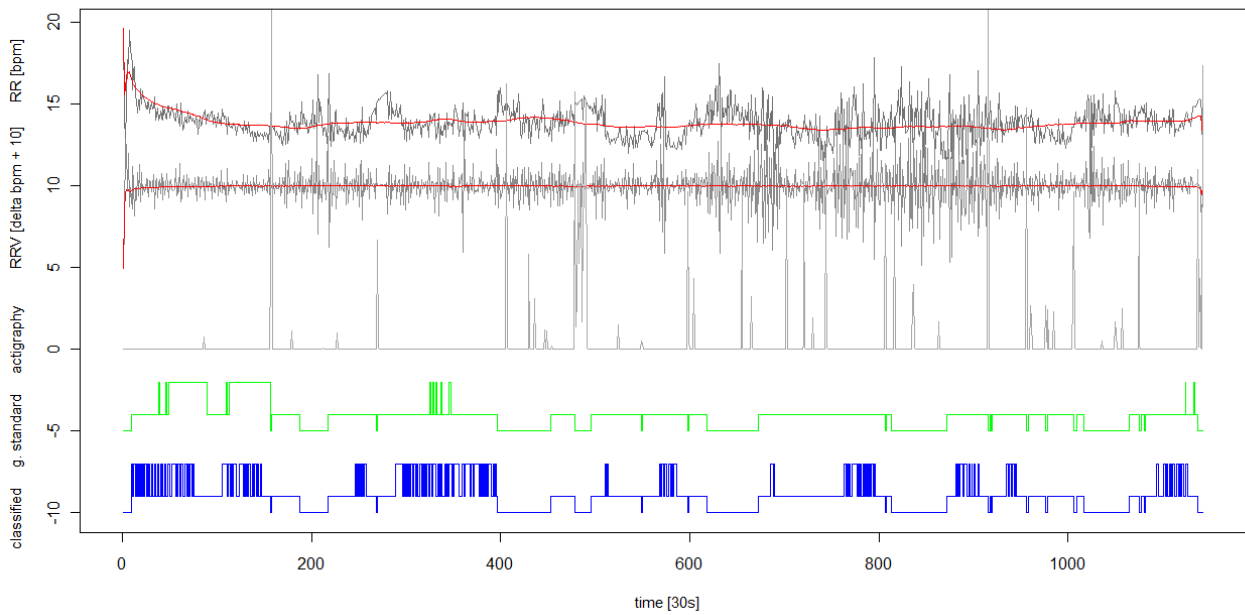
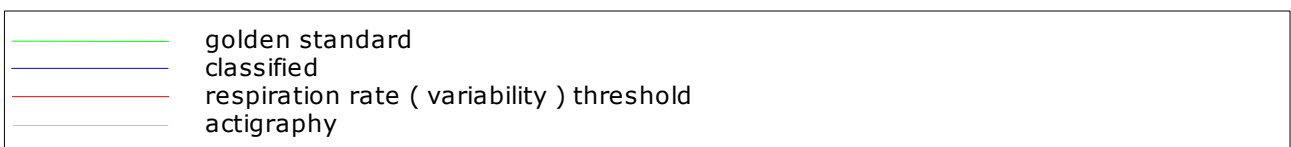


Diagram 75: Light sleep/deep sleep classification by module three for participant 308. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.



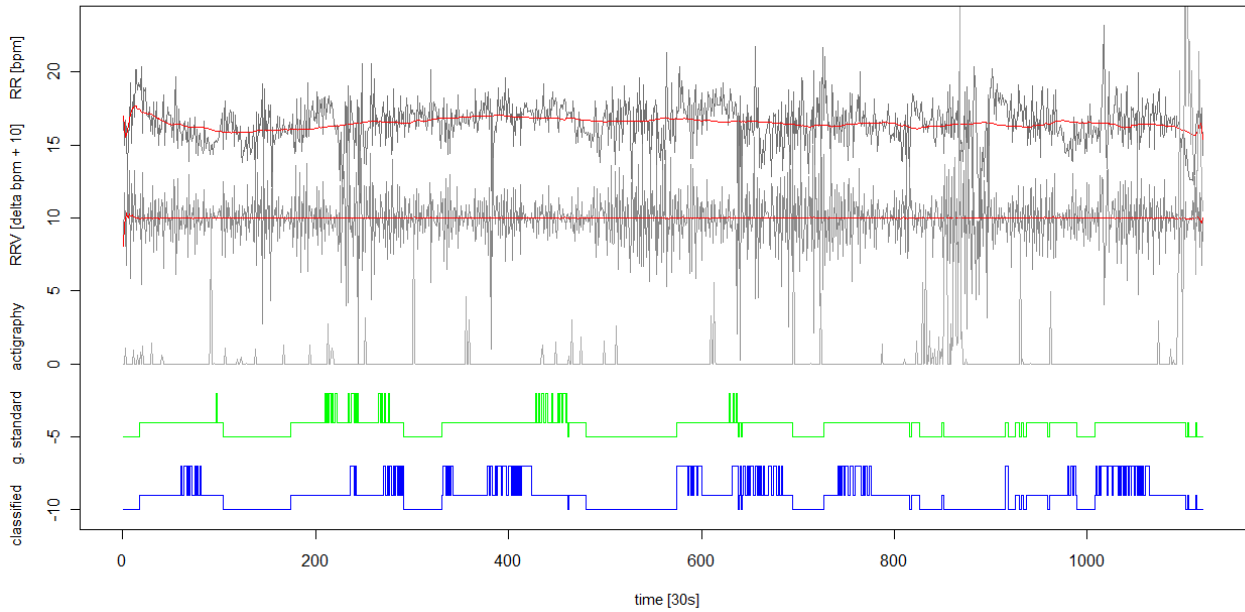


Diagram 76: Light sleep/deep sleep classification by module three for participant 309. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.

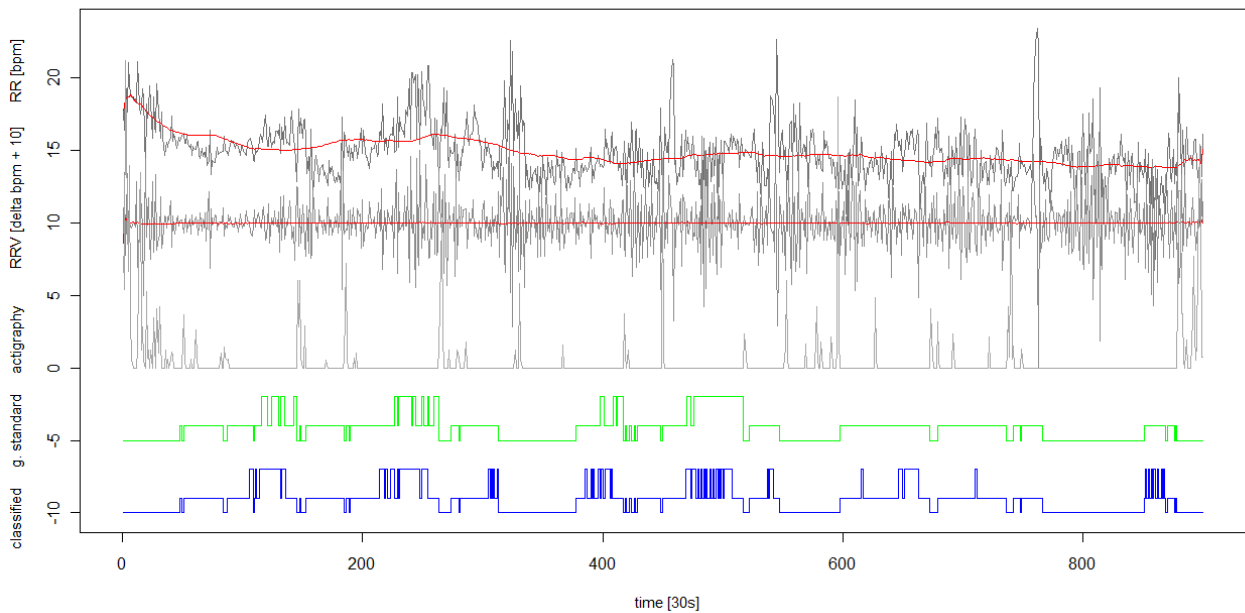
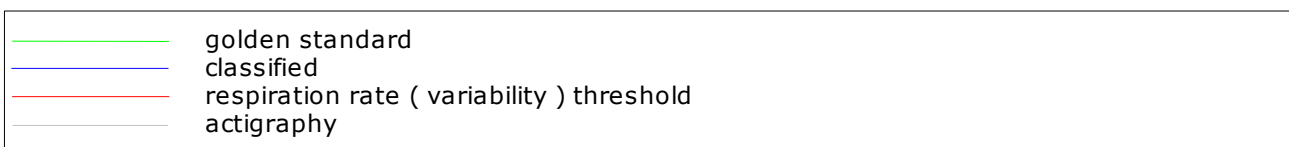


Diagram 77: Light sleep/deep sleep classification by module three for participant 310. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.



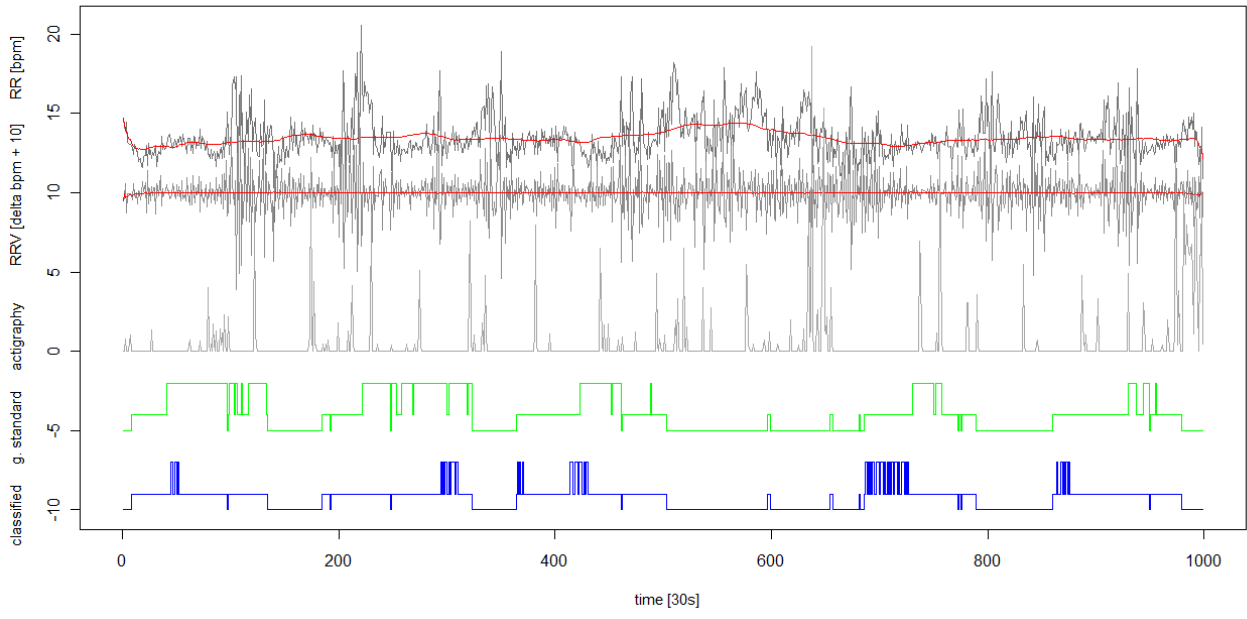
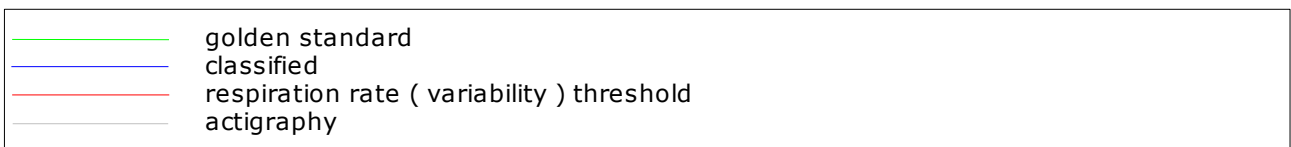


Diagram 78: Light sleep/deep sleep classification by module three for participant 311. Highest bars indicate deep sleep, middle bars light sleep, no bars indicate no NREM sleep. Whenever there exists body movement, given in grey, or the respiration rate (variability), given in grey, is respectively below or above the respective threshold, given in red, the epoch is classified as light sleep.



3-step hypnogram

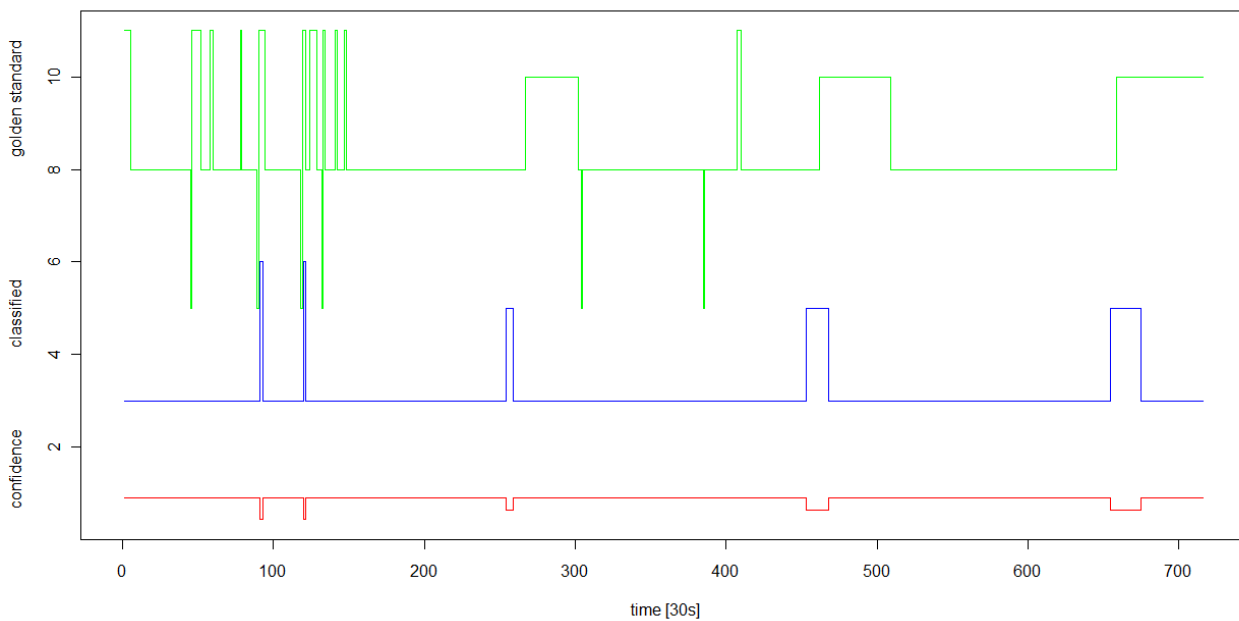


Illustration 79: Wake, REM and NREM sleep classification by the framework as a whole for participant 101. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.



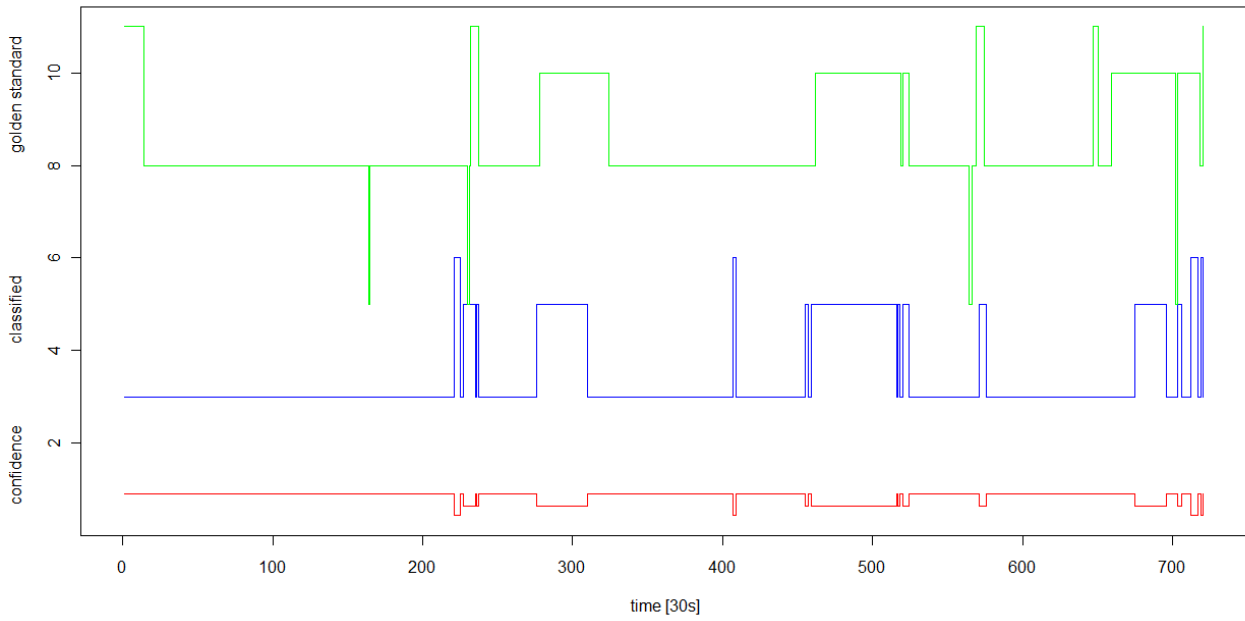


Diagram 80: Wake, REM and NREM sleep classification by the framework as a whole for participant 102. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.

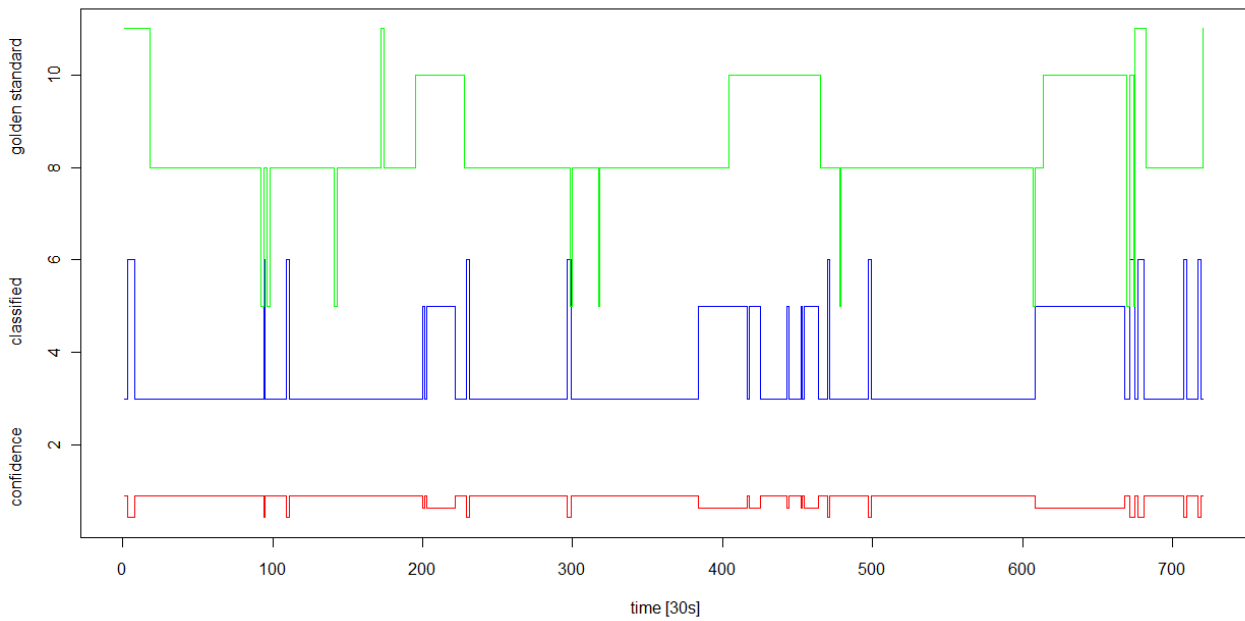


Diagram 81: Wake, REM and NREM sleep classification by the framework as a whole for participant 103. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.



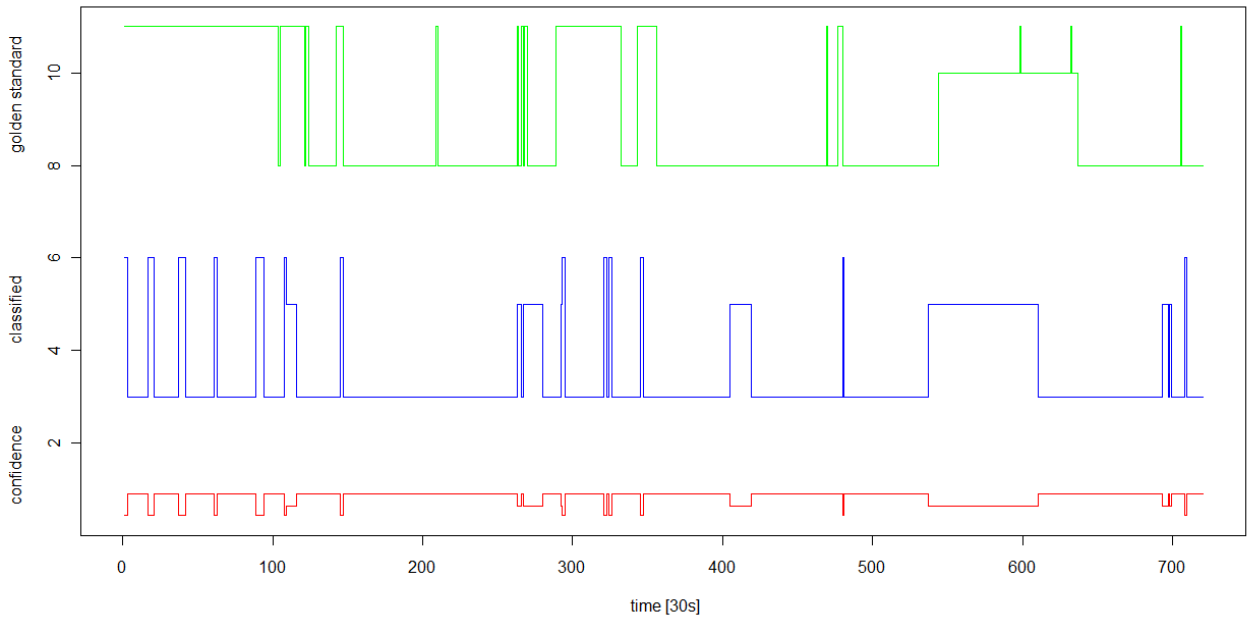


Diagram 84: Wake, REM and NREM sleep classification by the framework as a whole for participant 108. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.

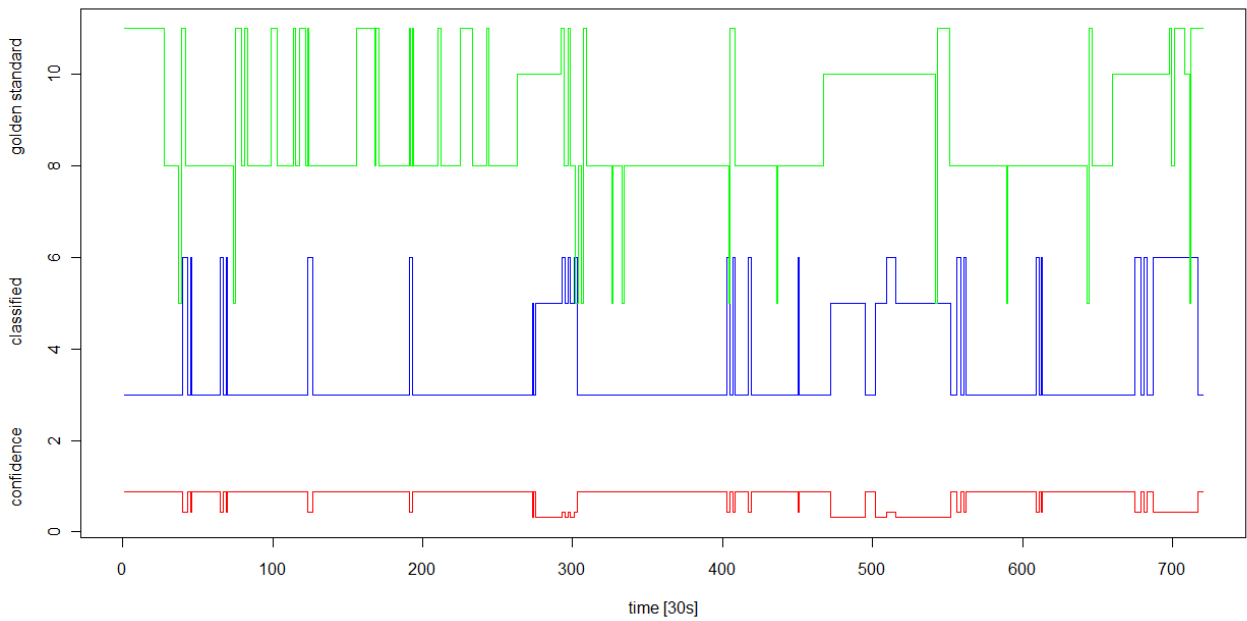


Diagram 85: Wake, REM and NREM sleep classification by the framework as a whole for participant 201. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.



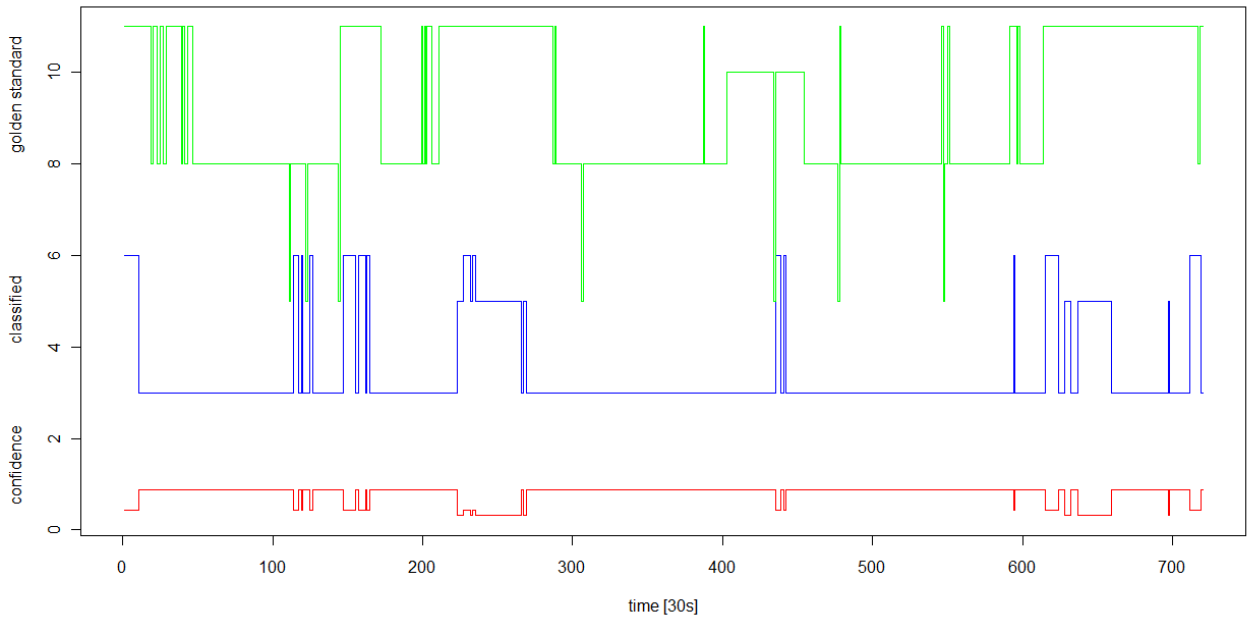


Diagram 86: Wake, REM and NREM sleep classification by the framework as a whole for participant 202. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.

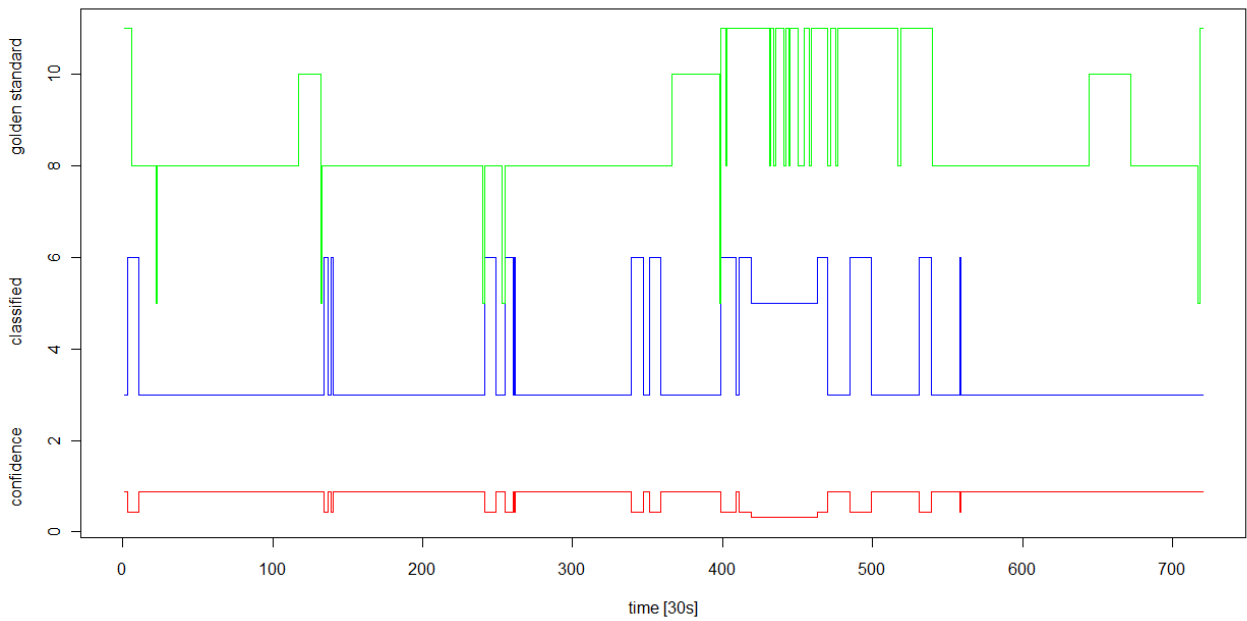


Diagram 87: Wake, REM and NREM sleep classification by the framework as a whole for participant 203. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.



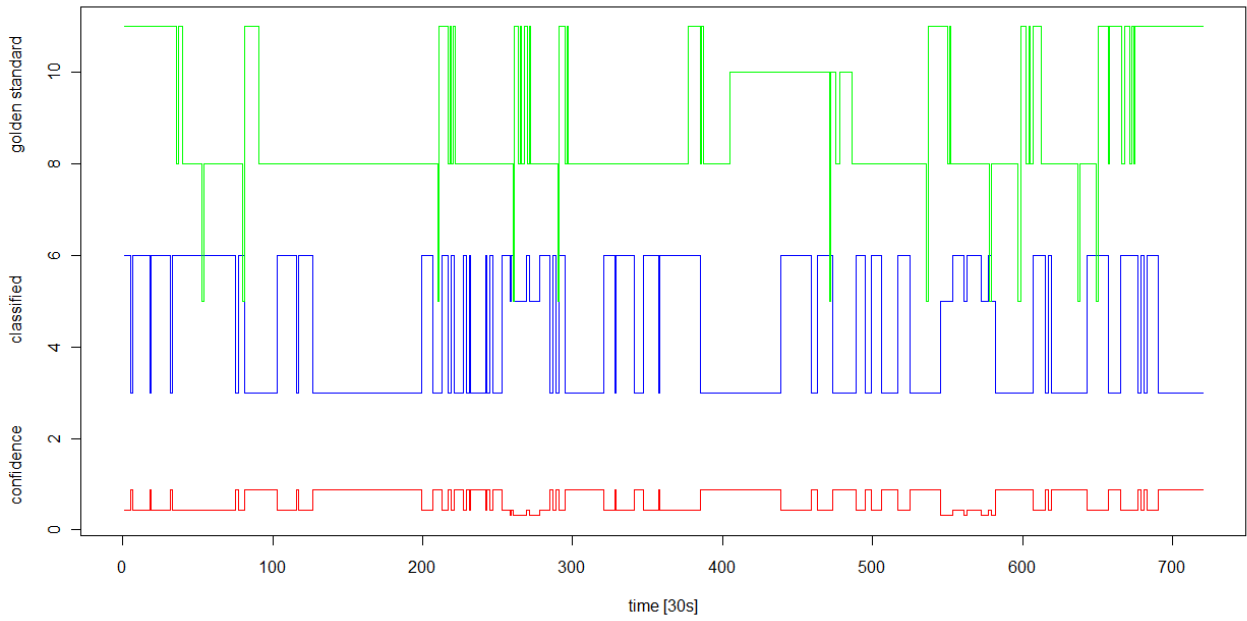


Diagram 88: Wake, REM and NREM sleep classification by the framework as a whole for participant 204. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.

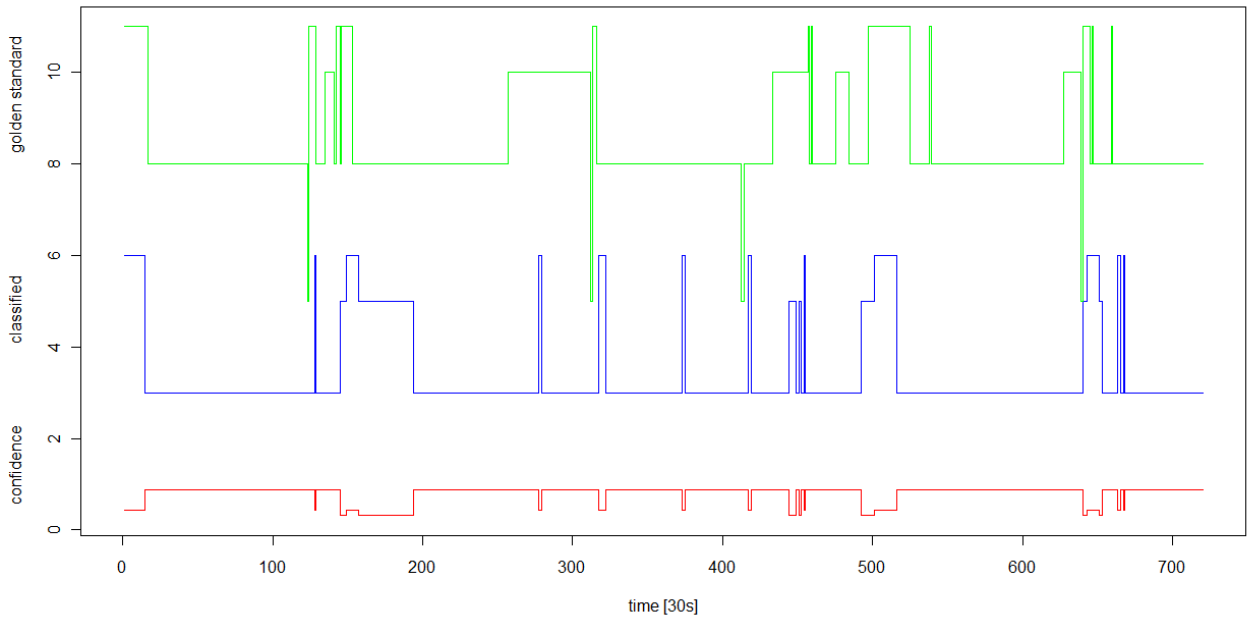
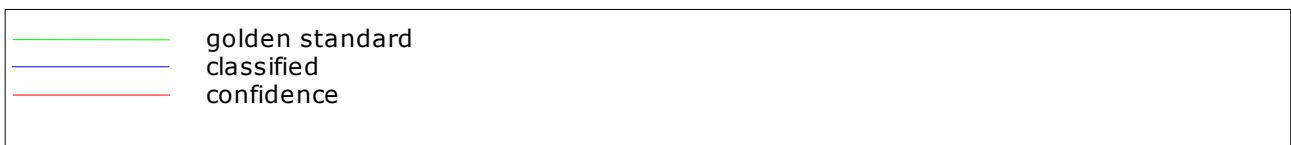


Diagram 89: Wake, REM and NREM sleep classification by the framework as a whole for participant 205. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.



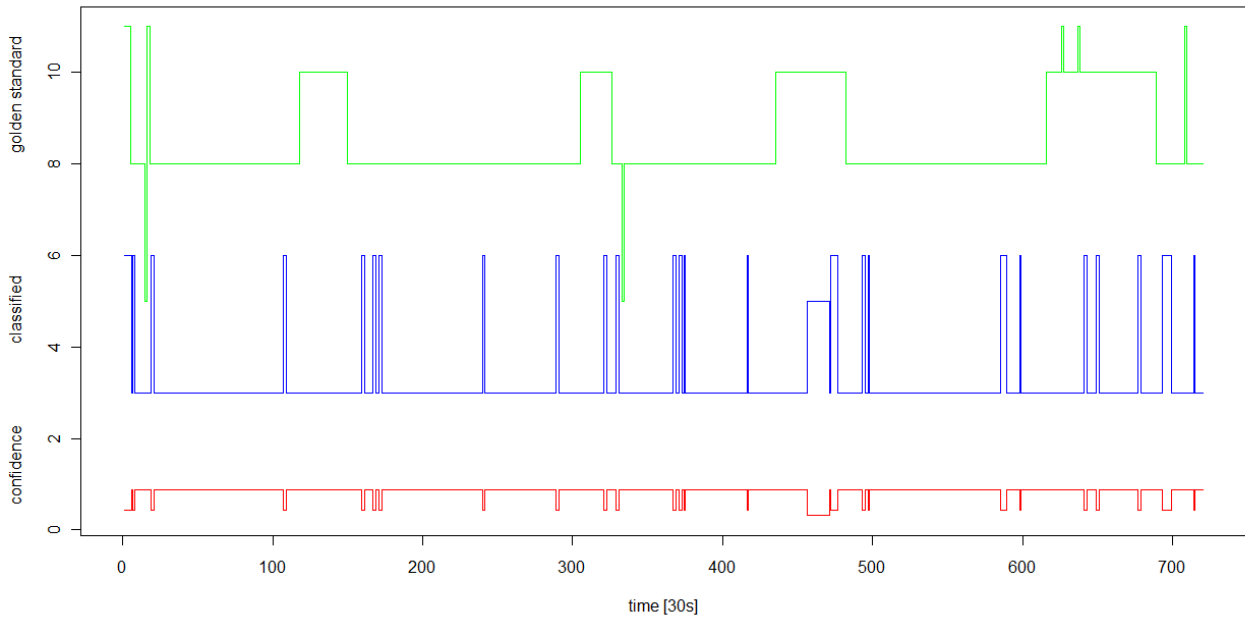


Diagram 90: Wake, REM and NREM sleep classification by the framework as a whole for participant 207. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.

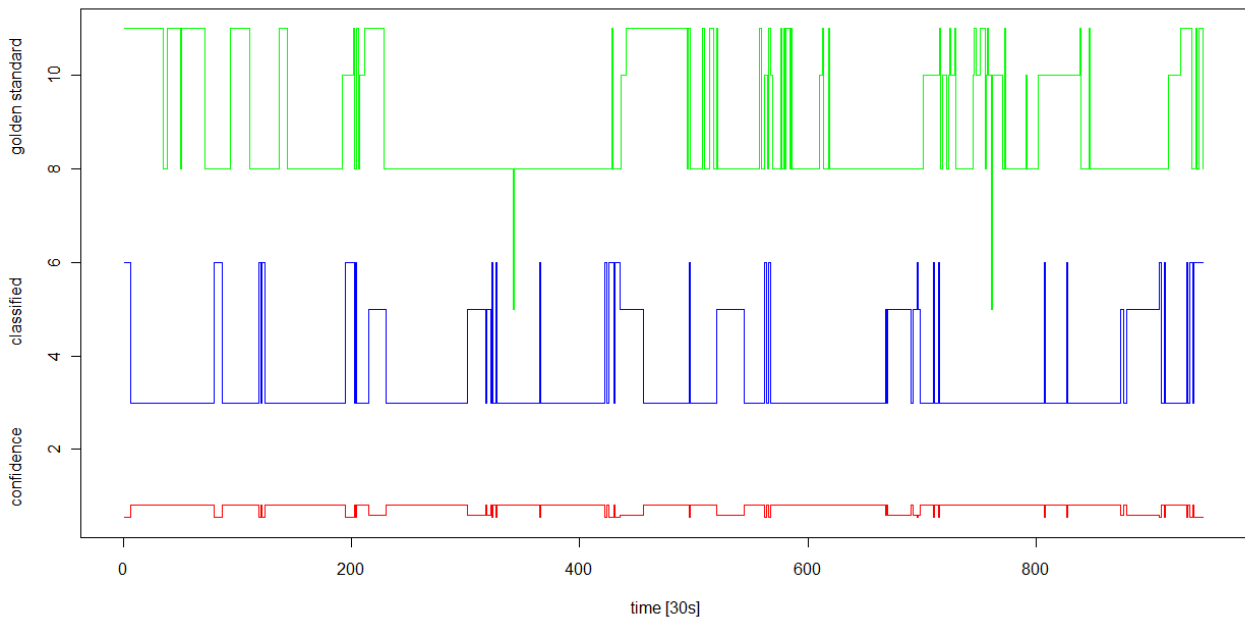
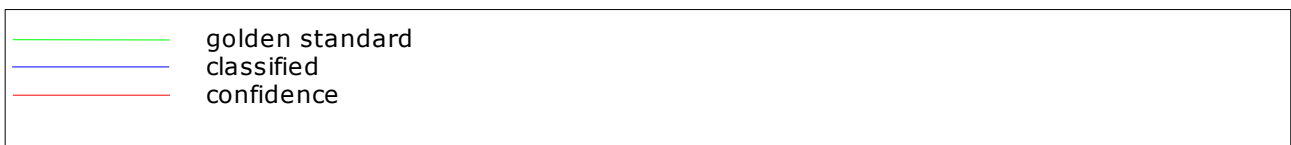


Diagram 91: Wake, REM and NREM sleep classification by the framework as a whole for participant 301. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.



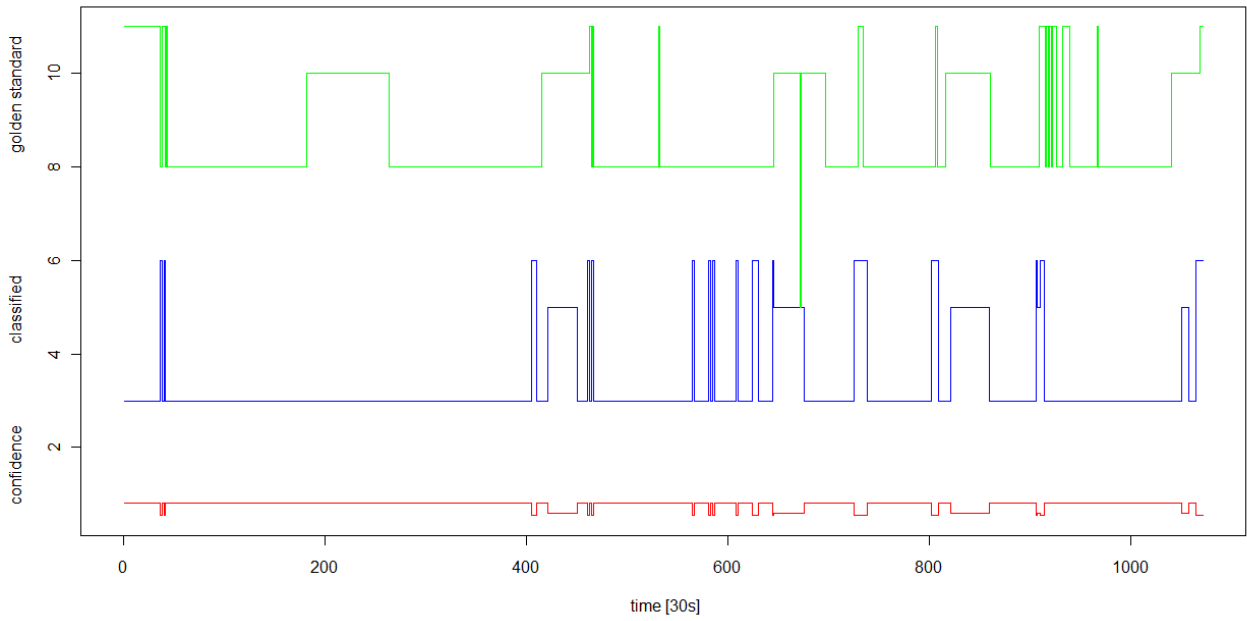


Diagram 92: Wake, REM and NREM sleep classification by the framework as a whole for participant 302. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.

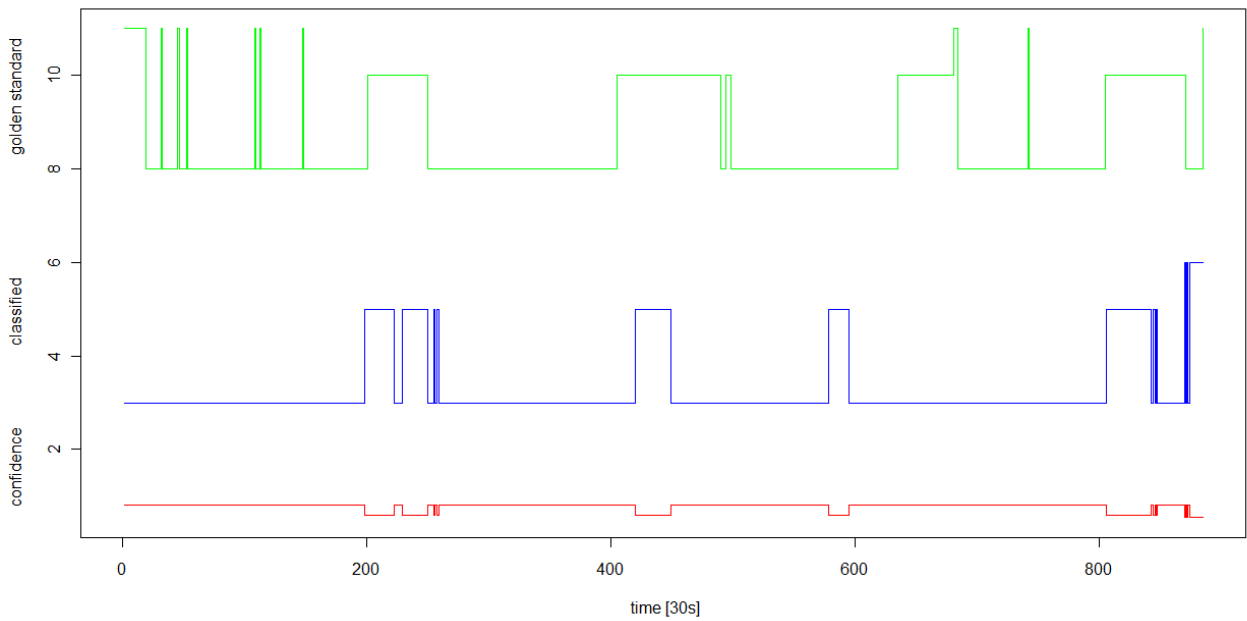


Diagram 93: Wake, REM and NREM sleep classification by the framework as a whole for participant 303. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.



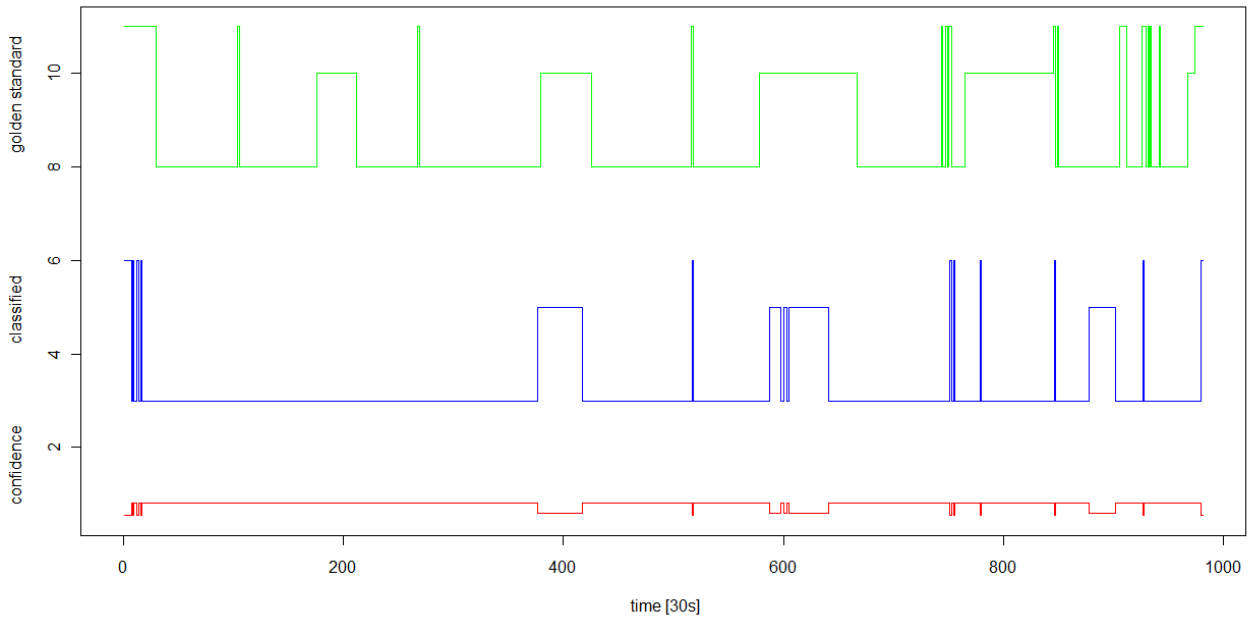


Diagram 94: Wake, REM and NREM sleep classification by the framework as a whole for participant 304. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.

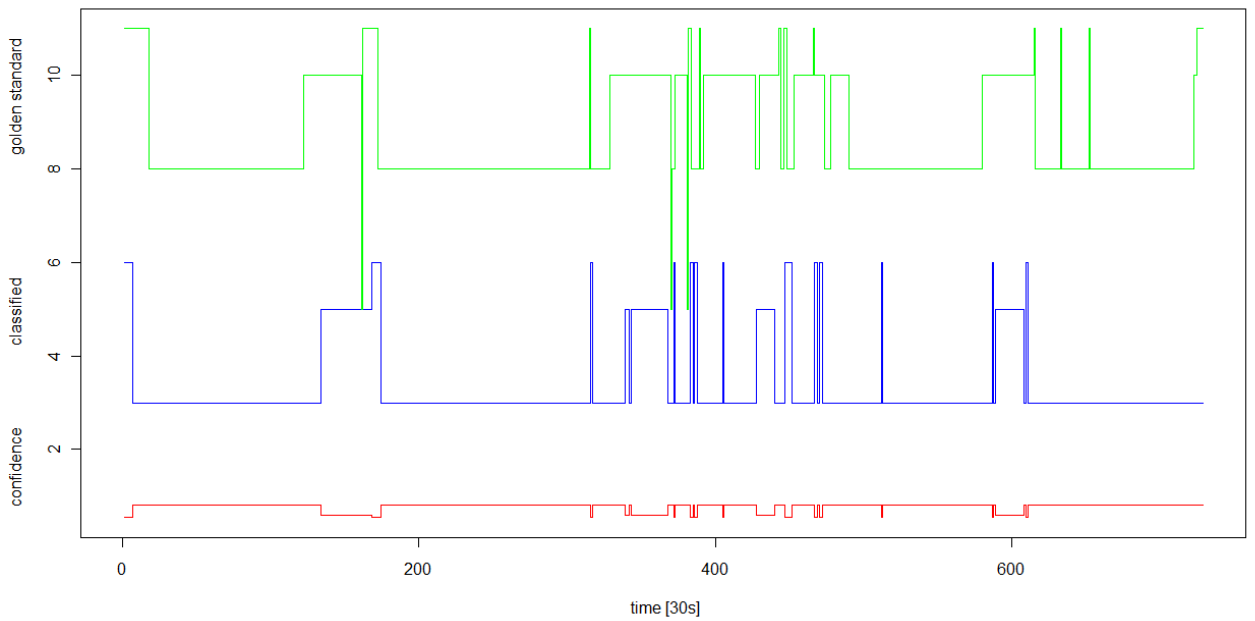


Diagram 95: Wake, REM and NREM sleep classification by the framework as a whole for participant 305. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.



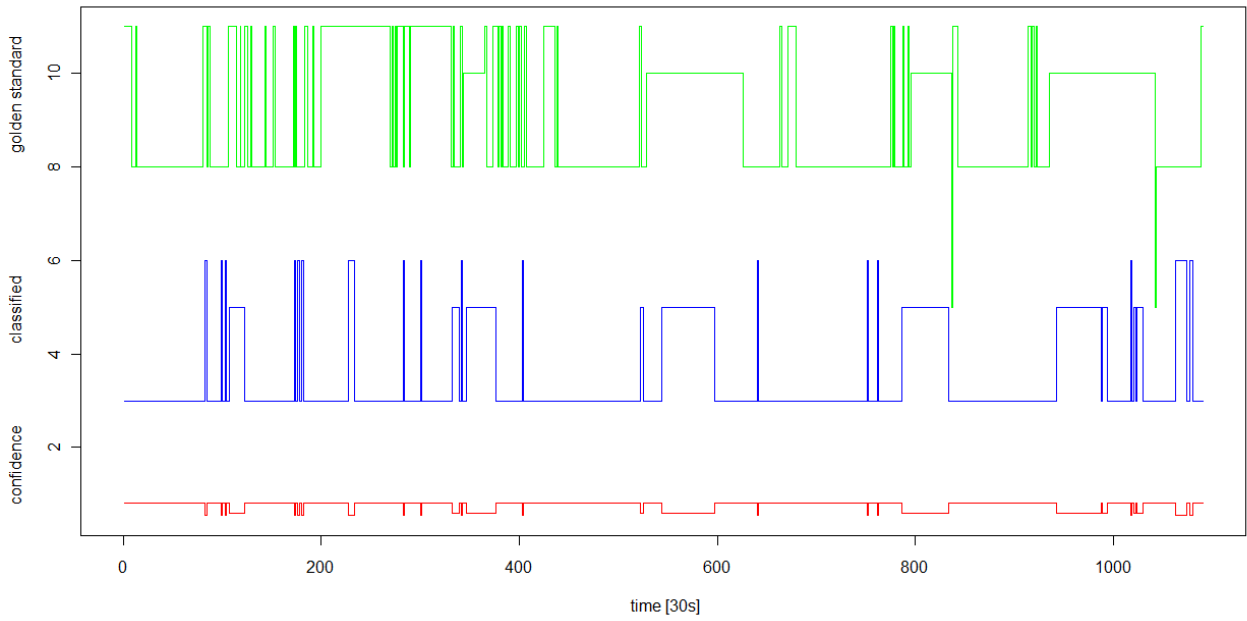


Diagram 96: Wake, REM and NREM sleep classification by the framework as a whole for participant 306. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.

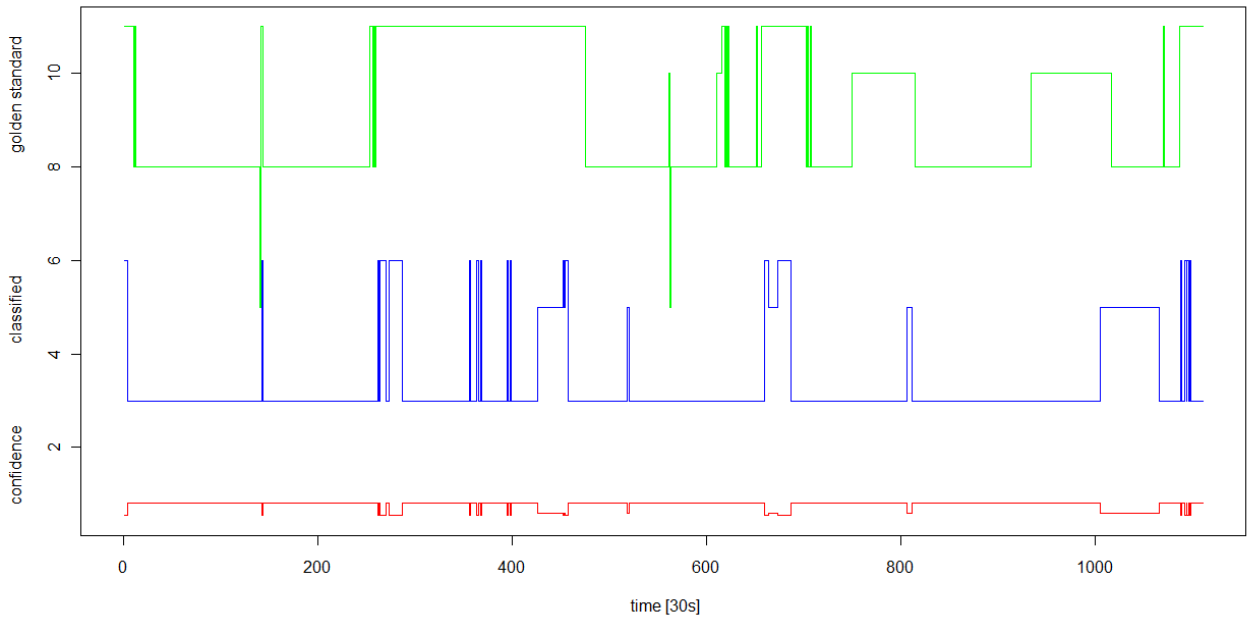


Diagram 97: Wake, REM and NREM sleep classification by the framework as a whole for participant 307. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.



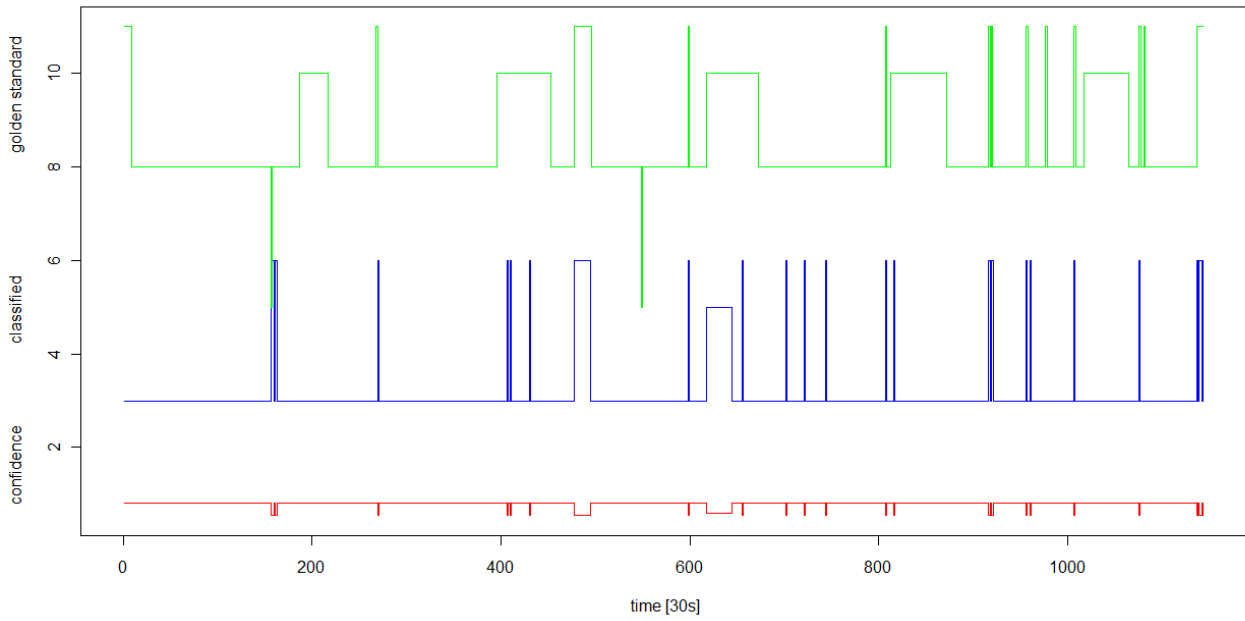


Diagram 98: Wake, REM and NREM sleep classification by the framework as a whole for participant 308. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.

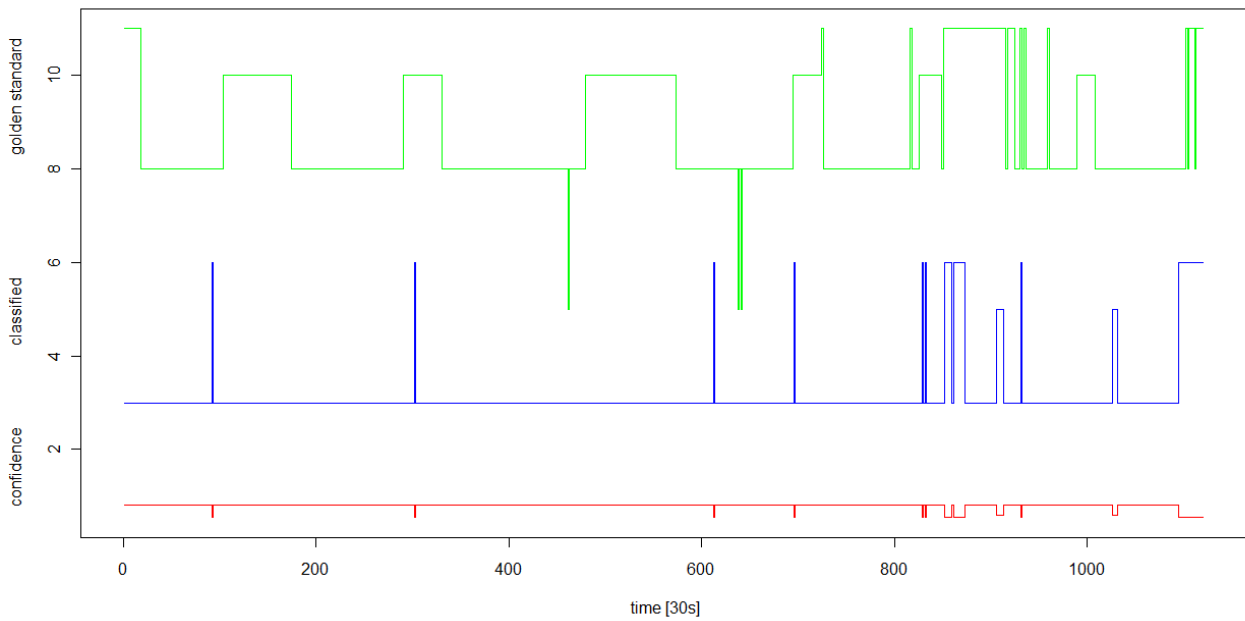


Diagram 99: Wake, REM and NREM sleep classification by the framework as a whole for participant 309. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.



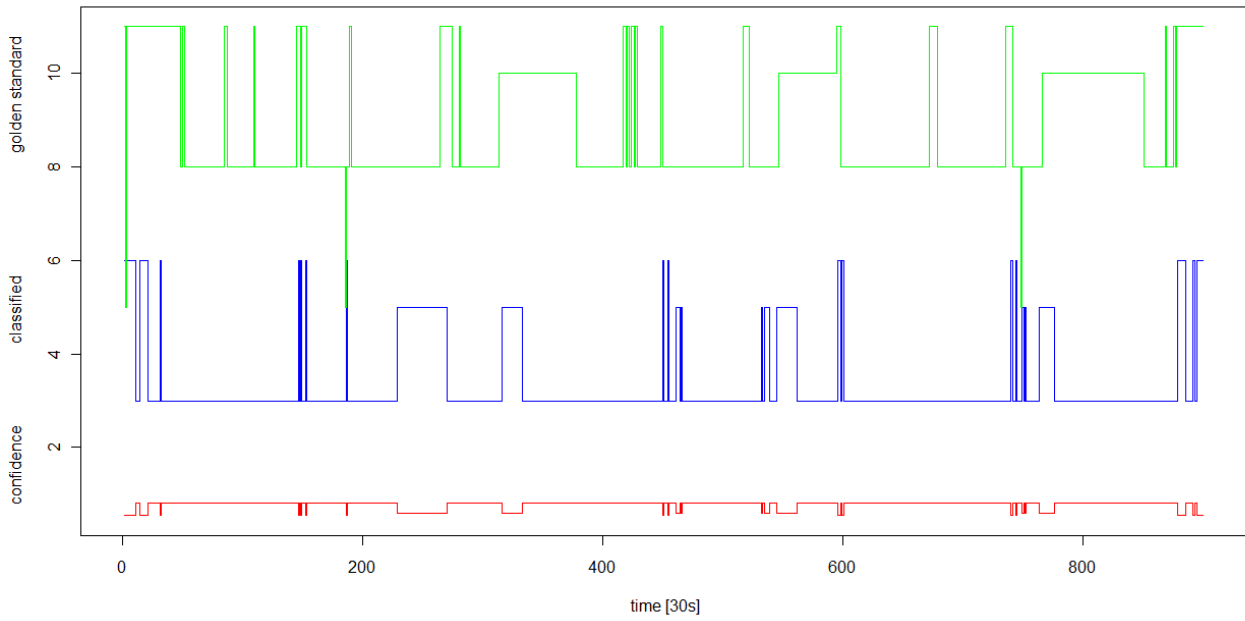


Diagram 100: Wake, REM and NREM sleep classification by the framework as a whole for participant 310. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.

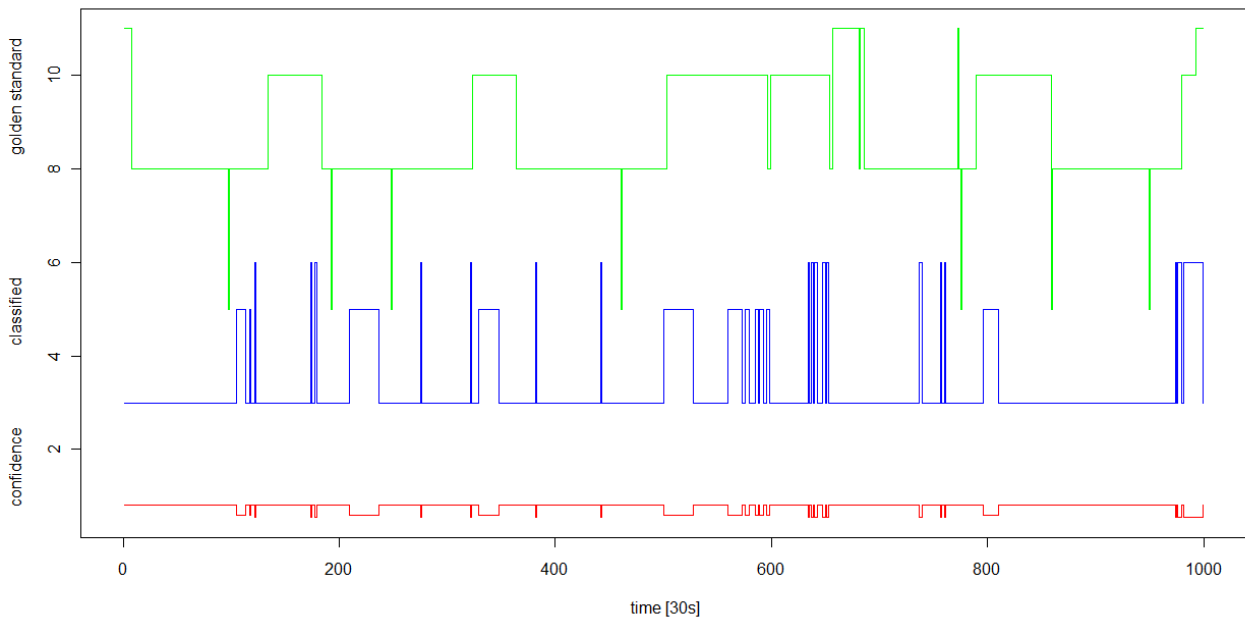


Diagram 101: Wake, REM and NREM sleep classification by the framework as a whole for participant 311. Highest bars indicate wake, middle bars indicate REM, lowest bars indicate NREM.

