Cuffless Blood Pressure Monitoring: Estimation of the Waveform and its Prediction Interval

by

Cederick Landry

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Mechanical and Mechatronics Engineering

Waterloo, Ontario, Canada, 2021

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Dr. Alberto Avolio Professor Emeritus Macquarie Medical School Macquarie University
Supervisors	Dr. Sean Peterson Professor Dept. of Mechanical and Mechatronics Engineering University of Waterloo
	Dr. Arash Arami Assistant Professor Dept. of Mechanical and Mechatronics Engineering University of Waterloo
Internal Members	Dr. Patricia Nieva Professor Dept. of Mechanical and Mechatronics Engineering University of Waterloo
	Dr. William Melek Professor Dept. of Mechanical and Mechatronics Engineering University of Waterloo
Internal-external Member	Dr. Richard Hughson Professor Emeritus Dept. of Kinesiology and Health Sciences University of Waterloo

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Cuffless blood pressure (BP) estimation devices are receiving considerable attention as tools for improving the management of hypertension, a condition that affects 1.13 billion people worldwide. It is an approach that can provide continuous BP monitoring, which is not possible with existing non-invasive tools. Therefore, it yields a more comprehensive picture of the patient's state. Cuffless BP monitoring relies on surrogate models of BP and the information encoded in alternative physiological measures, such as photoplethysmography (PPG) or electrocardiography (ECG), to continuously estimate BP. Existing models have typically relied upon pulse-wave delay between two arterial segments or other pulse waveform features in the estimation process. However, the models available in the literature (1) provide an estimation of the systolic BP (SBP), diastolic BP (DBP), and mean BP (MAP) only, (2) are validated solely in controlled environments, and (3) do not assign a confidence metric to the estimates. At this point, cuffless methods are not used by clinicians due to their inaccuracy, the validation inadequacy, and/or the unevaluated uncertainty of the existing methods.

The first objective of this thesis is to develop a cuffless modeling approach to estimate the BP waveform from ECG and PPG, and extract important BP features, such as the SBP, DBP, and MAP. Access to the full waveform has significant advantages over previous cuffless BP estimation tools in terms of accuracy and access to additional cardiovascular health markers (e.g., cardiac output), as well as potentially providing arterial stiffness. The second objective of this thesis is to validate cuffless BP estimation during activities of daily living, an uncontrolled environment, but also in more challenging physiological conditions such as during exercise. Such validation is important to increase confidence in cuffless BP monitoring, it also helps understand the limitation of the method and how they would affect clinical outcomes. Finally, in an effort to improve confidence in the cuffless BP estimation framework (third objective), a prediction interval (PI) estimation method is introduced. For potential clinical uses, it is imperative to assess the uncertainty of the BP estimate for acute outcome evaluation and it is even more so if cuffless BP is to be employed outside of the clinic.

In this thesis, user-specific nonlinear autoregressive models with exogenous inputs (NARX) are implemented using an artificial neural network (ANN) to predict the BP waveforms using ECG and/or PPG signals as inputs. To validate the NARX-based BP estimation framework during activities of daily living, data were collected during six-hours testing phase wherein the participants go about their normal daily living activities. Data are further collected at four-month and six-month time points to validate long-term performance. To broaden the range of BP in the training data, subjects followed a short

procedure consisting of sitting, standing, walking, Valsalva maneuvers, and static handgrip exercises. To evaluate the uncertainty of the BP estimates, one-class support vector machines (OCSVM) models are trained to cluster data in terms of the percentage of outliers. New BP estimates are then assigned to a cluster using the OCSVMs hyperplanes, and the PIs are estimated using the BP error standard deviation associated with different training data clusters. The OCSVM is used to estimate the PI for three BP model architectures: NARX models, feedforward ANN models, and pulse arrival time (PAT models). The three BP estimations from the models are fused using the covariance intersection fusion algorithm, which improves BP and PI estimates in comparison with individual model performance.

The proposed method models the BP as a dynamical system leading to better accuracy in the estimation of SBP, DBP and MAP when compared to the PAT model. Moreover, the NARX model, with its ability to provide the BP waveform, yields more insight into patient health. The NARX model demonstrates superior accuracy and correlation with "ground truth" SBP and DBP measures compared to the PAT models and a clear advantage in estimating the large range of BP. Preliminary results show that the NARX models can accurately estimate BP even months apart from the training. Preliminary testing suggests that it is robust against variabilities due to sensor placement. The employed model fusion architecture establishes a method for cuffless BP estimation and its PI during activities of daily living that can be used for continuous monitoring and acute hypotension and hypertension detection.

The NARX model, with its capacity to estimate a large range of BP, is next tested during moderate and heavy intensity exercise. Participants performed three cycling exercises: a ramp-incremental exercise test to exhaustion, a moderate and a heavy pseudorandom binary sequence exercise tests on an electronically braked cycle ergometer. Subject-specific and population-based NARX models are compared with feedforward ANN models and PAT (and heart rate) models. Population-based NARX models, when trained on 11 participants' three cycling tests (tested on the participant left out of training), perform better than the other models and show good capability at estimating large changes in MAP. A limitation of the approach is the incapability of the models to track consistent decreases in BP during the exercise caused by a decrease in peripheral resistance since this information is apparently not encoded in either the forehead PPG or ECG signals. Nevertheless, the NARX model shows good precision during the whole 21 minutes testing window, a precision that is increased when using a shorter evaluation time window, and that can potentially be even further increased if trained on more data.

The validation protocols and the use of a confidence metric developed in this thesis is of great value for such health monitoring application. Through such methodology, it is hoped that cuffless BP estimation becomes, one day, a well-established BP measurement method.

Acknowledgements

I would like to thank my supervisors, Sean Peterson and Arash Arami, for providing guidance and insight throughout this endeavor. Thank you to my examination committee members, Professors Alberto Avolio, Patricia Nieva, Richard Hughson, and William Melek, for investing the time to read this document, the valuable questions posed, and the suggested improvements. I would like to thank Arami's Lab members and Richard Hughson's Lab members for their many recommendations. I would like to give a special thanks my graduate school colleague Eric Hedge for his help with my research, especially for data collection.

Finally, to my wife, Anabel, thank you for supporting me in this amazing opportunity and the ones to come.

Table of Contents

List of Figures

.

List of Tables

List of Abbreviations

AAMI	Association for the Advancement of Medical Instrumentation
ABP	Arterial Blood Pressure
AIx	Augmentation Index
ANN	Artificial Neural Network
AR	Autoregressive Model
ARX	Autoregressive with Exogenous Inputs Model
BCG	Ballistocardiogram
BHS	British Hypertension Society
BP	Blood Pressure
BPW	Blood Pressure Waveform
CO	Cardiac Output
CVP	Central Venous Pressure
DBP	Diastolic Blood Pressure
ECG	Electrocardiography
ESH	European Society of Hypertension
GRU	Gated Recurrent Unit
HR	Heart Rate
L-H	Low-to-Heavy Exercise
L-M	Low-to-Moderate Exercise
LSE	Least Square Estimate
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error

MAP	Mean Arterial Pressure
MLP	Multilayer Perceptron
MSE	Mean Square Error
MTD	Mean Training Data
NARX	Nonlinear Autoregressive with Exogenous Inputs Model
OCSVM	One-Class Support Vector Machines
PAT	Pulse Arrival Time
PEP	Pre-ejection Period
PI	Prediction Interval
PP	Pulse Pressure
PPG	Photoplethysmography
PRBS	Pseudorandom Binary Sequence
PTT	Pulse Transit Time
PWV	Pulse Wave Velocity
RIA	Research Institute for Aging
RNN	Recurrent Neural Network
RTF	Return-to-Flow
SBP	Systolic Blood Pressure
SCG	Seismogram
SD	Standard Deviation
SV	Stroke Volume
SVM	Support Vector Machines

List of Symbols

$\widehat{\mathrm{BP}}$	Estimated Blood Pressure
$\dot{\mathrm{V}}\mathrm{O}_2$	Oxygen Uptake
$\dot{\mathrm{V}}\mathrm{O}_{2\mathrm{peak}}$	Peak Oxygen Uptake
σ_{T}	Standard Deviation Threshold
λ_i	Normalize Error Distribution
λ^{α}	Value of λ_i at the α^{th} percentile
μ_e	Mean Error
σ_e	Error Standard Deviation
C	Compliance
D	Damping Factor
Err_{σ}	Error on the Blood Pressure Standard Deviation
f	Frequency
I	Blood Flow
K	Blood Pressure Waveform Transfer Function Gain
K	Timestep
KL	Kullback–Leibler Divergence
M	Bias
$\mathrm{MAE}_{\Delta\mathrm{MAP}}$	Mean Absolute Error in Estimating Change in Mean Arterial Pressure
$\mathrm{MAE}_{\Delta\mathrm{P}}$	Mean Absolute Error in Estimating Change in Blood Pressure
MeRCI	Mean Rescaled Confidence Interval
nMAE	Normalize Mean Absolute Error
p	Null Hypothesis Significance testing

R	Resistance
r	Pearson Correlation Coefficient
R^2	Coefficient of Determination
t	Time
VT	Ventilatory Threshold
α	Comparison Significance Level
ΔP	Change in Blood Presssure
ν	One-Classe Support Vector Machines Regularization Parameter
P	Estimate Covariance
ξ	Mean Estimate
ω	Covariance Intersection Algorithm Weights

.

Chapter 1
Introduction

Hypertension, commonly known as high blood pressure (BP), is a condition that affects 1.13 billion people worldwide [1] and is the cause of 9.4 million deaths worldwide every year, which accounts for half of the deaths associated with stroke and heart disease [2]. Hypertension is a major cardiovascular risk factor that is treatable, yet the detection and control rates are low [3], [4]; only 46% of individuals with hypertension have their condition under control, primarily due to inadequate treatment [5].

Cuffless BP measurement devices are receiving considerable attention as tools to improve management of hypertension [6], since they require less cumbersome and expensive hardware [7] compared with cuff-based systems, which enables patient mobility, and eliminates the discomfort associated with the cuff [8]. Cuffless systems also have the potential to continuously monitor BP, which can be used to predict cardiovascular risks, such as short-term BP variability [9], and to better manage treatment [10]. Moreover, continuous BP monitoring can minimize false diagnoses due to isolated clinic hypertension (i.e., white-coat hypertension) [11].

Generally speaking, cuffless BP relies on the information encoded in alternative physiological measures, such as photoplethysmography or electrocardiography, to continuously estimate BP. This requires surrogate models of BP, which have typically relied upon pulse-wave delay between two arterial segments [6] or other pulse waveform features [12] in the estimation process. Recently, a few studies reported the use of end-to-end deep learning to handle feature extraction with the aim to cope with the difficulty of designing hand-crafted features to characterize BP [13].

Existing models for cuffless BP monitoring only provide systolic (BP peak) and diastolic (BP trough) estimations, with a couple of noted exceptions that also enables the estimation of the mean arterial pressure (MAP) [14], [15] or the pulse pressure [16]. However, considerable clinically-relevant data are lost in such methods. To date, no model of the dynamics of BP has been reported to estimate the complete BP waveform, which also contains additional information of potential interest to physicians [17]. Specifically, in the case of hypertension, the waveform provides information on the underlying disease and mechanisms leading to high blood pressure [18]. For example, information regarding arterial stiffness, which can be extracted from the BP waveform [60], can inform optimal hypertension treatment. Another example is pulse diagnosis, which utilizes the blood pressure waveform to identify different cardiovascular diseases and track pathological changes [19]. Moreover,

the complete BP waveform could give access to other important continuous cardiovascular health measures, such as cardiac output [20] and stroke volume [21].

Cuffless BP monitoring aims to: 1) render more accessible discrete BP measurements usually obtained with an arm cuff, and/or 2) offer convenient access to continuous BP measurements. The method should ideally not require recalibration, or at the very least, the calibration should remain valid for several hours [22], [23]. To enable access to the continuous BP, the system can be calibrated on each user, after which the BP calibration device is removed (e.g., Finapres® NOVA) and beat-to-beat BP estimation is obtained from a surrogate signal (e.g., photoplethysmography). While many studies have demonstrated continuous BP monitoring, validation is often performed for time scales on the order of minutes in controlled environments [24]–[26]. To the author's knowledge, there are only few noted exceptions in [15], [27], [28], wherein 24h BP estimation was performed by sampling BP every 30 minutes; however, their estimates during daily activities were determined to lack sufficient accuracy for clinical use, and as such the results focused on rest periods at night. Therefore, estimation of BP during activities is lacking and would be clinically valuable.

Another critical aspect of cuffless BP monitoring that has been overlooked thus far is the uncertainty associated with the estimation. The uncertainty in cuff-based measurements has been of concern for practitioners [29] which led to the development of methods for estimating uncertainty on such measurements [30]. The standard for expressing uncertainty for measurement devices [31] states that it should be applicable to a broad span of conditions. In practice, however, the uncertainty is obtained from a series of measurements under repeatable conditions, which do not necessarily capture every use case of the device. For instance, the British Hypertension Society and the Association for the Advancement of Medical Instrumentation (AAMI) have specific protocols for testing cuff-based BP measurement devices [32], which have been co-opted for cuffless BP estimation. However, the precision reported in specific tested situations, such as in-lab conditions, is unlikely to be reproducible in real-world conditions, which are potentially more challenging. Since continuous BP monitoring is likely to be employed outside of the clinic, it is imperative to assign a confidence metric to the estimates for outcome evaluation.

1.1 Motivation and Objectives

This Ph.D. thesis research aims to introduce a novel BP estimation strategy that overcomes the major hurdles associated with previously established techniques, namely the short prediction horizon (loosely captured in the ratio of required training time for a given prediction window) of deep machine learning

algorithms that renders them of questionable practical use [13], and the relatively poor performance of physiology-based models in comparison with medical standards [6]. More precisely, this work aims to combine knowledge in dynamical systems, physiology, and machine learning to develop a solution readily applicable to the real world. BP is a phenomenon with rich dynamics, yet existing models only provide SBP and DBP. It is hypothesized that by capturing subject-specific BP dynamics using data-driven methods that estimation accuracy will be improved and the prediction horizon will be extended. However, the flexibility offered by complex data-driven modeling can have the cost of decreasing the confidence in the model, as black-box models might lead to suspicion in the results due to the loss of input to output interpretability, which is of great importance to clinical providers [33]. Therefore, to be adopted and fully utilized in clinical settings, practitioners need to be exceptionally confident in the data-driven model outputs used for BP monitoring. Most efforts to date in the field of cuffless BP monitoring have focused on algorithm development, with very little consideration for increasing the confidence in the models. Thus, it is presumed that incorporation of a confidence metric into a data-driven model framework will increase the likelihood of clinical adoption of cuffless BP monitoring.

To address the identified critical gaps in the state-of-the-art of cuffless blood pressure monitoring, the following research objectives are proposed:

1. develop a data-driven model of BP dynamics to enable estimation of the complete BP waveform;
2. validate the model architecture in real-world scenarios;
3. incorporate an uncertainty measure into the modeling scheme.

1.2 Outline of the Thesis

The thesis is organized as follows. Chapter 2 presents a review of the literature pertinent to cuffless BP estimation. The core contributions of this thesis are presented in Chapters 3 to 6. In Chapter 3, the model architecture is introduced for accurate cuffless estimation of BP and tested for bed-ridden subjects. Chapter 4 presents a protocol to train the model architecture that enables BP estimation during activities of daily living from wearable sensors data as input. Chapter 5 deals with estimating the uncertainty associated with cuffless BP monitoring and improving BP estimates using estimation fusion. Chapter 6 considers the challenge of BP estimation during exercise using both subject-specific and population-based modeling approaches. Finally, Chapter 7 presents the major conclusions drawn from this body of work, in addition to providing some suggestions for future research.

Chapter 2
Literature Review

This chapter introduces the concepts and theories needed to design and validate a cuffless BP monitoring system. It is organized as follows: first, an overview of the cardiovascular system focusing on BP is presented with a description of some common physiological sensors. Second, data-driven models are explained in the context of this research project. Finally, a review of cuffless BP monitoring literature is presented.

2.1 Cardiovascular Physiology

This section provides a brief introduction to cardiovascular physiology with an emphasis on BP and how it can be monitored. It is divided as follows: (1) a high level overview of the cardiovascular system is presented; (2) BP is described, including internal and external factors that affect it; (3) physiological parameters that can be extracted from the BP waveform are discussed; and (4) the main categories of physiological sensors, with the emphasis on blood pressure, blood pulse, and electrocardiogram measurements are presented.

2.1.1 Cardiovascular System Overview[1]

The cardiovascular system is comprised of two circulatory loops, the pulmonary and the systemic circulations. The pulmonary system sends deoxygenated blood from the right ventricle of the heart to the lungs, wherein the blood is oxygenated. The oxygenated blood then returns back to the left atrium of the heart. The left atrium feeds the left ventricle of the heart, which then pumps the oxygenated blood into the systemic circulation. The oxygenated blood travels through the arterial tree, delivering oxygen to the capillary networks of various organs and limbs of the body. The veins transport the deoxygenated blood back to the heart's right atrium, which feeds the right ventricle, thus completing the circuit.

A two-element Windkessel model is a simplified version of the cardiovascular system that relates the blood flow to BP, the mechanical force driving blood flow, by the main cardiovascular properties: arterial compliance and peripheral resistance. The former is modeled as a capacitor and the latter as a resistor in a parallel electrical circuit, see Figure 2.1. The dynamical equation relating the flow, *I(t)*, to BP(*t*) is shown in equation (2.1.

[1] The content of this section is primarily from [9] and [10].

$$I(t) = \frac{BP(t)}{R} + C\frac{dBP(t)}{dt} \tag{2.1}$$

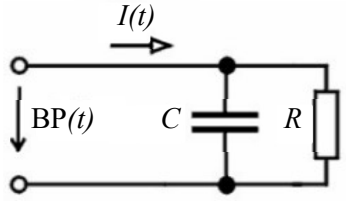

Figure 2.1 - Electrical analog of the 2-element Windkessel model

The blood flow in the cardiovascular system is pulsatile in nature (half sin wave during 2/5 of the cardiac cycle, elsewhere 0 in a Windkessel model). From equation (2.1, for a given *I(t)*, mean BP will increase with the peripheral resistance (*R*) and that the peak BP will decrease with an increase in arterial compliance (*C*). Other important variables characterizing the cardiovascular system are the stroke volume (SV) - the volume of blood ejected from the heart due to every contraction of the heart muscle, and the cardiac output (CO) - the amount of blood the heart pumps per minute. CO is usually expressed in litres per minute (L/min) and can be computed as CO = SV * heart rate [34]. The next subsection focuses on the physiology of BP and what affects its level.

2.1.2 Blood Pressure

The BP at various locations in the systemic circulatory system is shown in Figure 2.2, where the aorta is the artery fed by the left ventricle, and the vena cavae is the vein feeding the right atrium (returning the blood to the heart). The generic term blood pressure generally refers to arterial blood pressure (ABP), which is the complete waveform that is characterized by the change in pressure due to every heart contraction. This waveform has many characteristics, the most important of which are its maximum - the systolic (SBP), its minimum - the diastolic (DBP), the mean arterial pressure (MAP), and the pulse pressure (SBP-DBP). The normal ranges for SBP and DBP in the brachial artery (the location commonly measured by external BP cuffs; see Section 2.2.1.2) are 90-120 and 60-80 mmHg, respectively, in young, healthy normotensive persons but is elevated above this range in hypertension. [36]. The ABP is not to be confused with central venous pressure (CVP) that refers to the BP when the blood returns to the heart after flowing in the systemic circulation. A normal CVP is between 8 to 12 mmHg in supine position, and is correlated with how much blood is going back to the heart [37]. Although the BP differs for different locations in the body, Figure 2.2 highlights that the BP is high on the arterial side (red) and low on the venous side (blue).

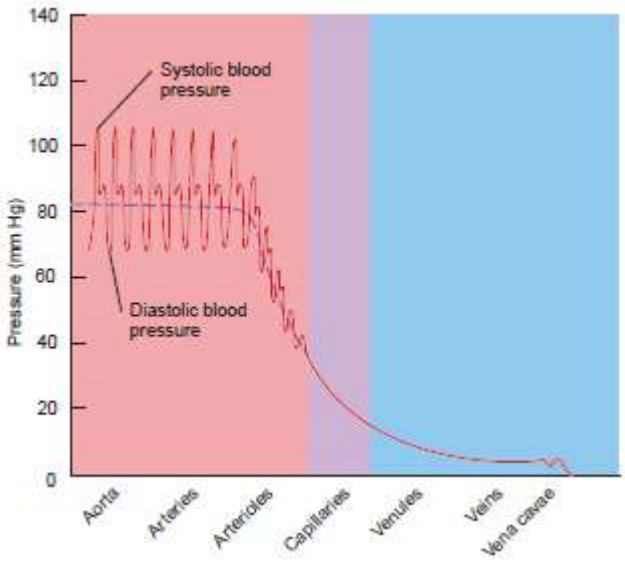

Figure 2.2 - Blood pressures in various parts of the cardiovascular system. The dashed line is the mean (average) blood pressure in the aorta, arteries, and arterioles. This figure was taken from *Human Physiology: An Integrated Approach* [35]. Note that pulse pressure should increase along the aorta, but this is not shown herein.

BP level varies between individuals, and also changes over time. The former is characterized by the difference in cardiovascular properties between individuals, where the latter comes from the states of the body, e.g., posture and activity level, over time. The factors influencing BP level are important considerations for a cuffless BP monitoring system.

2.1.2.1 Cardiovascular Properties Influencing Blood Pressure

According to the American Heart Association, there are five blood pressure categories, ranging from normal to hypertensive crisis [38]. Hypotension, or low BP, which is associated with persistent tiredness, is excluded from this categorization since treatment is not suggested as either possible or necessary [39] when it is not associated with organ disfunction. The two main cardiovascular properties affecting BP level are arterial stiffness and peripheral resistance of the cardiovascular system [40]. The latter describes how the arteries changes in diameter to a change in BP. The former describes how difficult it is to pump blood in the cardiovascular system.

Speculation continues as to whether an increase in arterial stiffness is a cause or a consequence of hypertension [41]. Nevertheless, this increase perturbs the natural mechanism, called impedance mismatch in the arteries, protecting the small capillaries from excessive pulsatility. This mechanism

changes the BP waveform as shown in Figure 2.3. In short, the pulse wave velocity (PWV) increases with stiffness, which increases progressively from the aorta to the large peripheral arteries. This, together with the tapering of the aortic diameter (impedance mismatch), induces wave reflections. These reflections are beneficial as they reduce the energy transmitted to the microcirculatory network, which is more fragile. With hypertension this stiffness gradient is reduced, see inversed, due proximal aortic stiffness increasing up to—and sometimes beyond—the level of aortic distal distal stiffness, leading to a higher PWV throughout the arteries. The reflected waves arrive at the central aorta in early systole and are superimposed on the incident pressure wave, a phenomenon which increases SBP, and therefore increases pulse pressure (SBP-DBP). It is to be noted that the complexity of different arterial sites and their varying distances from the aorta result in reflected waves with large phase differences, which tend to mitigate the cumulative effects of these waves on the central BP waveform [42]. Nevertheless, the overall BP waveform is a sum of reflected and forward traveling waves such that there is an increase in the observed aortic BP wave, as demonstrated in Figure 2.3, a condition that worsens with age and is correlated with arterial stiffness [41].

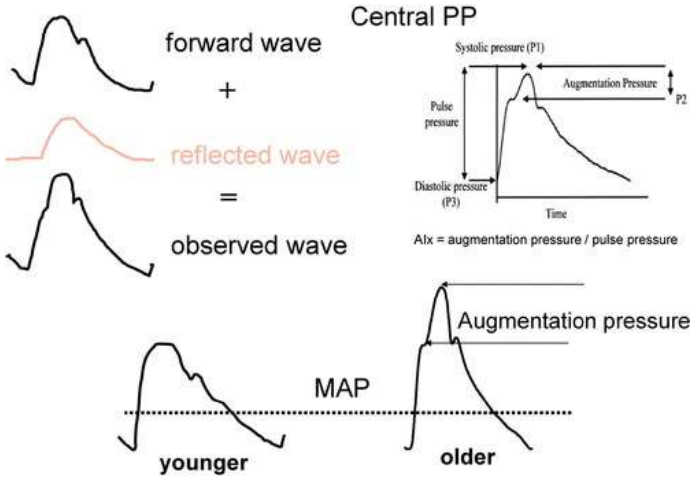

Figure 2.3 - Upper left: the summation of a forward and a backward wave is responsible for the total blood pressure (BP) curve. Upper right: a schematic representation of the aortic BP curve with the definitions of augmentation index (AIx) and augmentation pressure. Lower half: the same mean arterial pressure (MAP) may correspond to different BP curves in younger (left) and older (right) patients. However, the cross-sectional areas located under each curve remain equal [43].

As mentioned above, hypertension is characterized by an increase in arterial stiffness, which implies an increase in pulse pressure. Hypertension however is also characterized by an increase in peripheral resistance, which increases mean BP [40]. Increases in peripheral resistance and arterial stiffness therefore result in an increase in both DBP and SBP, and a significant change in the BP waveform shape (Figure 2.3). This increase is natural with age although it can be accentuated by genetics, lifestyle, medication, and nutrition [44].

2.1.2.2 States Influencing Blood Pressure

In this section, common states that cause BP to change that are encountered daily by humans are covered, which should be tractable by a cuffless BP measurement system. Herein, only body posture, motion, Valsalva maneuver, exercise and mental arithmetic are covered since they are encountered regularly by most individuals. Additional factors, such as ambient temperature, altitude, and noise are not covered; the reader is referred to [45] for additional details.

Body posture, such as sitting, standing, and supine, significantly changes BP [46]. It has been reported that significantly different BP measurements are encountered while sitting posture with and without crossed legs [47], showing how small details can change BP.

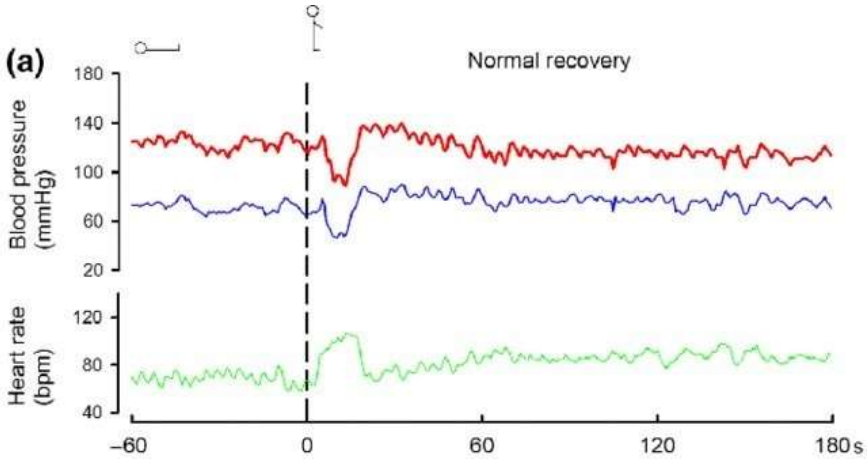

Figure 2.4 - Spectrum of normal and abnormal blood pressure (BP) and heart rate responses during the first 180 s after standing. Measurement of orthostatic BP with continuous noninvasive measurement in four subjects. The last 60 s in the supine position and 180 s of standing are illustrated. Dashed vertical lines indicate the onset of the active stand. Normal orthostatic BP recovery (a), initial orthostatic hypotension (systolic BP decrease >40 mmHg within 15 s of standing, with recovery to baseline within 30 s) [48].

The studies [46] and [47] were done for steady-state BP in different body postures. However, a larger variation in BP is observed during postural transitions, such as the supine-to-stand transition shown in Figure 2.4, where the red and blue lines represent the SBP and DBP, respectively. This case shows a normal orthostatic BP recovery; how fast the recovery happens is, however, depending on the health of the person. Orthostatic hypotension is the result in a slow recovery of the BP that might result in fainting [48].

Activity level, in general, significantly changes BP. For instance, a simple activity such as walking not only it changes BP and heart rate due to the effort, it also highly distorts the BP waveform in different ways, partially due to the acceleration of the blood in the body [49]. This effect is amplified while running, as shown in Figure 2.5, the BP waveforms during running are complex due to the interaction between the heart's pressure wave and the wave generated by the whole-body vertical movement. BP can be decreased with an asynchronous run, increased with a synchronous run, or highly variable beat-to-beat with unsynchronized running.

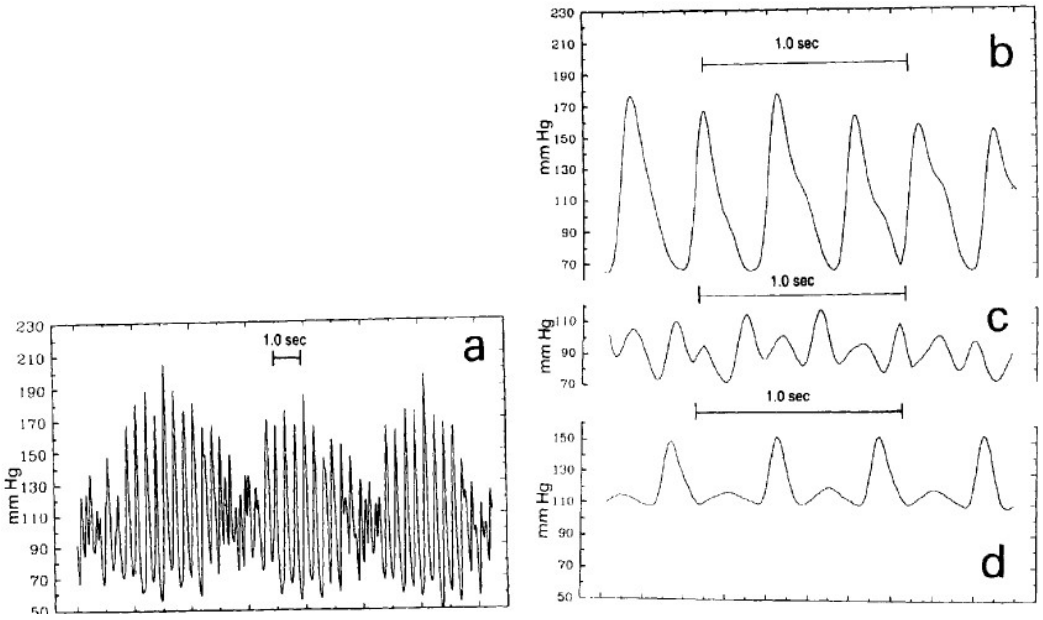

Figure 2.5 - Pressure wave recordings during running in one subject (a) (Compressed time scale) when heart rate and step rate were not synchronised, (b) when heart rate and step rate were synchronised, with heart rate and body wave in phase, (c) with synchronised heart and step rate, but with heart and body wave out of phase and (d) during rest after exercise. Time scale: 1 sec [49].

Static handgrip exercise increases BP drastically, as shown in Figure 2.6. This increase varies according to the percentage of maximum voluntary contraction [50]. There are two mechanisms involved in the cardiovascular response to a static handgrip exercise: central and peripheral neural mechanisms [50]. The former increases blood flow (increased cardiac output) and the latter refers to a vasoconstriction in some noncontracting tissues.

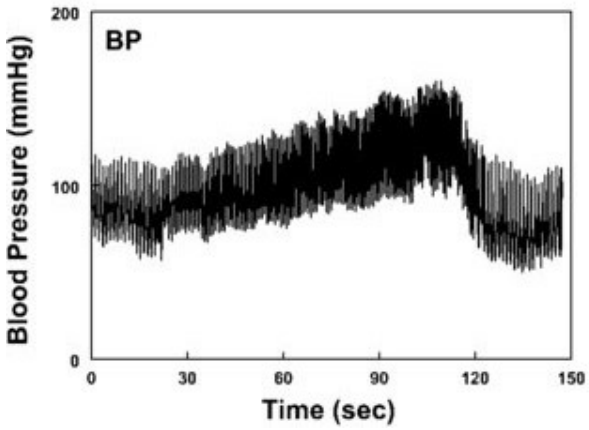

Figure 2.6 - Blood pressure from a normal subject during static handgrip exercise [51].

Mental arithmetic increases BP more than postural changes, but less than static handgrip exercise [52]. Interestingly, the physiological response of mental arithmetic is similar to static handgrip exercise; both BP and heart rate are correlated, which is not the case for postural change [52]. Although mental arithmetic won't be covered again in this thesis, any mental effort done during activities of daily living can change BP and should be tractable by a measurement device.

The Valsalva maneuver is a technique that can be used for the evaluation of heart murmurs. It is also valuable in the evaluation of left ventricular function and autonomic dysfunction [53]. The maneuver is performed as follow:

1- Deep inhalation and then the breath held.
2- The chest and stomach muscles are very tight, and the person bears down as though straining to initiate a bowel movement.
3- The position is held for a short time, usually about 10 seconds.
4- Breathing out forcibly to release the breath rapidly.

Although the Valsalva maneuver is not an activity per se, the BP response to the maneuver can resemble different activities experience daily, such as bowel movements or lifting heavy objects. A typical BP and heart rate response to a Valsalva maneuver is shown in Figure 2.7.

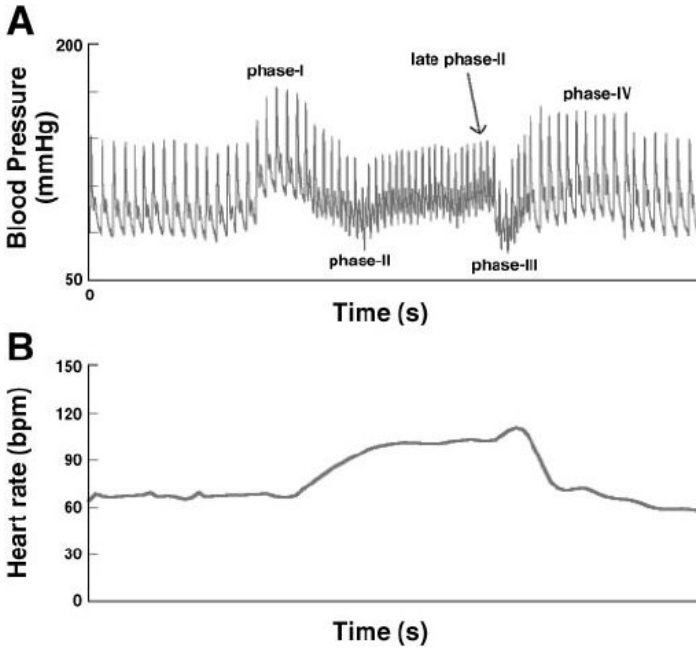

Figure 2.7 - A representative Valsalva maneuver: arterial blood pressure (BP; A) and heart rate (in beats/min; B). After inspiration, the subject blows against a resistance of 40 mmHg. This produces an initial brief increase in BP in phase I, a decrease in BP that is later restored in phase II, a brief decrease in BP with inspiration in phase III, and an overshoot in phase IV. Hypotension in phase II is associated with tachycardia, whereas hypertension in phase IV is associated with bradycardia [54].

2.1.3 Information in the Blood Pressure Waveform

As introduced in Chapter 1, the BP waveform contains information that is physiologically meaningful to clinicians. Herein, the pulse pressure amplification, arterial stiffness, and stroke volume and cardiac output are discussed.

2.1.3.1 Pulse Pressure Amplification

In 2007, a consensus document on the role of central (aortic) BP in arterial hypertension was published [55]. It concluded that there is "mounting evidence suggesting that central BP and indices correlate more closely with intermediate markers of cardiovascular risk than brachial BP". One such marker is the pulse pressure amplification, a measure of the difference in pulse pressure (SBP-DBP) between two arterial sites, generally between the brachial (or radial) artery and the aorta [56]. The method generally applies measurement of the radial BP waveform and then the use of a generalized

transfer function to recover the aortic BP [57]. The pulse pressure amplification is an independent predictor of all-cause and cardiovascular mortality [58], therefore an important parameter that can only be accessed through analysis of the complete waveform.

2.1.3.2 Arterial Stiffness

Although arterial stiffness cannot be computed directly from the waveform (in fact, it cannot be computed with any noninvasive method [59]), indirect indices that are correlated with arterial stiffness can be extracted [60]. The most commonly used surrogate of the arterial stiffness is the PWV, generally computed as $PWV=\Delta L/\Delta t$, where ΔL is the distance between two measuring sites, and Δt is the time it takes for the arterial pulse to travel from the proximal to the distal measuring site [61]. However, as mentioned, this method needs two measuring sites. From the waveform only, the time delay between the SBP and the reflected wave can be extracted and is highly correlated with the PWV. Another metric that can be extracted from the waveform is the radial artery augmentation index, which is the ratio of the reflected wave pressure and the SBP. The radial artery augmentation index is correlated with age, therefore with stiffness as well [60].

2.1.3.3 Stroke Volume and Cardiac Output

In this section, two different methods that compute CO and SV from the BP waveform are presented.

One method to compute CO from the BP waveform is based on the two-element Winkessel model [62], see Section 2.1. This linear model represents fairly well the relationship between CO and the aortic BP despite its reliance on many assumptions. One important parameter that can be extracted from this model is the time constant, τ, which represents the exponential decay of the BP during the diastolic interval (between SBP and DBP). This time constant is, however, difficult to extract from a real BP waveform due its superposition with the reflected wave. The technique proposed by Mukkamala et al. [20] is to extrapolate the BP waveform after all reflected wave effects and then fit an exponential to the extrapolation to compute the time constant. To extrapolate the BP waveform, an autoregressive with exogenous inputs model (see Section 2.1.3.3) is fit between the pulse pressure spike train and the actual BP waveform. Once the model parameters are found, the BP waveform response to a single pulse pressure spike train is simulated for a long period of time (longer than one heartbeat) and τ is computed. Finally, the ratio of mean BP and τ is computed, which is proportional to CO. The proportional CO may be calibrated, if desired, with a single, absolute invasive CO measurement.

Another method to compute CO is called ModelFlow, which is a nonlinear three-element Windkessel model that computes an aortic blood flow waveform from the arterial pressure wave [21], see Section 2.1 for the two-element model. The SV can be computed by integrating per beat the flow waveform and from the SV, and CO can then be computed by multiplying with the known heart rate.

2.2 Useful Physiological Sensors for Blood Pressure

This subsection focuses on identifying typical cardiovascular measurements used for cuffless BP monitoring. It is divided as follows: first, the different blood pressure measurement methods are discussed. Second, photoplethysmography is explained, followed by an explanation of electrocardiography.

2.2.1 Blood Pressure Measurement

In this section, the blood pressure measurement methods are discussed, which can be classified as invasive, non-invasive, and cuffless. Additional details on the latter are included in the cuffless blood pressure estimation section (Section 2.4).

2.2.1.1 Invasive Measurement

The gold standard method is invasive catheterization [63], which involves insertion of a catheter into the radial artery and pressure is measured by a pressure transducer [64]. This provides accurate, real-time, and direct measurement of the BP waveform, but tethers the patient to cumbersome equipment.

2.2.1.2 Arm Cuff-Based Measurement

Non-invasive BP monitoring techniques are widely used in healthcare and research for single time-point blood pressure measures. Auscultation is a BP measurement method that uses Korotkoff sounds to detect SBP and DBP [65]. The brachial artery is compressed with a cuff and the pressure is slowly decreased until the Korotkoff sounds are heard (SBP) and decreased again until no sound is heard (DBP).

Oscillometry generally detects the mean arterial pressure (MAP) from the pressure oscillation of the brachial cuff and uses algorithms, such as experimental ratios, to get the SBP and DBP (Figure 2.8) [66], [67]. In Figure 2.8, the envelope of the oscillation amplitude is extracted then the experimental ratios are used to compute the time at which SBP and DBP were measured with the cuff. These

algorithms, however, vary from machine to machine resulting in slightly different interpretations of the pressures [68].

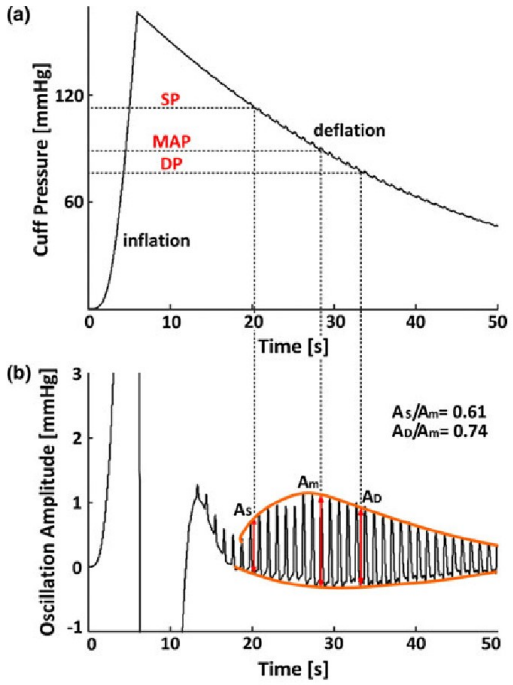

Figure 2.8 - Oscillometric method for non-invasive blood pressure measurement. (a) Cuff pressure during cuff inflation and deflation. (b) Estimation of SP (SBP), DP (DBP), and MAP (mean arterial pressure) from cuff pressure oscillations via fixed-ratio method and maximum oscillation amplitude [69].

Due to the different machine designs and algorithms used in oscillometric BP measurements, validation has always been of concern. In 1986, the Association for the Advancement of Medical Instrumentation (AAMl) published a standard for automated blood pressure measuring devices that included a protocol for the evaluation of device accuracy [70]. The ANSI/AAMI/ISO standard, referred herein simply as the AAMI standard, requires a minimum of 85 participants and a minimum of three paired tests compared with auscultation by a certified professional. The acceptable error distribution between the reference and test blood pressure value is characterized by a mean of 5 mmHg and standard deviation of 8 mmHg. In 1990, another protocol was developed, protocol by the British Hypertension Society (BHS), which used error-bands to quantify the error distribution of a medical grade automated blood pressure measuring devices [70]. This required that 60% of the cumulative readings fall within 5 mmHg, 85% within 10 mmHg, and 95% within 15 mmHg [71]. Later, the BHS protocol was replaced

by the European Society of Hypertension (ESH) protocol that requires 33 instead of 85 subjects [72]. It has, however, been discontinued because of concerns regarding its statistical power [73]. Finally, it has recently been chosen to design a new joint protocol of AAMI and ESH, in which mean and standard deviation are the preferred outcome for quantifying the error distribution of measurements [74].

One downside of arm cuff-based BP measurements is that SBP, DBP, and MAP are not measured during the same heartbeat due to the time delay between the measurement of each. This, however, is usually negligible for a person at rest. Another downside is isolated clinic hypertension, commonly referred to as white-coat hypertension [11]. It was shown that the use of a cuff increases anxiety, which subsequently creates a peak BP response in the measurement (range, 4-75 mm Hg systolic and 1-36 mm Hg diastolic) [75]. This can be minimized by consecutive measurements [76].

2.2.1.3 Finger Cuff-Based Measurement

Volume clamping measures the finger BP waveform by controlling a constant blood volume at the fingertip (see Figure 2.9) [77], [78]. A photoplethysmography sensor (discussed in 2.2.2) is used to measure the volume of blood and a cuff controls the pressure such that the volume of blood stays constant. The resulting controlled pressure is equal to the finger BP waveform.

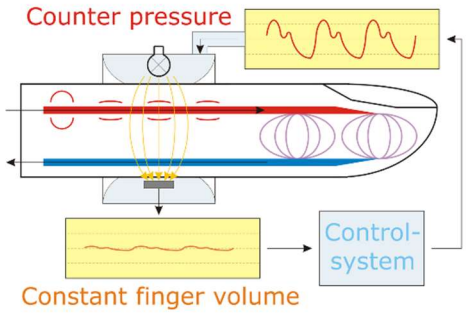

Figure 2.9 – Principle of volume clamp method for blood pressure measurement [79].

The finger BP waveform is, however, different from the brachial BP waveform, see Figure 2.10. To improve this measurement technique a generalized transfer function, followed by a level correction, and finally level calibration are applied to the finger BP [80]. Similar approaches can be used for recovering the aorta BP waveform [57].

As seen in Figure 2.10, the finger BP waveform is composed of higher frequency components than the brachial artery waveform. The relationship between the two waveforms can be modeled using a generalized transfer function from brachial to finger [81], shown in equation (2.2.

$$H(f) = K\frac{(1 + i\,f/f_0)^2}{1 + 2iD\,f/f_1 - (i\,f/f_1)^2} \tag{2.2}$$

where $i = \sqrt{-1}$. This model is equivalent to a gain (K), a second-order aperiodic high emphasis section at frequency f_0, followed by a second order underdamped low pass section at resonance frequency f_1, with a damping factor D < 1. The transfer function is forced to unity transfer at high frequencies by linking f_0 and f_1 using equation (2.3.

$$f_0 = f_1\sqrt{K} \tag{2.3}$$

The free parameters of the transfer function (f_1, K, and D) are found based on the best fit for each individual in the study. The parameters are then averaged across individuals to get the generalized transfer function. The inverse transfer function is used to find the brachial BP waveform from the measured finger waveform.

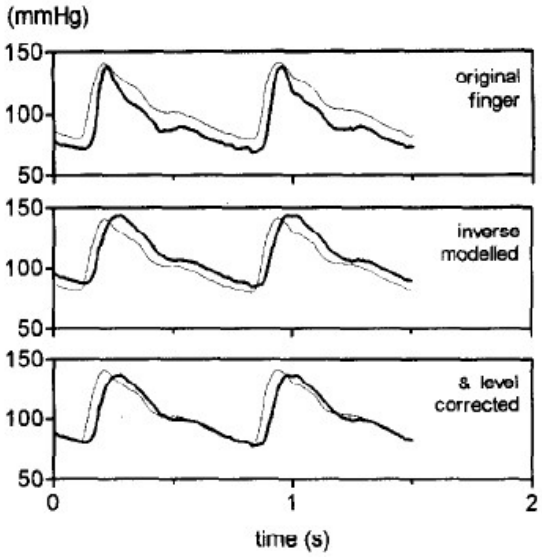

Figure 2.10 - Demonstration of the process of inverse modeling and level correction on a waveform of a 53-year-old subject. The original finger pressure (top panel) is distorted with respect to brachial (thin curve), and lower in level. After general inverse modelling (middle panel) the waveforms are nearly identical, but the level is still different. Level correction (bottom panel) shifts the waveform down by 7 mmHg in this case, after which it is nearly correct. The delay between waveforms was not modelled and, thus, not corrected, explaining the unchanged delays in the finger pressure derived waveforms [81].

The inverse generalized transfer function succeeds in removing the high frequency content but still overpredicts the mean BP due to neglecting frictional losses between the brachial artery and the finger

artery, which requires a pressure gradient to overcome. To compensate for those, a level correction is applied based on SBP and DBP [81] as shown in (2.4.

$$\Delta\mathrm{BP} = -13.3 - 0.194\mathrm{SBP} - 0.574\mathrm{DBP} \tag{2.4}$$

where the coefficients were found using linear regression based on the collected data. Since the level correction is applied on a beat-to-beat basis, it decreases not only the measurement bias, but also the variance.

The level correction increased the accuracy and precision of the measurement technique. However, the precision of the measurements does not meet the AAMI requirements. To improve the precision, a return-to-flow (RTF) calibration was developed. In short, an arm cuff inflates to stop the flow in the arm, then slowly deflates and return of the flow is measured using the finger cuff. This value of pressure corresponds to the SBP. This $\mathrm{P_{cuff-RTF}}$ can then be used in equation (2.5 [82] to compute a correction factor (Corr) to the BP waveform as a function of the finger SBP (FP_{SBP}) and the finger DBP (FP_{DBP}).

$$\mathrm{Corr} = 18.7 + 0.44 * \mathrm{P_{cuff-RTF}} - 0.36 * \mathrm{FP}_{SBP} - 0.34 * \mathrm{FP}_{DBP} \tag{2.5}$$

In equation (2.5, the coefficients of the equation were computed to minimize the error the the estimated waveform from their collected dataset. Although Corr only decreases the intra-individual measurement bias, it results in a decrease of the group-level error standard deviation (precision) [83].

An alternative to the BP measurement techniques mentioned above is cuffless BP monitoring, which is non-invasive and requires no pressurization mechanism [8]. These techniques are based on a surrogate model of the BP from other physiological sensors. Cuffless techniques are discussed in Section 2.4, but knowledge on data-driven modeling and on useful physiological sensors is required to understand better the different cuffless techniques.

2.2.2 Photoplethysmography

Photoplethysmography (PPG) is a low cost-optical technique that detects blood volume changes in the microvascular bed of tissues under the probe head [84]. A typical PPG sensor emits light at the tissue site with one or more LEDs. A photodiode measures the intensity of the non-absorbed light reflected or transmitted from the tissue [85]. A typical signal of a PPG sensor is pulsatile, where the main frequency of the AC component is related to the heart rate. The signal also has a quasi-DC component, which is related to tissues properties and average blood volume [86].

Heath monitoring using wearable sensors has gained attention in the past few years causing PPG to increase in popularity because of its low cost [87]. Most common monitoring applications are BP [12], pulse oximetry, heart rate, and screening of various cardiovascular-related diseases [85]. However,

many practical aspects need to be considered when using the sensor. For instance, one downside is that the signal is susceptible to motion artefacts, which are difficult to filter out [12]. Also, sensor placement affects the signals [84]. Using a PPG sensor to monitor heart rate is accurate, when monitoring average heart rate, but has been challenging on a beat-to-beat basis [86]. The main cause of error is the time for the pulse to travel from the heart to the arm, referred to as pulse transit time, which varies with posture and blood pressure, which can vary on a beat-to-beat basis. Another practical consideration when using a PPG sensor for cuffless BP monitoring is that the contact pressure affects the signal as observed in Figure 2.11 [88]. Despite all the practical challenges of using a PPG sensor for health monitoring, it is a relevant sensor to use when estimating BP from wearables, since they are highly correlated.

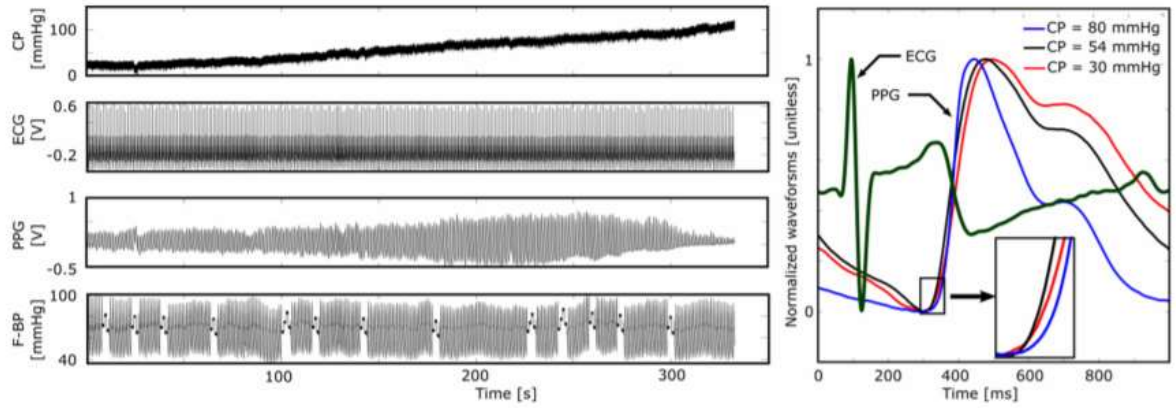

Figure 2.11 – Representative example of all measurements from a subject. As expected, the PPG amplitude and shape markedly varied with the contact pressure (CP) [88].

2.2.3 Electrocardiography

An electrocardiogram is a recording of the electrical signals generated by the heart. Electrocardiography (ECG) signals are generated by depolarization and repolarization of the heart muscles, which reflects the contraction of the heart [34]. However, the mechanical movement of the heart lags slightly its corresponding electrical signal as the contraction of a single cardiac muscle cell follows its action potential. The resulting pressure as a function of time in shown in Figure 2.12, wherein it can be observed that the pressure starts increasing with the QRS complex, but there is a time lapse between the peak pressure and the QRS complex.

ECG signals can be measured using many different methods. The most common configuration uses three electrodes, one positive, one negative, and one ground, which are to be positioned on the right arm, the left arm and the left leg. However, the electrodes are generally placed on the trunk close the

reference position. A three-lead ECG measures three bipolar leads (referred as ECG I, ECG II, and ECG III) depending on which probe is the ground.

Nowadays, ECG is used as a wearable as it is easily integrated into textiles [89] and is a signal commonly used in cuffless BP monitoring [6].

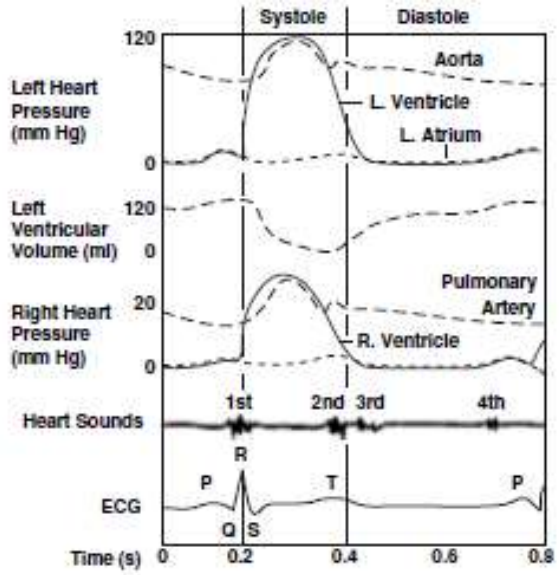

Figure 2.12 - When the ventricles depolarise, the ventricular muscles contract and rapidly builds up systolic pressure. When left ventricular pressure exceeds that of the aorta, the aortic valve opens and blood ejects out to circulate in the body [90].

2.2.4 Ballistocardiogram and Seismocardiogram

The ballistocardiogram (BCG) is a measurement of the recoil forces of the body in reaction to cardiac ejection of blood into the vasculature, which is generally measured with a force plate. The seismocardiogram (SCG) represents the local vibrations of the chest wall in response to the heartbeat, which is generally measure with accelerometers [91]. Although the exact physiological significance of the waveforms are unknown [91], they have been useful in estimating BP [92], [93].

2.3 Data-Driven Models

Most cuffless BP monitoring approaches use data-driven modeling to find the relationship between exogenous measurements and BP. Data-driven models can be used to tackle three main problems: regression, classification, and clustering. The two former are part of supervised learning problems,

where the model is built from data with known desired input and output pairs, and the latter is an unsupervised technique, where the desired output of the model is unknown [94]. Classifiers (classification models) and regression models are differentiated by their outputs. In regression, models are used to approximate continuous output variables, whereas classifiers approximate discrete output variables, taking the form of labels [95]. Clustering techniques are used to form groupings within data. The focus of this section is on regression models applied to dynamical systems, which is the approach taken to model BP, and on a clustering technique, the approach used to estimate the uncertainty of the model.

2.3.1 Autoregressive Models[2]

Autoregressive models (AR) are linear discrete-time models that allow the representation of the frequency domain of a dynamical system. AR models can be written using a discrete-time transfer function as seen in equation (2.6.:

$$y(k) = \frac{1}{A(z)} v(k) \tag{2.6}$$

but the most common notation for AR models is the difference equation (equation (**2.7**).

$$y(k) = -a_1 y(k-1) - \ldots - a_m y(k-m) + v(k) \tag{2.7}$$

where k is the time step (where the time $t = kT$, with T being the sampling period), y is the output times series, $y(k)$ is the k^{th} sample of that time series, $v(k)$ is the noise at instant k, z is the delay operator, $A(z)$ is the denominator polynomial of the transfer function associated with the poles of the system, and the number of coefficients m is the order of the system. An AR model without any noise input can model, for example, the free vibration of a structure in the discrete time domain.

The most widely used linear dynamic model for time series is the autoregressive with exogenous input (ARX) model. ARX models are extensions of AR models where an external input $u(k)$ is added. The complex notation is shown in equation (2.8.

$$y(k) = \frac{B(z)}{A(z)} u(k) + \frac{1}{A(z)} v(k) \tag{2.8}$$

Again, the most common notation for ARX models is the difference equation (equation (2.9).

$$y(k) = b_1 u(k-1) - \ldots - b_n u(k-n) - a_1 y(k-1) - \ldots - a_m y(k-m) + v(k) \tag{2.9}$$

[2] This subsection is based on [94].

where $B(z)$ is the polynomial of the numerator associated with the zeros of the system transfer function. There are other AR models such as AR moving average (ARMA) and ARMA with exogenous input (ARMAX) that will not be discussed here. ARMA and ARMAX are basically extensions of AR models that incorporate the noise dynamics [94].

AR models are general linear discrete time models, where the coefficients of the equation can be found from different methods. For instance, a known continuous time system dynamical equation can be discretized by replacing dt by Δt. However, the most common approach is to find the coefficients from data, which is a quadratic optimization problem in this case. For such quadratic problems, the most common technique employed is the ordinary least squares estimate (LSE). In the case of an ARX problem, the LSE equation is as shown in equation (2.10.

$$\hat{\theta} = (X^T X)^{-1} X^T y \tag{2.10}$$

where y is the vector of the data points of the output time series, X is the regression matrix, each row being all the $u(k-1),\ldots,u(k-n),y(k-1),\ldots,y(k-m)$ with respect to its $y(k)$, where k goes from 1 to the length of y, and $\hat{\theta}$ is the optimal estimate of the ARX coefficients $b_1,\ldots,b_n,a_1,\ldots,a_m$ in equation (2.9. The best estimate of y in the least square sense is $\hat{y} = X\hat{\theta}$. However, inverting the Hessian $(X^T X)$ requires large computer memory when the number of coefficients $b_1,\ldots,b_n,a_1,\ldots,a_m$ is large, since the size of $(X^T X)$ is $(m+n) \times (m+n)$. In that case, iterative methods, such as gradient descent (also called steepest descent) can be used [94]. The idea is to use the local gradient of the error function and update the coefficient of the equation such that it goes in the direction of the negative gradient. This process is done iteratively until an acceptable error is reached.

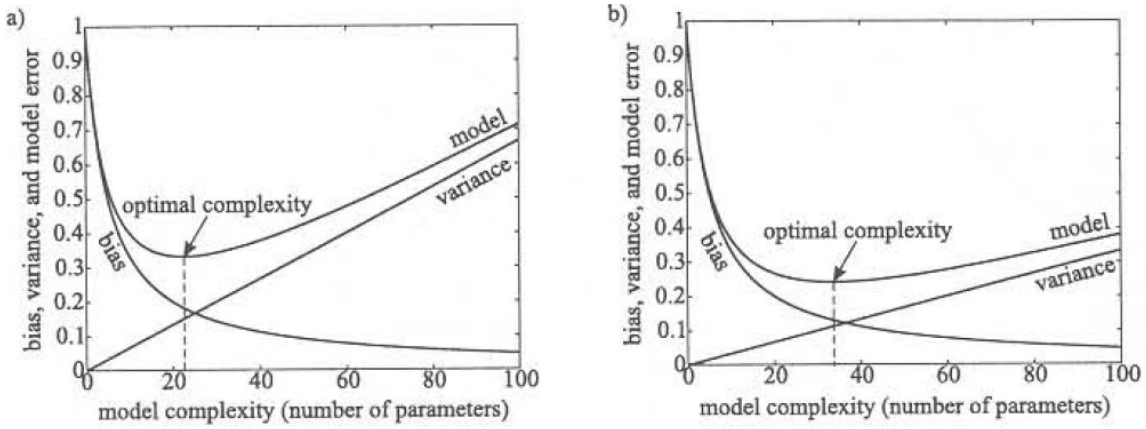

Figure 2.13 – Bias/variance tradeoff. The model error can be decomposed into bias and variance part: a) relatively high variance error, b) lower variance due to more data or less noise than in a) [94].

There are a variety of tools to aid in choosing the embedding dimensions (m and n) for a given system/dataset, such as autocorrelation, cross-correlation, and mutual information [96]; however successively embedding in higher dimensions and looking at consistency of the results is the method generally employed [96]. Model generalization power, which is directly related to embedding dimension, is a tradeoff between bias and variance, where one might want to choose a model complexity (embedded dimensions) that minimizes the mean square error (MSE) between any of the real system outputs and the AR output (Figure 2.13). In practice, a large bias refers to "underfitting", where the error is similar for the training data (data used for LSE) and the new data. A large variance is referred to "overfitting", meaning that the error for the training data will be small; however, the model will have a large error for a new dataset. For example, for a small noisy dataset being fitted with more parameters than needed the non-important parameters will be used to fit the noise. This leads to an estimate with large variance.

2.3.2 Artificial Neural Networks

The ARX models are limited to linear dynamics problems. However, the same architecture embeds nonlinear functions for building nonlinear ARX (NARX) models. The nonlinear functions can be implemented within an artificial neural networks (ANN), which is a universal function approximator [97]. One of the most commonly used feedforward ANN architectures is the multilayer-perceptron (MLP) shown in Figure 2.14 for a multi-input/multi-output configuration [98].

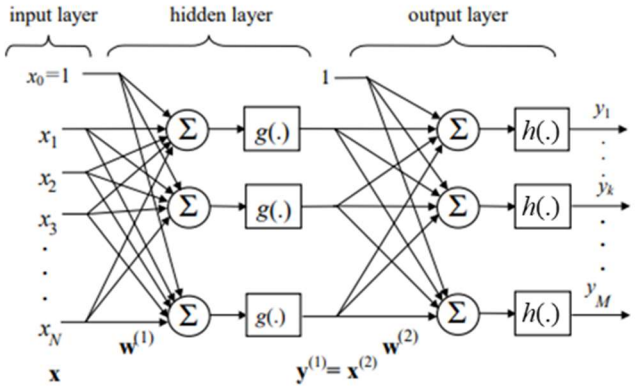

Figure 2.14 - A multi-layer perceptron with one hidden layer (adapted from [98]).

The outputs of the MLP are computed using equation (2.11).

$$y_k = g\left(\sum_{j=0}^{N_2} w_{kj}^{(2)} h\left(\sum_{i=0}^{N_1} w_{ji}^{(1)} x_i\right)\right) \tag{2.11}$$

where $w_{kj}^{(l)}$ denotes a weight of the MLP, which belongs to the k^{th} neuron in layer l, associated with the j^{th} neuron of the layer l-1. The fucntions $g(.)$ and $h(.)$ are called the activation function, which are sometimes different from one layer to the next; they are generally sigmoidal or hyperbolic tangent functions in the hidden layers and linear functions at the output layer for regression problems. In deep neural networks (more than one hidden layer), however, the rectified linear (ReLu) activation function is more common [99].

In ANN, the weights $w_{kj}^{(l)}$ are the coefficients of the equation that are modified to fit the training data. Error backpropagation is an adaptation of the (stochastic) gradient descent to ANN, where the weights $w_{kj}^{(l)}$ are updated at each iteration in the direction of the negative gradient of the loss function. However, due to the loss function of an ANN, the optimization would not usually guarantee a global minimum of the error norm since it can converge to local minima. It is not a problem for the engineering field, as long as this minimum achieves the error requirement prescribed *a priori*. See [94] for a complete derivation of error backpropagation for single and multiple hidden layer MLP models.

The learning rate η is the variable used to quantify the rate of change of the weights of the ANN proportional to the gradient of the error function at each iteration (equation (2.12)).

$$w_k = w_{k-1} - \eta \frac{\partial J_k}{\partial w_{k-1}} \tag{2.12}$$

where $\frac{\partial J_k}{\partial w_{k-1}}$ is gradient of the loss function computed from the backpropagation. A fixed learning rate would results in a tradeoff between fast convergence of the optimization and the model error at convergence. The former can be visualized by a change in the weights that is too large which results in missing the optimal value than is situated between w_k and w_{k-1}. For that reason, the learning rate is generally adaptive. There are many training methods that adapt the learning rate, such as the Levenberg-Marquardt, which is a combination of gradient-based methods and nonlinear least square, whereby convergence can be achieved in a lower number of iterations in comparison to stochastic gradient descent, but each iteration is more time consuming [94].

Another consideration for training an ANN is how the training data are fed to the model since it affects the training convergence. Common methods include one-shot batch adaptation, pure stochastic gradient descent, and minibatch adaptation. One-shot batch adaptation updates the weights over the

entire dataset, averaging over a large amount of information. It has generally faster convergence, but it is more susceptible of getting stuck in a local minima. In pure stochastic gradient descent (sample adaptation), on the other hand, the weights of the ANN are updated on every single instance of the dataset. Since it's based on one random data point, it's very noisy and may go off in a direction far from the batch gradient. However, the noisiness is needed in non-convex optimization, because it helps escaping local minima [100]. The disadvantage of sample adaptation is that it is inefficient and many iterations over the entire dataset are needed. The minibatch method is a compromise between injecting noise to each weight's update and achieving a relative speedy convergence. It can even be an optimal choice when combined with batch normalization techniques [101].

When training an ANN, one way to deal with the bias/variance tradeoff, including the previously discussed method, is to use cross-validation and early stopping. The idea is shown in Figure 2.15, where the ANN is trained only on a portion of the data and uses another portion to test the accuracy of the ANN on unseen data (generally called the validation dataset). The training can be stopped such that the error is minimized on the validation portion. One must be careful not to report the error on the validation portion as being the MSE of the ANN. This MSE is biased towards the validation data and minimized in the training process, so should provide an optimistic estimate of MSE. Generally, a third data split is used to really test the accuracy of the ANN.

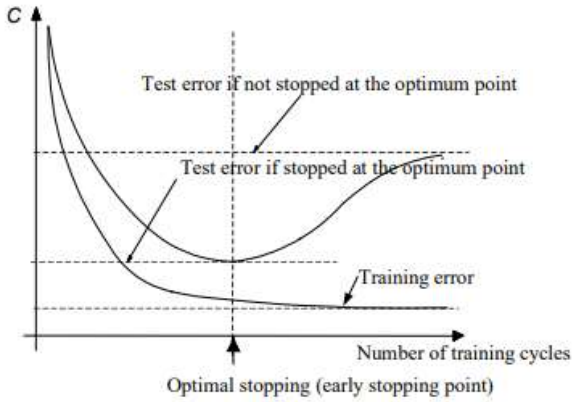

Figure 2.15 - Learning curves for the training and the test data (C= MSE) [98].

So far, only static ANNs were explained, meaning that there is no feedback between the input layer and the output layer of the ANN. When an ANN is dynamic, there is a feedback loop of the output layer to the input layer with a delay. These so-called recurrent neural network (RNN) are a form of NARX models [94]. Those models can be used in one-step-ahead prediction, where the prediction output $\hat{y}(k)$ is a function of the previous measurements $y(k-1), \dots, y(k-m)$. NARX models can also be used in

simulation mode, where the previous predicted outputs $\hat{y}(k-1), \ldots, \hat{y}(k-m)$ are used to predict the output $\hat{y}(k)$. Both, one-step-ahead prediction and simulation mode NARX models, can generally be trained as a static ANN, where the measurements are used in the input layer and the estimation output is compared with the measured input. However, simulation mode can sometimes benefit from a more sophisticated training method called backpropagation-through-time (BPTT). The idea is the same as backpropagation but the model is unfold with time (the output at one time step is connected to the input at the next time step), becoming similar to ANN with multiple hidden layers (called deep neural networks). However, BPTT is practically difficult because of the vanishing gradient problem [102] like deep neural networks. For instance, the hyperbolic tangent function have gradients in the range (0,1], and backpropagation computes gradients by the chain rule ($\frac{\partial J_k}{\partial w_{k-1}}$ in equation (2.12 depends on the derivative of the deeper layers). In an *n*-layer network, this means that the gradient decreases exponentially with *n* resulting in the early layers training very slowly and at one point the weights stop changing. Nevertheless, NARX model can have good simulation accuracy without BPTT when using the good inputs [103].

2.3.3 Support Vector Machines

Support vector machines (SVM), like ANNs, are data-driven models that tune their parameters according to a cost function. SVM are applicable to regression and classification problems, as well as outlier detection, which will be discuss in Section 2.3.3.1. In this section, a brief overview of classification using SVM is provided, which is then transposed to outlier detection using one-class SVM (OCSVM).

The general idea of SVM is to find the hyperplane that splits data into two classes. Since there are many of them, SVMs are trained to find the hyperplane with the optimal margin, i.e., the hyperplane with the largest distance to the closest datapoints, called support vectors [104]. The previous definition is called hard margin SVM since it considers that the two classes are perfectly separable. Most of the time, SVMs are used with soft margins, which permit some data to be within the margin or to be misclassified by the means of slack variables (see ξ_i in equation (2.14). The tradeoff between a large margin and misclassification rate is generally dealt with a tunable parameter such as C in C-SVM [105] or ν in ν-SVM [106]. This is equivalent to the bias-variance tradeoff discussed in Section 2.3.2.

SVM are a kernel-based learning algorithm [104], meaning that the input data is transformed, using a kernel, into a high-dimensional space (feature space), and the learning is performed in the feature

space. The most versatile kernel is the Gaussian kernel, since it is equivalent to projecting the data into an infinite dimensional space [104], is defined in equation (2.13.

$$k(x_1, x_2) = e^{-\|x_1 - x_2\|^2/c} \tag{2.13}$$

where c is a tunable parameter that controls how much overfit to data there is (larger c means more overfit), and x_1 and x_2 are the input vectors.

The hyperplane separating the data into two categories ($y = 1$ on one side and $y = -1$ on the other) is define as $w \cdot k(x_1, x_2) + b$, where w and b are the weights and the bias of the hyperplane to be learned from the data. This hyperplane has a soft margin ρ, in which some of the training data can lie (or even be misclassified) if their respective slack variables ξ_i>0.

Putting everything together for the ν-SVM algorithm [106], the hyperplane parameters can be found by selecting a value for the free parameter ν and by solving the primal optimization problem on the training data as follow:

$$\min_{w,\xi,\rho,b} \pi(w, \xi, \rho) = \frac{1}{2}\|w\|^2 - \nu\rho + \frac{1}{m}\sum_{i=1}^{m} \xi_i \tag{2.14}$$

subject to

$$y_i(w \cdot k(x_i) + b) \geq \rho - \xi_i \tag{2.15}$$

$$\xi_i \geq 0, \rho \geq 0 \tag{2.16}$$

where x is the input data, y the data label (1 or -1), m is the number of training datapoints, ξ_i is a slack variable, w and b are the parametrization of the hyperplane, ρ is the margin parameter, and v *is a* variable to be optimized (or chosen), which will be explained more in detail in the context of outlier detection (Section 2.3.3.1).

2.3.3.1 One-Class Support Vector Machines

When dealing with unlabelled data, OCSVM can be used to find the underlying distribution of the data. In this scenario, equation (2.16 is modified such that the data can be separated in two categories (1 or -1) according to the free parameter v, instead of separating labels (see equation (2.17)) [107].

$$\min_{w,\xi,\rho} \pi(w, \xi, \rho) = \frac{1}{2}\|w\|^2 + \frac{1}{m\nu}\sum_{i=1}^{m} \xi_i - \rho \tag{2.17}$$

subject to

$$(w \cdot k(x_i)) \geq \rho - \xi_i, \xi_i \geq 0 \tag{2.18}$$

Once solved, the decision function applied to new data points is the following:

$$f(x) = \text{sign}((w \cdot k(x_i) - \rho) \tag{2.19}$$

The primal formulation can become difficult to solve if the dimension of the input x is large. A solution to this curse of dimensionality is to solve the dual problem, which is derived by using the Lagrangian of the primal problem:

$$\min_{\alpha} \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j k(x_i, x_j) \tag{2.20}$$

subject to

$$0 \leq \alpha_i \leq \frac{1}{mv}, \sum_{i=1}^{m} \alpha_i = 1 \tag{2.21}$$

Once solved, the decision function applied to new data points is the following:

$$f(x) = \text{sign}((\sum_{i=1}^{n} \alpha_i \, k(x_i, x) - \rho)$$
$$\rho = \sum_{j=1}^{n} \alpha_j \, k(x_j, x_i) \tag{2.22}$$

where n is the number of support vectors, which represents the number of non-zero α's found in equation (2.20. The margin ρ, which acts as an offset in a one-class problem, can be recovered from any support vector x_i by summing proportionally the kernel with respect to x_j.

The free parameter *v has the following properties* [107]*:*

1. $(0 < \nu \leq 1)$
2. *v* is an upper bound on the fraction of outliers, that is, training points outside the estimated region.
3. *v* is a lower bound on the fraction of support vectors.

In other words, a small value of ν leads to fewer support vectors and, therefore, a smooth, crude decision boundary. Property 2 only holds if ρ is computed from equation (2.22. However, the margin ρ can be iteratively selected in order to have a known fraction of outliers in the training data [108].

2.4 Cuffless Blood Pressure Estimation

Cuff-based BP monitoring techniques (introduced in Section 2.2.1), are uncomfortable and disruptive [30], and consequently, negatively impact patient compliance, despite the fact that

continuous BP monitoring can result in better treatment management [10]. An alternative to cuff-based systems is cuffless BP monitoring, which requires less cumbersome and expensive hardware [7], enables mobility for the patient, and eliminates discomfort associated with the cuff-based monitoring [8]. Cuffless BP monitoring relies on the information encoded in an alternative physiological measure, generally PPG (see Section 2.2.2), coupled with a surrogate model of BP. In this section, three main cuffless BP estimation methods to estimate SBP and DBP are covered, namely: pulse transit time (PTT) based, feature extraction of the PPG waveform, and end-to-end machine learning. But first, how models are generated and calibrated will be introduced, since it is applicable to all BP estimation methods.

2.4.1 Model Generation, Calibration, and Monitoring Phase[3]

The first step of cuffless BP monitoring is the initialization phase in which a parametric model is selected and then calibrated using data. In this initialization context, calibration refers to fitting a model to data, e.g., linear regression or training an artificial neural network. Once calibration is done, the second and final step of cuffless BP is the monitoring part in which the calibrated model uses as input the measurement of the features and outputs an estimation of BP. Calibration data should include matching input-output pairs, e.g., measurements of relevant features of the surrogate signals and measurements of BP. This calibration process can be person-specific, population-based, or hybrid. Person-specific calibration relies solely on data from the individual for which the model will subsequently be used. Population-based models are calibrated on the data of many individuals and imply that the model should perform well on different individuals likely not included in the calibration data. Hybrid calibration is the process of calibrating a population-based model to a specific person.

When calibrating a BP model, the data need to cover a large range of BP in order to be well calibrated. In the case of population-based models, the range of BP generally comes from the spread in BP between individuals, i.e., hypotensive, normal BP, and hypertensive individuals. In the case of calibrating person-specific models, it is suggested to impose a change in BP through interventions, such that the calibrated model is able to generalize well in the monitoring phase in which the person's BP will change constantly depending on their activities of daily living.

[3] The content of this section is primarily from [109]

2.4.2 Pulse Transit Time

Pulse transit time (PTT) is the time delay for the pressure wave to travel between two arterial sites [6]. PTT is inversely related to BP as described by the Moens-Korteweg equation (equation (2.23), which relates the pulse wave velocity (PWV), discussed in Section 2.1.2.1, to arterial properties [110],

$$\mathrm{PWV} = \sqrt{\frac{Eh}{2r\rho}} \tag{2.23}$$

$$\mathrm{PTT} = l/\mathrm{PWV} \tag{2.24}$$

where E is the elastic modulus of the artery, h is its wall thickness, r is the vessel radius, and ρ is the blood density.

This rather simplified model is the base for most of the physical models used to develop the mathematical relationship between PTT and BP, in which an assumed function that relates E to BP is used. The most common physical models are logarithmic PTT, computed as [111]:

$$\mathrm{BP} = a\ln(\mathrm{PTT}) + b \tag{2.25}$$

and the inverse PTT, computed as [112]:

$$\mathrm{BP} = \frac{a}{\mathrm{PTT}} + b \tag{2.26}$$

Other relationships have been developed based on experimental PTT–BP data, such as linear PTT [113]:

$$\mathrm{BP} = a\mathrm{PTT} + b \tag{2.27}$$

and inverse quadratic PTT, computed as [114]:

$$\mathrm{BP} = \frac{a}{\mathrm{PTT}^2} + b \tag{2.28}$$

Some models even included other physiological parameters, such as inverse quadratic PTT with a linear heart rate (HR) term, computed as [115]:

$$\mathrm{BP} = \frac{a}{\mathrm{PTT}^2} + b\mathrm{HR} + c \tag{2.29}$$

where the user-specific coefficients a, b, and c are determined using regression techniques based on measurements of subject-specific PTT-BP pairs.

Typical PTT-based methods use a PPG combined with another signal, such as another PPG at a different peripheral site [116], [117], ECG (see Section 2.2.3) [16], [113], [118], a ballistocardiogram (see Section 2.2.4) [119]–[121], or a seismocardiogram (see Section 2.2.4) [91], [93], [122]. The time delay between the ECG waveform and the PPG waveform is called the pulse arrival time (PAT), which is equal to the sum of PTT and the pre-ejection period (PEP) of the cardiac cycle [6]. The PAT is often

used as a surrogate for PTT, the latter of which is not always practical. PAT has been defined in many ways over the years; the most common are the time delay between the ECG peak and the PPG foot, maximum slope, peak, or the PPG intersecting tangent method to name a few [6].

Estimating BP from PTT or PAT is physiologically meaningful, which explains the considerable efforts to date in pursuing these methods for cuffless BP estimation. However, using PTT or PAT means that at least one PPG sensor is usually employed and these sensors are known to be susceptible to motion artefacts [84], [123]. Even though, various definitions have been compared and the intersecting tangent method revealed the best PTT–BP correlation and robustness to artefact [6], extracting accurate PTT is still difficult. Moreover, there are no consensus as where PAT or PTT should be measured for BP estimation. For instance, in one study comparing ear PAT, toe PAT, finger PAT, ear-toe PTT, ear-finger PTT, and finger-toe PTT, toe PAT was significantly more correlated with SBP ($\rho \approx$ - 0.6) [124]. On the other hand, other studies shows that PAT is not an adequate surrogate for PTT as a marker of BP [125]. Therefore, there is a lot of variability in results in the literature. Another downside of PTT-based methods is that calibration models are user-specific, and are reliable for only minutes to hours [6].

2.4.3 PPG Waveform Features

The second set of cuffless BP methods uses a set of features extracted from the PPG waveform. The methods can be split into two different categories: physiological-based models and non-physiological-based models [12]. Although the difference is not clear cut, the former relies on extracting specific physiological parameters, such as instantaneous heart rate, and systolic and diastolic periods [12], whereas the latter do not rely solely on physiological parameters. For physiological-based methods, BP is then estimated using the extracted parameters and a parametric model, such as the Windkessel model [126], regression of long-term and short-term features [127], or the pulse transport theory-based model [128]. Despite physiological-based models being intuitive and computationally less expensive, those models are not common in the literature due to their inaccuracy and their reliance on frequent calibration [12].

Non-physiological-based models generally use a larger extracted feature pool from the PPG waveform. These features, which are not necessarily physiologically meaningful, are then fed into a machine learning-based algorithm to predict BP. Some common physiological features of the PPG waveform are [22]:

- Augmentation Index (AI): a measure of the wave reflection on arteries [129].

- Large Artery Stiffness Index (LASI): an indicator of the arterial stiffness [22].
- Inflection Point Area Ratio (IPA): can be considered as an indicator of the total peripheral resistance [129].

Other features, such as pulse amplitude [130], peak-to-onset interval [131], pressure index [132], heart rate and modified normalized pulse volume [8], along with other time and frequency domain features, can also be extracted from the PPG waveform [133]. There are too many existing features to make an exhaustive list. For instance, in [134], 153 features were extracted from the PPG waveform only. Moreover, these PPG features are sometimes combined with PTT to improve the accuracy [22], [24], [135]. Another set of common features used to estimate BP comes from the double time derivative of the PPG signal (sometimes called accelerated PPG), since the waveform is related to blood circulation [129].

Once selected features are extracted, they are fed into a machine learning algorithm to predict the BP. The different PPG waveform features methods used for cuffless BP monitoring found in the literature can be person-specific, population-based, or hybrid. The difficulty in this approach lies in the feature extraction from the morphology of PPG as it is highly susceptible to noise or artefacts [84], [123], but also because some morphologies do not fit in the typical mold [22]. This makes it difficult to apply the methods without manually tuning the algorithms. This is one motivation to used end-to-end learning, which is the topic of the next section.

2.4.4 End-to-End Machine Learning

End-to-end machine learning or deep learning, in contrast to typical machine learning, learns to extract the relevant features from raw inputs as well as learning their relationships [99]. Deep learning can therefore handle feature extraction and the learning, which copes with both the inadequacy of pre-determined (hand-crafted) features to fully characterize BP and the difficulty of extracting them using algorithms [13]. Many deep learning approaches to estimate BP have been developed and a recent review of the available models can be found in [13]. To the author's knowledge, all end-to-end models were developed for population-based or hybrid calibration.

The main characteristic of end-to-end BP models is that they use raw signals, such as ECG and PPG, as inputs. The various works in the field have tesed different deep learning architectures, such as convolutional neural network [136], long short-term memory network [137], auto-encoder [138], etc. Second, different inputs are used, such as 10 seconds of continuous ECG and PPG signals [139], single heartbeat PPG and its frequency domain information [140], PPG and personal information [141], etc.

Another difference between studies is the data used. The most common datasets are the MIMIC II [142] and the Queensland [143] databases, but some studies used their own collected data [141]. Nevertheless, the data used is heterogenous due to the difficulty of getting good data in databases. For instance, the number of subjects used in different studies varies widely from 15 [144] to 1557 [145].

The validation of end-to-end models are also largely heterogenous. Some studies split the data into training, validation, and testing [136], or used k-fold cross validation where all the data from different subjects are shuffled together [139], [145]. Those studies generally result in low error and high correlation estimation due to the fact that the models are already trained to estimate a given individual's BP, sometimes for only a couple of seconds [145]. This kind of method is part of the family of hybrid calibration models, since they use population data plus some of each individual's data to calibrate the model. Other studies used data in which some subject's data are left out for further validation [141]. This refers to population-based calibration and those models have larger estimation error but are more applicable to real world BP measurement.

2.4.5 Fusion of Models

The method of combining models is generally referred to as estimation fusion, information fusion, data fusion, of sensor fusion, among other names [11]. One well established definition of data fusion was provided by the Joint Directors of Laboratories (JDL), establishing five levels of data fusion [12]. Herein, the important levels to distinguish are the sensor level and the track level [13]. Sensor level fusion refers to the use of multiple sensors and combining their information to improve estimation performance. Sensor fusion has been employed previously to improve BP estimates. For instance, using multiple PPG sensors and a seismocardiogram [146], using multiple bioimpedance sensors [147] to better estimate PTT, and using ECG and two pulse pressure waveform sensors to extract more features for BP estimation [135], [148]. Track fusion, which combines multiple BP estimates to improve accuracy, has also been considered [149], but the uncertainty of the estimation has not been explored, which is a benefit of using track fusion algorithm [150].

2.5 Conclusion

BP is largely variable from one individual to another, due largely to differences in physiology (arterial stiffness and peripheral resistance), but it also varies continuously due to body posture and activity in general. Existing BP measurement techniques are cumbersome and do not provide continuous measurement of BP conveniently. An interesting alternative to measure BP is cuffless BP

monitoring. Existing models for cuffless BP monitoring, however, only provide SBP and DBP estimations, with a couple of noted exceptions that also enables the estimation of the mean arterial pressure (MAP) [103], [106] or the pulse pressure [87]. The estimation of the blood pressure waveform can give access to additional health markers such as arterial stiffness, pulse wave amplification, and stroke volume and cardiac output. Dynamical modeling approaches such as NARX with ANN are well suited for estimating the BP waveform due to the dynamic change in BP beat-by-beat. PPG and ECG signal have shown their potential as BP surrogate signals and end-to-end learning can help to cope with the difficulty of extracting features from them. Developing a cuffless BP monitoring technique that has the capabilities tracking SBP and DBP within the AAMI standard during activities of daily living, that gives access to the pulse pressure amplification, and that can monitor BP estimation uncertainty has significant value for diagnosing hypertension.

Chapter 3
Modeling the Blood Pressure Waveform[4]

In this chapter, a modeling approach is proposed to accurately estimate the blood pressure (BP) waveform and extract important BP features, such as systolic BP (SBP), diastolic BP (DBP), and mean arterial BP (MAP). Access to the full waveform has significant advantages over previous cuffless BP estimation tools in terms of accuracy and access to additional cardiovascular health markers (e.g., cardiac output [20], [21]), as well as potentially providing arterial stiffness [60] and identifying different cardiovascular diseases [18]. Nonlinear autoregressive models with exogenous input (NARX) are implemented using an artificial neural network (ANN) to predict the BP waveforms using electrocardiography (ECG), and/or photoplethysmography (PPG) signals as inputs. Once the ANN is trained on a time series, the BP monitoring can be removed and the future BP can be estimated using only the exogenous inputs (ECG and/or PPG). We hypothesize that modeling the dynamics of the BP and estimating its complete waveform using the proposed NARX model can improve the accuracy of SBP, DBP, and MAP estimations.

3.1 Method

3.1.1 Model Architecture

The ANN used by the NARX model was a multilayer Perceptron with one hidden layer of 10 neurons in every case. Hyperbolic tangent and linear activation functions were used for hidden neurons and output neurons, respectively. The NARX model was used to predict the complete time series of the BP, from which the SBP, DBP, and MAP were also extracted.

Three different exogenous inputs were explored: ECG, PPG, and both simultaneously. In its general form, the ANN inputs consisted of 1 and 2 time-step delays of the BP and 1 to *n* time-step delays of the exogenous input. The output layer generated the current BP estimate. The model architecture is

[4] Parts of this chapter have been adapted from the following publication:

C. Landry, S. D. Peterson, and A. Arami, "Nonlinear Dynamic Modelling of Blood Pressure Waveform: Towards an Accurate Cuffless Monitoring System," *IEEE Sens. J.*, vol. 20, no. 10, pp. 5368–5378, May 2020, doi: 10.1109/JSEN.2020.2967759.

summarized in Figure 3.1. A feedforward ANN was trained on the exogenous inputs and the BP measurements (Figure 3.1.a). Then, the trained ANN was used to estimate the BP on the test data, where a feedback loop with delay was implemented to provide the ANN with the estimates of the required delayed BPs (Figure 3.1b).

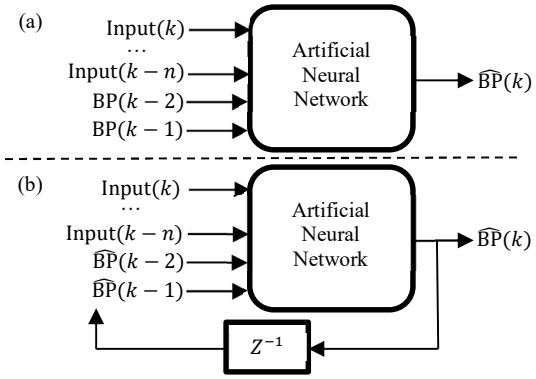

Figure 3.1 - Schematic of the NARX model used for a) training of the ANN as a feedforward ANN, and b) estimation and testing the BP using a recurrent ANN.

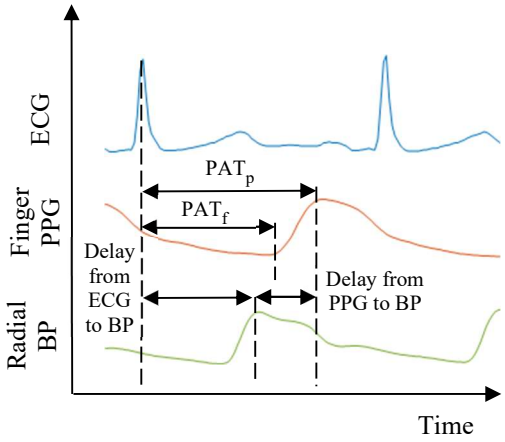

Figure 3.2 - Typical ECG, PPG, and BP signals illustrating the potential temporal delays between signals.

Figure 3.2 shows representative examples of typical ECG II, finger PPG, and radial artery BP waveforms, highlighting that a delay can appear between the peaks of the exogenous input and the BP pulse arrival at a peripheral measurement location. In this example, the ECG peak appears before the BP peak, whereas the PPG peak is located after due to the measurement site being located further down the arm (fingertip vs radial artery). Depending on the peripheral measurement site and the sampling frequency, the number 1 to *n* of time-step delays of the exogenous input can be adjusted to be larger

than the maximum delay between the BP peak and the input peak. This ensures that the peak of the exogenous input is fed into the ANN when predicting the BP peak.

3.1.2 Training

The delayed exogenous inputs and the delayed BP measurements were used as inputs for the training of the ANN (Figure 3.1.a). White Gaussian noise (Mean = 0 and Variance = 1 $mmHg^2$) was added to the BP time series to enrich the training data and improve the model robustness to input noise. Error backpropagation was used with the Levenberg-Marquardt optimization algorithm to minimize the norm of the difference between the predicted and measured values of the BP. A training stop condition was defined as an increase of the mean square error on the validation dataset in six consecutive epochs.

Two different data splits were used in training and testing the ANNs. Interval training mode, shown in Figure 3.3, used eight minutes at the beginning (A) and eight minutes at the end of the time series (C). The remaining data (B) were used for testing the accuracy of the trained model's estimation. Only 70% of the 16-minute training data were used for training the weights of the ANN, while 15% were used for validation (to check the training stop condition) and 15% for testing of estimation performance in a one-step-ahead prediction. The mean square error of the 15% test data was used as a threshold; if it was 20% higher than the mean square error on the training or the validation data, the ANN was retrained.

Predictive training mode only uses the first eight minutes of the dataset, part (A), and tests the model on Part B (Figure 3.3).

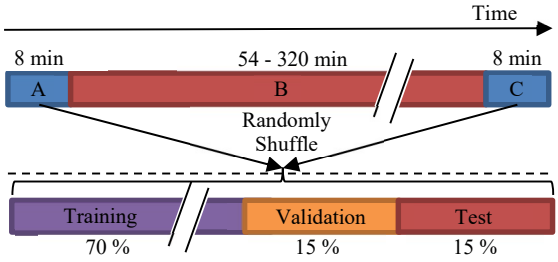

Figure 3.3 - Schematic of data split for the Interval training mode. Parts A and C were randomly split into training, validation, and test datasets. Part B is the sub dataset used for comparing the performance of the models. Only part A was used to train for the Predictive mode.

3.1.3 Subject Selection and Preprocessing

Patient data for this study were extracted from the MIMIC II database [151]. The database contains clinical signals, such as time-stamped nurse-verified physiological measurements, and vital signs time series obtained from a diverse population at an intensive care unit. The physiological waveforms from the MIMIC II database used in this study were obtained from the CareVue Clinical Information System (models M2331A and M1215A; Philips Healthcare, Andover, MA), which includes synchronized ECG (II, III or V), finger PPG, and invasive arterial BP from one of the radial arteries sampled at 125 Hz.

PhysioBank ATM was used to extract data from 15 different patients from the database to assess the performance of the NARX models [152]. The inclusion criterion for patients was the access to at least an hour of clean data, defined as a time span without abrupt changes in the ECG, PPG, and BP waveforms. The dataset lengths for the selected subjects ranged from 70 min to 336 min, with a mean and standard deviation of 160±84 min. The record numbers of the selected subjects are included in the Appendix.

Time points within each dataset that were corrupted by noise were removed from the dataset prior to model training. To find the noisy portions, double derivatives of the PPG and BP signals were computed. If the amplitude was outside +/- 5 standard deviations of the double derivative signals, the heartbeat was removed from the dataset. Outlier removal was double checked by a visual inspection of the data. It was observed that for subjects 3900975m and 3903773m, using +/- 5 standard deviation of the double derivative signals removed too much uncorrupted data. Therefore, 6 standard deviation was used for those subjects. Nevertheless, the removed time points generally spanned no more than a few heartbeats and may have arrived from temporary motion artifacts or minor signal loss/corruption. For all datasets employed no more than 7.2% of the data were removed for any subject, with an average and standard deviation of 2.6±2.2%.

The general architecture of the model (Figure 3.1.a, b) was adapted to the MIMIC II database, where specific measurement locations were used (three leads ECG, finger PPG, and radial artery BP). The delay between the ECG peak and the BP peak was smaller than 0.8 seconds. Zero to 99 time-step delays were used to ensure that the ECG peak was fed into the ANN when the output should be the BP peak. The NARX model with ECG inputs (ANN_{ECG}) used the following inputs to estimate BP(k):

$$ANN_{ECG} \text{ input layer: } BP(k\text{-}1), BP(k\text{-}2), ECG(k), \ldots, ECG(k\text{-}99) \quad (3.1)$$

PPG was measured further downstream than the BP (fingertip versus radial artery in the present study), resulting in a delay between the signals which is apparent in the measurement peaks (Figure 3.2). Thus, 1 to 100 future time-step delays of the PPG was used in the ANN (ANN_{PPG}) to estimate the BP(k):

$$ANN_{PPG} \text{ input layer: } BP(k\text{-}1), BP(k\text{-}2), PPG(k\text{+}101), \ldots, PPG(k\text{+}1) \tag{3.2}$$

Finally, the ANN model used to estimate $BP(k)$ from the ECG and PPG waveforms ($ANN_{ECG+PPG}$) was a combination of the two previous models with the following input layer:

$$\begin{aligned} ANN_{ECG+PPG} \text{ input layer: } & BP(k\text{-}1), BP(k\text{-}2), ECG(k), \ldots, ECG(k\text{-}99), \\ & PPG(k\text{+}101), \ldots, PPG(k\text{+}1) \end{aligned} \tag{3.3}$$

3.1.4 Pulse Arrival Time Model

Subject-specific PAT models were implemented to facilitate direct comparison with this established cuffless BP monitoring method. Specifically, SBP and DBP estimates were computed as [118]:

$$SBP = a_1 PAT + b_1 HR + c_1 \tag{3.4.a}$$

$$DBP = a_2 PAT + b_2 HR + c_2 \tag{3.4.b}$$

where the linear coefficient a_n, b_n, and c_n were patient-specific. The coefficients were determined using the Least Square Estimate method for each training dataset [118]. The model was implemented with both PAT_f and PAT_p, computed as the time difference between the ECG peak and the foot PPG or its peak respectively (Figure 3.2). The heart rate (HR) was computed from the ventricular contraction peaks (R-R interval) of the ECG signal. The MAP estimates were computed as [14]:

$$MAP = (SBP + 2\,(DBP))/3 \tag{3.5}$$

3.1.5 Data Analysis

For each NARX model, a cross-correlation analysis was performed to quantify any delay between the predicted and the measured BP. The mean and variance of the models' error, which can describe its bias and precision, were assessed via scatterplots and Bland-Altman plots, where ±1.96 SD was used for a 95% confidence interval. The SBP, DBP, MAP and the BP waveform estimations for the four different models (three NARX and one PAT) were benchmarked against BP error standards from the Association for the Advancement of Medical Instrumentation (AAMI) protocol, which requires that a device must achieve within the limits of 5 mmHg bias and 8 mmHg precision [153]. The mean of the subjects' mean error and standard deviation error were used for the comparison. Furthermore, the mean absolute error (MAE) and the Pearson correlation coefficient (r) were used to compare the models.

Moreover, different visual comparisons of the measured BP and estimated BPs are made by plotting the SBP (SBP has a larger variance than DBP and MAP) to understand how well the model can estimate changes in BP on different time scales (beat-to-beat, minute-scale, and hour-scale). The CPU training time and the training epochs were compared for each NARX model.

Friedman tests (α=0.05) was used to identify any statistical difference at the group level (comparing more than two distributions). If a difference was observed at the group level, the Friedman test was followed by Wilcoxon rank sum tests (p=0.05) with Bonferroni corrections to identify any statistical differences between the models. Those statistical tests are non-parametric, and there is no underlying assumption regarding the distributions of the tested samples. Comparing four different models (three NARX and one PAT), we used six pair-wise comparisons on each metric (e.g., mean errors) based on the data of 15 individuals.

3.2 Results

3.2.1 Model Performance Overview

Representative samples of predicted BP waveforms using ANN_{ECG} in both training modes at early and late times of segment (B), see Figure 3.3, are benchmarked against the measured BP in Figure 3.4 for three subjects with qualitatively different waveform characteristics. For all 15 subjects, no delay was observed in the closed-loop predicted waveform based on the cross correlation analysis, as can be observed in Figure 3.4. The other models (ANN_{PPG} and $ANN_{ECG+PPG}$), while not presented in this figure, yield similar results characteristics.

The ANN_{ECG} error evolution and overlayed linear regression over a period of 55 minutes (the shortest sample length across all subjects; subject 3000860m) is presented for all 15 subjects in Figure 3.5. The red lines are the subjects with the largest positive or negative error at the end of the interval. This figure shows the error evolution obtained for models trained under both training modes. It can be observed that the Interval training mode has lower model drift compared to the Predictive training mode, which is consistent with qualitative observations in Figure 3.4. Moreover, the drift of the error varies from subject to subject.

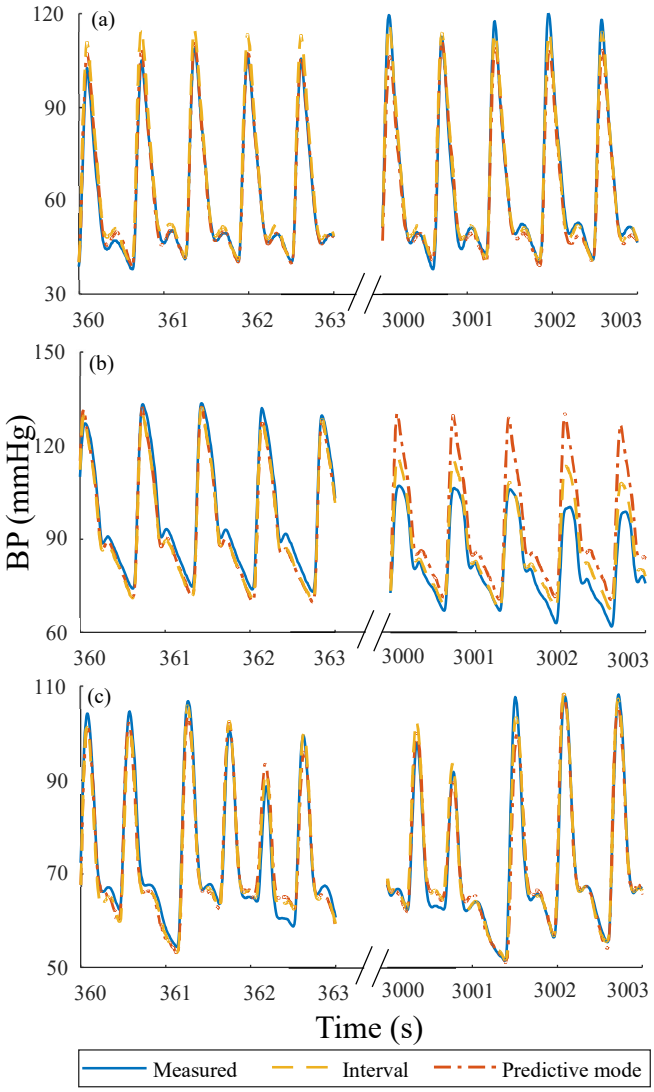

Figure 3.4 - Comparison of actual versus predicted BP waveforms for subject (a) 3002090m, (b) 3000480m, and (c) 3000989m, see Appendix for details. The BP waveforms show considerable qualitative differences in shape and heart rate variability.

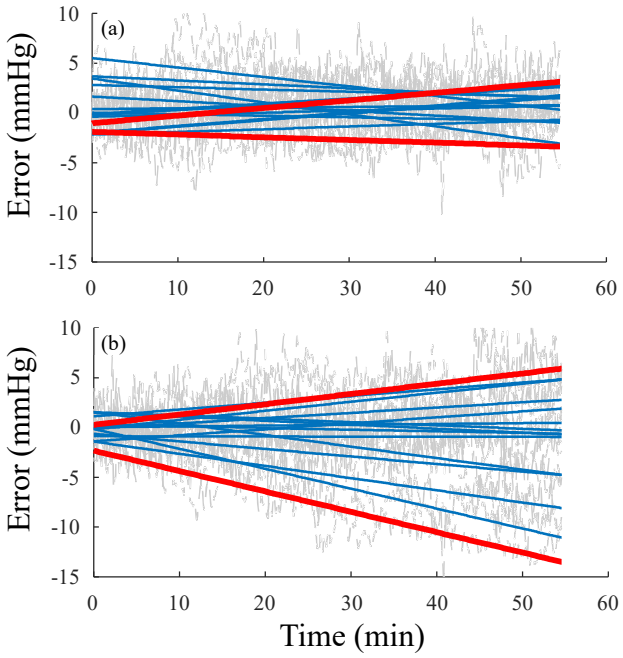

Figure 3.5 - Filtered BP error and linear fit for the 15 subjects for (a) Interval training, and (b) Predictive training for the first 55 min of testing data for each subject. Grey dashed lines are raw data, blue lines are linear regressions for each subject, and red lines are linear regressions of the subjects with the highest error at the end of the interval, (positive or negative).

The mean and standard deviation of the absolute error averaged over all subjects is presented in Figure 3.6. Time is normalized by the total signal length for each subject, which is used to bin the data prior to averaging. For the Interval training mode the expected absolute error is relatively constant over the entire prediction time; on the contrary, error tends to increase over time when Predictive training is employed. The standard deviation of error is similarly constant for Interval training and increases for Predictive training over time.

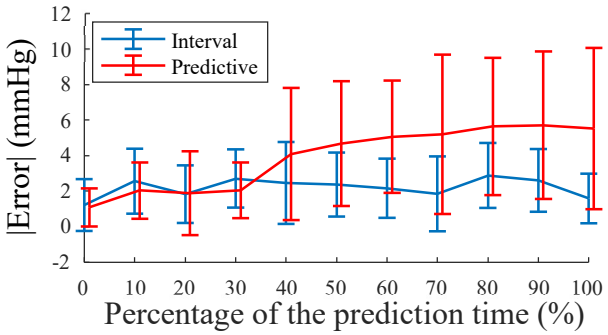

Figure 3.6 - Mean of the 15 subjects mean absolute error over time with the error bars representing the standard deviation. For each subject the mean absolute error is taken over 20 s of estimation.

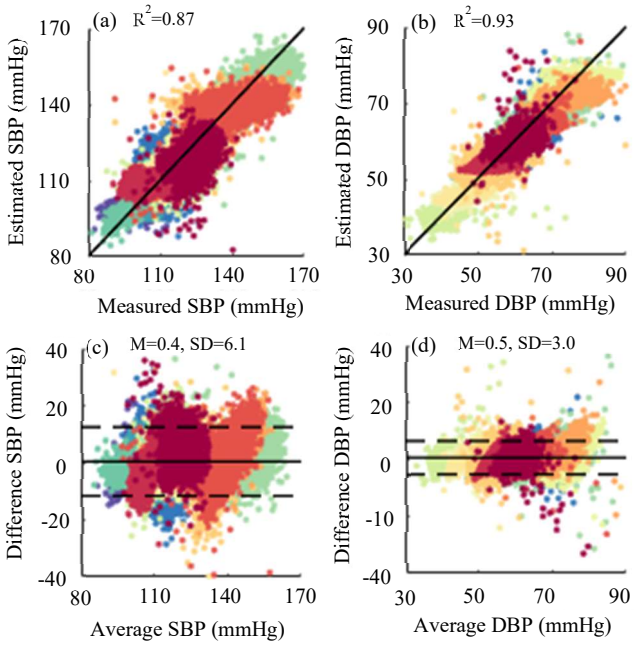

Figure 3.7 - Scatterplots of (a) SBP and (b) DBP for Interval training mode. The solid line is the regression line through all data Bland–Altman plots of (c) SBP and (d) DBP. The solid line and dashed lines represent fixed bias (*M*) and *M* ± 1.96 standard deviation range, respectively. Average = (estimated + measured)/2, Difference = measured – estimate. Subjects are color coded as defined in the Appendix.

Figure 3.7 presents scatterplots of the measured versus estimated SBP and DBP extracted from the BP waveform estimates with ANN_{ECG} for all subjects for Interval training. Bland-Altman plots of the SBP and DBP are also presented to highlight the bias and variance in the estimates. The data were sampled at equal intervals; 3000 data points (part B) for each subject were used, which is the largest

common number of data points for the subjects. The linear fit between estimated and measured SBP and DBP showed high coefficients of determination, R^2=0.87 and R^2=0.93, respectively. SBP and DBP were accurately estimated, where the errors fall within the 95% confidence interval of 0.4±12.0 mmHg and 0.5±5.8 mmHg, respectively.

Similarly, scatterplots and Bland-Altman plots for Predictive training mode are presented in Figure 3.8. The estimated SBP and DBP exhibit good correlation with the measurements, R^2=0.76 and R^2=0.86, respectively. Furthermore, SBP and DBP were estimated with 95% of confidence within -1.2±17.8 mmHg and -0.6±8.4 mmHg, respectively.

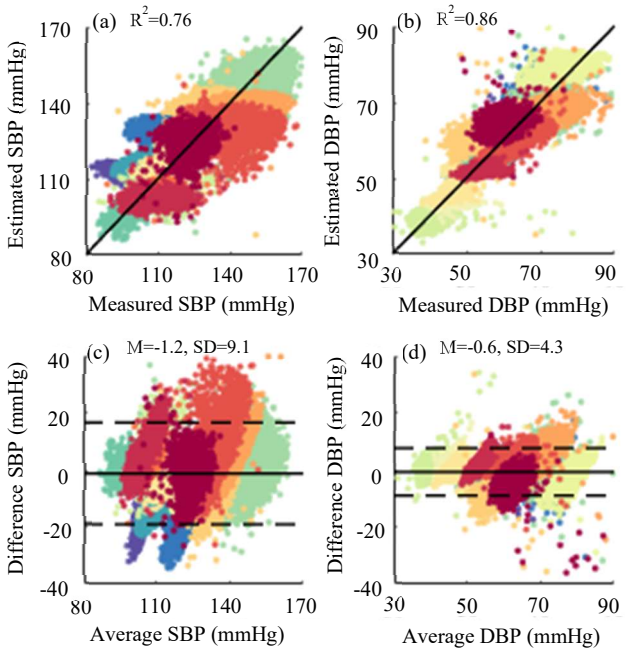

Figure 3.8 - Scatterplots of (a) SBP and (b) DBP for Predictive training mode. The solid line is the regression line through all data Bland–Altman plots of (c) SBP and (d) DBP. The solid line and dashed lines represent fixed bias (*M*) and *M* ± 1.96 standard deviation range, respectively. Average = (estimated + measured)/2, Difference = measured – estimate. Subjects are color coded as defined in the Appendix.

3.2.2 Model Comparison

To compare both PAT models (PAT_p and PAT_f) the mean value of the intra-subject mean error, the standard deviation of error, MAE, and Pearson correlation coefficient (r) for the SBP, DBP, and MAP

estimates were used in both training modes (Interval or Predictive). There were no statistical differences between any of the distributions (Wilcoxon test $p \geq 0.45$).

The expected value and standard deviation of the MAE computed across the 15 subjects for SBP, DBP, and MAP in both training modes are shown in Figure 3.9 for PAT_p and PAT_f. The results exhibit mild differences, with PAT_p yielding lower mean MAE in all cases. Given the slightly better performance, PAT_p was used as a reference to compare the NARX models within this section.

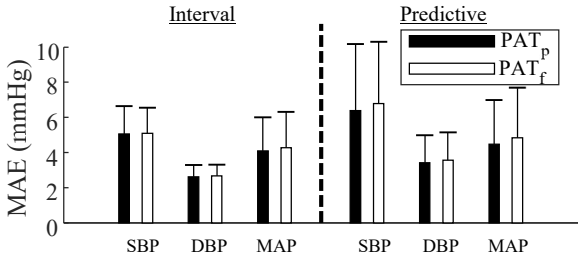

Figure 3.9 - Comparison of mean absolute error between different PAT models estimates and the BP measurements. The results are shown for SBP, DBP, and MAP for both the Interval and Predictive training modes. Bars represent the mean of all subjects and the error bars show ± standard deviation.

The absolute value of the mean error, $|\mu_e|$, (expected value and standard deviation) computed on the 15 subjects is presented in Figure 3.10.a for SBP, DBP, MAP, and the full BP waveform (BPW) for both training modes and all input signals considered. For the PAT model, the three outputs (SBP, DBP, and MAP estimates) are shown. The absolute errors were used to enable plotting in one quadrant, which is computed after performing all other calculations. The horizontal dashed line in Figure 3.10.a-b indicates the AAMI requirement limit. All models meet the AAMI requirement of less than 5 mmHg bias. When comparing the bias results for the four models, there is no group-level statistical difference (Friedman test $p \geq 0.282$).

The standard deviation of the error, σ_e, of each model with its expected value and standard deviation computed on the 15 subjects is shown in Figure 3.10.b. All expected standard deviations meet the AAMI standard deviation requirement of 8 mmHg and are at least one standard deviation lower than the AAMI limit. Moreover, $ANN_{ECG+PPG}$ estimates of the BPW when trained with the Interval mode was statistically different from the ANN_{PPG} model (Wilcoxon test $p=0.010$).

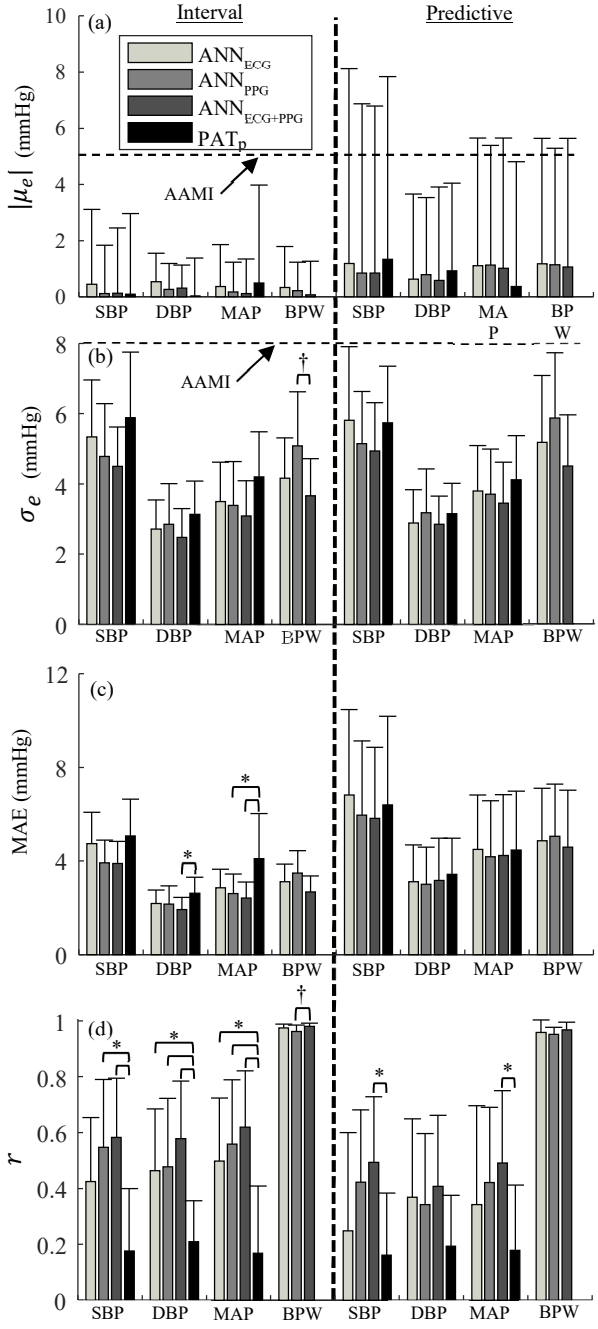

Figure 3.10 - Comparison of (a) the absolute value of the mean error (b) standard deviation of the error (c) mean absolute error, and (d) the Pearson correlation coefficient between different models estimates and the BP measurements. The results are shown for SBP, DBP, MAP and BPW for the Interval training and the Predictive training modes. Bars represent the mean of all subjects and the error bars show ± standard deviation. Legend: * p<0.05/6, † p<0.05/3.

The MAE is computed on the 15 subjects and illustrated in Figure 3.10.c. The MAE is lower for all NARX models compared to the PAT model, except for ANN_{ECG} estimates of SBP, when trained in the Predictive mode. Moreover, the $ANN_{ECG+PPG}$ always has the lowest absolute error among the models; it is significantly lower than the PAT model for DBP and MAP when trained with the Interval mode (Wilcoxon test $p<0.007$). ANN_{PPG} also has a significantly lower MAE for DBP than the PAT model when trained with interval mode (Wilcoxon test $p=0.006$).

Figure 3.10.d presents the average Pearson correlation coefficient, r, between the estimated and measured SBP, DBP, MAP, and BPW for all 15 subjects. The correlation coefficients obtained with the NARX models are always higher than the PAT model (Wilcoxon test $p\leq0.005$). Moreover, the $ANN_{ECG+PPG}$ trained with the Interval mode statistically outperformed the ANN_{PPG} for the BPW prediction (Wilcoxon test $p=0.006$).

Changes in BP happen over different time scales as shown in Figure 3.11. This figure compares the representative samples of measured SBP with the estimations on the test data obtained by different models tuned using Interval training. Figure 3.11.a shows that NARX models are more suitable than PAT for estimating large beat-to-beat BP variability observed for this subject. The MAEs (mean ± standard deviation) for this specific subject are 2.34±2.80, 2.11±2.77, 2.06±2.54, and 3.25±4.17 mmHg for ANN_{ECG}, ANN_{PPG}, $ANN_{ECG+PPG}$, and PAT, respectively.

Figure 3.11.b presents 10-beat moving averaged results of each model that better highlight the Mayer wave, which is characterized by a fluctuation in blood pressure related to vascular sympathetic activity at a frequency ranging from 0.03 to 0.15 Hz depending on the subject [154]. The selected exemplar subject has large Mayer wave compared to the others in this dataset. The NARX models are more accurate than PAT for estimating BP with a large Mayer wave and yield to the same results for smaller ones. The MAEs on the 10-beat moving averaged estimates for this subject are 3.91±4.90, 2.59±3.21, 3.05±3.81, and 4.20±5.01 mmHg for ANN_{ECG}, ANN_{PPG}, $ANN_{ECG+PPG}$, and PAT respectively.

Figure 3.11.c shows 300-beat moving average (approximately 5 min) results to highlight slow changes in BP. This subject exhibited the largest slow changes in mean SBP (from 108 to 134 mmHg) among the subjects selected for this study. The MAEs on the 300-beat moving averaged data for this subject are 4.36±4.41, 3.45±4.37, 3.55±4.26, and 4.55±5.54 mmHg for ANN_{ECG}, ANN_{PPG}, $ANN_{ECG+PPG}$, and PAT, respectively. Even though the slow changing SBP estimation exhibits relatively large errors, the NARX SBP estimates are highly correlated with the measured one. The Pearson

correlation coefficients are 0.72, 0.85, 0.91 and -0.57 for ANN_{ECG}, ANN_{PPG}, $ANN_{ECG+PPG}$, and PAT, respectively.

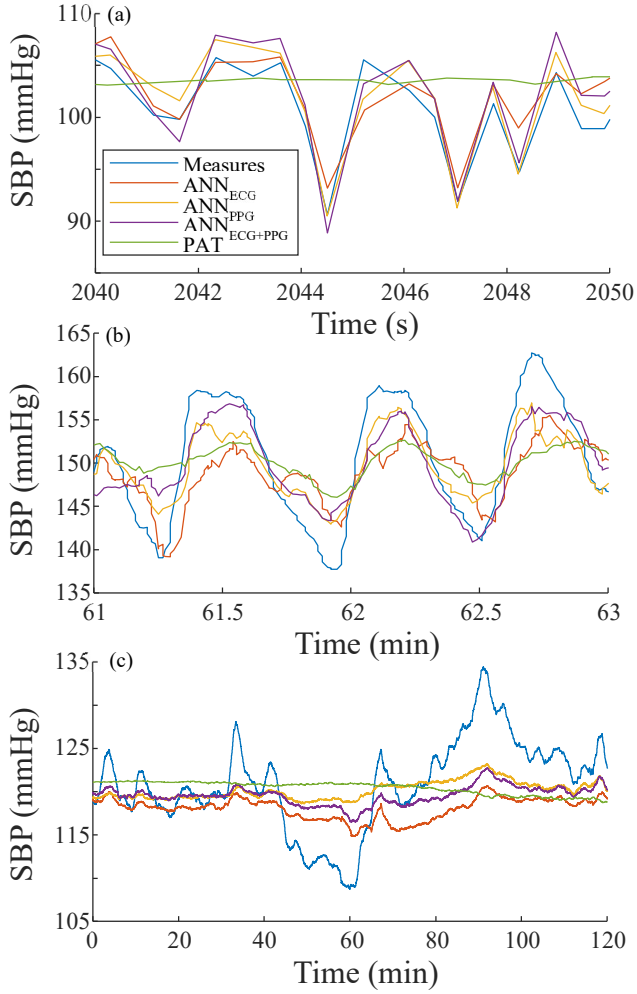

Figure 3.11 - Comparison of actual versus the four model estimations of SBP for subject (a) 3000989m, (b) 3001689m, and (c) 3901339m, showing BP changes on different time scales. In (a), the SBP is plotted for each heartbeat. In (b), the SBP is averaged over 10 heart beats. In (c), the SBP is average over 300 heart beats.

Slow changing SBP can be well modeled as it is observed in Figure 3.12. The MAEs on the 300-beat moving averaged data for this subject are 5.28±4.02, 2.20±2.61, 2.39±2.18, and 4.27±3.36 mmHg and Pearson correlation coefficients of 0.64, 0.85, 0.92 and 0.56 for ANN_{ECG}, ANN_{PPG}, $ANN_{ECG+PPG}$, and PAT, respectively.

A source of error for the models is that the SBP range covered in the training data was different from the one in the test data. The difference in the maximum SBP in the training data for a subject

compared to their test data was -2.76 ± 4.97 mmHg, and the difference in the minimum SBP was 5.21 ± 4.77 mmHg. The SBP in the test data that was outside the SBP training range was in average 1.2 ± 1.7% and ranged from 0 to 5.1 %.

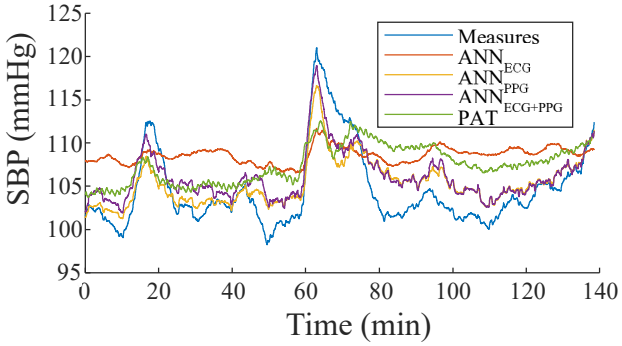

Figure 3.12 - Comparison of actual versus the four model estimations of SBP for subject (a) 3903867m. The SBP is average over 300 heartbeats.

The training CPU time and the number of epochs required for convergence of the NARX models' weights are compared for each input and training mode depicted in Table 3.1. The times presented are using MATLAB 2016b on a computer housing four i7-3770 3.4 GHz cores and 16 GB RAM.

Table 3.1 - Training average CPU time and the number of epochs for every input and training mode. Legend: * p<0.05/3 in comparison with ANN_{ECG}, † p<0.05/3 in comparison with ANN_{PPG}.

Input	Interval training time (mean ± SD min)	Interval training epochs (mean ± SD)	Predictive mode training time (mean ± SD min)	Predictive mode training epochs (mean ± SD)
ANN_{ECG}	6.8 ± 3.7	58 ± 32	2.0 ± 1.2	36 ± 23
ANN_{PPG}	13.2 ± 10.6*	123 ± 100*	4.2 ± 2.4*	78 ± 47*
$ANN_{ECG+PPG}$	14.6 ± 7.6*	44 ± 33†	4.3 ± 0.6*	20 ± 3†*

3.3 Discussion

The model architecture presented in this chapter enables the estimation of the complete BP waveform only using the ECG or peripheral PPG signals. To the best of our knowledge, this is the first such demonstration. The estimation of the BP waveform, our main contribution, can be used by clinicians to gain additional insights into patient health conditions; for instance, access to an estimation

of the cardiac output [20], stroke volume [21], and arterial stiffness, which provides information on the mechanisms leading to high blood pressure [18] and can provide a guideline for enhanced hypertension treatment [60]. Moreover, this work introduces a novel approach to modeling BP, wherein it is considered as a dynamical system that changes over time as a function of the heart signal (ECG and/or PPG), and thus offers a new cuffless BP monitoring methodology, which has contributed to improving the accuracy of the SBP, DBP, and MAP estimations.

The NARX model is easy to implement and train; only simple pre-processing, such as adding white noise to the BP training data and data normalization, are employed. No complicated feature extraction and computations are needed as in other machine learning based studies [155], [156]. Furthermore, the model is flexible to input signal type, so long as it carries information about BP, making the applicability less dependent on the hardware available. In addition, the low complexity of the ANN used in the model architecture (one hidden layer of 10 neurons) makes the training relatively fast, which has been an issue with other deep learning models [12]. Note that while for this study we performed some preprocessing to remove erroneous outlier segments, in practice this step is not necessary, as the model will re-establish fidelity once the input signal recovers.

The method developed herein could be used in practice in two scenarios. In scenario 1, the exogenous input data are collected using a typical three-lead ECG and a finger-mounted PPG. The BP waveform for training the model is collected using a volume clamping method [77], [78]. Once eight minutes of BP data are collected, BP measurement equipment is removed, and the dataset can be used to train the ANN (Predictive training mode). Next, future ECG and/or PPG data are used as input to the trained ANN to estimate BP. In scenario 2, after the first eight-minute data collection including BP, ECG and/or PPG is collected for several hours, followed by an additional eight minutes of BP data (Interval training mode). Both eight minutes datasets are used to train the ANN, where the BP in the multi-hour interval can be retrieved using the trained ANN and the ECG and/or PPG data.

3.3.1 Interpretation of the Results

The good performance of the NARX models employed with the proposed input layer herein indicates that BP can be modeled with a nonlinear second-order system (two time-step delays of the BP), with some history of an exogenous input. This nonlinear second-order model is able to estimate a variety of BP waveform shapes (Figure 3.4.a, b), which are known to be descriptive of physiological factors, such as arterial stiffness, as well as measurement location [157], and is robust to changes in BP magnitude and heart rate over the time scale of a single beat (Figure 3.4.c). Longer time-scale variations

in the vascular system, such as blood redistribution during digestion, can result in drift of the NARX model estimates. This was observed in the case of Predictive training mode, and is also found in the literature for PAT/PTT models [118], [158], wherein recalibration was used to compensate for that drift. Herein it was shown that longer time-scale changes in the cardiovascular system can be accommodated by training the NARX on two temporally separated portions of a time series and estimating the BP in the interval. The improved accuracy of this training mode over the Predictive training mode has practical implications, as it would enable clinicians to correct recorded data post hoc.

The two training modes affect the model outputs differently. Interval training reduced model bias for every output estimation (SBP, DBP, MAP, and BPW). This suggests that the PAT model can also benefit from this type of training and post hoc correction can improve the overall performance. The standard deviation of the error is not much influenced by the training mode. This suggests that the difference in MAE between both modes is mainly due to the model mean error drift for the Predictive training mode. Finally, Interval training results exhibit better overall correlation to the measured BP. This implies that if precise information is to be mined from the BP waveform for additional health monitoring insights, Interval training is preferred.

The SBP and DBP of the subjects ranged between 80 and 170 mmHg and from 30 to 90 mmHg, respectively, which spans the range of hypotension to hypertension for individuals at rest [8], [159]. The accuracy of the present model was not dependent on the amplitude of the BP; there are no indications that higher BPs result in any performance degradation given the nonlinear capacity of the NARX model; this is not necessarily the case for linear models [111].

Different definitions of PAT have been used in the literature for estimating BP [6]. Some studies have compared different definitions, but the results are inconclusive [160]–[162]. Nevertheless, to have a fair comparison for the NARX models, two PAT definitions, PAT_p and PAT_f, were implemented. The results show no statistical difference between the two PAT approaches, but PAT_p yielded a slightly better overall accuracy for BP estimates on this dataset, which is in accordance with some of the previous work on other datasets [159], [160].

The prediction of the SBP and DBP with the NARX model is not made at the expense of predicting all the waveforms when compared with the PAT model. On the contrary, the NARX models have lower error and higher correlation with the measured BP when compared with the PAT model, with the exception of SBP estimation based on ECG data when using Predictive training mode. The $ANN_{ECG+PPG}$ exhibited a significantly lower MAE on the DBP than the PAT model when trained using Interval mode. Moreover, the correlation between the DBP signals and the estimates using all NARX models, and

correlation for the SBP using the ANN_{PPG} and the $ANN_{ECG+PPG}$ are significantly higher when compared to the PAT model. The computed MAP from the complete waveform estimates is significantly better correlated to the measured MAP for all NARX model when trained with the Interval method and leads to better MAE using the ANN_{PPG} and $ANN_{ECG+PPG,}$. Moreover, $ANN_{ECG+PPG}$ is trending towards a lower MAP error standard deviation than the PAT model (p-value = 0.009). $ANN_{ECG+PPG}$ always yields the best results in terms of model bias, error standard deviation, mean absolute error, and correlation coefficient; it is significantly better than ANN_{PPG} for prediction of the waveform in terms of precision and correlation when trained using Interval mode.

It was shown that BP changes over different time scales can be well modeled by NARX models, mostly for beat-to-beat scale (Figure 3.11.a) and the minute-scale (Figure 3.11.b). As for the hour-scale, ANN_{ECG} seems to be performing poorly for all subjects (Figure 3.11.c and Figure 3.12). However, for ANN_{PPG} and $ANN_{ECG+PPG}$, the accuracy of estimations varies from one subject to another. The difference in results can be attributed to the difference in the training data of subjects. For example, the training data from subject 3903867m (Figure 3.12) had the SBP ranging from 95 to 135 mmHg while the trained model was used to estimate SBP ranging from 89 to 137 mmHg. This can explain why ANN_{PPG} and $ANN_{ECG+PPG}$ are predicting the maximum SBP but failed to predict the minimum SBP due to lack of similar data in the training set. Similarly, the training data from subject 3901339m (Figure 3.11.c) ranged from 109 to 132 mmHg but the trained model was used to estimate SBP with a larger range of 97 to 147 mmHg, leading to both poor estimations in the high BP and low BP range. This was the case for most of the current subjects, where the range of SBP in training data was generally smaller than the one in the test data. The range of training SBP combined with the edge effect, i.e., less training data closer to the extremum values resulting in poorer model performance close to the extremum, can explain why, in Figure 3.7, even though the 95% confidence interval is quite narrow, the difference of estimates can be quite large (> 60 mmHg); depending on the training data, the model can perform poorly on unseen high BP and/or low BP data. This indicates a limitation of the current study that can be addressed in the future by collecting rich training data, such as by including measurements of BP in relaxed (to decrease its BP) and exercise (to increase the BP) conditions.

Using two sensors may not always be possible or convenient; for example, PPGs suffer from motion artifact [84], [123], so they may be less desirable for ambulatory monitoring. If a single sensor is to be implemented, ANN_{ECG} generally leads to better estimation of the DBP and the BPW, where its BP estimation is not statistically different from the $ANN_{ECG+PPG}$. The ANN_{PPG} showed better results in terms of SBP estimation, though not statistically different from the other input models. Physiologically,

the results can be explained by the ECG, being the electrical signal for the contraction of the heart, thus better represent the dynamics of the BP, resulting in accurate BPW estimates. The better mapping of changes in blood volume to the BP with ANN_{PPG} can be attributed to a higher maximum cross correlation of BP with the PPG signals than with the ECG signal. which could lead to a better estimation of the peaks.

The present NARX model differs from other machine learning models for BP estimation in how much data are used for testing. In the present literature, 15% to 30% of the data are used for testing the model [163], [14]. Herein, 64% to 90% of the data were used for testing on different subjects; the goal was to build a model out of a relatively small amount of training data to estimate the BP for long time intervals.

Adding hidden layers in the multilayer Perceptron [163] or using a deep recurrent neural network [164], can lead to longer training time and potentially better performance if sufficient training data are available. However, the complexity of the model in those cases was justified by the attempt of building a universal model, i.e., a model that could suit anyone, which was not the objective herein. In this study, the model was subject-specific and trained on a small amount of data, and only required one hidden layer of 10 neurons. Increasing the number of hidden neurons to 20 resulted in no decrease in the MAE of the estimated BPW. The MAE was increased when using 5 neurons in the hidden layer though. The NARX architecture was able to model the nonlinear behavior of the BP with only eight minutes of training data. It is interesting to compare training CPU time for every input and training mode as it can help the user in choosing the NARX configuration depending on the application. For example, the model is applicable for physicians, where four hours of data could be retrieved within a single visit. In that case, ANN_{ECG} could be preferred since the training time is significantly smaller than the others.

The general model architecture was adapted for the MIMIC II database, in which the measurement locations and the sampling frequency of 125 Hz were fixed. The number of time-step delays of the exogenous input was then adapted such that the pulse wave delay for each input was shorter than the maximum input delay. Nevertheless, the size of the input can be changed to account for a different pulse wave delay if one is using a different peripheral location of the sensors. Similarly, the sampling frequency can be changed as long as all the information about the BP is within the input layer. If a higher sampling frequency is to be used, one must be careful in using two consecutive BP time-step delays as feedback for the ANN as the dynamics of the BP could become buried in the noise.

It was observed that adding white noise to the BP training time series improved the robustness of the model when being used in the recurrent fashion (estimation configuration). Although the effect of

noise amplitude was not extensively studied here, preliminary work showed that a 4-fold change in the standard deviation of the noise, ranging from half to twice, has no considerable effect on the reported results.

Data corrupted by motion artifacts were removed in this study as they diminished the accuracy of the estimation of the NARX model. Nevertheless, the model could be used with some degree of motion artifact with a risk of a decline in the estimation accuracy. It is, however, not possible to quantify the loss of accuracy as even the ground truth BP was noisy. In order to increase the robustness to such noise, the model could be trained on data with motion artifact, or filters could be applied to the input signal [12].

In this study, all training windows were eight minutes long; model performance as a function of training window size was not assessed. The aim was to minimize the training duration and maximize the span between required training sessions. Eight minutes was selected as an estimate of a feasible window for real-world applications. The results were acceptable when compared with PAT results; in future work the training duration could be optimized for each subject and accuracy requirement. In this study, the model estimated the BP for one to 5.2 consecutive hours without recalibration, depending on the subject data availability, because the aim was to maximize the time span of the prediction. However, the accuracy of the estimation will be a function of the length of the time span between the training sections, i.e., a shorter testing interval will lead to a more accurate result. This is true for Predictive training mode, but also applicable to Interval training mode, as it would minimize the interpolation needed from the model.

3.4 Conclusion

It was shown that the complete BP waveform can be predicted from an ECG and/or PPG time series with an NARX model using ANN. The SBP, DBP and MAP estimation accuracy is better than a PAT model when compared on the same 15 subjects from the MIMIC II database. Using eight minutes of training data, the model was able to estimate the BP for one to 5.2 consecutive hours without recalibration, depending on the subject. With the availability of the complete BP waveform, this model can help physicians extract additional details of the physiological condition of a subject and better monitor changes in time. The additional information will enable the estimation of other metrics, such as cardiac output and stroke volume of the patient; this is left as future work.

Chapter 4
Estimation of Blood Pressure during Activities of Daily Living[5]

In this chapter, subject-specific NARX models are trained to estimate the BP waveform time series from electrocardiography (ECG) and forehead photoplethysmography (PPG) input signals during activities of daily living. To broaden the range of BP in the training data, subjects followed a short procedure consisting of sitting, standing, walking, Valsalva maneuvers, and static handgrip exercises. This chapter evaluates two different components of the method.

First, the efficacy of the cuffless BP method for continuous BP estimation during activities of daily living is evaluated when varying training data segments of the procedure to understand better the effect of each intervention. Second, the performance of three different NARX models are compared with three pulse arrival time (PAT) models. We hypothesize that the developed method for cuffless BP estimation during activities of daily living can be used for continuous BP monitoring and, therefore, for acute hypotension and hypertension detection.

4.1 The Training Procedure

The main focus of this section is to investigate the effect of the various training protocol sections on the accuracy of the model estimations throughout the day.

4.1.1 Method

4.1.1.1 Measurements

Data were collected on 3 young healthy males [age 27 ± 3.5 yr (mean ± SD)] of varied fitness levels, ranging from sedentary to regularly active, and free of cardiovascular and peripheral vascular disease.

[5] Parts of this chapter have been adapted from the following publications:

C. Landry, E. T. Hedge, R. L. Hughson, S. D. Peterson, and A. Arami, "Cuffless Blood Pressure Estimation for Activities of Daily Living," presented at the Conf Proc IEEE Eng Med Biol Soc, Jul. 2020.

C. Landry, E. T. Hedge, R. L. Hughson, S. Peterson, and A. Arami, "Accurate Blood Pressure Estimation during Activities of Daily Living: A Wearable Cuffless Solution," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2021, doi: 10.1109/JBHI.2021.3054597.

All participants signed informed consent forms before participating in the study. Study protocols and procedures were approved by the University of Waterloo, Clinical Research Ethics Committee (ORE#41490) and conformed with the Declaration of Helsinki.

Arterial blood pressure, see Figure 4.3, was continuously measured at 100 Hz using finger PPG (Portapres; Finapres Medical Systems, the Netherlands) [81]. Pressure measures were height adjusted to heart level and the arterial pressure waveform was calculated using BeatScope™ 1.1a software. The "finger to brachial: level corrected," option was employed, which passes the finger pressure signal through a transfer function and a regression-based level correction to estimate the brachial artery BP [81].The middle finger of the non-dominant hand was used for measuring the finger BP, and the arm was put into a sling to minimize motion artefact. Each time the Portapres turns on, it performs a calibration that lasts anywhere from 3 to 11 minutes. Time points during these initial calibrations, as well as during periodic calibrations throughout the day, were removed from the data.

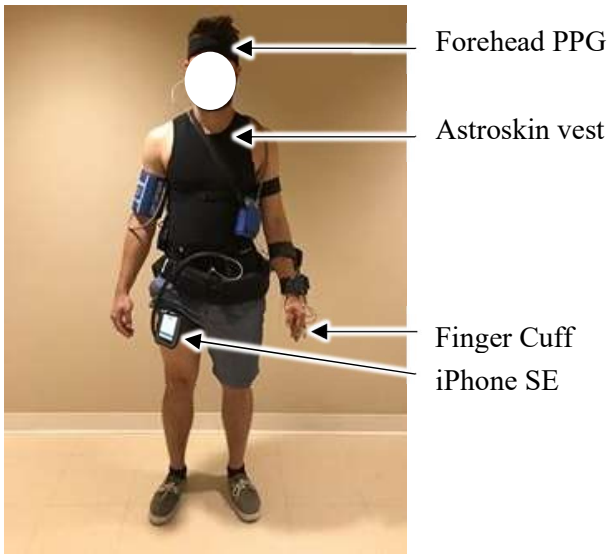

Figure 4.1 – A participant with the sensors used during monitoring

The Astroskin wearable body metrics vest (Carré Technologies Inc., Canada) was used for ECG, PPG, and tri-axial acceleration measurements [165], see Figure 4.3 for sample signals. Each participant was fitted with the correct size based on thoracic and abdominal circumference measurements. Astroskin hip accelerometers, recording at 64 Hz, were secured in the vest pocket as indicated by the manufacturer. The ECG signal was recorded at 256 Hz from 3 electrodes embedded into the vest (2 located on the chest, and another located lower on the right side of the rib cage). The PPG was recorded at 64 Hz on the center of the forehead and was held in place with an elastic headband. The PAT is

calculated internally at 256 Hz. The sensor placement can be seen in Figure 4.1. Note that the iPhone SE on Figure 4.1 was not installed in this study, and only used in the next study (see Section 4.2.1.1).

4.1.1.2 Protocol

Testing was performed at the Schlegel-University of Waterloo Research Institute for Aging (RIA). Participants had access to the RIA Research Apartment to move around, take breaks, and have lunch while their bio-signals were continuously recorded (Figure 4.2.a-c). The participants performed their office work in the Vascular Aging and Brain Health Lab (Figure 4.2.d).

Figure 4.2 - The a) living room, b) kitchen, and c) bedroom in the RIA Research Apartment, and d) office and lab space in the Vascular Aging and Brain Health Lab.

The testing session consisted of three one-hour windows of continuous BP monitoring separated by one hour. When data were not being collected the Portapres was turned off to relieve pressure from the subject's finger. A 15 min procedure, used as training data for the model, was performed at the beginning of the data collection. This procedure was designed to change the participant's BP within a range that is usually experienced during activities of daily living. The sequence of activities included in the procedure is shown in Table 4.1. Five minutes of seated rest at the end of the day was also recorded and included in the model training.

Between the initial training and the seated rest at the end of the day (4.67 hours) the participants were free to do as they pleased, excluding vigorous exercise. There were no structured activities, though participants were encouraged to go for a walk from time to time, which included going up and down stairs between two floors, to ensure the collection was not only on sedentary data. Moreover, it was

suggested to every participant that they lie down at some point during the day to increase the range of measured BPs.

4.1.1.3 Model Architecture

The NARX architecture using an Artificial Neural Network (ANN) was adapted from Chapter 3 for this study. The ANN used a multilayer Perceptron with one hidden layer of 20 neurons in every case. Hyperbolic tangent and linear activation functions were used for the hidden layer and the output layer, respectively. The time between the ECG peak and the DBP was set to be 0.2 seconds to synchronize the devices. In addition, the PPG signal was temporally shifted such that there was no delay between the foot of the ECG and PPG signals (see PPG_d in Figure 4.3). Temporally shifting the signals with respect to one another (ECG, PPG_d, and BP) was performed in order to minimize the ANN input layer; that is, phase shifting allowed the ECG and PPG peaks to be always in the data fed to the input layer when estimating SBP for all individuals. That approach was selected instead of having an input layer large enough to cover different ranges of PAT (ECG to BP and ECG to PPG phase delays) [166]. The specific time shifts employed were for convenience and do not impact the estimation.

Table 4.1 – Duration of each activity included in the procedure

Segment #	Duration (min)	Activity
1	0.5	Sit
2	0.25	Seated Valsalva
3	0.75	Sit
4	1	Stand
5	0.25	Standing Valsalva
6	0.75	Stand
7	1	Walk
8	1	Sit
9	1	Seated handgrip
10	1	Sit
11	0.5	Stand
12	1	Standing handgrip
13	1	Stand
14	5	Sit

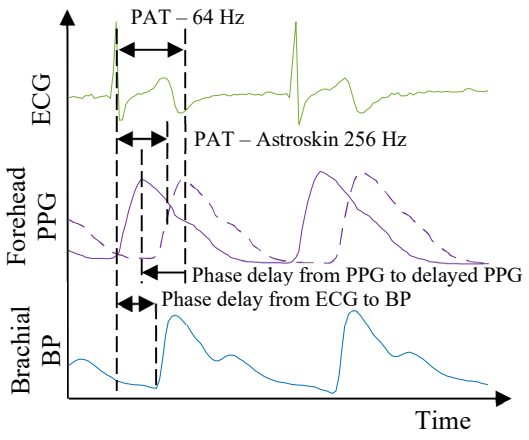

Figure 4.3 - Representative ECG, PPG (dashed), delayed PPG (solid), and BP signals illustrating the potential temporal delays between signals.

The input of the ANN consisted of 0 to 18 time-step delays of the ECG and the PPG_d sampled at 64 Hz. The NARX model also used the BP at time-steps *k*-1 and *k*-2 to estimate BP at time point *k*. The model architecture is summarized in Figure 4.4. A feedforward ANN was trained on the exogenous inputs and the BP measurements (Figure 4.4.a). Then, the trained ANN was used to estimate the BP on the test data, where a feedback loop with delay was implemented to provide the ANN with the estimates of the required delayed BPs (Figure 4.4.b).

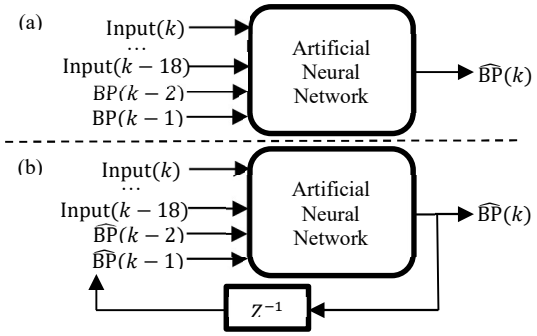

Figure 4.4 - Schematic of the NARX model used for a) training of the ANN as a feedforward ANN, and b) estimating and testing the BP using a recurrent ANN. The inputs are ECG and PPG_d.

4.1.1.4 Training

The delayed exogenous inputs and the delayed BP measurements of a given participant were used as inputs for training a personalized ANN (Figure 4.4.a). Zero mean white Gaussian noise (standard deviation = 1 $mmHg^2$) was added to the BP time series to enrich the training data and improve the model robustness to input noise. Error backpropagation was used with the Levenberg-Marquardt optimization algorithm to minimize the norm of the difference between the predicted and measured values of the BP. A training stop condition was defined as an increase of the mean squared error on the validation dataset in six consecutive epochs.

The data split for training and testing the ANN is shown in Figure 4.5, where the 15-minute protocol at the beginning (A) and 5-minute seated at the end of the time series (C) were used for training. To explore the effect of the training protocol on BP estimation fidelity, 6 different permutations of the training protocol were considered (other data from the training sessions were removed from the set). The 6 training permutations are shown in Table 4.2 with their respective included segment numbers from Table 4.1. Segments number 3, 8, 10, and 14 contain transitions from one activity to another at the beginning of the segment. Those transitions were removed (15 seconds) when investigating different training protocol sections. It is also noted whether the training set included the 5-minute seated data at the end of the day (C).

Table 4.2 – Permutations of the training data employed

Permutation	Segment #	Training on Portion (C)
Train1	**1 to 14**	**No**
Train 2	**1 to 14**	**Yes**
Train 3	**1, 2, 3, 8, 9, 10, 14**	**Yes**
Train 4	**1, 3, 8, 9, 10, 14**	**Yes**
Train 5	**1, 2, 3, 8, 10, 14**	**Yes**
Train 6	**14**	**Yes**

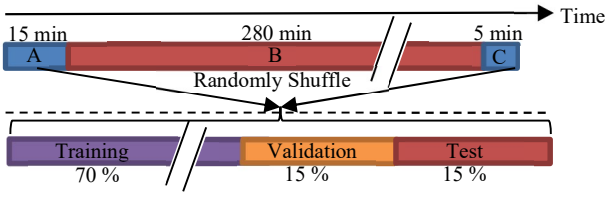

Figure 4.5 - Schematic of data split for training and testing the models. Parts A and C were randomly split into training, validation, and test datasets for training the ANN, stopping the training, and acceptation/rejection of the trained ANN, respectively. Part B is the sub-dataset used for comparing the performance of the models.

Only 70% of the training data were used for tuning the weights of the ANN, while 15% were used for validation (to check the training stop condition) and 15% for testing estimation performance in a one-step-ahead prediction. The mean square error of the 15% test data was used as a threshold for retraining; that is, if it was 20% higher than the mean square error on the training or the validation data, the ANN was retrained.

4.1.1.5 Validation

The portion (B) of the full dataset (see Figure 4.5) was used for testing the estimation accuracy of the model. Both the full data set in (B) and the dataset with the walking and lying down/sleeping segments removed (noWnoS) were tested to observe any difference in accuracy for different activities of daily living. The walking and lying down segments were detected via the Astroskin accelerometer.

4.1.1.6 Data Analysis

Predicted SBP and DBP time series were compared with the measured values from the Portapres. The SBP and DBP error time series were compared for each participant. The mean absolute errors (MAE) on the complete dataset (B) were used as a metric for comparing the model performance for each participant. The mean and the standard deviation of the MAE of the 3 participants' noWnoS data were compared for the 6 training permutations and benchmarked against Train1 on the complete dataset (B). The 3 training permutations with the lowest MAE were then investigated more in-depth. The errors were separated into bins according to the distribution of the noWnoS BP data over the course of the day. The bins were defined for each participant as -3 to -2, -2 to -1, -1 to 0, 0 to 1, 1 to 2, and 2 to 3 standard deviations (SD) away from the mean BP of the noWnoS data. The MAEs were then compared for each bin. Finally, the mean and variance of the model's error, which describes its bias and precision,

were assessed via scatterplots and Bland-Altman plots, where ±1.96 SD was used for a 95% confidence interval.

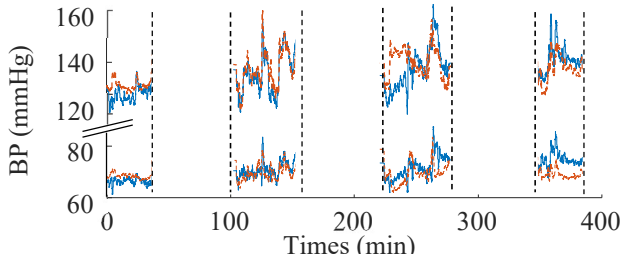

Figure 4.6 - Estimated (dashed) and the measured values (solid) for SBP (top) and DBP (bottom). Gaps in data are when the Portapres was off.

4.1.2 Results

4.1.2.1 Model Performance Overview of the Full Dataset

The extracted SBP and DBP from the predicted BP waveform are compared with the measured values in Figure 4.6 for one participant. The BP estimation has an error of 0.93±8.32 mmHg for the SBP and -1.68±6.96 mmHg for the DBP, resulting in MAEs of 6.3 and 5.2 mmHg, respectively. At the beginning of the third segment of data, a large discrepancy is observed between the measured and predicted BP. This discrepancy coincides with the person lying down, as detected by the Astroskin accelerometer.

BP estimation error evolution is shown in Figure 4.7 when using Train1 for each participant. For the complete dataset, the MAE on the SBP estimates for each participant (BP01, BP02, and BP03) is 11.3, 8.4, and 6.3 mmHg, respectively and the MAE on the DBP is 9.8, 6.4, and 5.2 mmHg, respectively. It can be observed that between 45% (see grey boxes in Figure 4.7) and 60% both BP02 and BP03 have a large error due to the participants lying down. This is also reflected in Figure 4.6 at the beginning of the third estimation segement.

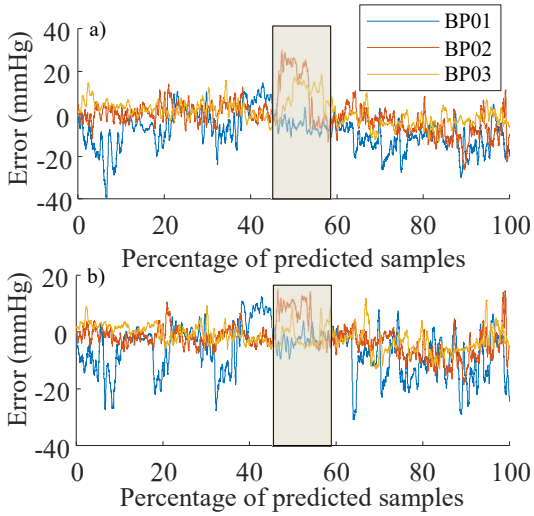

Figure 4.7 - Estimation error evolution for a) SBP and b) DBP

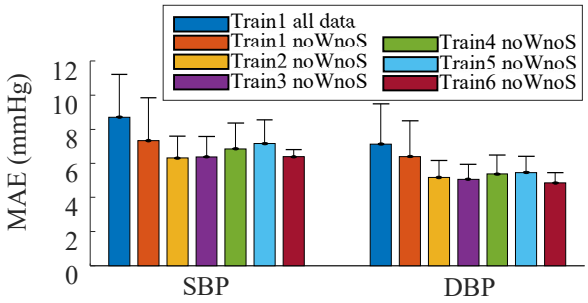

Figure 4.8 - Mean and standard deviation of the MAE of the 3 participants for SBP and DBP. The comparison is made for the 6 training data splits on noWnoS data and benchmarked against Train1 on all data.

4.1.2.2 Training Comparison

Figure 4.8 compares the MAEs for different training data splits on the noWnoS data and is benchmarked against the MAE on the complete dataset using Train1. It can be observed that by removing the walking and sleeping segments, there is an average MAE reduction of 1.4 mmHg when using Train1. Train2, Train3 and Train6 have the lowest mean MAE, with values of 6.3, 6.4, and 6.4 mmHg, respectively, for SBP and 5.2, 5.1, and 4.9 mmHg, respectively, for DBP.

The MAE on the 3 participants according to the distribution on their SBP and DBP throughout the day is shown in Figure 4.9. The results are sorted in bins according to numbers of standard deviation (SD) away from the mean BP of noWnoS data.

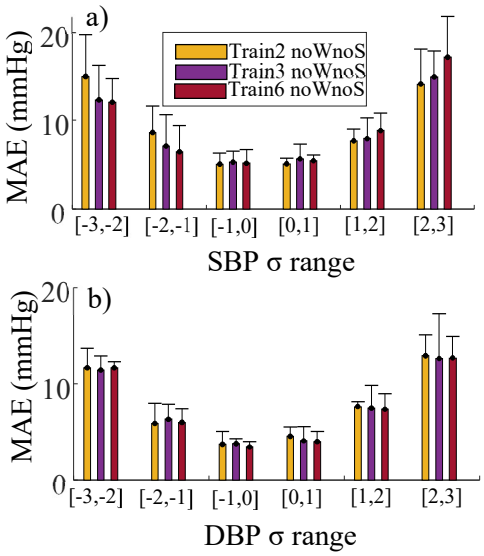

Figure 4.9 - Mean and standard deviation of the MAE for the 3 participants with respect to the distribution of the BP throughout the day. a) SBP, and b) DBP.

4.1.2.3 Best Estimation Results

Figure 4.10 presents scatterplots of the measured versus estimated SBP and DBP extracted from Train2 noWnoS model estimates for the 3 subjects. Bland-Altman plots of the SBP and DBP are also presented to highlight the bias and variance in the estimates. The linear fit between the estimated and measured SBP and DBP have coefficients of determination of R^2=0.69 and R^2=0.60, respectively. SBP and DBP were accurately estimated, where the errors fall in 95% confidence interval of -0.8±15.34 mmHg and 0.17±14.18 mmHg, respectively.

Table 4.3 compares the results of the Train2 noWnoS model for the three subjects with the requirements according to the AAMI protocol. It shows that the model achieves the bias and precision requirements for DBP and is near the AAMI requirement for the SBP precision.

Table 4.3 - Error comparison between the proposed model estimates and the AAMI requirements for the three subjects

	Mean error (mmHg)	**Standard deviation error (mmHg)**
SBP	**-0.80**	**8.67**
DBP	0.17	7.09
AAMI	**5.0**	**8.0**

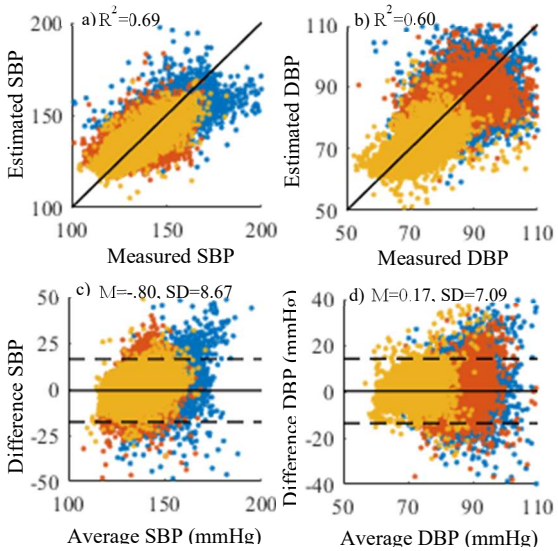

Figure 4.10 - Scatterplots of a) SBP and b) DBP. The solid line is the regression line through all data. The coefficient of determination is shown near each regression line. Bland–Altman plots of c) SBP and d) DBP. The solid line and dashed lines represent fixed bias (*M*) and *M* ± 1.96 standard deviation range, respectively. Average = (estimated + measured)/2, Difference = measured – estimate. Subjects are color-coded.

4.1.3 Discussion

The model architecture used in this section enables estimating the complete BP waveform using information encoded in ECG and PPG signals. The NARX architecture performance using an ANN was shown to be dependent on inclusion of informative activities in the training data. Using a 15 min protocol that changes the BP over a wide range and 5 min of sitting at the end of the day as training data gives access to a continuous estimation of the BP for at least 280 min. Moreover, the performance of the model is enhanced on a larger range of BPs. It was shown that the use of MAE as a criterion for continuous estimation of the BP can bias the estimation as most of the time during the day the BP is close to an average value. It is important to have a model that can detect fluctuations in the BP even though they may happen briefly as short-term BP variability is often associated with cardiovascular risk factors [9].

It was shown that using a training protocol that does not incorporate the supine position of the participant leads to large prediction errors, up to 20 mmHg, when lying down (Figure 4.7). This may indicate that the relationship between the input (ECG and PPG) and the output (BP) changes as a

function of the global body posture. This may be the case for standing and sitting as well, but the Astroskin accelerometers used in the experiment did not give enough information to distinguish between these latter body postures throughout the day. It could explain why training on the complete protocol (Train2) performed slightly better than using only the sitting protocol (Train3). It was only slightly better since the amount of data of people standing and not walking during the day was probably negligible. Nevertheless, it would be interesting to compare the performance on both sitting and standing BPs when using one model trained on the full protocol versus 2 models trained on sitting and standing protocols, respectively.

The data when walking was removed because the waveform was highly variable. This variability was probably partially due to the acceleration of the blood in the arm and finger. Therefore, it was hard to judge if the reconstructed brachial BP was still accurate. Regardless, the model could not predict accurately the BP while walking. More details are available in Sections 4.2.2 and 4.2.3.

A large range of BPs was tested for each individual (approximately 40 mmHg for the DBP and 70 for the SBP). A dependency between the accuracy of the model and the magnitude of the BP was observed (Figure 4.9). The advantage of using a protocol that broadens the BP range when collecting training data is clearly observed on the high extremum values of the SBP estimates, where the Train2 protocol has the lowest error. On the lower range of BPs, the model has a large estimation error, but the model has good performance on the higher range. This is aligned with the developed training protocol, as it was designed mostly to increase the BP. Though the participants were told to relax during the last 5 min of the protocol, decreasing BP is not as easy as increasing it. Relaxing while standing was also not part of the training protocol, which can explain the large error of Train2 protocol at low values of SBP. Again, for hypertension detection, the lower range of BP is not as important as the higher range, as long as a representative low-range BP can be estimated.

Large variations of the coefficient of determination between estimated and measured BP have been reported (R^2=0.194-0.846), which highly depends on the modelling approach [8]. Herein, the R^2 values for the SBP (R^2=0.69) and DBP (R^2=0.60) indicate a high correlation between the NARX model predictions and the measurements (Figure 4.10), in comparison with the literature. It has to be noted that, herein, the measured BP comes from the Portapres device, which has its own bias and precision. Therefore, the model error is possibly lower than the reported one as the results are compared with a BP measurement that includes its own error.

It was shown that the accuracy of the DBP estimation meets the AAMI requirements. By comparing with the AAMI, we find that the error of the NARX model is sufficiently accurate for the DBP

estimation to be classified as a medical device, but the SBP prediction needs further improvement. However, it must be noted that the AAMI standard was designed for intermittent blood pressure monitoring and may not be applicable for continuous monitoring. Nevertheless, it is suggested in *The Handbook of Cuffless Blood Pressure Monitoring* that a preliminary study resulting in a MAE lower than 7 mmHg on the SBP and DBP estimates should move forward [167]. Herein, 6.3 and 5.2 mmHg were achieved for SBP and DBP, respectively.

4.1.4 Conclusion

It was shown that a NARX model using an ANN can predict the complete BP waveform using ECG and PPG as inputs during activities of daily living. The accuracy of the model on the estimated DBP meets the AAMI requirement, whereas the SBP has an error standard deviation that is slightly too large. The results on the 3 participants of this study indicate the potential of using a protocol that changes the BP over a large range during the collection of training data for the NARX model for accurately predicting BP during activities of daily living. The next section improves on this one by repeating the protocol at the end of the day to increase the performance of the NARX model and further validation on additional subjects. Moreover, an accelerometer on the leg is be added to classify the position of the participants.

4.2 The model inputs

In this section, the protocol designed in Section 4.1 is modified to improve the richness of the BP data used for training the NARX model. Different preprocessing steps on the model's inputs (ECG and PPG) are explored. The BP estimates from the user-specific NARX models are compared with PAT models. The main contributions of this section are two-fold; first, the combination of a training procedure with a NARX architecture realizes accurate continuous BP estimation during activities of daily living, regardless of the user's position (sitting or standing); and second, the trained model remains accurate over months if one-point-calibration is used for SBP and DBP. The accuracy of the BP estimates is demonstrated over a broad range of BPs for each user.

4.2.1 Method

4.2.1.1 Measurements

Data were collected on five young healthy individuals (four males, age 28 ± 6.6 yrs) of varied fitness levels, ranging from sedentary to regularly active, and free of cardiovascular and peripheral vascular

disease. All participants signed informed consent forms before participating in the study. Study protocols and procedures were approved by the University of Waterloo (ORE#41490), Clinical Research Ethics Committee and conformed with the Declaration of Helsinki.

Arterial blood pressure, see Figure 4.3, was continuously measured at 100 Hz using finger PPG and the Astroskin wearable body metrics vest was used for ECG, PPG, and tri-axial acceleration measurements (see Section 4.1.1.1), see Figure 4.3 for sample signals. In this study, an iPhone SE was added and fixed to the thigh to record the orientation at 64 Hz to identify the body posture. The sensor placement can be seen in Figure 4.1.

The data from the three measurement devices (Portapres, Astroskin and iPhone) were store in their respective internal memory for offline processing.

Table 4.4 – Duration of each activity in the 15-minute training procedure

Segment	Duration (min)	Activity
1	1	Sit-to-stand and stand-to-sit every 10 s
2	2	Stand
3	0.25	Standing Valsalva
4	0.75	Stand
5	1	Standing handgrip
6	1	Stand
7	1	Walk
8	1	Sit
9	0.25	Seated Valsalva
10	0.75	Sit
11	1	Seated handgrip
12	5	Sit

4.2.1.2 Protocol

Testing was performed at the Schlegel-University of Waterloo RIA (see Section 4.1.1.2). Testing sessions consisted of six consecutive 50-minute windows and a final 30-minute window of continuous BP monitoring. These were separated by 10-minute intervals when BP data were not being collected (the Portapres was turned off) to relieve pressure on the subject's finger. A 15-minute procedure was performed at the beginning of the data collection (in the first 50-minute window) and again six hours later, at the end of the final 30-minute testing window; the two procedures were used for model training.

The training procedure was designed to change the participant's BP within a range usually experienced during activities of daily living. The sequence of activities included in the procedure is shown in Table 4.4. During the procedures, the participants were asked not to talk, and the experimenter guided them using a chronometer. For the Valsalva maneuvers, the participants were asked to push as hard as they could for 15 seconds (though the participants were told to release if they felt dizzy; none did). The static handgrip exercises had the participants squeezing a rolled towel at maximum voluntary contraction for one minute.

During the six hours of testing the participants were free to do as they pleased, excluding lying down and vigorous exercise. There were no structured activities, though participants were encouraged to walk at least once during each 50-minute data collection window, which included going up and down the stairs to ensure a wide range of BP in the testing data.

The protocols for the five participants started between 7:00am and 9:00am, and finished between 3:00pm and 5:00pm. The data collection steps were as follows: i) The participant read and signed the consent form. ii) The steps in the training procedure were thoroughly explained. iii) Thoracic and abdominal circumferences were measured, and the appropriate Astroskin vest size were fitted to the participant. Gel was applied between the skin and each ECG electrode, and elastic bands were fitted at the thoracic and abdominal level. iv) The Portapres was fitted to the participant's non-dominant hand with an appropriate cuff size. v) The iPhone SE was fixed to the participant's thigh. vi) Data collection initiated on all the equipment and the signals were individually validated (signals were synchronized offline). vii) The chronometer was started, and the six and a half hours of data collection protocol began.

4.2.1.3 Model Architecture

A NARX architecture based on a multilayer Perceptron detailed in Section 4.1.1.3. The NARX model used exogenous inputs and the BP at time-steps k-1 and k-2 to estimate BP at time point k. Each exogenous input consisted of 0 to 18 time-step delays (19 samples) of the signals sampled at 64 Hz. In this study, the NARX input signals are the PPG_d, ECG, and their z-normalized counterparts across the 19 time-delayed samples, denoted by PPG_{norm}, and ECG_{norm}. Three different exogenous input combinations for the ANN were explored: i) ECG and PPG (NARX); ii) ECG and PPG_{norm} (NARX – PPG_{norm}); and iii) ECG_{norm} and PPG_{norm} (NARX – $ECG_{norm}PPG_{norm}$).

Estimating BP from the NARX model is a two-step process comprising training and testing. For training, a feedforward ANN was trained for each participant on the three sets of exogenous inputs and the BP measurements ((4.1).

$$\widehat{BP}(k) = f(BP(k\text{-}1), BP(k\text{-}2), Input(k), \ldots, Input(k\text{-}18)) \quad (4.1)$$

where $\widehat{BP}$ is the estimated BP.

For testing, the trained ANN was used to estimate the BP on the six-hour test data, wherein a feedback loop with delay using $\widehat{BP}$ from the previous two time steps was implemented to provide the ANN with the required BPs ((4.2).

$$\widehat{BP}(k) = f(\widehat{BP}(k\text{-}1), \widehat{BP}(k\text{-}2), Input(k), \ldots, Input(k\text{-}18)) \quad (4.2)$$

4.2.1.4 Training

The delayed exogenous inputs and the delayed BP measurements were used as inputs for training a user-specific ANN. The training was performed on the two 15-minute procedures at the beginning and at the end of the six-hour data collection. Zero mean white Gaussian noise with a standard deviation (SD) of 1 mmHg was added to the BP time series to enrich the training data and improve the model robustness to input noise. Error backpropagation was used with the Levenberg-Marquardt optimization algorithm to minimize the norm of the difference between the estimated ($\widehat{BP}$) and measured values of the BP. Only 70% of the training data were used for tuning the ANN weights, while 15% were used for internal validation (to check the training stop condition) and 15% for testing estimation performance during the procedures in a one-step-ahead prediction. The mean square error of this internal test data was used as a threshold for retraining; if it was 20% higher than the mean square error on the training or the validation data, the ANN was retrained.

4.2.1.5 Validation

The six hours of data between the two training procedures were used for testing the estimation accuracy of the models. The Portapres BP waveform was distorted due to walking resulting in a large variability from beat to beat (Figure 4.11). The variability might have been caused in part by acceleration of blood in the arm and finger, and by the variations in peripheral vascular resistance with leg muscle contraction and relaxation. Therefore, from the six hours of test data, all the data collected while walking were removed for the initial comparison of the different models. The test datasets were then divided into three subsets: i) sitting data only, ii) standing data only, and iii) walking data only, to observe any differences in accuracy for those three activities of daily living. The walking segments were detected via the Astroskin vertical accelerometer that was virtually aligned with subjects' torso, using the gravity vector measurement during upright standing. The sitting and standing data were detected using the orientation of the phone.

4.2.1.6 Pulse Arrival Time Model

User-specific PAT models were implemented to facilitate direct comparison with an established cuffless BP monitoring method. According to a recent study [115], PAT models wherein BP is assumed to be linear with respect to PAT are the most commonly used due to their robustness, though logarithmic models are also common [111]. In [115], the most accurate model was found to be a quadratic inverse dependency between BP and PAT with a linear dependency on heart rate (HR). In order to estimate SBP and DBP we thus implemented three PAT models, linear PAT (PAT_{Lin}) [113]:

$$SBP = a_1 PAT + b_1 \quad (4.3.a)$$

$$DBP = a_2 PAT + b_2 \quad (4.3.b)$$

logarithmic PAT (PAT_{Log}), computed as [111]:

$$SBP = a_3 \ln(PAT) + b_3 \quad (4.4.a)$$

$$DBP = a_4 \ln(PAT) + b_4 \quad (4.4.b)$$

And inverse quadratic PAT with linear HR term (PAT_{InvHR}), computed as [115]:

$$SBP = a_5/PAT^2 + b_5 HR + c_1 \quad (4.5.a)$$

$$DBP = a_6/PAT^2 + b_6 HR + c_2 \quad (4.5.b)$$

where the user-specific coefficients a_n, b_n, and $c_n$were determined using a least squares estimation for each training dataset [113]. Two PAT calculation algorithms were used; the first computed as the time difference between the ECG peak and the PPG peak at 64 Hz (PAT – 64 Hz); the second was computed using the Astroskin's embedded algorithm, in which PAT is defined as the time difference between the ECG peak and the PPG 50% amplitude (50% of foot to peak), see Figure 4.3. The latter was post-processed at 256 Hz, and thus had improved temporal resolution (PAT – Astroskin 256 Hz).

4.2.1.7 Mean Training Data Model

User-specific mean training data (MTD) models were implemented for baseline comparison. In this case, SBP and DBP estimations were equal to their respective means in the training data.

4.2.1.8 Data Analysis

For each model configuration (3 NARX, 4 PAT, 1 MTD model), the mean error (μ_{Err}), the error SD (σ_{Err}), the mean absolute error (MAE), the Pearson correlation coefficient (r), and the estimated BP SD error (Err_σ) were computed for the test dataset (six hours) for each participant. The Pearson correlation coefficient (r) between BP and the three PAT relationship (PAT, ln(PAT), $1/PAT^2$) were

computed. The NARX model and the PAT model with the lowest MAE were then investigated more in-depth and compared with the MTD model. To visualize how the models performed in estimating changes in BP (ΔP), the mean absolute value of ΔP_{model} - $\Delta P_{portapres}$ was used ($MAE_{\Delta P}$), where ΔP is defined as the BP deviation from the mean. The errors were separated into bins representing the distribution of the BP data in the test dataset over the day. Each bin covers a range of 5 mmHg, and the bins indicate deviations from mean BP of the participant by 5*n* to 5(*n*+1) mmHg, where *n* ranged from -5 to 5 for SBP and -4 to 4 for DBP. The MAEs were then compared for each bin using Wilcoxon rank sum tests (α=0.05) with Bonferroni corrections to identify any statistical differences between the models. The MAEs for the sitting data were compared with the MAEs obtained on standing and walking data. The MAE was also normalized by the SD of the BP of the respective user-specific dataset. The normalized MAE (nMAE) was used to provide a fair comparison between the datasets, as it represents the relative MAE with respect to the change in BP, which varies considerably across datasets. The MAE for sitting, standing and walking data were also binned according to the method described above for more in-depth comparisons.

4.2.1.9 Long-term Validity

User-specific cuffless BP estimation is usually more accurate than uncalibrated models, but fails to maintain long term validity [6]. It is crucial for long term use to have an indication of when retraining or recalibration is needed. Even if the NARX model is trained on compact, but rich, data (see Sec. II.B), we are interested in evaluating its long-term validity using one-point calibration; that is, to determine if the model can be periodically calibrated and remain accurate on data acquired months apart.

Data were further collected four months apart for two of the participants and six months apart for one of them. The data collection protocol for these tests was identical to the one described in II.B, except: i) there was no monitoring of the leg position, and ii) there were three one-hour windows of BP monitoring separated by one hour.

BP was estimated for these datasets using the models for the participants described in II.D. The estimated SBP and DBP were then one-point-calibrated using the mean of the last minute of the sitting data at the end of the first procedure. The estimated DBP and SBP were offset such that their estimated mean values matched their respective mean values of the Portapres BP measurements during the last minute of sitting data. The same calibration method was applied to the PAT model. The results were compared using the MAE, the nMAE and the μ_{Err} over different days for the NARX and the PAT models. The distribution of $MAE_{\Delta P}$ was also compared.

4.2.2 Results

Representative samples of estimated BP waveforms using NARX – PPG_{norm} are benchmarked against the measured BP in Figure 4.11 for the sitting, standing and walking cases for one subject. Figure 4.11.a,b demonstrate agreement between the measured and estimated waveforms for sitting and standing, respectively. In Figure 4.11.c, the measured BP waveform looks distorted due to walking, which was observed in every participant's data. Therefore, it was decided to remove the walking data from the analysis (otherwise mentioned). The computed MAE across the five subjects for the complete BP waveform are 6.18±0.73 mmHg and 5.85±0.73 mmHg with and without the walking data, respectively, showing the model's good agreement in estimating the complete BP waveform.

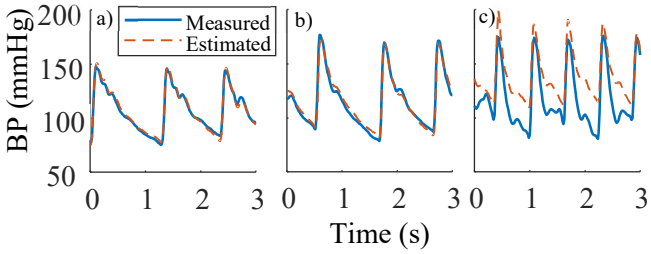

Figure 4.11 - Comparison of the actual versus estimated BP waveforms from NARX - PPG_{norm} for one representative participant during a) sitting, b) standing, and c) walking.

The estimated and measured BPs and their filtered signals (61 datapoints centered moving average) during the six-hour test data are shown for one participant in Figure 4.12. In this case, with walking data included, the BP estimation has an error of -2.96±8.42 mmHg for the SBP and -1.99±6.70 mmHg for the DBP, resulting in MAEs of 6.79 and 5.31 mmHg, respectively. When removing the walking data, the error is -3.96±6.98 mmHg for the SBP and -2.80±5.55 mmHg for the DBP, resulting in MAEs of 6.35 and 4.90 mmHg, respectively.

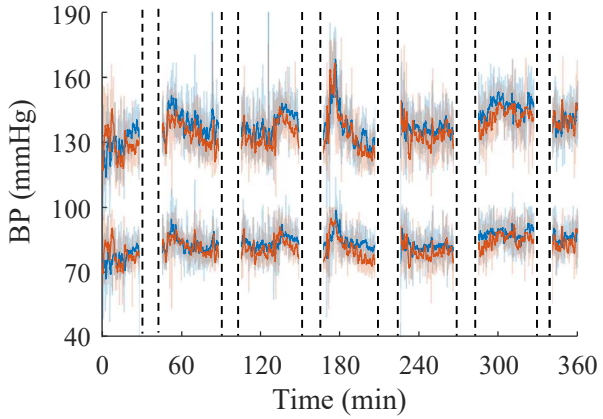

Figure 4.12 - Measured (blue) and estimated (orange; NARX - PPG_{norm}) SBP (top) and DBP (bottom) for the six-hour validation window. The original signals appear transparent and the filtered signals are solid lines. Gaps in data are when the Portapres was off.

4.2.2.1 Model Comparison

The computed Pearson correlation coefficient (*r*) across the five subjects between the PAT relationship (PAT, ln(PAT), and $1/PAT^2$) and BP are shown in Table 4.5.

Table 4.5 – Pearson correlation coefficient (*r*) across the five subjects between the PAT relationship and BP

Feature	SBP	DBP
PAT – Astroskin 256 Hz	-0.35±0.07	-0.12±0.07
ln(PAT) – Astroskin 256 Hz	-0.35±0.07	-0.12±0.07
$1/PAT^2$ – Astroskin 256 Hz	0.34±0.08	0.12±0.07
PAT – 64 Hz	0.01±0.08	0.12±0.11
ln(PAT) – 64 Hz	0.00±0.08	0.12±0.11
$1/PAT^2$ – 64 Hz	0.02±0.08	-0.11±0.11

The expected value and SD of the MAE computed across the five subjects for SBP and DBP are shown in Figure 4.13.a. Included in the plot are data points presenting the respective MAE for each subject. The results exhibit differences in mean MAE between the models, ranging from 6.21 mmHg for NARX – PPG_{norm} (best) to 7.58 for the MTD model (worst) for SBP, and from 5.03 mmHg for PAT_{Log} – 64 Hz (best) to 6.16 for NARX – $ECG_{norm}PPG_{norm}$ (worst) for DBP. The IEEE cuffless wearable standard limit [167] for preliminary studies is also shown in the figure. All models meet the standard except for the MTD model for SBP.

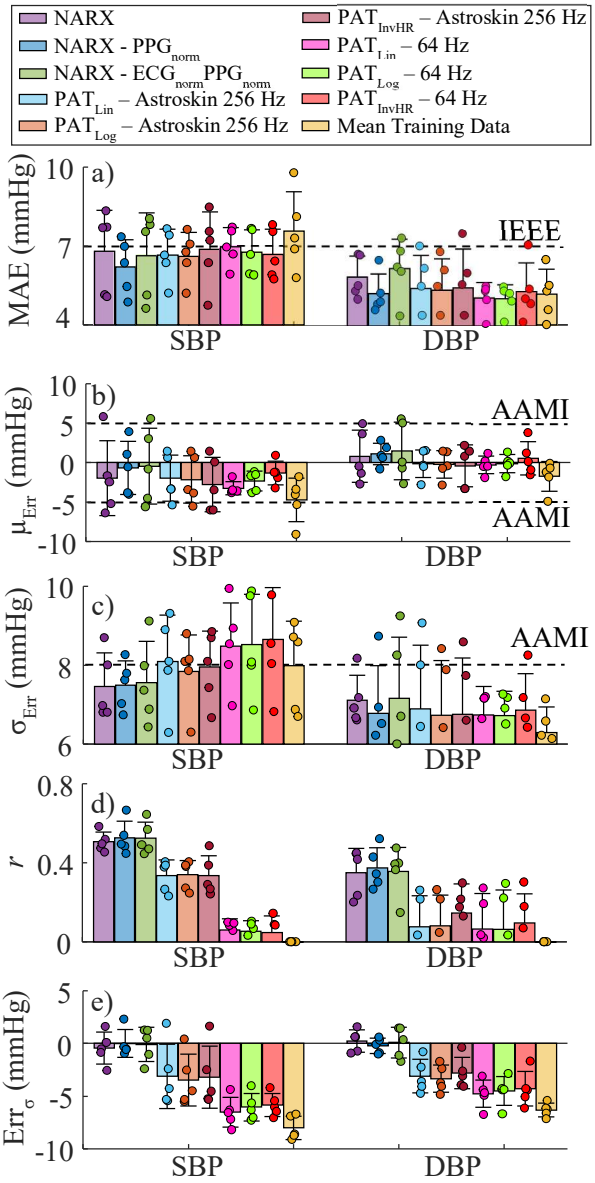

Figure 4.13 - Comparison of a) mean absolute error, b) mean error, c) SD of the error, d) Pearson correlation coefficient between different model estimates and the BP measurements, and e) the error between the daily SD of the measured BP and the estimated BP for SBP (left) and DBP (right). Bars represent the mean of all subjects and the error bars show ± SD. Each data point represents one participant. The IEEE cuffless wearable and the AAMI standard limits are also shown where applicable.

The mean error, μ_{Err}, (expected value and SD) computed for the five subjects is presented in Figure 4.13.b for both SBP and DBP. The horizontal dashed lines in Figure 4.13.b-c indicates the AAMI

requirement limit [153]. All models meet the AAMI requirement of less than 5 mmHg bias for SBP and DBP.

The SD of the error, σ_{Err}, of each model is shown in Figure 4.13.c. Only the expected SDs of the NARX models, PAT_{Log}, and PAT_{InvHR} – Astroskin 256 Hz meet the AAMI SD requirement of 8 mmHg for SBP and all models meet the standard for DBP.

Figure 4.13.d presents the average Pearson correlation coefficient, *r*, between the estimated and measured SBP and DBP for all five participants. The correlation coefficients obtained with the NARX models are higher than the other models.

Figure 4.13.e shows the difference between the actual BP SD and the estimated BP SD (Err_{σ}). The three NARX models exhibit the lowest Err_{σ}.

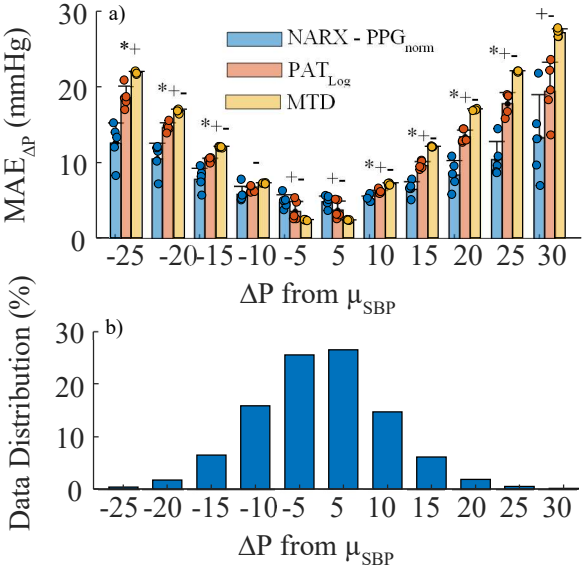

Figure 4.14 - a) Mean and SD of the $MAE_{\Delta P}$ for all participants, and b) the distribution of the aggregated datasets, against the distribution of the ΔP from μSBP measured throughout the day. In a) bars represent the mean of all subjects and the error bars show ± SD. Each data point represents one participant. Legend: (*) NARX vs PAT, (+) NARX vs MTD, and (–) PAT vs MTD ($p<0.05/3$)

The $MAE_{\Delta P}$ as a function of BP deviation from the mean measured during the experiments is shown in Figure 4.14.a for SBP and Figure 4.15.a. for DBP; the data distributions when aggregating the data of all participants are included in Figure 4.14.b and Figure 4.15.b for SBP and DBP, respectively. The $MAE_{\Delta P}$ are binned in increments of 5 mmHg from -25 mmHg to 30 mmHg for SBP and -20 mmHg to 25 mmHg for DBP. The range of BP shown is the minimum range spanned by all the participants. The

NARX – PPG_{norm} MAE is significantly lower than PAT_{Log} – Astroskin 256 Hz MAE for large positive and negative ΔP for both SBP and DBP. PAT_{Log} and NARX – PPG_{norm} are significantly different from MTD model in every bin except for the NARX [-10,-5] bin (Wilcoxon $p<0.05/3$). The bin widths were selected for clarity of presentation. There is, however, sufficient data density to improve the resolution. For example, for a bin width of 1 mmHg, the MAE for NARX – PPG_{norm} is significantly lower than PAT_{Log} and MTD from 7 mmHg to 24 mmHg, and -10 mmHg to -21 mmHg for SBP and from 7 mmHg to 18 mmHg, and -6 mmHg to -12 mmHg for DBP (Wilcoxon $p<0.05$). This represents 24% and 23% of the data for SBP and DBP, respectively.

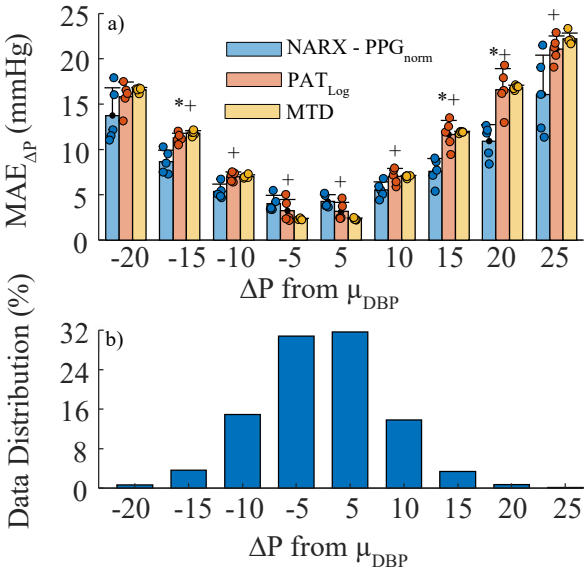

Figure 4.15 - a) Mean and SD of the $MAE_{\Delta P}$ for all participants, and b) the distribution of the aggregated datasets, against the distribution of the ΔP from μDBP measured throughout the day. In a), bars represent the mean of all subjects and the error bars show ± SD. Each data point represents one participant. Legend: (*) NARX vs PAT, (+) NARX vs MTD, and (–) PAT vs MTD ($p<0.05/3$)

The MAEs obtained using NARX - PPG_{norm} and PAT_{Log} – Astroskin 256 Hz were compared for sitting, standing, and walking data in Figure 4.16.a, with SBP to the left and DBP on the right. While the error in BP estimation is slightly larger during standing, when normalized by the SD of the BP (nMAE), the error is lower for the standing data (Figure 4.16.b). This suggests larger variance of BP during standing postures. Figure 4.16 also shows that the NARX – PPG_{norm} trends towards outperforming the PAT_{Log} – Astroskin 256 Hz in term of nMAE for SBP (Wilcoxon $p=0.095$). The

MAE on the walking data is large in comparison with standing and sitting data. However, this is partly caused by the large variation in BP, which is evident in the nMAE, which is similar for walking, standing, and sitting, for SBP, though not for DBP.

The $MAE_{\Delta P}$ as a function of the ΔP deviation from the mean SBP measured during the experiments is shown in Figure 4.17 for the sitting, standing, and walking data for NARX - PPG_{norm}. The bins span the intersection of the ranges of both datasets. The results are similar for sitting and standing; however, differences can be seen for the [-25,-20] bin (Wilcoxon p=0.095). There is also a significant difference for the [5,10] and the [10,15] bins between sitting and walking data (Wilcoxon p<.05/3).

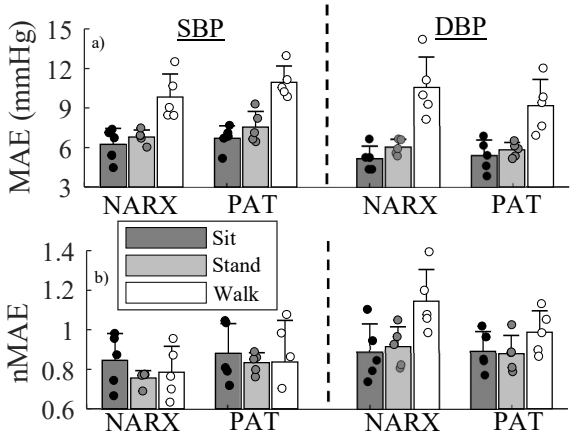

Figure 4.16 - Comparison of NARX - PPG_{norm} and PAT_{Log} – Astroskin 256 Hz models during sitting, standing, and walking for SBP (left), and DBP (right) using a) mean absolute error, and b) the normalized mean absolute error. Bars represent the mean of all subjects and the error bars show ± SD. Each data point represents one participant.

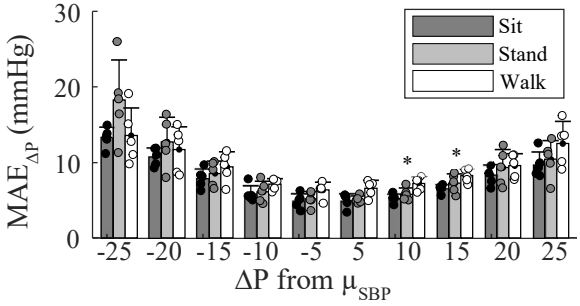

Figure 4.17 - Mean and SD of the $MAE_{\Delta P}$ for the five participants for NARX - PPG_{norm}, plotted with respect to the distribution of the ΔP from μ_{SBP} measured throughout the day. Bars represent the mean of all subjects and the error bars show ± SD. Each data point represents one participant. Legend: (*) Sit vs Walk (p<0.05/3)

4.2.2.2 Long-Term Performance

Figure 4.18 shows the SBP MAE, nMAE, and the mean error (μ_{Err}) of the NARX – PPG_{norm} and PAT_{Log} – Astroskin 256 Hz over the span of months for two participants. Once the NARX model is one-point calibrated, the MAE is under 7.1 mmHg for all cases, and always lower than the PAT model. The nMAE is lower for 4 months and 6 months, showing more variance in the SBP data during those days. The MAE is largely a function of model's bias, which is quite arbitrary, as seen in Figure 4.18.c, since it depends on the initial bias of the uncalibrated model. Once the bias of the model is removed, using the one-point calibration, the NARX - PPG_{norm} performs similarly over the range of ΔP for different days for both participants, see Figure 4.19.

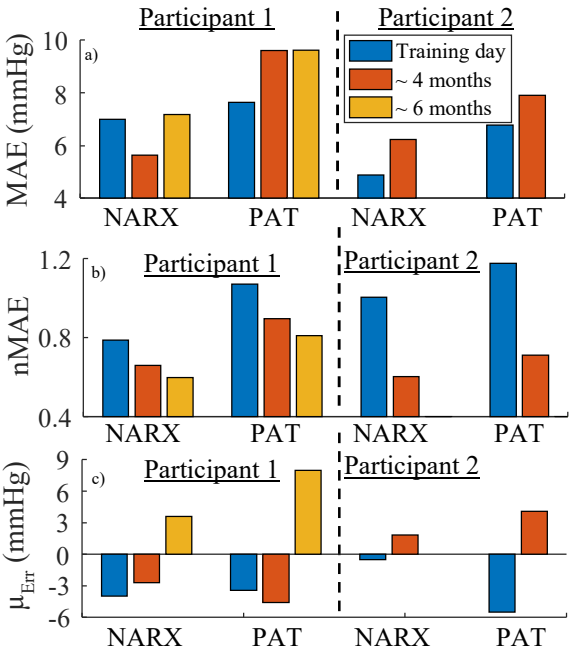

Figure 4.18 - Comparison of a) mean absolute error, b) normalized mean absolute error, and c) mean error for different days. The results are comparing NARX - PPG_{norm} and PAT_{Log} – Astroskin 256 Hz for SBP. The model was trained on Training day, and calibrated at the beginning of the Testing day "~ 4 months" and "~ 6 months".

4.2.3 Discussion

This section presents an effective approach to continuously estimate BP over a wide range of variability for each individual. To the best of our knowledge, this is the first demonstration of tracking beat-to-beat BP during activities of daily living. Our proposed approach uses a NARX architecture trained on the data from a deliberately designed procedure. The procedure induces BP variability in the

training data and enables accurate BP estimation during daily activity. The proposed method demonstrated a clear advantage over a PAT model on estimating a large range of BPs with low error. The NARX model requires no complex signal preprocessing, other than the normalization of the input, unlike in feature extraction-based procedures, where accurate and reliable physiological feature extraction can be difficult due to widely varying signal morphologies between individuals [22]. Training of the ANN used by the NARX model is quite fast due to its shallow architecture [166].

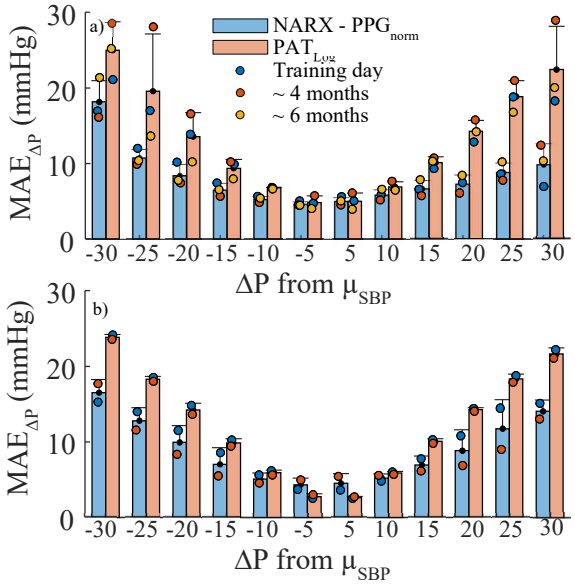

Figure 4.19 - Mean and SD of the $MAE_{\Delta P}$ for NARX - PPG_{norm} and PAT_{Log} – Astroskin 256 Hz for different days for a) participant 1, and b) participant 2. The results are plotted with respect to the distribution of the ΔP from μSBP measured throughout the day. Bars represent the mean of all days and the error bars show ± SD. Each data point represents one day.

Though not emphasized here, this architecture estimates the complete BP waveform (see Figure 4.11) with an expected MAE of 5.85±0.73 mmHg, which is lower than the SBP MAE and higher than the DBP MAE, similar to [166]. The BP waveform contains additional information of potential interest to physicians [17]. Specifically in the case of hypertension, the waveform provides information on the underlying disease and mechanisms leading to high blood pressure [18]. For example, pulse diagnosis utilizes the blood pressure waveform to identify different cardiovascular diseases and follow pathological changes [19]. Moreover, the complete BP waveform could give access to other important cardiovascular health measures, such as cardiac output [20] and stroke volume [21].

This study considered important parameters in developing a user-specific continuous cuffless BP measurement method: i) the range of BP experienced by a user, ii) the time span of estimation, iii) the applicability of the calibration procedure (training) in a real world application, and iv) treating each participant's data separately when analyzing the results. The range of BP covered in this study represents a realistic range, since it was acquired during the participants' activities of daily living. Although a large range of BP was covered during testing (at least ΔP of 55 mmHg for SBP and 45 mmHg for DBP), the BP SD was relatively low (mean of 8 mmHg and 6.3 mmHg for SBP and DBP, respectively), which is not surprising since the participants were healthy individuals [168], most of the time sitting at a desk. Some models have been previously developed to estimate large ranges of BP using a cold pressor test [25], reporting expected μ_{Err} of 0.04 and 0.01 mmHg for SBP and DBP, respectively, and expected σ_{Err} of 5.0 and 3.7 mmHg for SBP and DBP, respectively. However, they were only tested on a short time span (13 minutes), with 10-fold cross-validation (no test data). Herein, each participants' BP estimation was tested on six hours of data. Other studies have shown the potential of estimating BP for a long time span (months) [148], reporting expected MAEs for SBP of 6.56, 6.66, 6.70, 6.90 mmHg for four different models (compared with 6.21 mmHg in the present study), but the models were trained on a complete day of data collection. Herein, six hours of BP estimation, with the potential of extending to months using one-point calibration, is realized by employing only 30 minutes of training data, acquired during a designed procedure, making our method a more practical solution. In contrast to PAT models, the NARX model cannot be trained with only a few heartbeats. Thirty minutes of training data seem to be a practical compromise between the required length of training samples and acceptable estimation accuracy. For practical reasons, we employed the integrated forehead PPG of the Astroskin vest. However, previous work has demonstrated that the NARX model works well with a finger PPG [166], therefore, it is likely that the model accuracy will not be affected much by PPG placement. We acknowledge the small sample size of our study; nevertheless, the $MAE_{\Delta P}$ of the NARX architecture showed to be significantly less than the $MAE_{\Delta P}$ of the PAT approach (Figure 4.14 and Figure 4.15).

It was chosen not to include lying down in our definition of activities of daily living, since the model was previously shown to be accurate in the supine position using bed ridden patients [166]. Therefore, we focused on other activities to minimize the training requirement of the NARX model. Walking data was analyzed separately from the rest of the activities due to its highly variable waveform (Figure 4.11.c). This variability was partially due to the acceleration of the blood in the arm and finger. Distorted BP waveforms caused by body motion have been observed in the literature [49]. Those

waveforms are much more complex due to the interaction between the heart's pressure wave and the wave generated by the whole body vertical movement. A preliminary attempt to include vertical acceleration as input to the ANN (19 time-steps delay) showed promising results in terms of estimating the distorted waveform during walking, but resulted in larger error when estimating sitting and standing BP; furthermore, sometimes the model became unstable when the acceleration data were too different from the training data. The training data only covered two times one minute of walking data, which was likely not sufficient to cover the whole spectrum of walking frequencies and phases between acceleration and BP observed during the activities of daily living. Future work will include training on more walking data to account for the movement related variations of BP, and potentially employing separate NARX models for sitting/standing and walking.

The NARX model with the ECG and PPG_{norm} inputs yielded the smallest MAE (Figure 4.13.a), indicating the relevance of relative PPG signal to the BP estimation, similar to normalized PPG waveforms in [25]. While the non-normalized PPG inputs should lead to the same results, due to the similar information content, they are more sensitive to the sensor placement and can increase the estimation error. This hypothesis is solidified further by the error from the NARX (non-normalized) model on a different day (not shown herein), which was large as a result of different sensor placement (difference in peak-to-peak PPG amplitude from one day to another). In contrast, NARX – $ECG_{norm}PPG_{norm}$ performance varied across the participants (Figure 4.13.b-c).

We acknowledge, that the present study met neither the AAMI nor the IEEE standards in terms of the number of subjects and the range of tested BP. However, it should be noted that these standards are not directly applicable to continuous cuffless blood pressure monitors; the AAMI is designed for cuff-based intermittent blood pressure monitors and IEEE does not cover all aspects, such as model's calibration for instance, of cuffless blood pressure measurement [167]. This is further supported by this study, where the MTD model showed that applying the standards are likely inadequate for continuous BP estimation using an individualized model trained the same day (or calibrated). It would be difficult to think of a less predictive model than the MTD model, and still, it passes the AAMI standard and nearly meets the IEEE standard. This can be explained by the distribution of BP during the activities of daily living (Figure 4.14.b and Figure 4.15.b), where more than 50% of the daily BP lies between +/- 5 mmHg of the mean BP (SBP and DBP). Pearson correlation coefficient was used to capture the ability of a model to estimate BP changes in Figure 4.13.d. From that figure, it is clear that the NARX models outperform the PAT and MTD models in estimating changes in BP. This is further supported by the estimated BP SD, which is accurately estimated by the NARX models (Figure 4.13.e). Accurate BP

SD estimation is important for a BP monitoring device as it is a predictor of cardiovascular events, stroke, and mortality [168]. The performance of the NARX model is good at low acquisition frequency, which is not true for the PAT models. The Pearson correlation coefficient was practically zero between all PAT relationships sampled at 64 Hz and SBP, whereas it increases to about 0.35 when sampled at 256 Hz (positive or negative depending on the relationship), see Table II. The NARX performance at low acquisition frequency is interesting from a wearables perspective since it would allow reducing the power consumption and memory requirement of the device [169].

The distributions of $MAE_{\Delta P}$ in Figure 4.14 and Figure 4.15 show the capability of the NARX model to estimate BP varying in a large range. While not observed with the MAE, the model overestimates low BP and underestimates high BP values. The significantly lower $MAE_{\Delta P}$, for both high and low BP values, represents 24% and 23% of the data for SBP and DBP, respectively. This is in the case of activities of daily living (office work) where the ΔP distribution SD of the whole dataset was 7.74 mmHg and 6.17 mmHg for SBP and DBP, respectively. These low SDs explain why the NARX benefit over the PAT or MTD models are not remarkable when looking at the aggregated data (Figure 4.13.a-d). For larger BP variations, the difference between models is expected to increase. This suggests that the NARX architecture may be more beneficial for cuffless BP monitoring where large changes in BP are observed, such as in sports wearables [170], but more importantly in a population with high risk of cardiovascular events [168], [171]. For example, patients with a daytime SBP SD greater than 15 mmHg are at increased risk of developing cardiovascular events, independent of their absolute BP levels [171]. This emphasizes the importance of capturing BP changes during activities of daily living as yielded by the NARX model, in contrast to the MTD and PAT models (Figure 4.13.e, Figure 4.14, and Figure 4.15).

The NARX model performed well estimating BP during sitting and standing (Figure 4.16). The MAE was on average larger on the standing data, the nMAE showed that the larger error was due to more variation in BP. Comparing the performances by looking at the distribution of error as a function of the ΔP away from the mean (Figure 4.17), we observed that the NARX model performs similarly on both standing and sitting. For the lowest ΔP bin [-25,-20 mmHg], it appears the standing BP is not as well estimated as in the sitting. This may be due to the small size of the standing dataset, 6.8% of the sitting dataset size. Therefore, the standing MAE on the samples may not represent the MAE of the population due to its small size. For example, the $MAE_{\Delta P}$ in the [-25,-20 mmHg] bin for the five subjects is computed on sample sizes ranging from 3 to 15, whereas the sitting datasets MAEs are made on samples of 16 to 130. The NARX model showed poor MAE in estimating BP during walking (Figure

4.16.a). Raw data inspection suggests that DBP is largely distorted by walking, whereas SBP is less affected (Figure 4.11.c). This is further supported by the walking nMAE of SBP, which has similar values to the sitting and standing data, whereas the DBP nMAE is much larger than its counterparts (Figure 4.16.b). Moreover, the SBP $MAE_{\Delta P}$ is similar to the standing data showing the model capability to capture changes in SBP during walking (Figure 4.17). We thus postulate that the DBP measurements were largely distorted by the body acceleration and could not be estimated without acceleration information, whereas the SBP walking data, less affected by the acceleration, were difficult to estimate due to the large variance in BP generated by the heart's pressure wave resulting from the exercise. Therefore, the large MAE of the NARX model during walking is explained by the large distribution of BP for the SBP and it is explained by a poor model performance for DBP as shown by the nMAE (Figure 4.16.b).

Preliminary results show that the NARX model can be used to estimate BP over a long timescale if periodically calibrated (Figure 4.18). Within the present study, the duration over which the NARX model accurately predicts the BP without recalibration was not explored. However, no degradation of accuracy was observed over the six-hour estimation window, suggesting that the NARX model can accurately estimate a complete day of activities of daily living, extending this to the time scale of months when one-point calibrations are applied. Such calibration could be performed with an arm cuff in a real scenario. The calibration cannot remove the model bias completely as the bias of the uncalibrated model is unknown due to the variability of baseline BP on different days. Nevertheless, one-point-calibration resulted in a smaller bias to the uncalibrated model. This explains why in Figure 4.18.c, the mean error μ_{Err} can be either positive or negative. This also explains why the PAT models result in a different mean error than the NARX models, even though the models are calibrated on the same data (they have different initial biases).

Comparing the $MAE_{\Delta P}$ distribution over ΔP shows that there is no degradation of the model performance over months (Figure 4.19). The distribution of the data points in each bin is narrow, as shown by the small SD of each bin, indicating similar MAE for the different days for participant 1 (Figure 4.19.a), whereas participant 2 model performance was better for the data 4 months apart (Figure 4.19.b).

It needs to be acknowledged that the "ground truth" BP measurements are from the Portapres, which has an inherent bias and precision error [81]. The model calibration across days may be needed to match the Portapres measurements and not the actual BP. Cuff placement variations may bias the measurements differently from one day to another [172], which requires a recalibration. Arterial

pressure estimates obtained via the finger volume clamp method have been reported to artificially drift over an extended period of monitoring due to changes in local factors induced by clamping the finger, such as altered finger blood flow, edema, and/or changes in finger temperature [172]. This can lead to erroneous BP measurements, which is usually compensated by switching fingers for the BP measurement [172]. However, this can also lead to a difference in BP measurement between the two fingers [173]. Before adopting the NARX approach for medical purposes, it should be validated with catheterized participants. Nevertheless, the work presented in this section shows the great potential of this approach in estimating BP with a relatively large variability range and demonstrates its potential for long-term and multi-session monitoring.

4.2.4 Conclusion and future work

This work demonstrated the advantage of using a NARX architecture in estimating a large range of SBP and DBP. BP of five participants was continuously monitored during their activities of daily living and the BP waveform was estimated and validated over six hours for each individual. The NARX model outperformed the established PAT models in estimating both low and high ranges of BP. Preliminary results showed the NARX model can maintain its performance for up to a 6-month horizon. Future steps will include monitoring more participants over months for a thorough analysis of the model performance on short, medium, and long timescales. Moreover, more investigations will be required on BP measurements during walking.

Chapter 5
A Fusion Approach to Improve Accuracy and Estimate Prediction Interval[6]

In this chapter, a one-class support vector machine approach is investigated to estimate the prediction interval (PI) of BP estimates from three different model architectures. The PI is practical confidence metric concerned with the accuracy at which the observed target value itself can be predicted. Therefore, since continuous BP monitoring is likely to be employed outside of the clinic, the PI would be of high clinical value when it comes to outcome evaluation. Next, a fusion algorithm is used to combine BP estimates from those models using their estimated PI. We hypothesize that (1) estimation fusion will increase the BP estimation accuracy and (2) The PI can be used to identify high quality data for clinical uses.

5.1 Method

5.1.1 Measurements

The data from the five healthy participants presented in Chapter 4 was used in this chapter (see Section 4.1.1.1 for Measurements and Section 4.2.1.2 for Protocol).

5.1.2 BP Estimation Models

NARX: The NARX that uses the ECG and a z-normalized PPG as input signals, the details of which, along with network training descriptions, can be found in Section 4.2.1.3. To reduce the estimation variance, five NARX models were bootstrap-ensembled [174], herein referred to as NARX.

ANN: Using the same ANN architecture, models were developed to estimate directly the SBP (ANN_{Sys}) and the DBP (ANN_{Dias}) from ECG and the z-normalized PPG signal (no BP estimates in the input layer). A similar model architecture is employed in [22]. Again, five ANN models were bootstrap-ensembled [174], herein referred to as ANN_{Sys}(ANN_{Dias}).

[6] Parts of this chapter have been adapted from the following submitted article:

C. Landry, S. D. Peterson, A. Arami, "A Fusion Approach to Improve Cuffless Blood Pressure and Prediction Interval Estimation," submitted to *IEEE Journal of Biomedical and Health Informatics* (August).

PAT: According to Chapter 4, the common logarithmic PAT model (PAT_{Log}) [111] was the most accurate in estimating SBP during activities of daily living. The PAT was computed using the Astroskin's embedded algorithm, in which PAT is defined as the time difference between the ECG peak and 50% of the PPG amplitude (50% of foot-to-peak), see Section 4.2.1.6. PAT was post-processed at 256 Hz, resulted in improved temporal resolution compared to extracting PAT from the outputted 64 Hz data.

5.1.3 Prediction Interval Estimation

A PI estimation model was trained independently for each individual's SBP and DBP. The model architecture consisted of multiple one-class support vector machines (OCSVMs) [175], trained on the two 15-minute training sets, using Gaussian kernels and a regularization parameter ν=0.5 ($0 < \nu \leq 1$), where a small value of ν leads to fewer support vectors and, therefore, a smooth, crude decision boundary [107]. Each OCSVM was trained on a different percentage of outliers from 0% to 99% in 1% increments. The 100 OCSVM hyperplanes were used to group data into 101 clusters, each spanning 10 hyperplanes (*i.e.*, between the 0% and the 9% hyperplanes, between the 1% and the 10% hyperplanes, etc. Note that from 91% to 99%, the clusters were defined as inside the X% hyperplane, instead of from X% to X+9%). The data outside the 0% hyperplane was defined as the outlier cluster. A visual representation of the clustering method is shown in Figure 5.1 with increments of 20% outliers for an arbitrary 2-dimensional input. For each cluster of training data, the BP estimation (SBP or DBP) error distribution was calculated, and the error standard deviation (σ_{Err}) was associated with the uncertainty of that cluster. Therefore, any new datapoint assigned to a cluster was associated with that cluster's uncertainty, herein referred to as the PI, in the testing phase. Clustering by intervals of 10 hyperplanes was used to increase the resolution of the estimated PI and ensure that σ_{Err} was calculated on a sufficient number of datapoints.

Ten different inputs to the OCSVM were tested by varying the length of the input time series (ECG and PPG) in increments of 0.3s (segment lengths were $\tau = 0.3n$ seconds long, where $n \in \{1,2,3,4,5\}$) with (w/ BP) and without (w/o BP) the estimated BP waveform of length τ. The BP waveform was provided only for NARX. For ANN_{Sys} and PAT_{Log}, the estimated SBP (or DBP) was provided instead.

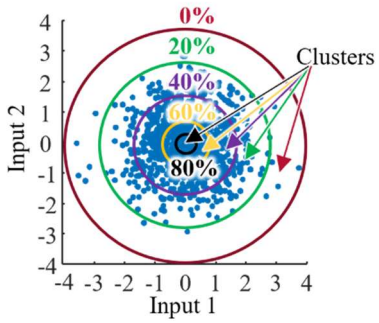

Figure 5.1 - Algorithm schematic featuring arbitrary data. The shown percentages indicate the number of outliers associated with the projections of the one-class support vector machine models' hyperplanes on the two-dimensional input space (circles in this example). The clusters are upper and lower bounded by the hyperplanes (except inside 80% and outside 0%).

5.1.4 Estimation Fusion

The covariance intersection [150] method was used to fuse the BP estimates from the different models. For consistent estimates, $A_i = \{\xi_{a_i}, P_{a_i}\}$, where $i = 1$ to n number of estimates, ξ and P are the mean and covariance of the estimate, respectively, the mean and covariance of the fused estimate C can then be determined as

$$P_c^{-1} = \omega_1 P_{a_1}^{-1} + \cdots + \omega_n P_{a_n}^{-1} \tag{5.1}$$

$$\widehat{\xi_c} = P_c\left(\omega_1 P_{a_1}^{-1}\widehat{\xi_{a_1}} + \cdots + \omega_n P_{a_n}^{-1}\widehat{\xi_{a_n}}\right) \tag{5.2}$$

where $\sum_{i=1}^{n} \omega_i = 1$ and ω_i are determined through an optimization process to minimize P_c.

In this chapter, ξ refers to the estimated BP and P is the PI estimated by the OCSVM model. Covariance intersection was used to fuse NARX and $ANN_{Sys}(ANN_{Dias})$ (NARX + $ANN_{Sys}(ANN_{Dias})$), PAT_{Log} with other models (NARX + PAT_{Log} and $ANN_{Sys}(ANN_{Dias})$ + PAT_{Log}), and for the fusion of all three models (NARX + $ANN_{Sys}(ANN_{Dias})$ + PAT_{Log}).

5.1.5 Validation and Data Analysis

The six hours of data between the two training procedures were used for testing the PI estimates of the models. For the NARX model, the estimated BP waveform was supplied to the OCSVM for training. For $ANN_{Sys}(ANN_{Dias})$ and PAT_{Log} models, only the estimated SBP/DBP was fed to the OCSVM.

For each of the 10 PI model configurations, the MeRCI score [176], defined as:

$$\text{MeRCI}^{\alpha} = \frac{1}{N}\sum_{i=1}^{N} \lambda^{\alpha}\sigma_i \quad (5.3)$$

was computed for the test dataset for each participant, where N is the number of SBP(DBP) datapoints, σ is the BP estimated PI from the OCSVM model, and λ^{α} is obtained by first evaluating all ratios $\lambda_i = |\hat{y}_i - y_i|/\sigma_i$ and then extracting the α^{th} percentile, in this case, 99.7%, of the λ_i distribution. The MeRCI score represents a mean scaled PI that contains α% of the data; thus, a smaller MeRCI score represents a better PI estimation. The MeRCI scores were calculated only for the BP estimates associated with a PI; *i.e.*, if the BP estimate is classified outside the 0% hyperplane, the estimate is discarded.

The PI model configurations with the lowest MeRCI scores for SBP(DBP) were further investigated for each of the BP models. The Pearson correlation coefficients (r) were calculated for each participant to examine the linear relationship between the σ_{Err} of each cluster in the training and of the test data. The expected σ_{Err} ± standard deviation (SD) in each cluster was also plotted against the percentage of outliers in the cluster.

The performance of the three BP models and the four fused models were analyzed using the mean error (μ_{Err}), σ_{Err}, mean absolute error (MAE), Pearson correlation coefficient, MeRCI score, and mean PI computed for the six-hour test data for each participant. To visualize the models' performance in estimating changes in BP (ΔP), the mean absolute value of $\Delta P_{\text{model}} - \Delta P_{\text{portapres}}$ was used ($MAE_{\Delta P}$), where ΔP is defined as the BP deviation from the mean and the subscript indicates whether the data are from the model or the Portapres device. The errors were separated into bins representing the distribution of the BP in the test dataset over the day. Each bin spans a range of 5 mmHg, and the bins indicate deviations from the mean BP of the participant by $5m$ to $5(m+1)$ mmHg, where m ranged from -6 to 6 for SBP and -4 to 4 for DBP. The difference of each bin with analogous bins from PAT_{Log} was computed as a function of |ΔP|. The MAEs were compared for each bin using a 2-way ANOVA test followed by a t-test with Bonferroni corrections to identify statistical differences across the models.

A standard deviation threshold (σ_T) was used to filter the data, such that all BP estimates associated with a PI higher than σ_T were removed from the dataset. For the remaining BP estimates (PI < σ_T), the percentage of data removed as a function of the absolute estimation error was computed for different σ_T. The σ_{Err} and the percentage of data kept were computed for the six-hour test data to observe the tradeoff between the two. To visualize the error in estimating ΔP for the remaining data (PI < σ_T) in comparison with the initial dataset, the $MAE_{\Delta P}$ as a function of |ΔP| was computed for σ_T = 8 and 9 mmHg. To understand better the PI estimation performance, the percentage of data removed in each

|ΔP| bin was computed. To assess the performance of the fused models during activities of daily living, σ_{Err} was computed for different activities (sitting, standing, and walking), and compared with the value when using σ_T = 8 mmHg.

5.2 Results

Herein only the results for SBP are presented for brevity. DBP results are included in the Appendix B.

The MeRCI scores (mean±SD) across the five subjects exhibit differences between the models, ranging from 24.4 for τ=0.9 w/ BP (best) to 30.3 for τ=0.3 w/o BP (worst) for NARX, from 24.9 for τ=0.9 w/ BP (best) to 28.4 for τ=1.5 w/o BP (worst) for ANN_{Sys}, and from 28.0 for τ=0.6 w/ BP (best) to 33.1 for τ=0.3 w/o BP (worst) for PAT_{Log}.

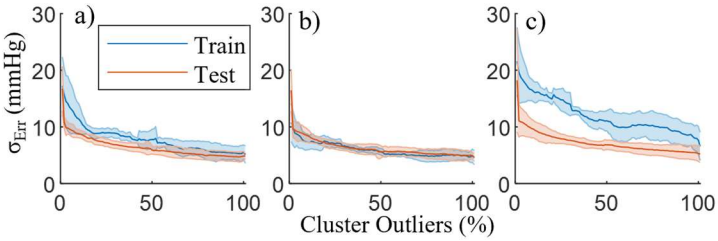

Figure 5.2 - The SBP training and testing data standard deviation of error in every cluster for a) NARX, b) ANN_{Sys}, and c) PAT_{Log}. Lines represent the mean of all subjects, and the error bars show ± SD.

The MeRCI score for PAT_{Log} using the τ=0.9 w/ BP model was 28.5, only slightly higher than its best configuration (τ=0.6 w/ BP); therefore, we use the PI model τ=0.9 w/ BP for the three BP models for simplicity of comparison for the remainder of the analysis. The training and testing clusters' σ_{Err} as a function of the clusters' percentage of outliers (lower bound) are shown in Figure 5.2 for the NARX, ANN_{Sys}, and PAT_{Log} models. Note that estimates of σ_{Err} on the training data are the model output on predicting the testing data σ_{Err} (in blue) compared to its real values (in red). The over/underestimation of σ_{Err} in each cluster reflects the bias of the three different models in predicting the testing data σ_{Err}. The change in σ_{Err} for the training data compared to the test data was 17%, -22% and 82%, for NARX, ANNSys, and PAT_{Log}, respectively. Note that the training σ_{Err} at 0% of outliers is not defined, as there are no training data in this cluster. In the test data, the data points outside the 0% hyperplane are outliers and represent 3.9±2.0%, 3.3±1.3% ,and 4.9±4.2% of the BP estimates for

the NARX, ANN_{Sys}, and PAT_{Log} models, respectively. The Pearson correlation coefficients between the training and testing σ_{Err} of each cluster are r=0.88±0.11, r=0.81±0.09, and r=0.86±0.13 for the NARX, ANN_{Sys}, and PAT_{Log} models, respectively. It can be observed in Figure 5.2 that the model overestimates, on average, the σ_{Err} of the test data by 21% and 68% for NARX and PAT_{Log}, respectively, and underestimates it by 3% for ANN_{Sys}.

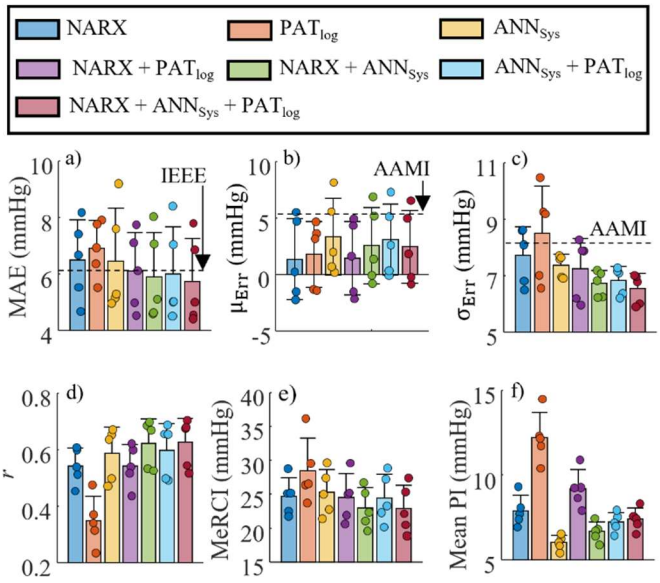

Figure 5.3 - Comparison of a) mean absolute error, b) mean error, c) standard deviation of the error, d) Pearson correlation coefficient between different model estimates and the BP measurements, e) MeRCI score, and f) mean estimated prediction interval. Bars represent the mean of all subjects, and the error bars show ± SD. Each data point represents one participant. The IEEE cuffless wearable and the AAMI standard limits are also shown in subplots a), and b-c), respectively.

The mean and SD of the MAE, μ_{Err}, σ_{Err}, r, MeRCI score, and mean of the PI during activities of daily living computed across the five subjects are shown in Figure 5.3. The results exhibit differences in mean MAE between the models, ranging from 5.74 mmHg for NARX + ANN_{Sys} + PAT_{Log} (best) to 6.91 for PAT_{Log} (worst). The IEEE cuffless wearable standard limit [167], which all models meet, is also shown. The horizontal dashed lines in Figure 5.3.b indicate the AAMI requirement limit for μ_{Err} [153], which is met by all models. All models meet the σ_{Err} AAMI standard except for PAT_{Log}. It is noted that the σ_{Err} is under 8 mmHg for all participants except for NARX, PAT_{Log}, and NARX + PAT_{Log}. The average Pearson correlation coefficient between the estimated and measured SBP for all

five participants obtained with the fusion of at least the two ANN models (NARX and ANN_{Sys}) are higher than the other models. The MeRCI score shows that model fusion enhances PI estimation when compared to the three initial non-fused models. ANN_{Sys} has the lowest mean PI, but a larger MeRCI score than the other fused models, indicating that it underestimates the PI. The difference between $MAE_{\Delta P}$ of each model and the $MAE_{\Delta P}$ of PAT_{Log} (in percentage) as a function of absolute BP deviation from the mean is shown in Figure 5.4.a; the data distributions when aggregating the data of all participants are included in Figure 5.4.b. The range of SBP shown is the minimum range spanned by all participants.

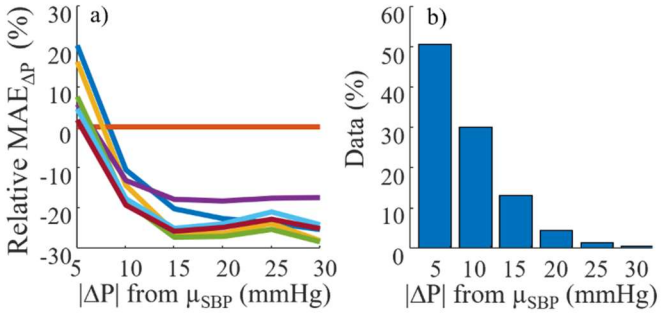

Figure 5.4 - Mean of SBP $MAE_{\Delta P}$ relative to $MAE_{\Delta P}$ – PAT_{Log} for all participants, and b) the distribution of data (all participants aggregated) against the distribution of $|\Delta P|$ from μ_{SBP} measured throughout the day. The $MAE_{\Delta P}$ are binned in increments of 5 mmHg from -30 mmHg to 30 mmHg then grouped according to their absolute value.

The percentage of data removed is plotted against the absolute estimation error in Figure 5.5.a for the NARX + ANN_{Sys} + PAT_{Log} model. The tradeoff between the amount of data retained and the σ_{Err} when varying σ_T is shown in Figure 5.5.b.

$MAE_{\Delta P}$ as a function of $|\Delta P|$ is plotted in Figure 5.6.a for the NARX + ANN_{Sys} + PAT_{Log} model using σ_T = 8 and 9 mmHg. The percentages of data removed in each $|\Delta P|$ bin are plotted in Figure 5.6.b. When using the OCSVM model of PI to remove data with large expected error SD (PI ≥8 or 9 mmHg), all expected $MAE_{\Delta P}$ values are lower than their original values. In every case, the BP estimates by the fused model at large $|\Delta P|$ have higher expected error than BP estimates closer to the mean BP. There were, however, statistical differences when comparing each bin; randomly removing the same amount of data in the dataset did not result in the observed reduction in $MAE_{\Delta P}$.

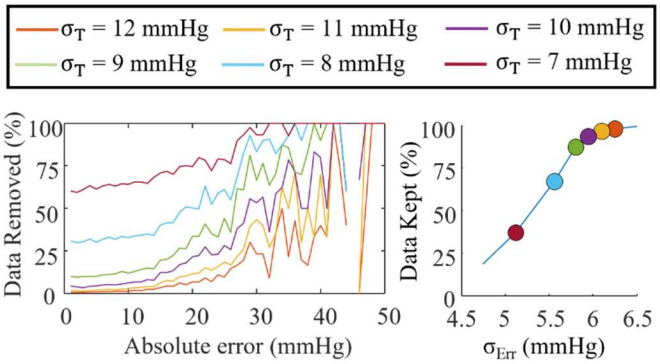

Figure 5.5 - For the NARX + ANN_{Sys} + PAT_{Log} model, a) percentage of the data removed as a function of the absolute error using different σ_T. The results are calculated from all the test data from every participant and grouped in increments of 1 mmHg. b) Percentage of data kept as a function of the error standard deviation when varying the threshold on the prediction interval. The line was computed by varying the σ_T, where specific values are marked with circles according to the legend.

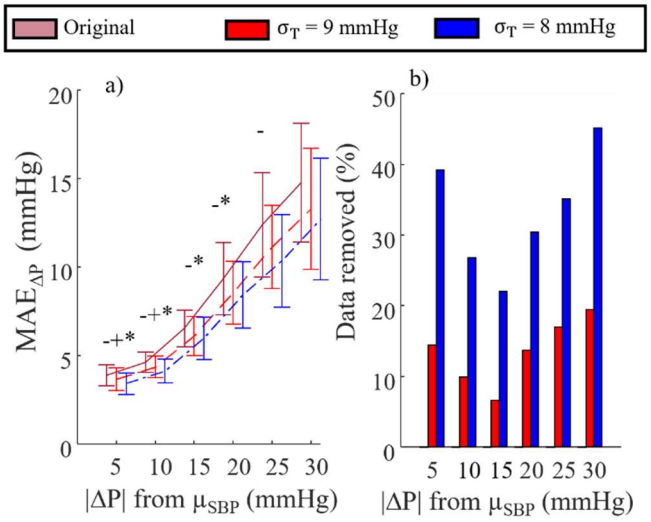

Figure 5.6 - For the NARX + ANN_{Sys} + PAT_{Log} model, a) mean and standard deviation of $MAE_{\Delta P}$ for all participants, and b) percentage of data removed from each bin, against the $|\Delta P|$ from μ_{SBP} measured throughout the day. In a) lines represent the mean of all subjects and the error bars show ± SD. Legend: (*) Original vs σ_T = 9 mmHg, (-) Original vs σ_T = 8 mmHg, and (+) σ_T = 8 mmHg vs σ_T = 9 mmHg ($p<0.05/3$)

The σ_{Err} for the NARX + ANN_{Sys} + PAT_{Log} model is shown in Figure 5.7.a when estimating BP during sitting, standing, and walking. The σ_{Err} is shown for original beat-by-beat estimation and for the σ_T = 8 mmHg case; the percentage of data kept for each activity is shown in Figure 5.7.b. It can be seen that estimating BP during walking is difficult. The expected σ_{Err} of NARX + ANN_{Sys} + PAT_{Log} is 10.3 mmHg, which is lower by 0.6 mmHg, 1.6 mmHg, and 3.9 mmHg, than the NARX, ANN_{Sys}, and PAT_{Log}, models, respectively. Moreover, the expected σ_{Err} decreases to 7.95 mmHg when using a σ_T = 8 mmHg threshold on the PI, and the expected mean error is down to 1.7 mmHg. The previous results show that using the fusion approach presented herein results in BP estimates during walking achieving the AAMI standard 29% of the time (Figure 5.7.b). Using σ_T = 9 mmHg increases the error SD to 8.3 mmHg while 53% of data are kept. Estimating BP while sitting and standing is an easier task, as can be observed with their respectively lower σ_{Err} values in comparison with walking. However, they still benefit by using a σ_T = 8 mmHg threshold, which decreases their σ_{Err} by 12% and 14%, respectively, while keeping 71% and 64% of the BP estimates, respectively.

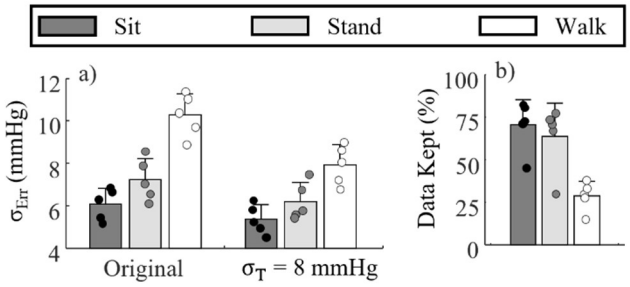

Figure 5.7 - Comparison of error standard deviation for the original NARX + ANN_{Sys} + PAT_{Log} model and the same model using a σ_T = 8 mmHg during sitting, standing, and walking, and b) percentage of data kept (prediction interval < 8 mmHg). Bars represent the mean of all subjects and the error bars show ± SD. Each data point represents a participant.

5.3 Discussion

This work was motivated by two needs: (1) to improve cuffless BP estimation, and (2) to provide a measure of the degree of confidence a practitioner should have in the BP estimates. The PI model architecture using OCSVM can be applied to different BP models, from simple PAT models to more complex NARX models. The covariance intersection fusion algorithm is also generic since it does not rely on any assumptions on the variables to be fused; only their means and covariances are needed. The method developed herein is not limited in terms of the number of models that can be fused. However,

the covariance intersection algorithm will become more complex to optimize as the number of models increases [177]. Herein, we used a small dataset specifically designed to train subject-specific BP models; however, the method is also applicable to population-based models. The one modification that would be needed for Big Data is to add kernel approximation techniques to minimize the training time of the OCSVMs [178]. Further studies are required to assess the fusion of population-based models.

In this study, we have shown that the PI estimation varied depending on the inclusion of estimated BP as an input to the OCSVMs, and the length of the input signals (ECG, PPG, and BP). The obtained MeRCI scores, indicating how well the PI represented the observed error [176], demonstrate the importance of using estimated BP as an input to the PI model.

High Pearson correlation coefficients for the NARX, ANN_{Sys}, and PAT_{Log} models have been observed between the estimated PI and the actual PI for each cluster. This shows the validity of using the training data clusters' σ_{Err} as a surrogate for the PI when estimating BP, as is typically done in clustering-based PI computations [179]. The over/underestimation of σ_{Err} in each cluster can be explained by the difficulty of estimating the training dataset and by how each model is trained. The 15-minute procedures performed by the participants to train the models were designed to create large BP variability in a short amount of time. This makes it a short but difficult time series to estimate for PAT_{Log}, which has only two tunable parameters, in comparison with the ANN_{Sys} (780 parameters) and NARX (820 parameters) models. NARX has more parameters, but it is trained in a feedforward fashion, and predicts as a recurrent ANN, which makes the prediction of the training data more difficult compared to ANN_{Sys}. Apart from over/underestimating, all models exhibit negative trends in Figure 5.2, which shows that the further away from the center of the training dataset is a BP estimate, the more likely the estimation error will be large. This highlights the importance of using an outlier detection algorithm, such as OCSVM, to cluster data.

The estimated σ_{Err} of the BP was smaller when fusing all three models (Figure 5.3.c). On average, the NARX + ANN_{Sys} + PAT_{Log} model decreased σ_{Err} by 1.2 mmHg, 0.8 mmHg, and 2.0 mmHg over the individual NARX, ANN_{Sys}, and PAT_{Log} models, respectively, with a consistent decrease for every participant. The μ_{Err}, however, is slightly higher when compared with NARX and PAT_{Log} due to the fusion with ANN_{Sys}, which is more biased than other estimators (Figure 5.3.b). Nevertheless, the MAE is the lowest when fusing all the models (Figure 5.3.a). A simple approach to decrease the bias of the fused models would be to correct its mean estimation to that of the least biased model (PAT_{Log}). However, a larger sample size would be required to validate that PAT_{Log} is the least biased.

Using PAT_{Log} in the fusion process is beneficial for the activities of daily living dataset because the model is good at estimating small change in BP (Figure 5.4.a); for this dataset, as observed in Figure 5.5.b, about 50% of the BP fluctuates between ±5 mmHg of the mean BP. The NARX + ANN_{Sys} + PAT_{Log} model benefits from this and performed similarly to PAT_{Log} in that range of ΔP. Fusion of the three models is beneficial up to a ΔP of ±10 mmHg, where NARX + ANN_{Sys} + PAT_{Log} model outperformed the other models for 80% of the data (except PAT_{Log} for ΔP of ±5 mmHg), explaining its best overall MAE. This result, however, will be highly dependant on the distribution of BP that is to be estimated. With a larger change in BP, NARX + ANN_{Sys} might be more suited for the task.

The estimation of the PI is best when using both NARX + ANN_{Sys} and NARX + ANN_{Sys} + PAT_{Log} models based upon the MeRCI scores. The mean PI is, however, lower for NARX + ANN_{Sys}. An estimation of how much the PI is over/underestimated can be computed by dividing the MeRCI score by the mean PI, which should give 3 for an ideal PI when computing the MeRCI score with 99.7% of the λ_i distribution (99.7% of the data of a normal distribution should be covered by ±3 SD). For the NARX + ANN_{Sys} model, this ratio is 3.5, thus, on average the PI is underestimated by 16%. For the NARX + ANN_{Sys} + PAT_{Log} model, the mean underestimation is 4%. This computation is only a crude estimation of the goodness of the PI since it only considers the mean PI and MeRCI score, and not the PI estimation accuracy at different ranges of PI and ΔP. Nevertheless, it was shown that the estimated PI could be effectively used to remove large errors from the BP estimation time series. However, it comes with a compromise between removing BP estimates with low and high errors for different σ_T, as shown in Figure 5.5.a. This compromise is, however, successful in decreasing σ_{Err} as σ_T is decreased (Figure 5.5.b). Moreover, the PI model is successful in decreasing σ_{Err} for every individual.

One might expect that OCSVMs cluster the data by range of BP, therefore, when using a σ_T = 8 (or 9) mmHg threshold on the PI it would remove only estimates at both ends of the BP range. However, this is not the case, as observed in Figure 5.6.a, the PI estimation model decreased the error for the whole range of BP, and removed data from every |ΔP| bin (Figure 5.6.b).

An ideal PI model would, when selecting a σ_T, result in no individual σ_{Err} larger than σ_T in every activity/posture. This was not the case with the NARX + ANN_{Sys} + PAT_{Log} model as it could not reach 8 mmHg for every participant while walking, although it reaches that level of error on average and for every participant while sitting and standing. This shows that the approach presented herein is a viable solution for cuffless BP estimation to be used in clinical settings as the BP estimates with a PI lower than 8 mmHg will meet the AAMI standard independent from the conditions in which the BP is

measured. In the worst-case scenario, there are no high-quality BP estimates (PI < 8 mmHg), but at least erroneous estimates would not be mistakenly used for diagnosis.

5.4 Conclusion

This chapter examined the estimation fusion of up to three state-of-the-art models to achieve better cuffless BP estimation accuracy. We emphasized the importance of the prediction interval in the process of fusion, also in the context of clinical use of the estimated BP data, for which high quality estimates are essential. For that purpose, a clustering approach using one-class support vector machines was developed to estimate the prediction interval independent of the model architecture.

It was shown that the prediction interval model can cluster the BP estimates in terms of their quality, which was quantified by the error standard deviation and used to estimate the prediction interval. The prediction interval model was used with the three different BP estimation models (NARX, ANN_{Sys}, and PAT_{Log}), and the results were fused using the covariance intersection fusion algorithm. The fusion of the three models showed improvement in BP estimation accuracy as well as in its estimated prediction interval.

Chapter 6
Blood Pressure Estimation During Moderate and Heavy Intensity Exercise

In this chapter, the estimation of BP during moderate and heavy intensity cycling exercise is investigated. The performance of the estimation is studied using MAP which is a good indicator of perfusion in the body [180]. Population-based and subject-specific model accuracies are compared during exercise in which BP is changing over a wide range in a short time interval. Based on the observations in Chapter 4 we hypothesize that modeling the complete BP waveform will lead to a more precise estimation of MAP than when directly using ECG and PPG features or the entire signals.

6.1 Methods

6.1.1 Data Collection

Eighteen young, healthy adults with no known musculoskeletal, respiratory, cardiovascular, or metabolic conditions participated in the study. The data used in this chapter are a subset of the dataset detailed in [181], during which BP was monitored as part of an investigation into the oxygen uptake kinetics in humans. Briefly, arterial BP was measured using finger PPG (NOVA; Finapres Medical Systems, the Netherlands), and ECG and PPG were measured via the Astroskin wearable body metrics vest (Carré Technologies Inc., Canada). All data were stored for processing at 64 Hz. In 6 of the participants the BP measurement device could not recalibrate during the exercise or automatically turned off and had to restart the initial calibration process. As a result, 12 participants were used for this study (7 males, 5 females; age: 27 ± 6 yr). The study was approved by the University of Waterloo Research Ethics committee (ORE #32164) and conducted in accordance with the Declaration of Helsinki.

Participants visited the laboratory on four separate occasions to perform a ramp-incremental exercise test and two different pseudorandom binary sequence (PRBS) exercise tests [182], [183], in which 21 minutes of data were collected. Each exercise session was separated by at least 48 hours. Participants were instructed to arrive for testing at least two hours postprandial and abstain from alcohol, caffeine, and vigorous exercise in the 24 hours preceding each test. To ensure reliable data collection, each participant was fitted with the correct Astroskin vest size based on thoracic and

abdominal circumference measurements. The middle finger or the thumb of the left hand was used for measuring finger BP, and the hand and forearm were rested on a warm beanbag placed on a platform set at heart level to minimize both motion artefact and poor perfusion in the finger. All exercise tests were performed in an environmentally controlled laboratory on an electronically braked cycle ergometer (Lode Excalibur Sport, Lode B.V., Groningen, Netherlands). Participants were instructed to maintain cadence at 60 revolutions per minute for all exercise tests.

On the first visit, 5 min of seated rest was followed by a ramp-incremental exercise test to exhaustion (Max test; 25 W baseline for 4 min followed by a 25 W/min ramp to exhaustion and a recovery period for nominally 21 minutes). The Max test was terminated when the cadence dropped below 55 revolutions per minute despite strong verbal encouragement. The remainder of the 21 minute Max test was used as a recovery period during which data continued to be collected; the duration of the recovery period varied between participants depending on each individual's time to exhaustion (ramp up time varied between 7.8 min and 12.1 min). The Max test was used to determine each participant's ventilatory threshold (VT) [184], peak oxygen uptake ($\dot{V}O_{2peak}$), and the work rates used in the design of the PRBS exercise tests (Figure 6.1). $\dot{V}O_{2peak}$ was defined as the highest $\dot{V}O_2$ computed from a 20 s moving average during the exercise test. $\dot{V}O_2$ at VT was estimated by visual inspection as previously described in [184]. Work rates at 90% VT, and the midpoint between VT and $\dot{V}O_{2peak}$ (referred to as Δ50%) were estimated by left-shifting the $\dot{V}O_2$ response by each individual's mean response time to align the $\dot{V}O_2$ and work rate profiles [185]. In visits 2 and 3, participants performed one of two different PRBS exercise tests in a randomized order. Work rates systematically alternated in the two PRBS exercise tests between 25 W and 90% VT (low-to-moderate; L-M), and 25 W and Δ50% (low-to-heavy; L-H). In this study, the 90% VT WR was 106 +/- 8 W and Δ50% WR was 185 +/- 25 W. A single PRBS was composed of 15 units, each 30 seconds in duration, totaling 7.5 min. Each complete PRBS testing session consisted of a 3.5-min warmup (the last 3.5 min of the 7.5-min PRBS), and then two full repetitions of the PRBS for a total of 18.5 min of continuous cycling per session. Also included were an initial two-minute resting period and a 30 second of recovery period for a total testing time of 21 minutes for each dataset.

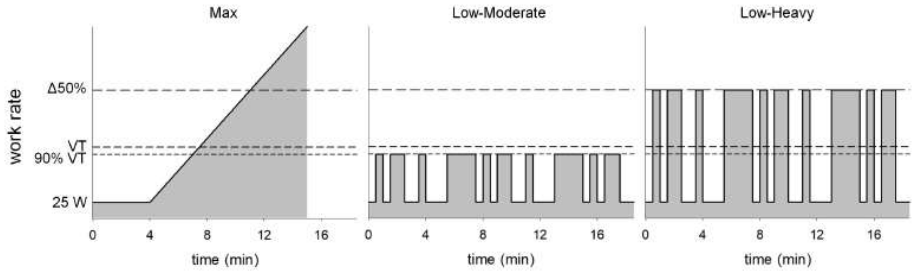

Figure 6.1 - Schematic of the incremental ramp and two pseudorandom binary sequence (PRBS) cycling tests. PRBS protocols were designed using work rates at 90% ventilatory threshold (VT), and the midpoint between VT and $\dot{V}O_{2peak}$ (Δ50%) using participant-specific VT and $\dot{V}O_{2peak}$ determined from the ramp test [181].

6.1.2 BP Estimation Models

$NARX_{pop}$ and $NARX_{sub}$: A NARX architecture based on a multilayer Perceptron artificial neural network (ANN) was developed. The ANN was composed of two hidden layers of 40 neurons each with a batch normalization layer before each fully connected hidden layer. Hyperbolic tangent and linear activation functions were used for the hidden layers and the output layer, respectively. The NARX model used exogenous inputs and the estimated BP at time steps *k-1* and *k-2* to estimate BP at time *k*. The exogenous inputs consisted of 0 to 18 time-step delays (19 samples) of the signals sampled at 64 Hz. The NARX input signals were ECG and a Z-normalized PPG signal, the details of which can be found in Chapter 4. The trained NARX model were used to estimate the BP waveform, from which MAP was extracted for further analysis. To reduce the estimation variance, five NARX models were ensemble-averaged [174], herein (depending on the model training) referred to as $NARX_{pop}$ or $NARX_{sub}$, which are population-based and subject-specific models, respectively.

ANN_{pop}: Using the same ANN architecture as NARX, models were developed to directly estimate the MAP from ECG and the Z-normalized PPG signal (no BP estimates in the input layer). For this ANN, 64 samples from each signal were used as inputs to the ANN. The 64 samples started from the ECG peak (QRS peak) up to the sample 63 time steps ahead. A similar model architecture was employed in [22]. Again, five ANN models were ensemble-averaged [174], herein referred to as ANN_{pop}.

ANN_{sub}: Using the same inputs as ANN_{pop}, a single hidden layer ANN was developed to be trained on smaller datasets. The ANN was a multilayer Perceptron with one hidden layer of 20 neurons in every

case. Hyperbolic tangent and linear activation functions were used for hidden layer and output neurons, respectively. The ANN_{sub} directly estimates MAP.

PAT-HR: User-specific pulse arrival time with heart rate (PAT-HR) models were implemented to facilitate direct comparison with an established cuffless BP monitoring method. The inverse PAT with linear HR term was computed as [115]:

$$\text{MAP} = a/\text{PAT} + b\text{HR} + c \tag{6.1.a}$$

where the user-specific coefficients a, b, and c were determined using a least-squares estimation for each training dataset [113]. PAT was computed as the time difference between the ECG peak and the PPG peak at 64 Hz ; HR was computed from the R-R interval in the ECG with resolution of 64 Hz.

$NARX_{corr}$: Subject-specific ANN models were trained to predict the error of the NARX BP waveform estimation using the 0 to 18 time-step delays Z-normalized PPG signal as input. This estimated error was subtracted from the NARX BP estimation. The goal of this step was to understand better to what extent the input signals still contain information about BP.

6.1.3 Model Training

In this study, separate population-based and subject-specific models were trained. In the former, leave-one-subject-out cross-validation was performed, where a model was trained on 11 subjects and tested on the one left out; this process was performed iteratively for the 12 participants. In the latter, a model was trained for participant on one cycling exercise test with the others used for testing.

Table 6.1 – Hyperparameters for training the different deep neural networks

	$NARX_{pop}$	**$NARX_{sub}$**	**ANN_{pop}**
Number of epochs	100	500	2500
Initial learning rate	0.01	0.01	0.01
Learning rate drop factor	0.9	0.95	0.95
Learning rate drop period (epochs)	1	1	100
Minibatch size	640	640	64

The $NARX_{pop}$, $NARX_{sub}$, and the ANN_{pop} networks were trained using the Adam optimizer, whereas the ANN_{sub} and the $NARX_{corr}$ were trained using error backpropagation with the Levenberg-Marquardt

optimization algorithm. The hyperparameters for the Adam optimizer were different for each model architecture and are shown in Table 6.1. During training, the data was shuffled every epoch and each minibatch does not contain a sequence of data; it contains random data as for training a regular feedforward ANN.

6.1.4 Validation and Data Analysis

To compare the performance of the different architectures using population-based and subject-specific trainings, the models were trained on the Max test and tested on the L-M and L-H exercise tests. To understand the effect of training data on population-based NARX models, training on different permutations of data were performed and tested on the L-M and L-H tests (in a leave-one-subject-out fashion): training on six random participants' Max tests ($NARX_{pop}$ – Max 6 participants), training on L-M or L-H only ($NARX_{pop}$ – L-M or $NARX_{pop}$ – L-H), and training on all the tests (Max, L-M, and L-H tests; $NARX_{pop}$ – All). Finally, to understand to what extent the performance of a model based on PPG and ECG can be increased, $NARX_{sub}$ models were trained on L-M, L-H ($NARX_{sub}$ – L-M (Train) or $NARX_{sub}$ – L-H (Train)) or all the tests ($NARX_{sub}$ – All (Train)) and validated on L-M(L-H). Here, the goal was to deliberately test the NARX model on its training data (or a portion of them) to see how well the model can learn the BP dynamics. Finally, $NARX_{corr}$ was trained to estimate the $NARX_{sub}$ estimation error, which was then was subtracted from the BP estimation to observe the best BP estimation performance that can be obtained solely from PPG and ECG. This final step was showing how much BP information was left in both PPG and ECG signals after modeling BP using the NARX model. For instance, if the error could be modeled perfectly from the PPG and ECG signals, it would mean that the NARX model could not extract all the relevant information from those, whereas if none of the error could be modeled, the NARX extracted all the information from the input signals.

The performance of the different models was analyzed using the mean error (μ_{Err}), error SD (σ_{Err}), Pearson correlation coefficient (r), coefficient of determination of the unbiased models (R^2), and the Kullback-Leibler (KL) divergence of the unbiased models. The unbiased models were computed as $\widehat{MAP} - \mu_{\mathrm{Err}}$. The equation for computing R^2 is shown in equation (6.2 and the KL divergence in (6.3. The KL divergence was used as a metric to compare the unbiased MAP and estimated unbiased MAP distributions. It shows how well the model covers a large range of BP.

$$R^2 = 1 - \frac{\sum_i (\mathrm{MAP}_i - \widehat{MAP}_i - \mu_{\mathrm{Err}})^2}{\sum_i (\mathrm{MAP}_i - \overline{MAP}_i)^2} \tag{6.2}$$

$$KL = \ln \frac{\sigma_{MAP}}{\sigma_{\widehat{MAP}}} + \frac{{\sigma_{MAP}}^2}{2{\sigma_{\widehat{MAP}}}^2} - \frac{1}{2} \quad (6.3)$$

To visualize model performance in estimating changes in MAP (ΔMAP), the mean absolute value of $\Delta MAP_{model} - \Delta MAP_{Nova}$ was used ($MAE_{\Delta MAP}$), where ΔMAP is defined as the MAP deviation from the mean and the subscript indicates whether the data are from the model or the Nova device. The errors were split into bins representing the distribution of MAP in the test dataset. Each bin spans a range of 5 mmHg, and the bins indicate deviations from the mean MAP of the participant by 5*m* to 5(*m*+1) mmHg, where *m* ranges from -6 to 5. The MAEs were compared for each bin using a Wilcoxon signed rank test with Bonferroni corrections to identify statistical differences across the models. The maximum cross-correlation values between the estimated BP waveform error and the input signals ($r_{err\text{-}PPG}$, and $r_{err\text{-}ECG}$) were computed to get a sense of the unmodeled relationship between the exogenous input signals and BP. Group averaged BP response to the L-M(L-H) exercise test was compared with the group averaged BP estimation to visualize trends in the estimation error. Heart rate (HR), the peaks of the PPG waveform (PPG_{peak}), and the inverse of the pulse arrival time (1/PAT), which are all common BP features that correlate with BP [22], [115], were computed to visualize the trends throughout the tests.

6.2 Results

Herein only the results for L-H are presented for brevity. L-M results are included as supplementary material in Appendix C.

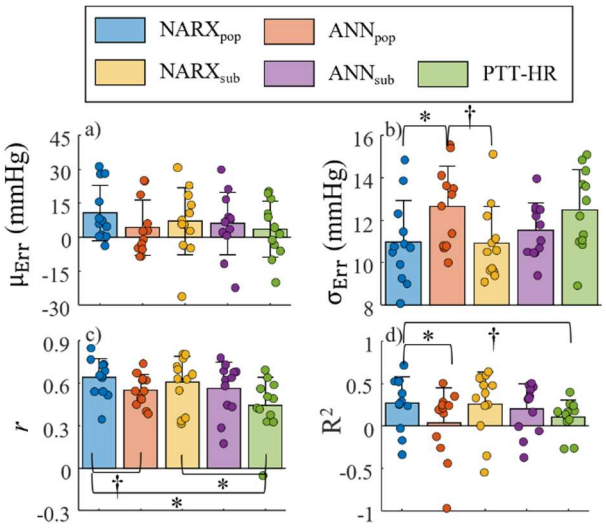

Figure 6.2 - Comparison of a) mean error, b) SD of the error, c) Pearson correlation coefficient, and d) coefficient of determination between different model estimates and the BP measurements. Bars represent the mean of all subjects and the error bars show ± SD for the L-H test. Each data point represents one participant. Legend: Wilcoxon with (*) p<0.05/10 and (†) p<0.01.

The means and SDs of μ_{Err}, σ_{Err}, r, and R^2 during the L-H test computed across the 12 subjects are shown in Figure 6.2 for the different models. The SD of μ_{Err} is large indicating that the BP models for some participants are biased. The results exhibit differences in mean σ_{Err} between the models, ranging from 10.91 mmHg for NARX$_{sub}$ (best) to 12.66 mmHg for ANN$_{pop}$ (worst). As a reference, the MAP SD across the 12 participants is 13.4±1.9 mmHg (mean±SD) during the L-H test. Due to the distribution of results, only the NARX$_{pop}$ σ_{Err} is found to be statistically lower than the ANN$_{pop}$ ($p<0.05/10$), but NARX$_{sub}$ trends toward a statistical difference ($p<0.01$). The average Pearson correlation coefficient between the estimated and measured MAP is higher for NARX$_{pop}$ and statistically different from PAT-HR ($p<0.05/10$) and trends toward a statistical difference with ANN$_{pop}$ ($p<0.01$). Also, NARX$_{sub}$ is significantly higher r than PTT-HR. Similarly, R^2 is the largest for both NARX$_{pop}$ and statistically different from ANN$_{pop}$ ($p<0.05/10$) and trended toward a statistical difference with PAT-HR ($p<0.01$).

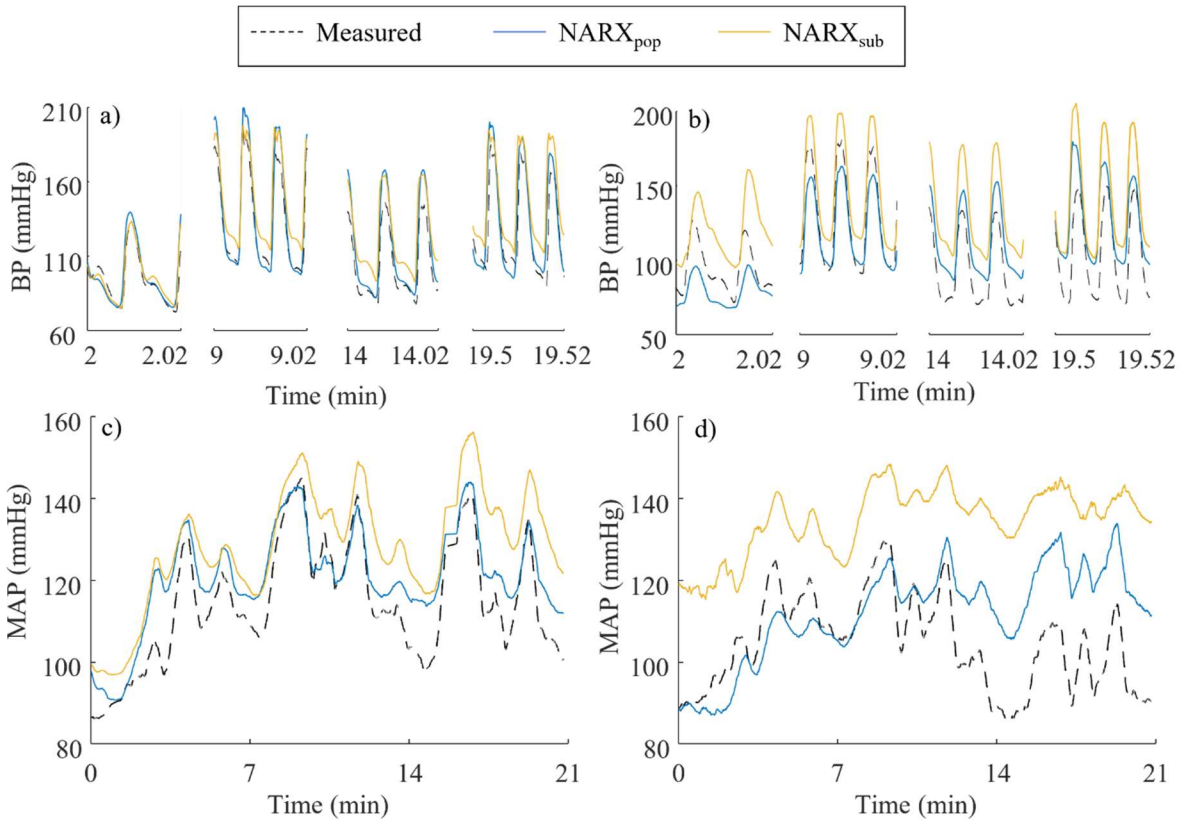

Figure 6.3 - Comparison of actual versus estimated BP waveforms and MAP (after applying a moving average). The results are shown for the subject with the lowest, a) and c), and highest, b) and d), SD of the error for the L-H test.

The estimated BP time series from $NARX_{pop}$ and $NARX_{sub}$ are shown for the subjects with the lowest and the highest σ_{Err} in Figure 6.3. In Figure 6.3.a-b, the estimated BP waveforms are shown at rest (2 min), during the first PRBS high intensity (Δ50%) section (9 min), during the low intensity (25 W) section between the two PRBS (14 min), and at the last high intensity (Δ50%) section (19.5 min). In Figure 6.3.c-d, MAP is shown as a function of time for the complete 21 minutes, including the initial resting portion and the complete L-H exercise test.

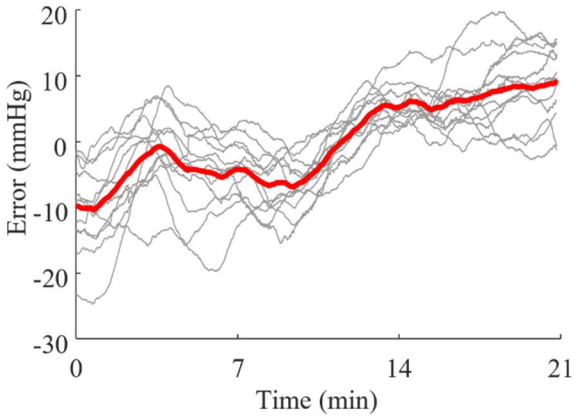

Figure 6.4 - Filtered BP error for the 12 subjects for the L-H cycling test. Grey lines represent individual subject data and the red line is the mean of all subjects.

The $NARX_{pop}$ error (with removed bias) evolution over the 21-minute testing period is presented for all 12 participants in Figure 6.4. The red line is the mean of all participants. It can be observed that the model in the majority of subjects underestimates the change in MAP up to approximately the end of the first PRBS (13 min) and overestimates it for the rest of the test, which is consistent with qualitative observations in Figure 6.3.

The means and SDs of σ_{Err}, r, R^2, and KL-divergence during the L-H test computed across the 12 subjects are shown in Figure 6.5 for $NARX_{pop}$ trained on different data. In general, training the model on only half of the data ($NARX_{pop}$ – Max 6 participants) leads to poorer outcomes, although there is no statistical difference due to the large variance in the results, it trends toward statistical significance ($p<0.05$). Training on the Max tests ($NARX_{pop}$) results in higher r and significantly lower KL divergence than training the model on L-H tests ($NARX_{pop}$ – L-H). On average, training on all tests ($NARX_{pop}$ – All) leads to larger R^2 and smaller σ_{Err} compared to the other models.

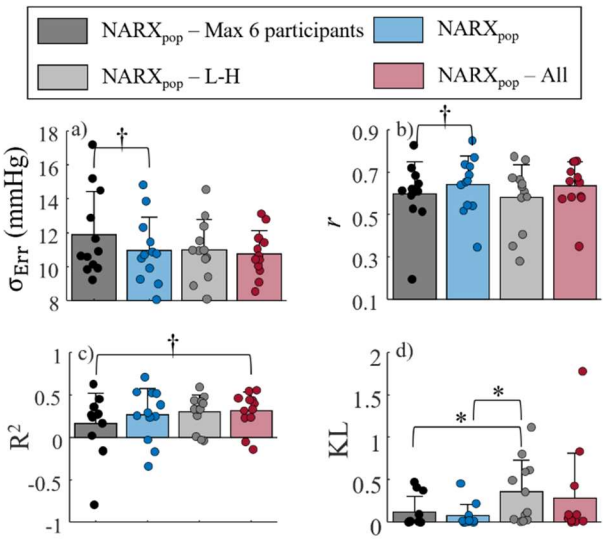

Figure 6.5 - Comparison of a) SD of the error, b) Pearson correlation coefficient, c) coefficient of determination, and d) KL divergence between different model estimates and the BP measurements distributions. Bars represent the mean of all subjects and the error bars show ± SD for the L-H test. Each data point represents one participant. Legend: Wilcoxon with (*) p<0.05/6 and (†) p<0.05.

The capability of predicting changes in MAP by NARX$_{pop}$ is shown in Figure 6.6 for the different training datasets. As shown by the p-values obtained from the Wilcoxon tests, using the data of all the tests as training data leads to lower MAE in the small range of ΔMAP without affecting the model's capability to estimate BP at large ΔMAP. Similarly, training on only the L-H tests leads to significantly lower MAE at low ΔMAP, but the model trained on the Max tests is significantly better at estimating large positive ΔMAP. Even though there are no significant differences between training on six or 11 participant's Max tests when comparing every ΔMAP bin, the model trained on 11 participants has a significantly lower mean ΔMAP error curve.

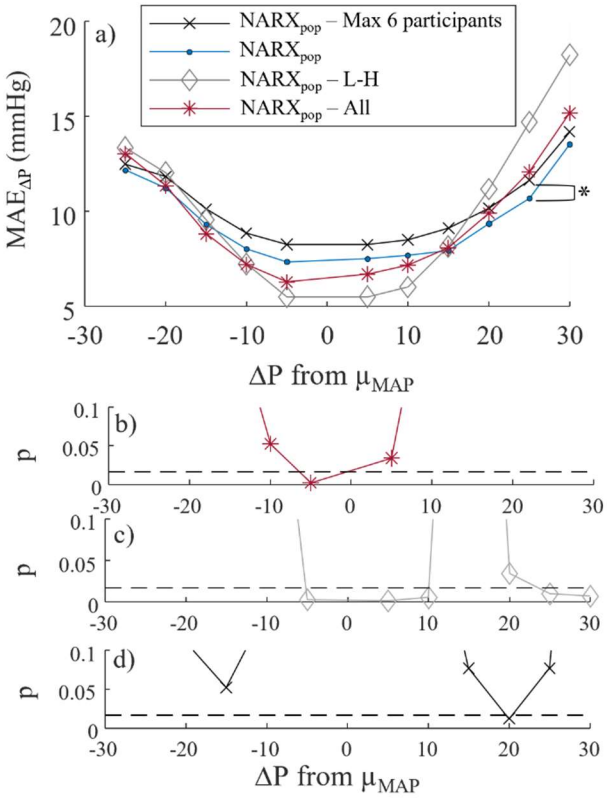

Figure 6.6 - a) Mean of $MAE_{\Delta P}$ for all participants against the distribution of ΔMAP from μMAP measured throughout the L-H test. Wilcoxon tests' p-value for b) $NARX_{pop}$ vs $NARX_{pop}$ – All, c) $NARX_{pop}$ vs $NARX_{pop}$ – L-H, and d) $NARX_{pop}$ vs $NARX_{pop}$ – Max 6 participants. Symbols represent the p-values (or $MAE_{\Delta P}$ value) and the location on the abscissa indicates the bin over which the statistical test was performed.

The means and SDs of σ_{Err}, r, R^2, $r_{err\text{-}PPG}$, and $r_{err\text{-}ECG}$ during the L-H test computed across the 12 subjects are shown in Figure 6.7 for $NARX_{pop}$, $NARX_{sub}$ – L-H (Train), $NARX_{sub}$ – All (Train), and $NARX_{corr}$. These results indicate that when a model can extract more information from the PPG and ECG signals (lower $r_{err\text{-}PPG}$ and $r_{err\text{-}ECG}$) it is better at estimating BP (lower σ_{Err} and higher r). It can also be observed that the employed NARX architectures are not able to extract all the information from the PPG and ECG even when the test data are included in the training set, though this helps the estimation ($NARX_{sub}$ – All (Train)).

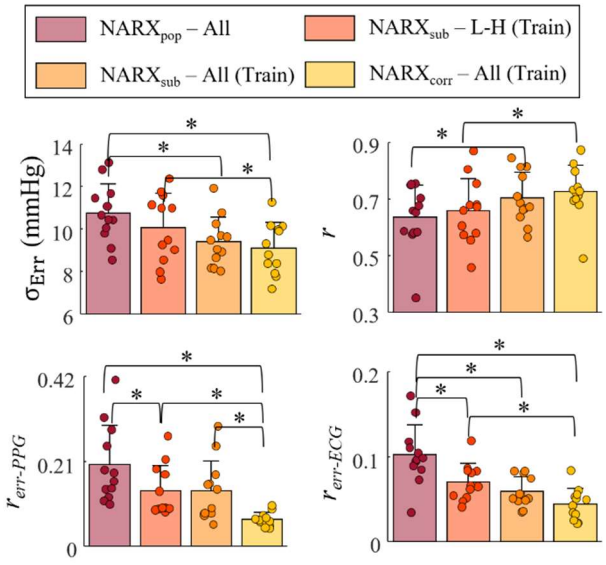

Figure 6.7 - Comparison of a) SD of the error, and b) Pearson correlation coefficient between different model estimates and the BP measurements, and the maximum cross-correlation between c) BP and PPG, and d) BP and ECG signals. Bars represent the mean of all subjects and the error bars show ± SD for the L-H test. Each data point represents one participant. Legend: (*) Wilcoxon with p<0.05/6.

The mean of the 12 participants' ΔMAP during the L-H test is shown in Figure 6.8.a. It can be seen that the measured ΔMAP during the second PRBS is lower than the first one, which is not the case when BP is estimated with the NARX models. Figure 6.8.b shows HR, PPG_{peak} and 1/PAT, which are all features that keep increasing throughout the test. This can explain why the NARX models do not predict the lower BP during the second PRBS.

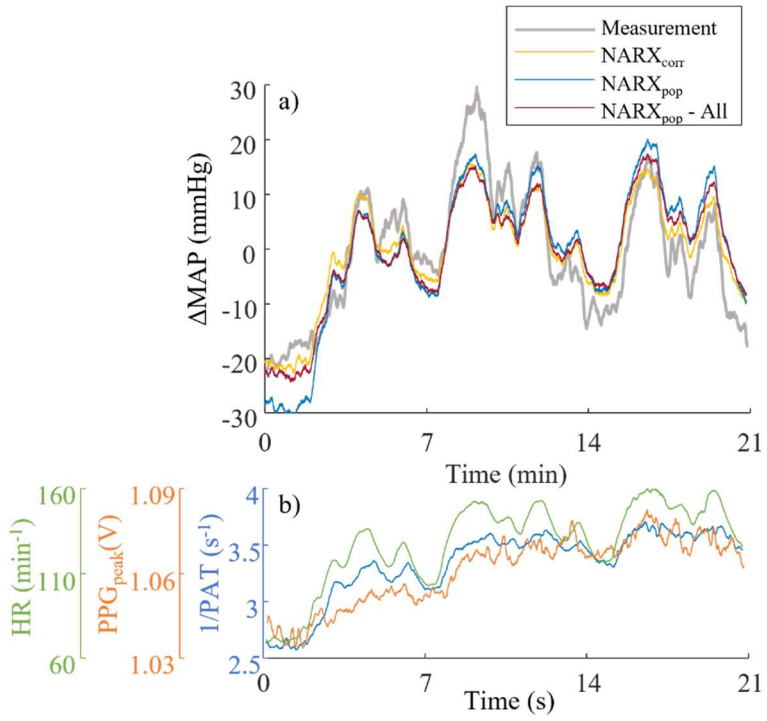

Figure 6.8 – Mean of the 12 participants a) measured and estimated ΔMAP and b) common features that correlate with BP for the L-H exercise test.

6.3 Discussion

This work looked into the generalizability of BP estimation during moderate and heavy cycling exercise, which, to the best of the author's knowledge, is a physiological condition for which BP estimation has never previously been attempted. Most cuffless BP estimation studies have used exercise as a means to increase BP prior to the estimation; they only estimated BP during the recovery period and not during the exercise [186]–[188]. One noted exception is the estimated BP during two minutes of low, medium, and heavy cycling exercise [147], wherein the models were subject-specific, trained per specific exercise, and no cross-validation was performed (only training data was analyzed). Herein, the intensities of the exercise were subject-specific and based on VT and $\dot{V}O_{2peak}$, which is a reliable way to quantify intensity [182], [184].

In this work, deep neural networks were trained to estimate the complete BP waveform from PPG and ECG signals. Final hyperparameters were chosen based on multiple iterations. For each iteration, the training data was split into 70% training and 30% validation to ensure that the models were not overfit. Once the final configuration was chosen, the models were trained on 100% of the training data

and tested as explained in the Method section. It is worth mentioning that hyperparameters shown in Table 6.1 did not affect the results likely because 1) the PPG and ECG signals do not contain enough information to overfit BP (see $NARX_{Corr}$ in Figure 6.7), and 2) the precision of the BP measurement device is not perfect [189]. It was shown in [189] that the Finapres device during Max tests had error in SBP and DBP of -2.7±10 mmHg and -2.6±10 mmHg at 180 W (average Δ50% herein), respectively, and increased to 0.7±15 mmHg and -3.3±13 mmHg at maximum work rate.

Herein, different model architectures are compared to better understand BP estimation results. For most cuffless BP studies, the results can be compared against the AAMI standard [153], however, it might be set to high since it was not intended for heavy intensity exercise. Therefore, no standard is directly applicable to judge the goodness of results. Instead, a well understood subject-specific PAT-HR model is used as a baseline comparison. This model leads to relatively good BP estimates on both the training (not shown here) and test data, and even better results on the training data when using PPG_{peak} as a feature. However, PPG_{peak} is not a robust feature on a day-to-day basis, probably due, partly, to sensors repositioning. For that reason, it is expected that temporal features are more robust than amplitude features when using linear models such as PAT-HR. However, when using ANNs, the Z-normalized PPG input signal is the solution to track BP robustly based on the PPG (and ECG) waveform. This goes in hand with the observation in Chapter 4 that the Z-normalized PPG signal improved estimation of BP during activities of daily living on another day than the one on which the training data were collected.

During cycling exercise, it is shown in Figure 6.2 that both $NARX_{sub}$ and ANN_{sub} performed similarly at estimating BP, although mean σ_{Err}, r, and R^2 were lower for $NARX_{sub}$ than ANN_{sub}. The fact that $NARX_{pop}$ is able to perform well in comparison to ANN_{pop} may be due to its capacity to rely more on the dynamics of the BP, which are relatively similar between subjects, and less on exogenous signals (PPG and ECG) features that can exhibit larger inter-subject variability.

As shown in Figure 6.3, Figure 6.4, and Figure 6.8, one factor that affects σ_{Err} the most is the difference in BP responses during the first and the second PRBS. For instance, the participant with the lowest σ_{Err} in Figure 6.3.c had approximately no change in BP response between the two PRBS (-4 mmHg peak MAP), whereas the participant with the largest σ_{Err} in Figure 6.3.d had a drop in BP response of approximately 20 mmHg. This drop of BP between the large peaks of the PRBSs is on average 12 mmHg for the L-H test (Figure 6.8.a). This phenomenon can be explained by a decrease in total peripheral resistance (TPR) due to factors such as increased skin blood flow during exercise [190].

The decrease in TPR was much smaller during the L-M test (see Appendix C); it was almost neglectable on the timescale of 21 min. This means that BP during exercise is influenced by prior exercise and is not predictable based on PPG and ECG alone; $NARX_{corr}$ could not correct the error drift and, therefore, those signals do not contain enough information about the change in TPR. Moreover, previous exercise has different effects on future BP depending not only on intensity, but also duration [190]. The employed NARX architecture, which has only a short memory of two time steps at 64 Hz and, therefore, can only learn faster dynamics, was not able to learn these longer timescale dynamics. Though not predicted by the models in this study, the longer timescale variations in TPR might be learnable dynamics. Since the timescale is long, however, occuring over minutes and potentially even affected by prior activities during the course of the day (morning vs evening) [190], it may be practically challenging to train such model. Long short-term memory (LSTM) and gated recurrent unit (GRU) architectures, which are known for their temporal memory, are left for future work.

$NARX_{pop}$ BP estimation improves by training on more participants' data and also by training on different cycling test types (Figure 6.5). Training the population-based model on only the L-H test to estimate L-H test data resulted in poorer results than $NARX_{pop}$ (- Max) and $NARX_{pop}$ – All, showing first that the model can generalize to different types of exercise and, second, that the model benefits from training on Max due to the large range of BP covered within that test. The latter is shown in Figure 6.6 where $NARX_{pop}$ was significantly better than $NARX_{pop}$ – L-H at estimating the upper range of ΔMAP. Based on the results shown in Figure 6.6, $NARX_{pop}$ can track relatively well changes in BP during cycling exercise, but cannot track the absolute value of BP (Figure 6.2). However, those results are based on a small amount of data for a population-based model. Adding more participant data for training the model might be a solution to tracking absolute BP values.

Figure 6.7 shows that the NARX model architecture could not extract all the PPG and ECG information if provided with the necessary training data. This is shown in $r_{err\text{-}PPG}$ and $r_{err\text{-}PPG}$ of $NARX_{sub}$ – All (Train) and $NARX_{corr}$ – All (Train), which are statistically different. Even though σ_{Err} from both models are statistically and meaningfully different (9.4 vs 8.5 mmHg), it is not realistic to base a model on the error since it cannot be known *a priori*. Figure 6.7 shows instead the expected best performance of the NARX model architecture that could be achieved. Although it cannot be tested herein, it is hypothesized that $NARX_{pop}$ – All could reach the performance of $NARX_{sub}$ – All (Train) if provided with data of a larger and more evenly distributed demographic. Even with this best performance, however, the model does not track the slow change in BP dynamics observed between the two PRBSs

(Figure 6.8). On the other hand, BP tends to show a larger rise in BP at the onset of exercise and linearly decrease over time [190], [191] depending on the exercise intensity. Therefore, it might be argued that the NARX model could estimate BP changes during exercises in a time window where the decrease in TPR is small. For instance, one PRBS would be more reliable data to measure the performance of the model in estimating BP during heavy intensity exercise, where the time window is shorter than the timescale of the change in TPR. As an example, $NARX_{pop}$ - All σ_{Err} would drop from 10.8 mmHg to 8.9 mmHg when considering only the second PRBS. Assuming that $NARX_{sub}$ – All (Train) would be the best achievable performance, its σ_{Err} on the second PRBS is 7.9 mmHg, which is lower than the requirement by the AAMI standard [153]. In the case of moderate exercise, the change in TPR is small during a time window of 21 min, where $NARX_{pop}$ – All's σ_{Err} is already 8.1 mmHg and drops to 7.2 mmHg with $NARX_{sub}$ – All (Train). Therefore, the NARX model performance is a function of the intensity of the exercise and would require recalibration periodically based on the exercise intensity.

In this study PPG was measured on the forehead, which is known to provide high-quality and reliable signals [192]–[194]. However, PPG signals acquired from the forehead have been reported to be less affected by vasoconstriction than the fingers [192], [193], [195], [196], which might have limited the precision of the BP models herein as the change in TPR was not capture on the forehead, but might be measurable from the fingers. Future studies should investigate the estimation of BP during considerable changes in TPR using a finger or toe PPG, from which vasomotion can be clearly observed in comparison to the ear PPG [197]. Another study has reported a correlation between ballistocardiography signal and TPR [198], however, this would likely be an impractical sensor to employ during exercise. Moreover, changes in TPR obtained by slow breathing and breath holding are not in the same range as observed herein and might not correlate at that point.

It should be noted that the ground truth BP measurements used in this study have limited the accuracy of the estimations. Volume clamping during high intensity exercise are not as reliable as in a resting posture [189]. This is clearly observed with the participants removed from this study, where the BP measurement device could not recalibrate during exercise or would automatically turn off and restart the initial calibration process. The precision of the device might, therefore, explain the relatively low R^2 computed for the BP estimation models. For instance, if the device has a precision of 10 mmHg during heavy exercise conditions, assuming that the measurement error is independent of the PPG and ECG signals, a model estimating BP perfectly can only achieve an R^2 of 0.44 when the SD of BP is 13.4 mmHg ($R^2 = 1 - \frac{\sum_i (\mathrm{MAP}_i - \widehat{MAP}_i - \mu_{Err})^2}{\sum_i (\mathrm{MAP}_i - \overline{MAP}_i)^2} = 1 - \frac{10^2}{13.4^2} = 0.44$, see equation (6.2). This also explains

why R^2 is lower during L-M (see Appendix C), since the MAP SD is only 9.1 mmHg. Therefore, a possible improvement to this study would be an invasive BP measurement to better ground the data.

6.4 Conclusion

This chapter explored the feasibility of estimating BP during moderate and heavy intensity exercise. It was shown that the NARX model architecture can track large changes in MAP with both subject-specific and population-based training. This architecture was shown to better model BP during moderate and heavy intensity cycling exercise than a PAT-HR linear model and an ANN using the entire PPG and ECG waveforms as input. It was shown that a limitation of estimating BP from PPG and ECG is the change in total peripheral resistance that happens during exercise (and is a function of intensity) and cannot be modeled using those signals. Although the NARX model is a dynamical model, it does not have sufficiently long memory to capture this change in total peripheral resistance that happens over a timescale on the order of minutes. Nevertheless, population-based NARX was able to track changes in MAP with a reasonable precision during moderate and heavy intensity exercise, which is increased when estimating on a smaller time window (e.g. 21 minutes for moderate and 8.5 minutes for heavy). Therefore, it can be concluded that the population-based NARX model can estimate BP during moderate and heavy intensity exercise, which was achieved by training on the data of only 11 participants. Moreover, its precision can be improved with more data and by using finger PPG as an input to measure changes in TPR.

Chapter 7
Conclusions and Recommendations

The main findings of this thesis are summarized and recommendations for future work are given.

7.1 Conclusions

In this thesis, cuffless BP monitoring was explored, an approach that can give a continuous measurement of the BP, which is not possible with existing non-invasive tools, and therefore gives a more complete picture of the patient's cardiovascular state. This approach requires a surrogate model of BP, which has typically relied upon pulse-wave delay between two arterial segments or other pulse waveform features in the estimation process. Typical approaches provide estimates of only systolic and diastolic blood pressure as a function of time. However, considerable clinically-relevant data are lost by neglecting the BP waveform. Moreover, most efforts to date in the field of cuffless BP monitoring have been in validating algorithms in a laboratory setting and with little to no consideration to algorithm uncertainty.

In this thesis, critical gaps in cuffless BP monitoring were studied using machine learning in the pursuit of several objectives, which were laid out in Section 1.1 and are repeated here:

4. develop a data-driven model of BP dynamics to enable estimation of the complete BP waveform;
5. validate the model architecture in real-world scenarios;
6. incorporate an uncertainty measure into the modeling scheme.

The main findings, as they relate to these objectives, are summarized as follows.

In Chapter 3, a NARX model was introduced; this is the first BP estimation model that gives access to the full BP waveform (compared to just SBP, DBP, and mean BP), which can provide clinicians with additional salient information (Objective 1), such as cardiac output, stroke volume, and arterial stiffness. Moreover, the proposed model improved the accuracy of SBP, DBP, and mean BP estimations when compared to state-of-the-art pulse arrival time models. This NARX approach is the first to consider BP as a dynamical system instead of the typical approach where static features are used to model the BP, which leads to a more accurate model.

In Chapter 4, the first validation of a cuffless BP method during activities of daily living was presented. A 15-minute training procedure was introduced, which captures a wide range of BP observed during activities of daily living that, combined with the NARX model, showed accurate continuous BP estimation, regardless of the user's position (Objective 2). Moreover, the trained NARX model remains accurate over months if a one-point calibration is employed prior to each use. A major contribution of this chapter is the validation and the applicability of the protocol used for the developed cuffless BP monitoring system. The developed solution was validated as a wearable system, which could potentially be prescribed by clinicians. With only a 30-minute training session and a daily calibration, accurate continuous BP estimation up to 6 months is realizable. This is attractive in comparison to the majority of the methods in the literature that either require several days of data collection to be able to build a model or cannot track BP changes due to the lack of model complexity. Moreover, the accuracy of the BP estimates over a broad range of BPs for each user in an uncontrolled environment validates the method as a truly viable wearable solution.

In Chapter 5, it is argued that it is imperative to assign a confidence metric to cuffless BP estimates for outcome evaluation, since it is likely to be employed outside of the clinic. As such, the prediction interval (PI), which is a practical confidence metric concerned with the prediction accuracy of the target value, would be of high clinical relevance. In Chapter 5, a method for estimating PI independently of the BP model architecture was developed. A fusion algorithm was then used to combine three established BP models (using their estimated PI). The results showed (1) the benefit of estimation fusion on the BP and the PI estimations accuracy, and (2) how PI can be used to identify high-quality data. This shows that the approach presented is a viable solution for cuffless BP estimation to be used outside the clinic as the BP estimates with a PI higher than a predetermined quality level can be discarded. In the worst-case scenario, there are no high-quality BP estimates, but at least erroneous estimates would not be mistakenly used for diagnosis.

In Chapter 6, the first attempt to estimate BP during low and heavy intensity exercise is presented. Even though the experiments were performed in an environmentally controlled laboratory on an electronically braked cycle ergometer, they should be representative of real-life exercises where motion artefacts are not dominant such as cycling, cross-country skiing, or aerobic weightlifting. In Chapter 6, it was shown that the NARX model could track large changes in BP precisely when both trained in a subject-specific or population-based manner. The largest source of error that could not be tracked by the NARX model using PPG and ECG as input was the change in BP due to a decrease in peripheral

resistance over the course of the exercise. This might limit the time window on which the model can be used without drifting too much. It was shown that within a 21-minute moderate pseudorandom binary sequence (PRBS) exercise test or within one heavy PRBS exercise test of 8.5 minutes, the change in BP due to the decrease in peripheral resistance was small enough and led to acceptable model precision. Since this change in peripheral resistance depends on the intensity of the exercise, the model could be precise continuously for low intensity exercise or if periodically calibrated for moderate and heavy intensity exercise.

7.2 Recommendations

Validation protocols for cuffless BP monitoring are lacking. IEEE is working on a standard for validating cuffless approaches, but it is still an ongoing process [167]. For that reason, the results reported in the literature are difficult to compare. The main reasons are inhomogeneity in the calibrations used (Section 2.4), datasets employed (database versus in-house data collection), and methods for reporting the results.

The results from population-based models are easier to report since cuff-based measurement standards, such as the AAMI, are readily applicable. However, most results reported in the literature used the AAMI protocol's mean and standard deviation of error criteria (see Section 2.2.1.2), but do not apply the requirements in terms of number of participants and range of BP that should be covered by the validation protocol. Therefore, the results are only applicable to the dataset on which the model has been validated. I suggest that future work on cuffless BP estimation should use data in a different setting to validate models. For instance, if training a model on the MIMIC database, one should collect data in a different environment to validate the applicability of the model in a realistic environment.

The problems with person-specific and hybrid calibrations are the same. First and foremost, the calibration of those models has an expiry date, that is not covered by the AAMI protocol. When validating one of those models, the error needs to be reported along with a time frame for which the periodic calibration is needed. This, along with testing different models with different data, makes it difficult to compare results. The last problem that will be covered herein is the assessment of results for person-specific and hybrid calibrations. Many different studies report their results by aggregating data from all participants. This leads to a high correlation between estimated and measured BP ($\rho > 0.9$). However, this is an artefact coming from the different BP levels across individuals; the correlation only reflects the difference in the initial BP calibration measurement between the different individuals

and not the capacity of a model in tracking change in BP in a given individual. This is applicable to this thesis work as well. For instance, in Chapter 3, high correlation coefficients between estimated and measured BP were presented in Figure 3.8, in order to be comparable to other studies, but then the study went on with subject-specific results instead of aggregated results. Cuffless BP studies should at least compare with other models on the same data. This, however, has limitations as far as all models from the literature cannot be reproduced in each study. Future standards for evaluating calibrated cuffless BP model should include a definition to normalize the reported error by the difficulty of estimating a given dataset. This would facilitate comparisons between different studies.

The main limitation of the collected data used in Chapters 4, 5, and 6 is the moderate accuracy of BP reference signal. As explained in Section 2.2.1.3, a finger cuff BP measurement system (Portapres and NOVA; Finapres Medical Systems, the Netherlands) has its own bias and variance, which does not meet the AAMI standard requirement (although the NOVA system uses RTF calibration, which improves the precision; see Section 2.2.1.3). Therefore, the developed model estimation error can reflect not only the error in estimating BP, but also the error of the measurement device itself. In future work, catheterization of participants should be performed and compared with finger cuff measures. Since training subject-specific models using catheterization is not a viable solution, the study should focus on the impact of the accuracy of the training data on the estimation of the BP model.

The precision of the estimation of BP during exercise was limited by the change in peripheral resistance, which could not be captured by the NARX model based on PPG and ECG. Although we suggested in Chapter 6 to use a deep neural network architecture with longer memory (e.g., LSTM), it is in the author's opinion that the timescale in question is too long and variable for such network to learn the underlying dynamics caused by the changing peripheral resistance. Future studies should investigate integrating a physiological measure that correlates with peripheral resistance or change in peripheral resistance. For instance, measuring skin blood flow with a doppler probe might lead to good results [190], however, this is not a convenient sensor to use. Skin temperature is a more practical sensor that would easily be integrated into a cuffless BP monitoring system, and, based on the data of a few participants in Chapter 6, the temperature measured with some models of the Astroskin vest (Carré Technologies Inc., Canada) increased throughout the exercise and might be inversely correlated with the change in peripheral resistance. This is left for future work.

Chapters 4 and 5 were limited to estimating BP during fairly sedentary activities of daily living, which are applicable to office work, for instance. However, the performance of the NARX model in

estimating BP for a more active daily life, such as during a day of a construction worker, will drastically decrease (expected by the performance during walking). Nevertheless, it is in the author's opinion that, by using the PI, good BP estimation data could be selected, which would be representative of the actual BP distribution and, therefore, useful to a clinician. Future work should investigate the performance of the NARX and PI models during more active daily lives.

The current approach to diagnose hypertension uses cuff-based BP measurements every 20 minutes for 24 hours. Therefore, future studies should explore estimating BP during the night to close the loop, such that the current approach can be replaced by cuffless continuous measurement of the BP to give a more accurate picture of a patient's state. It was shown in Chapter 3 that estimating BP using NARX is feasible in the supine position (bed-ridden patients). However, practical challenges need to be addressed for nighttime estimation using the NARX model. There is a well-known drop in BP while sleeping, so collecting data to train the NARX model that contains a similar drop without disturbing sleep might be challenging. Ideally, the 15-minute training protocol developed in Chapter 4 could be enhanced such that it encompasses such BP drop, leading to a 24-h estimation based solely on this training protocol. One potential solution would be to add deep breathing in the supine position to reach a lower BP level to the protocol [109].

This thesis covered a large range of inter-individual BP from hypo- to hypertensive cases in Chapter 3. However, Chapter 3 was limited to bed-ridden patients and, therefore, only small BP variabilities had to be estimated. Future studies should investigate the effect of inter-individual BP level on the accuracy of the NARX model in predicting BP variability during activities of daily living. It will be of high importance to assess the NARX model performance on different sex and ethnicities before the deployment of such a cuffless BP monitoring system for clinical uses.

In this thesis, PI was used as a confidence metric for the BP estimation. Future work should investigate explainable AI [199]–[201] as a means to increase confidence in BP models. Physiological models, such as PTT-based models, have gained a lot of attention over the years because of their simplicity and interpretability. Extracting physiological parameters using explainable AI would be an interesting way to gain confidence in BP models without sacrificing accuracy. Another interesting approach that could be used for the NARX model specifically is using interpretable nonlinear functions instead of ANN. This way, the interpretability of the relationship from input to output would not be lost in the ANN weights and could be interpreted using nonlinear dynamics techniques.

This thesis was certainly limited in terms of participants. Nevertheless, it is in the author's opinion the methodology expressed by this thesis in terms of validation protocols and using PI (or other confidence metrics) for such health monitoring application should be pursued by other scholars. Through such methodology, it is hoped that cuffless BP estimation might be called, one day, cuffless BP measurement.

Bibliography

[1] "Hypertension." https://www.who.int/westernpacific/health-topics/hypertension (accessed Jun. 25, 2020).

[2] "WHO | Q&As on hypertension," *WHO*. http://www.who.int/features/qa/82/en/index.html (accessed Nov. 05, 2018).

[3] V. L. Burt *et al.*, "Prevalence of Hypertension in the Us Adult Population. Results from the Third National Health and Nutrition Examination Survey, 1988-1991," *Hypertens. Dallas Tex 1979*, vol. 25, no. 3, pp. 305–313, Mar. 1995.

[4] Meeting of Experts on Accurate and Affordable Blood Pressure Measuring Devices for Office/Clinic Use in Low Resource Settings, "Affordable Technology: Blood Pressure Measuring Devices for Low Resource Settings," Geneva : World Health Organization, Geneva, 2005. Accessed: Dec. 04, 2018. [Online]. Available: http://apps.who.int/iris/handle/10665/43115

[5] R. Merai *et al.*, "CDC Grand Rounds: A Public Health Approach to Detect and Control Hypertension," *Morb. Mortal. Wkly. Rep.*, vol. 65, no. 45, pp. 1261–1264, Nov. 2016, doi: 10.15585/mmwr.mm6545a3.

[6] R. Mukkamala *et al.*, "Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Theory and Practice," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 8, pp. 1879–1901, Aug. 2015, doi: 10.1109/TBME.2015.2441951.

[7] T. S. Schoot, M. Weenk, T. H. van de Belt, L. J. Engelen, H. van Goor, and S. J. Bredie, "A New Cuffless Device for Measuring Blood Pressure: A Real-Life Validation Study," *J. Med. Internet Res.*, vol. 18, no. 5, May 2016, doi: 10.2196/jmir.5414.

[8] K. Matsumura, P. Rolfe, S. Toda, and T. Yamakoshi, "Cuffless Blood Pressure Estimation Using Only a Smartphone," *Nat. Sci. Rep.*, vol. 8, no. 1, Dec. 2018, doi: 10.1038/s41598-018-25681-5.

[9] R. C. Hermida *et al.*, "2013 Ambulatory blood pressure monitoring recommendations for the diagnosis of adult hypertension, assessment of cardiovascular and other hypertension-associated risk, and attainment of therapeutic goals (summary). Joint recommendations from the International Society for Chronobiology (ISC), American Association of Medical Chronobiology and Chronotherapeutics (AAMCC), Spanish Society of Applied Chronobiology, Chronotherapy, and Vascular Risk (SECAC), Spanish Society of Atherosclerosis (SEA), and Romanian Society of Internal Medicine (RSIM)," *Clin. E Investig. En Arterioscler.*, vol. 25, no. 2, pp. 74–82, Jun. 2013, doi: 10.1016/j.arteri.2013.03.002.

[10] J. A. Staessen *et al.*, "Task Force Ii: Blood Pressure Measurement and Cardiovascular Outcome," *Blood Press. Monit.*, vol. 6, no. 6, pp. 355–370, Dec. 2001.

[11] J. A. Staessen, L. Beilin, G. Parati, B. Waeber, and W. White, "Task Force Iv: Clinical Use of Ambulatory Blood Pressure Monitoring. Participants of the 1999 Consensus Conference on Ambulatory Blood Pressure Monitoring," *Blood Press. Monit.*, vol. 4, no. 6, pp. 319–331, Dec. 1999.

[12] G. Wang, M. Atef, and Y. Lian, "Towards a Continuous Non-Invasive Cuffless Blood Pressure Monitoring System Using PPG: Systems and Circuits Review," *IEEE Circuits Syst. Mag.*, vol. 18, no. 3, pp. 6–26, thirdquarter 2018, doi: 10.1109/MCAS.2018.2849261.

[13] P. C.-P. Chao, C.-C. Wu, D. H. Nguyen, B.-S. Nguyen, P.-C. Huang, and V.-H. Le, "The Machine Learnings Leading the Cuffless PPG Blood Pressure Sensors into the Next Stage," *IEEE Sens. J.*, pp. 1–1, 2021, doi: 10.1109/JSEN.2021.3073850.

[14] M. Simjanoska, M. Gjoreski, M. Gams, and A. Madevska Bogdanova, "Non-Invasive Blood Pressure Estimation from ECG Using Machine Learning Techniques," *Sens. Basel*, vol. 18, no. 4, Apr. 2018, doi: 10.3390/s18041160.

[15] Y. Zheng, B. P. Yan, Y. Zhang, and C. C. Y. Poon, "An Armband Wearable Device for Overnight and Cuff-Less Blood Pressure Measurement," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 7, pp. 2179–2186, Jul. 2014, doi: 10.1109/TBME.2014.2318779.

[16] Z. Tang *et al.*, "A Chair–Based Unobtrusive Cuffless Blood Pressure Monitoring System Based on Pulse Arrival Time," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 5, pp. 1194–1205, Sep. 2017, doi: 10.1109/JBHI.2016.2614962.

[17] I. Moxham, "Understanding Arterial Pressure Waveforms," *South. Afr. J. Anaesth. Analg.*, vol. 9, no. 1, pp. 40–42, Feb. 2003, doi: 10.1080/22201173.2003.10872991.

[18] M. F. O'Rourke, "Arterial Pressure Waveforms in Hypertension," *Minerva Med.*, vol. 94, no. 4, pp. 229–250, Aug. 2003.

[19] R. Velik, "An Objective Review of the Technological Developments for Radial Pulse Diagnosis in Traditional Chinese Medicine," *Eur. J. Integr. Med.*, vol. 7, no. 4, pp. 321–331, Aug. 2015, doi: 10.1016/j.eujim.2015.06.006.

[20] R. Mukkamala, A. T. Reisner, H. M. Hojman, R. G. Mark, and R. J. Cohen, "Continuous Cardiac Output Monitoring by Peripheral Blood Pressure Waveform Analysis," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 459–467, Mar. 2006, doi: 10.1109/TBME.2005.869780.

[21] M. P. M. Harms *et al.*, "Continuous Stroke Volume Monitoring by Modelling Flow from Non-Invasive Measurement of Arterial Pressure in Humans Under Orthostatic Stress," *Clin. Sci.*, vol. 97, no. 3, pp. 291–301, Sep. 1999, doi: 10.1042/cs0970291.

[22] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuffless Blood Pressure Estimation Algorithms for Continuous Health-Care Monitoring," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 4, pp. 859–869, Apr. 2017, doi: 10.1109/TBME.2016.2580904.

[23] "Valencell Wearable and Hearable Sensor White Papers," *Valencell*. https://valencell.com/white-papers/ (accessed Jul. 03, 2020).

[24] X.-R. Ding, Y.-T. Zhang, J. Liu, W.-X. Dai, and H. K. Tsang, "Continuous Cuffless Blood Pressure Estimation Using Pulse Transit Time and Photoplethysmogram Intensity Ratio," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 5, pp. 964–972, May 2016, doi: 10.1109/TBME.2015.2480679.

[25] W.-R. Yan, R.-C. Peng, Y.-T. Zhang, and D. Ho, "Cuffless Continuous Blood Pressure Estimation From Pulse Morphology of Photoplethysmograms," *IEEE Access*, vol. 7, pp. 141970–141977, 2019, doi: 10.1109/ACCESS.2019.2942936.

[26] I. Sharifi, S. Goudarzi, and M. B. Khodabakhshi, "A novel dynamical approach in continuous cuffless blood pressure estimation based on ECG and PPG signals," *Artif. Intell. Med.*, vol. 97, pp. 143–151, Jun. 2019, doi: 10.1016/j.artmed.2018.12.005.

[27] Y. Zheng, B. P. Yan, Y. Zhang, C. M. Yu, and C. C. Y. Poon, "Wearable cuff-less PTT-based system for overnight blood pressure monitoring," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2013, pp. 6103–6106. doi: 10.1109/EMBC.2013.6610945.

[28] Y. Zheng, C. C. Y. Poon, B. P. Yan, and J. Y. W. Lau, "Pulse Arrival Time Based Cuff-Less and 24-H Wearable Blood Pressure Monitoring and its Diagnostic Value in Hypertension," *J. Med. Syst.*, vol. 40, no. 9, p. 195, Jul. 2016, doi: 10.1007/s10916-016-0558-6.

[29] S. Hansen and M. Staber, "Oscillometric blood pressure measurement used for calibration of the arterial tonometry method contributes significantly to error," *Eur. J. Anaesthesiol. EJA*, vol. 23, no. 9, pp. 781–787, Sep. 2006, doi: 10.1017/S0265021506000688.

[30] S. Lee, H. R. Dajani, S. Rajan, G. Lee, and V. Z. Groza, “Uncertainty in Blood Pressure Measurement Estimated Using Ensemble-Based Recursive Methodology,” *Sensors*, vol. 20, no. 7, Art. no. 7, Jan. 2020, doi: 10.3390/s20072108.
[31] I. BIPM, I. IFCC, and I. IUPAC, “OIML ‘Guide to the Expression of Uncertainty in Measurement’ International Organization for Standardization,” *Geneva Switz.*, 1995.
[32] E. O’Brien and N. Atkins, “A comparison of the British Hypertension Society and Association for the Advancement of Medical Instrumentation protocols for validating blood pressure measuring devices: can the two be reconciled?,” p. 6.
[33] M. A. Ahmad, C. Eckert, and A. Teredesai, “Interpretable Machine Learning in Healthcare,” in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, New York, NY, USA, Aug. 2018, pp. 559–560. doi: 10.1145/3233547.3233667.
[34] D. U. Silverthorn, *Human Physiology: An Integrated Approach*, 5 edition. San Francisco: Pearson, 2009.
[35] G. J. Tortora and B. H. Derrickson, *Principles of Anatomy and Physiology*, 12 edition. Hoboken, NJ: Wiley, 2008.
[36] “Low Blood Pressure Symptoms, Chart, Causes, and Treatments,” *MedicineNet*. https://www.medicinenet.com/low_blood_pressure/article.htm (accessed Mar. 15, 2019).
[37] P. Shah and M. A. Louis, “Physiology, Central Venous Pressure,” in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2018. Accessed: Feb. 04, 2019. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK519493/
[38] “Understanding Blood Pressure Readings,” *www.heart.org*. https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings (accessed Aug. 25, 2021).
[39] S. Wessely, J. Nickson, and B. Cox, “Symptoms of low blood pressure: a population study.,” *Br. Med. J.*, vol. 301, no. 6748, pp. 362–365, Aug. 1990, doi: 10.1136/bmj.301.6748.362.
[40] M. F. O’Rourke, “From theory into practice: arterial haemodynamics in clinical hypertension,” *J. Hypertens.*, vol. 20, no. 10, pp. 1901–1915, Oct. 2002.
[41] M. E. Safar *et al.*, “Interaction Between Hypertension and Arterial Stiffness,” *Hypertension*, vol. 72, no. 4, pp. 796–805, Oct. 2018, doi: 10.1161/HYPERTENSIONAHA.118.11212.
[42] A. Noordergraaf, *Circulatory System Dynamics*. Elsevier, 2012.
[43] M. E. Safar *et al.*, “Hypertension and Vascular Dynamics in Men and Women With Metabolic Syndrome,” *J. Am. Coll. Cardiol.*, vol. 61, no. 1, pp. 12–19, Jan. 2013, doi: 10.1016/j.jacc.2012.01.088.
[44] familydoctor org editorial staff, “High Blood Pressure - Best Way To Lower Blood Pressure,” *familydoctor.org*. https://familydoctor.org/condition/high-blood-pressure/ (accessed Aug. 26, 2021).
[45] R. D. Brook, A. B. Weder, and S. Rajagopalan, “‘Environmental hypertensionology’ the effects of environmental factors on blood pressure in clinical practice and research,” *J. Clin. Hypertens. Greenwich Conn*, vol. 13, no. 11, pp. 836–842, Nov. 2011, doi: 10.1111/j.1751-7176.2011.00543.x.
[46] İ. Eşer, L. Khorshid, Ü. Y. Güneş, and Y. Demir, “The effect of different body positions on blood pressure,” *J. Clin. Nurs.*, vol. 16, no. 1, pp. 137–140, 2007, doi: 10.1111/j.1365-2702.2005.01494.x.
[47] R. Pinar, N. Sabuncu, and A. Oksay, “Effects of crossed leg on blood pressure,” *Blood Press.*, vol. 13, no. 4, pp. 252–254, Jan. 2004, doi: 10.1080/08037050410000903.
[48] V. K. van Wijnen *et al.*, “Noninvasive beat-to-beat finger arterial pressure monitoring during orthostasis: a comprehensive review of normal and abnormal responses at different ages,” *J. Intern. Med.*, vol. 282, no. 6, pp. 468–483, 2017, doi: 10.1111/joim.12636.

[49] M. O'rourke, A. Avolio, V. Stelliou, J. Young, and D. E. Gallagher, "The rhythm of running: can the heart join in?," *Aust. N. Z. J. Med.*, vol. 23, no. 6, pp. 708–710, 1993, doi: https://doi.org/10.1111/j.1445-5994.1993.tb04732.x.

[50] G. A. Fontana, T. Pantaleo, F. Bongianni, F. Cresci, R. Manconi, and P. Panuccio, "Respiratory and cardiovascular responses to static handgrip exercise in humans," *J. Appl. Physiol.*, vol. 75, no. 6, pp. 2789–2796, Dec. 1993, doi: 10.1152/jappl.1993.75.6.2789.

[51] J. M. Stewart, I. Taneja, and M. S. Medow, "Reduced central blood volume and cardiac output and increased vascular resistance during static handgrip exercise in postural tachycardia syndrome," *Am. J. Physiol.-Heart Circ. Physiol.*, vol. 293, no. 3, pp. H1908–H1917, Sep. 2007, doi: 10.1152/ajpheart.00439.2007.

[52] K. M. Rose *et al.*, "Blood pressure and pulse responses to three stressors: associations with sociodemographic characteristics and cardiovascular risk factors," *J. Hum. Hypertens.*, vol. 18, no. 5, pp. 333–341, May 2004, doi: 10.1038/sj.jhh.1001677.

[53] R. A. Nishimura and A. J. Tajik, "The Valsalva Maneuver and Response Revisited," *Mayo Clin. Proc.*, vol. 61, no. 3, pp. 211–217, Mar. 1986, doi: 10.1016/S0025-6196(12)61852-7.

[54] J. M. Stewart, M. A. Medow, B. Bassett, and L. D. Montgomery, "Effects of thoracic blood volume on Valsalva maneuver," *Am. J. Physiol.-Heart Circ. Physiol.*, vol. 287, no. 2, pp. H798–H804, Aug. 2004, doi: 10.1152/ajpheart.01174.2003.

[55] E. Agabiti-Rosei *et al.*, "Central blood pressure measurements and antihypertensive therapy: a consensus document," *Hypertens. Dallas Tex 1979*, vol. 50, no. 1, pp. 154–160, Jul. 2007, doi: 10.1161/HYPERTENSIONAHA.107.090068.

[56] A. P. Avolio *et al.*, "Role of Pulse Pressure Amplification in Arterial Hypertension," *Hypertension*, vol. 54, no. 2, pp. 375–383, Aug. 2009, doi: 10.1161/HYPERTENSIONAHA.109.134379.

[57] M. KARAMANOGLU, M. F. O'ROURKE, A. P. AVOLIO, and R. P. KELLY, "An analysis of the relationship between central aortic and peripheral upper limb pressure waves in man," *Eur. Heart J.*, vol. 14, no. 2, pp. 160–167, Feb. 1993, doi: 10.1093/eurheartj/14.2.160.

[58] J. Hashimoto, Y. Imai, and M. F. O'Rourke, "Monitoring of Antihypertensive Therapy for Reduction in Left Ventricular Mass:," *Am. J. Hypertens.*, vol. 20, no. 11, pp. 1229–1233, Nov. 2007, doi: 10.1016/j.amjhyper.2007.07.010.

[59] P. Segers, E. R. Rietzschel, and J. A. Chirinos, "How to Measure Arterial Stiffness in Humans," *Arterioscler. Thromb. Vasc. Biol.*, vol. 40, no. 5, pp. 1034–1043, May 2020, doi: 10.1161/ATVBAHA.119.313132.

[60] W. W. Nichols, "Clinical Measurement of Arterial Stiffness Obtained from Noninvasive Pressure Waveforms," *Am. J. Hypertens.*, vol. 18, no. 1 Pt 2, pp. 3S-10S, Jan. 2005, doi: 10.1016/j.amjhyper.2004.10.009.

[61] S. Laurent *et al.*, "Expert consensus document on arterial stiffness: methodological issues and clinical applications," *Eur. Heart J.*, vol. 27, no. 21, pp. 2588–2605, Nov. 2006, doi: 10.1093/eurheartj/ehl254.

[62] M. J. Bourgeois, B. K. Gilbert, G. Von Bernuth, and E. H. Wood, "Continuous determination of beat to beat stroke volume from aortic pressure pulses in the dog.," *Circ. Res.*, vol. 39, no. 1, pp. 15–24, Jul. 1976, doi: 10.1161/01.RES.39.1.15.

[63] B. H. McGhee and E. J. Bridges, "Monitoring Arterial Blood Pressure: What You May Not Know," *Crit. Care Nurse*, vol. 22, no. 2, pp. 60–64, 66–70, 73 passim, Apr. 2002.

[64] L. H. Lehman, M. Saeed, D. Talmor, R. Mark, and A. Malhotra, "Methods of Blood Pressure Measurement in the ICU," *Crit. Care Med.*, vol. 41, no. 1, pp. 34–40, Jan. 2013, doi: 10.1097/CCM.0b013e318265ea46.

[65] Perloff D *et al.*, "Human Blood Pressure Determination by Sphygmomanometry.," *Circulation*, vol. 88, no. 5, pp. 2460–2470, Nov. 1993, doi: 10.1161/01.CIR.88.5.2460.
[66] B. S. Alpert, D. Quinn, and D. Gallick, "Oscillometric Blood Pressure: A Review for Clinicians," *J. Am. Soc. Hypertens.*, vol. 8, no. 12, pp. 930–938, Dec. 2014, doi: 10.1016/j.jash.2014.08.014.
[67] G. Drzewiecki, R. Hood, and H. Apple, "Theory of the Oscillometric Maximum and the Systolic and Diastolic Detection Ratios," *Ann. Biomed. Eng.*, vol. 22, no. 1, pp. 88–96, Feb. 1994.
[68] P. S. Lewis, "Oscillometric measurement of blood pressure: a simplified explanation. A technical note on behalf of the British and Irish Hypertension Society," *J. Hum. Hypertens.*, vol. 33, no. 5, pp. 349–351, May 2019, doi: 10.1038/s41371-019-0196-9.
[69] J. Liu, J.-O. Hahn, and R. Mukkamala, "Error mechanisms of the oscillometric fixed-ratio blood pressure measurement method," *Ann. Biomed. Eng.*, vol. 41, no. 3, pp. 587–597, Mar. 2013, doi: 10.1007/s10439-012-0700-7.
[70] E. O'Brien and N. Atkins, "A comparison of the British Hypertension Society and Association for the Advancement of Medical Instrumentation protocols for validating blood pressure measuring devices: can the two be reconciled?," *J. Hypertens.*, vol. 12, no. 9, pp. 1089–1094, Sep. 1994.
[71] E. O'Brien *et al.*, "Short report: An outline of the revised British Hypertension Society protocol for the evaluation of blood pressure measuring devices," *J. Hypertens.*, vol. 11, no. 6, pp. 677–679, Jun. 1993.
[72] E. O'Brien *et al.*, "European Society of Hypertension International Protocol revision 2010 for the validation of blood pressure measuring devices in adults," *Blood Press. Monit.*, vol. 15, no. 1, pp. 23–38, Feb. 2010, doi: 10.1097/MBP.0b013e3283360e98.
[73] B. A. Friedman, B. S. Alpert, D. Osborn, L. M. Prisant, D. E. Quinn, and J. Seller, "Assessment of the validation of blood pressure monitors: a statistical reappraisal," *Blood Press. Monit.*, vol. 13, no. 4, pp. 187–191, Aug. 2008, doi: 10.1097/MBP.0b013e3283071a64.
[74] G. S. Stergiou *et al.*, "A Universal Standard for the Validation of Blood Pressure Measuring Devices," *Hypertension*, vol. 71, no. 3, pp. 368–374, Mar. 2018, doi: 10.1161/HYPERTENSIONAHA.117.10237.
[75] G. Mancia *et al.*, "EFFECTS OF BLOOD-PRESSURE MEASUREMENT BY THE DOCTOR ON PATIENT'S BLOOD PRESSURE AND HEART RATE," *The Lancet*, vol. 322, no. 8352, pp. 695–698, Sep. 1983, doi: 10.1016/S0140-6736(83)92244-4.
[76] M. B. Schulze, A. Kroke, M. M. Bergmann, and H. Boeing, "Differences of blood pressure estimates between consecutive measurements on one occasion: Implications for inter-study comparability of epidemiologic studies," *Eur. J. Epidemiol.*, vol. 16, no. 10, pp. 891–898, Oct. 2000, doi: 10.1023/A:1011020823807.
[77] B. P. Imholz, W. Wieling, G. A. van Montfrans, and K. H. Wesseling, "Fifteen Years Experience with Finger Arterial Pressure Monitoring: Assessment of the Technology," *Cardiovasc. Res.*, vol. 38, no. 3, pp. 605–616, Jun. 1998.
[78] K. H. Wesseling, D. B. Wit, V. D. G. M.A. Hoeven, V. J. Goudoever, and J. J. Settels, "Physiocal, Calibrating Finger Vascular Physiology for Finapres," *Homeost. Health Dis.*, vol. 36, no. 2–3, pp. 67–82, 1995.
[79] "Continuous noninvasive arterial pressure," *Wikipedia*. Sep. 29, 2018. Accessed: Nov. 05, 2018. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Continuous_noninvasive_arterial_pressure&oldid=861734601
[80] I. Guelen *et al.*, "Finometer, finger pressure measurements with the possibility to reconstruct brachial pressure," *Blood Press. Monit.*, vol. 8, no. 1, pp. 27–30, Feb. 2003.

[81] P. Gizdulich, A. Prentza, and K. H. Wesseling, “Models of brachial to finger pulse wave distortion and pressure decrement,” *Cardiovasc. Res.*, vol. 33, no. 3, pp. 698–705, Mar. 1997, doi: 10.1016/s0008-6363(97)00003-5.
[82] W. J. Bos, J. van Goudoever, G. A. van Montfrans, A. H. van den Meiracker, and K. H. Wesseling, “Reconstruction of brachial artery pressure from noninvasive finger pressure measurements,” *Circulation*, vol. 94, no. 8, pp. 1870–1875, Oct. 1996, doi: 10.1161/01.cir.94.8.1870.
[83] I. Guelen *et al.*, “Validation of brachial artery pressure reconstruction from finger arterial pressure,” *J. Hypertens.*, vol. 26, no. 7, pp. 1321–1327, Jul. 2008, doi: 10.1097/HJH.0b013e3282fe1d28.
[84] J. Allen, “Photoplethysmography and its application in clinical physiological measurement,” *Physiol. Meas.*, vol. 28, no. 3, pp. R1–R39, Mar. 2007, doi: 10.1088/0967-3334/28/3/R01.
[85] M. Ghamari, “A review on wearable photoplethysmography sensors and their potential future applications in health care,” *Int. J. Biosens. Bioelectron.*, vol. 4, no. 4, 2018, doi: 10.15406/ijbsbe.2018.04.00125.
[86] D. Biswas, N. Simões-Capela, C. Van Hoof, and N. Van Helleputte, “Heart Rate Estimation From Wrist-Worn Photoplethysmography: A Review,” *IEEE Sens. J.*, vol. 19, no. 16, pp. 6560–6570, Aug. 2019, doi: 10.1109/JSEN.2019.2914166.
[87] A. A. Alian and K. H. Shelley, “Photoplethysmography,” *Best Pract. Res. Clin. Anaesthesiol.*, vol. 28, no. 4, pp. 395–406, Dec. 2014, doi: 10.1016/j.bpa.2014.08.006.
[88] A. Chandrasekhar, M. Yavarimanesh, K. Natarajan, J.-O. Hahn, and R. Mukkamala, “PPG Sensor Contact Pressure Should Be Taken Into Account for Cuff-Less Blood Pressure Measurement,” *IEEE Trans. Biomed. Eng.*, vol. 67, no. 11, pp. 3134–3140, Nov. 2020, doi: 10.1109/TBME.2020.2976989.
[89] G. Acar, O. Ozturk, A. J. Golparvar, T. A. Elboshra, K. Böhringer, and M. K. Yapici, “Wearable and Flexible Textile Electrodes for Biopotential Signal Monitoring: A review,” *Electronics*, vol. 8, no. 5, Art. no. 5, May 2019, doi: 10.3390/electronics8050479.
[90] R. Acharya U, Ed., *Advances in cardiac signal processing*. Berlin ; New York: Springer, 2007.
[91] O. T. Inan *et al.*, “Ballistocardiography and Seismocardiography: A Review of Recent Advances,” *IEEE J. Biomed. Health Inform.*, vol. 19, no. 4, pp. 1414–1427, Jul. 2015, doi: 10.1109/JBHI.2014.2361732.
[92] C.-S. Kim, A. M. Carek, R. Mukkamala, O. T. Inan, and J.-O. Hahn, “Ballistocardiogram as Proximal Timing Reference for Pulse Transit Time Measurement: Potential for Cuffless Blood Pressure Monitoring,” *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2657–2664, Nov. 2015, doi: 10.1109/TBME.2015.2440291.
[93] A. K. Verma, R. Fazel-Rezai, J. M. Zanetti, and K. Tavakolian, “Preliminary Results for Estimating Pulse Transit Time Using Seismocardiogram,” *J. Med. Devices*, vol. 9, no. 2, Jun. 2015, doi: 10.1115/1.4030124.
[94] O. Nelles, *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Springer Science & Business Media, 2013.
[95] N. Castle, “Regression vs. Classification Algorithms.” https://www.datascience.com/blog/regression-and-classification-machine-learning-algorithms (accessed Mar. 21, 2019).
[96] M. Small, *Applied Nonlinear Time Series Analysis: Applications In Physics, Physiology And Finance*. Wspc, 2005.
[97] B. C. Csáji and H. T. Eikelder, “Approximation with Artificial Neural Networks,” Eötvös Loránd University, Budapest, Hungary, 2001.
[98] G. Horváth, “Chapter 4 Neural Networks in System Identification,” p. 36.

[99] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
[100] R. Ge, F. Huang, C. Jin, and Y. Yuan, "Escaping From Saddle Points — Online Stochastic Gradient for Tensor Decomposition," in *Proceedings of The 28th Conference on Learning Theory*, Jun. 2015, pp. 797–842. Accessed: Oct. 08, 2021. [Online]. Available: https://proceedings.mlr.press/v40/Ge15.html
[101] D. Masters and C. Luschi, "Revisiting Small Batch Training for Deep Neural Networks," *ArXiv180407612 Cs Stat*, Apr. 2018, Accessed: Oct. 08, 2021. [Online]. Available: http://arxiv.org/abs/1804.07612
[102] S. Hochreiter, "The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions," *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, vol. 06, no. 02, pp. 107–116, Apr. 1998, doi: 10.1142/S0218488598000094.
[103] J. M. P. Menezes and G. A. Barreto, "Long-term time series prediction with the NARX network: An empirical evaluation," *Neurocomputing*, vol. 71, no. 16, pp. 3335–3343, Oct. 2008, doi: 10.1016/j.neucom.2008.01.030.
[104] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.
[105] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.
[106] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New Support Vector Algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, May 2000, doi: 10.1162/089976600300015565.
[107] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the Support of a High-Dimensional Distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001, doi: 10.1162/089976601750264965.
[108] "Train support vector machine (SVM) classifier for one-class and binary classification - MATLAB fitcsvm." https://www.mathworks.com/help/stats/fitcsvm.html#bt8z1o_-2 (accessed Sep. 15, 2021).
[109] R. Mukkamala and J.-O. Hahn, "Initialization of Pulse Transit Time-Based Blood Pressure Monitors," in *The Handbook of Cuffless Blood Pressure Monitoring: A Practical Guide for Clinicians, Researchers, and Engineers*, J. Solà and R. Delgado-Gonzalo, Eds. Cham: Springer International Publishing, 2019, pp. 163–190. doi: 10.1007/978-3-030-24701-0_10.
[110] X. Ding, "Machine Learning Techniques," in *The Handbook of Cuffless Blood Pressure Monitoring: A Practical Guide for Clinicians, Researchers, and Engineers*, J. Solà and R. Delgado-Gonzalo, Eds. Cham: Springer International Publishing, 2019, pp. 139–162. doi: 10.1007/978-3-030-24701-0_9.
[111] J. Muehlsteff, X. L. Aubert, and M. Schuett, "Cuffless Estimation of Systolic Blood Pressure for Short Effort Bicycle Tests: The Prominent Role of the Pre-Ejection Period," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2006, pp. 5088–5092. doi: 10.1109/IEMBS.2006.260275.
[112] C. C. Young, J. B. Mark, W. White, A. DeBree, J. S. Vender, and A. Fleming, "Clinical evaluation of continuous noninvasive blood pressure monitoring: Accuracy and tracking capabilities," *J. Clin. Monit.*, vol. 11, no. 4, pp. 245–252, Jul. 1995, doi: 10.1007/BF01617519.
[113] M. Y.-M. Wong, C. C.-Y. Poon, and Y.-T. Zhang, "An Evaluation of the Cuffless Blood Pressure Estimation Based on Pulse Transit Time Technique: a Half Year Study on Normotensive Subjects," *Cardiovasc. Eng.*, vol. 9, no. 1, pp. 32–38, Mar. 2009, doi: 10.1007/s10558-009-9070-7.

[114] P. Fung, G. Dumont, C. Ries, C. Mott, and M. Ansermino, “Continuous noninvasive blood pressure measurement by pulse transit time,” in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Sep. 2004, vol. 1, pp. 738–741. doi: 10.1109/IEMBS.2004.1403264.

[115] J. M. Bote, J. Recas, and R. Hermida, “Evaluation of blood pressure estimation models based on pulse arrival time,” *Comput. Electr. Eng.*, vol. 84, p. 106616, Jun. 2020, doi: 10.1016/j.compeleceng.2020.106616.

[116] X. Ding *et al.*, “Continuous Blood Pressure Measurement From Invasive to Unobtrusive: Celebration of 200th Birth Anniversary of Carl Ludwig,” *IEEE J. Biomed. Health Inform.*, vol. 20, no. 6, pp. 1455–1465, Nov. 2016, doi: 10.1109/JBHI.2016.2620995.

[117] P. M. Nabeel, S. Karthik, J. Joseph, and M. Sivaprakasam, “Arterial Blood Pressure Estimation From Local Pulse Wave Velocity Using Dual-Element Photoplethysmograph Probe,” *IEEE Trans. Instrum. Meas.*, vol. 67, no. 6, pp. 1399–1408, Jun. 2018, doi: 10.1109/TIM.2018.2800539.

[118] F. S. Cattivelli and H. Garudadri, “Noninvasive Cuffless Estimation of Blood Pressure from Pulse Arrival Time and Heart Rate with Adaptive Calibration,” in *Proc. BSN*, Jun. 2009, pp. 114–119. doi: 10.1109/BSN.2009.35.

[119] C. Kim, A. M. Carek, O. Inan, R. Mukkamala, and J. Hahn, “Ballistocardiogram-Based Approach to Cuff-Less Blood Pressure Monitoring: Proof-of-Concept and Potential Challenges,” *IEEE Trans. Biomed. Eng.*, pp. 1–1, 2018, doi: 10.1109/TBME.2018.2797239.

[120] D. D. He, E. S. Winokur, and C. G. Sodini, “An Ear-Worn Vital Signs Monitor,” *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2547–2552, Nov. 2015, doi: 10.1109/TBME.2015.2459061.

[121] S. L.-O. Martin *et al.*, “Weighing Scale-Based Pulse Transit Time is a Superior Marker of Blood Pressure than Conventional Pulse Arrival Time,” *Sci. Rep.*, vol. 6, no. 1, Art. no. 1, Dec. 2016, doi: 10.1038/srep39273.

[122] E. J. Wang *et al.*, “Seismo: Blood Pressure Monitoring using Built-in Smartphone Accelerometer and Camera,” in *Proc. CHI*, Montreal QC, Canada, 2018, pp. 1–9. doi: 10.1145/3173574.3173999.

[123] Y. Maeda, M. Sekine, and T. Tamura, “Relationship Between Measurement Site and Motion Artifacts in Wearable Reflected Photoplethysmography,” *J. Med. Syst.*, vol. 35, no. 5, pp. 969–976, Oct. 2011, doi: 10.1007/s10916-010-9505-0.

[124] R. C. Block *et al.*, “Conventional pulse transit times as markers of blood pressure changes in humans,” *Sci. Rep.*, vol. 10, no. 1, Art. no. 1, Oct. 2020, doi: 10.1038/s41598-020-73143-8.

[125] G. Zhang, M. Gao, D. Xu, N. B. Olivier, and R. Mukkamala, “Pulse arrival time is not an adequate surrogate for pulse transit time as a marker of blood pressure,” *J. Appl. Physiol.*, vol. 111, no. 6, pp. 1681–1686, Dec. 2011, doi: 10.1152/japplphysiol.00980.2011.

[126] A. D. Choudhury, R. Banerjee, A. Sinha, and S. Kundu, “Estimating blood pressure using Windkessel model on photoplethysmogram,” in *Conf Proc IEEE Eng Med Biol Soc*, Aug. 2014, pp. 4567–4570. doi: 10.1109/EMBC.2014.6944640.

[127] R. Kondo, M. S. Bhuiyan, H. Kawanaka, and K. Oguri, “Separate Estimation of Long- and Short-Term Systolic Blood Pressure Variability from Photoplethysmograph,” in *Conf Proc IEEE Eng Med Biol Soc*, Aug. 2014, pp. 1851–1854. doi: 10.1109/EMBC.2014.6943970.

[128] S. Huang, P. Hung, C. Hong, and H. Wang, “A New Image Blood Pressure Sensor Based on PPG, RRT, BPTT, and Harmonic Balancing,” *IEEE Sens. J.*, vol. 14, no. 10, pp. 3685–3692, Oct. 2014, doi: 10.1109/JSEN.2014.2329676.

[129] M. Elgendi, “On the Analysis of Fingertip Photoplethysmogram Signals,” *Curr. Cardiol. Rev.*, vol. 8, no. 1, pp. 14–25, Feb. 2012, doi: 10.2174/157340312801215782.

[130] E. C.-P. Chua, S. J. Redmond, G. McDarby, and C. Heneghan, "Towards Using Photo-Plethysmogram Amplitude to Measure Blood Pressure During Sleep," *Ann. Biomed. Eng.*, vol. 38, no. 3, pp. 945–954, Mar. 2010, doi: 10.1007/s10439-009-9882-z.
[131] D.-G. Jang, S.-H. Park, and M. Hahn, "Enhancing the Pulse Contour Analysis-Based Arterial Stiffness Estimation Using a Novel Photoplethysmographic Parameter," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 1, pp. 256–262, Jan. 2015, doi: 10.1109/JBHI.2014.2306679.
[132] H. Shin and S. D. Min, "Feasibility study for the non-invasive blood pressure estimation based on ppg morphology: normotensive subject study," *Biomed. Eng. OnLine*, vol. 16, no. 1, p. 10, Jan. 2017, doi: 10.1186/s12938-016-0302-y.
[133] V. R. Nafisi and M. Shahabi, "Intradialytic hypotension related episodes identification based on the most effective features of photoplethysmography signal," *Comput. Methods Programs Biomed.*, vol. 157, pp. 1–9, Apr. 2018, doi: 10.1016/j.cmpb.2018.01.012.
[134] F. Mouney, T. Tiplica, J.-B. Fasquel, M. Hallab, and M. Dinomais, "A New Blood Pressure Estimation Approach Using PPG Sensors: Subject Specific Evaluation over a Long-term Period," in *Science and Technologies for Smart Cities*, Cham, 2021, pp. 45–63. doi: 10.1007/978-3-030-76063-2_4.
[135] F. Miao *et al.*, "A Novel Continuous Blood Pressure Estimation Approach Based on Data Mining Techniques," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 6, pp. 1730–1740, Nov. 2017, doi: 10.1109/JBHI.2017.2691715.
[136] S. Baek, J. Jang, and S. Yoon, "End-to-End Blood Pressure Prediction via Fully Convolutional Networks," *IEEE Access*, vol. 7, pp. 185458–185468, 2019, doi: 10.1109/ACCESS.2019.2960844.
[137] Md. S. Tanveer and Md. K. Hasan, "Cuffless blood pressure estimation from electrocardiogram and photoplethysmogram using waveform based ANN-LSTM network," *Biomed. Signal Process. Control*, vol. 51, pp. 382–392, May 2019, doi: 10.1016/j.bspc.2019.02.028.
[138] S. Shimazaki, S. Bhuiyan, H. Kawanaka, and K. Oguri, "Features Extraction for Cuffless Blood Pressure Estimation by Autoencoder from Photoplethysmography," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2018, pp. 2857–2860. doi: 10.1109/EMBC.2018.8512829.
[139] C. Yan *et al.*, "Novel Deep Convolutional Neural Network for Cuff-less Blood Pressure Measurement Using ECG and PPG Signals," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2019, pp. 1917–1920. doi: 10.1109/EMBC.2019.8857108.
[140] G. Slapničar, N. Mlakar, and M. Luštrek, "Blood Pressure Estimation from Photoplethysmogram Using a Spectro-Temporal Deep Neural Network," *Sensors*, vol. 19, no. 15, Art. no. 15, Jan. 2019, doi: 10.3390/s19153420.
[141] S. Shimazaki, H. Kawanaka, H. Ishikawa, K. Inoue, and K. Oguri, "Cuffless Blood Pressure Estimation from only the Waveform of Photoplethysmography using CNN," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2019, pp. 5042–5045. doi: 10.1109/EMBC.2019.8856706.
[142] M. Saeed *et al.*, "Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): A public-access intensive care unit database," *Crit. Care Med.*, vol. 39, no. 5, pp. 952–960, May 2011, doi: 10.1097/CCM.0b013e31820a92c6.
[143] D. Liu, M. Görges, and S. A. Jenkins, "University of Queensland Vital Signs Dataset: Development of an Accessible Repository of Anesthesia Patient Monitoring Data for Research," *Anesth. Analg.*, vol. 114, no. 3, pp. 584–589, Mar. 2012, doi: 10.1213/ANE.0b013e318241f7c0.

[144] H. Eom *et al.*, "End-To-End Deep Learning Architecture for Continuous Blood Pressure Estimation Using Attention Mechanism," *Sensors*, vol. 20, no. 8, Art. no. 8, Jan. 2020, doi: 10.3390/s20082338.
[145] M. Panwar, A. Gautam, D. Biswas, and A. Acharyya, "PP-Net: A Deep Learning Framework for PPG-Based Blood Pressure and Heart Rate Estimation," *IEEE Sens. J.*, vol. 20, no. 17, pp. 10000–10011, Sep. 2020, doi: 10.1109/JSEN.2020.2990864.
[146] V. G. Ganti, A. M. Carek, B. N. Nevius, J. A. Heller, M. Etemadi, and O. T. Inan, "Wearable Cuff-Less Blood Pressure Estimation at Home via Pulse Transit Time," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 6, pp. 1926–1937, Jun. 2021, doi: 10.1109/JBHI.2020.3021532.
[147] F. Heydari, M. P. Ebrahim, J.-M. Redoute, K. Joe, K. Walker, and M. Rasit Yuce, "A chest-based continuous cuffless blood pressure method: Estimation and evaluation using multiple body sensors," *Inf. Fusion*, vol. 54, pp. 119–127, Feb. 2020, doi: 10.1016/j.inffus.2019.07.001.
[148] F. Miao, Z.-D. Liu, J.-K. Liu, B. Wen, Q.-Y. He, and Y. Li, "Multi-Sensor Fusion Approach for Cuff-Less Blood Pressure Measurement," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 1, pp. 79–91, Jan. 2020, doi: 10.1109/JBHI.2019.2901724.
[149] Z.-D. Liu, J.-K. Liu, B. Wen, Q.-Y. He, Y. Li, and F. Miao, "Cuffless Blood Pressure Estimation Using Pressure Pulse Wave Signals," *Sensors*, vol. 18, no. 12, Art. no. 12, Dec. 2018, doi: 10.3390/s18124227.
[150] S. Julier and J. K. Uhlmann, "General Decentralized Data Fusion with Covariance Intersection," in *Handbook of Multisensor Data Fusion*, 2nd ed., CRC Press, 2009.
[151] M. Saeed *et al.*, "Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): A public-access intensive care unit database," *Crit. Care Med.*, vol. 39, no. 5, pp. 952–960, May 2011, doi: 10.1097/CCM.0b013e31820a92c6.
[152] Goldberger Ary L. *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000, doi: 10.1161/01.CIR.101.23.e215.
[153] "Electronic or Automated Sphygmomanometers. ANSI/AAMI Standard SP10-192, ed. Arlington, VA, USA, 1993."
[154] M. A. Cohen and J. A. Taylor, "Short-term cardiovascular oscillations in man: measuring and modelling the physiologies," *J. Physiol.*, vol. 542, no. Pt 3, pp. 669–683, Aug. 2002, doi: 10.1113/jphysiol.2002.017483.
[155] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuff-Less High-Accuracy Calibration-Free Blood Pressure Estimation Using Pulse Transit Time," in *2015 IEEE International Symposium on Circuits and Systems*, May 2015, pp. 1006–1009. doi: 10.1109/ISCAS.2015.7168806.
[156] E. Monte-Moreno, "Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques," *Artif. Intell. Med.*, vol. 53, no. 2, pp. 127–138, Oct. 2011, doi: 10.1016/j.artmed.2011.05.001.
[157] F. Y. Liang, S. Takagi, R. Himeno, and H. Liu, "Biomechanical Characterization of Ventricular-Arterial Coupling During Aging: A Multi-Scale Model Study," *J. Biomech.*, vol. 42, no. 6, pp. 692–704, Apr. 2009, doi: 10.1016/j.jbiomech.2009.01.010.
[158] A. Jadooei, O. Zaderykhin, and V. I. Shulgin, "Adaptive Algorithm for Continuous Monitoring of Blood Pressure Using a Pulse Transit Time," in *2013 IEEE XXXIII International Scientific Conference Electronics and Nanotechnology*, Apr. 2013, pp. 297–301. doi: 10.1109/ELNANO.2013.6552042.
[159] C. C. Y. Poon and Y. T. Zhang, "Cuff-Less and Noninvasive Measurements of Arterial Blood Pressure by Pulse Transit Time," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, Jan. 2005, pp. 5877–5880. doi: 10.1109/IEMBS.2005.1615827.

[160] Y. Choi, Q. Zhang, and S. Ko, "Noninvasive cuffless blood pressure estimation using pulse transit time and Hilbert–Huang transform," *Comput. Electr. Eng.*, vol. 39, no. 1, pp. 103–111, Jan. 2013, doi: 10.1016/j.compeleceng.2012.09.005.
[161] D. Špulák, R. Čmejla, and V. Fabián, "Parameters for mean blood pressure estimation based on electrocardiography and photoplethysmography," in *2011 International Conference on Applied Electronics*, Sep. 2011, pp. 1–4.
[162] I. cheol Jeong, J. Wood, and J. Finkelstein, "Using individualized pulse transit time calibration to monitor blood pressure during exercise," *Stud. Health Technol. Inform.*, vol. 190, pp. 39–41, 2013.
[163] Y. Kurylyak, K. Barbe, F. Lamonaca, D. Grimaldi, and W. Van Moer, "Photoplethysmogram-Based Blood Pressure Evaluation Using Kalman Filtering and Neural Networks," in *2013 IEEE International Symposium on Medical Measurements and Applications*, Gatineau, QC, May 2013, pp. 170–174. doi: 10.1109/MeMeA.2013.6549729.
[164] P. Su, X. Ding, Y. Zhang, J. Liu, F. Miao, and N. Zhao, "Long-Term Blood Pressure Prediction with Deep Recurrent Neural Networks," in *2018 IEEE EMBS International Conference on Biomedical Health Informatics*, Mar. 2018, pp. 323–328. doi: 10.1109/BHI.2018.8333434.
[165] R. Villar, T. Beltrame, and R. L. Hughson, "Validation of the Hexoskin wearable vest during lying, sitting, standing, and walking activities," *Appl. Physiol. Nutr. Metab.*, vol. 40, no. 10, pp. 1019–1024, Jun. 2015, doi: 10.1139/apnm-2015-0140.
[166] C. Landry, S. D. Peterson, and A. Arami, "Nonlinear Dynamic Modeling of Blood Pressure Waveform: Towards an Accurate Cuffless Monitoring System," *IEEE Sens. J.*, vol. 20, no. 10, pp. 5368–5378, May 2020, doi: 10.1109/JSEN.2020.2967759.
[167] W. J. Verberk, "Design of Clinical Trials to Validate Cuffless Blood Pressure Monitors," in *The Handbook of Cuffless Blood Pressure Monitoring: A Practical Guide for Clinicians, Researchers, and Engineers*, J. Solà and R. Delgado-Gonzalo, Eds. Cham: Springer International Publishing, 2019, pp. 203–224. doi: 10.1007/978-3-030-24701-0_12.
[168] Hansen Tine W. *et al.*, "Prognostic Value of Reading-to-Reading Blood Pressure Variability Over 24 Hours in 8938 Subjects From 11 Populations," *Hypertension*, vol. 55, no. 4, pp. 1049–1057, Apr. 2010, doi: 10.1161/HYPERTENSIONAHA.109.140798.
[169] Y. Athavale and S. Krishnan, "Biosignal monitoring using wearables: Observations and opportunities," *Biomed. Signal Process. Control*, vol. 38, pp. 22–33, Sep. 2017, doi: 10.1016/j.bspc.2017.03.011.
[170] D. R. Seshadri, C. Drummond, J. Craker, J. R. Rowbottom, and J. E. Voos, "Wearable Devices for Sports: New Integrated Technologies Allow Coaches, Physicians, and Trainers to Better Understand the Physical Demands of Athletes in Real time," *IEEE Pulse*, vol. 8, no. 1, pp. 38–43, Jan. 2017, doi: 10.1109/MPUL.2016.2627240.
[171] G. Parati and H. Schumacher, "Blood pressure variability over 24 h: prognostic implications and treatment perspectives. An assessment using the smoothness index with telmisartan–amlodipine monotherapy and combination," *Hypertens. Res.*, vol. 37, no. 3, Art. no. 3, Mar. 2014, doi: 10.1038/hr.2013.145.
[172] Ristuccia Heather L., Grossman Paul, Watkins Lana L., and Lown Bernard, "Incremental Bias in Finapres Estimation of Baseline Blood Pressure Levels Over Time," *Hypertension*, vol. 29, no. 4, pp. 1039–1043, Apr. 1997, doi: 10.1161/01.HYP.29.4.1039.
[173] Imholz B P *et al.*, "Feasibility of ambulatory, continuous 24-hour finger arterial pressure recording.," *Hypertension*, vol. 21, no. 1, pp. 65–73, Jan. 1993, doi: 10.1161/01.HYP.21.1.65.
[174] J. G. Carney, P. Cunningham, and U. Bhagwan, "Confidence and prediction intervals for neural network ensembles," in *IJCNN'99. International Joint Conference on Neural Networks.*

Proceedings (Cat. No.99CH36339), Jul. 1999, vol. 2, pp. 1215–1218 vol.2. doi: 10.1109/IJCNN.1999.831133.
[175] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Process.*, vol. 99, pp. 215–249, Jun. 2014, doi: 10.1016/j.sigpro.2013.12.026.
[176] M. Moukari, L. Simon, S. Picard, and F. Jurie, "n-MeRCI: A new Metric to Evaluate the Correlation Between Predictive Uncertainty and True Error," *ArXiv190807253 Cs Stat*, Aug. 2019, Accessed: Apr. 09, 2021. [Online]. Available: http://arxiv.org/abs/1908.07253
[177] S. J. Julier and J. K. Uhlmann, "Using covariance intersection for SLAM," *Robot. Auton. Syst.*, vol. 55, no. 1, pp. 3–20, Jan. 2007, doi: 10.1016/j.robot.2006.06.011.
[178] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," 2007.
[179] A. Zarnani, S. Karimi, and P. Musilek, "Quantile Regression and Clustering Models of Prediction Intervals for Weather Forecasts: A Comparative Study," *Forecasting*, vol. 1, no. 1, Art. no. 1, Dec. 2019, doi: 10.3390/forecast1010012.
[180] D. DeMers and D. Wachs, "Physiology, Mean Arterial Pressure," in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2021. Accessed: Oct. 19, 2021. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK538226/
[181] R. Amelard, E. T. Hedge, and R. L. Hughson, "Temporal prediction of oxygen uptake dynamics from wearable sensors during low-, moderate-, and heavy-intensity exercise," *ArXiv210509987 Cs*, May 2021, Accessed: Oct. 13, 2021. [Online]. Available: http://arxiv.org/abs/2105.09987
[182] R. L. Hughson, D. A. Winter, A. E. Patla, G. D. Swanson, and L. A. Cuervo, "Investigation of VO2 kinetics in humans with pseudorandom binary sequence work rate change," *J. Appl. Physiol.*, vol. 68, no. 2, pp. 796–801, Feb. 1990, doi: 10.1152/jappl.1990.68.2.796.
[183] E. T. Hedge and R. L. Hughson, "Frequency domain analysis to extract dynamic response characteristics for oxygen uptake during transitions to moderate- and heavy-intensity exercises," *J. Appl. Physiol.*, vol. 129, no. 6, pp. 1422–1430, Dec. 2020, doi: 10.1152/japplphysiol.00503.2020.
[184] W. L. Beaver, K. Wasserman, and B. J. Whipp, "A new method for detecting anaerobic threshold by gas exchange," *J. Appl. Physiol.*, vol. 60, no. 6, pp. 2020–2027, Jun. 1986, doi: 10.1152/jappl.1986.60.6.2020.
[185] D. A. Keir, D. H. Paterson, J. M. Kowalchuk, and J. M. Murias, "Using ramp-incremental V̇O2 responses for constant-intensity exercise selection," *Appl. Physiol. Nutr. Metab.*, vol. 43, no. 9, pp. 882–892, Sep. 2018, doi: 10.1139/apnm-2017-0826.
[186] A. Esmaili, M. Kachuee, and M. Shabany, "Nonlinear Cuffless Blood Pressure Estimation of Healthy Subjects Using Pulse Transit Time and Arrival Time," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 12, pp. 3299–3308, Dec. 2017, doi: 10.1109/TIM.2017.2745081.
[187] B. Ibrahim and R. Jafari, "Cuffless Blood Pressure Monitoring from an Array of Wrist Bio-Impedance Sensors Using Subject-Specific Regression Models: Proof of Concept," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 6, pp. 1723–1735, Dec. 2019, doi: 10.1109/TBCAS.2019.2946661.
[188] Q. Liu, C. C. Y. Poon, and Y. T. Zhang, "Time–frequency analysis of variabilities of heart rate, systolic blood pressure and pulse transit time before and after exercise using the recursive autoregressive model," *Biomed. Signal Process. Control*, vol. 6, no. 4, pp. 364–369, Oct. 2011, doi: 10.1016/j.bspc.2011.03.009.
[189] P. Gizdulich, B. P. M. Imholz, A. H. van den Meiracker, G. Parati, and K. H. Wesseling, "Finapres tracking of systolic pressure and baroreflex sensitivity improved by waveform filtering," *J. Hypertens.*, vol. 14, no. 2, pp. 243–250, Feb. 1996.

[190] A. Faisal, K. R. Beavers, and R. L. Hughson, “O2 uptake and blood pressure regulation at the onset of exercise: interaction of circadian rhythm and priming exercise,” *Am. J. Physiol.-Heart Circ. Physiol.*, vol. 299, no. 6, pp. H1832–H1842, Dec. 2010, doi: 10.1152/ajpheart.00762.2010.
[191] J. R. MacDonald *et al.*, “Hypotension following mild bouts of resistance exercise and submaximal dynamic exercise,” *Eur. J. Appl. Physiol.*, vol. 79, no. 2, pp. 148–154, Jan. 1999, doi: 10.1007/s004210050488.
[192] P. D. Mannheimer, M. P. O’Neil, and E. Konecny, “The Influence of Larger Subcutaneous Blood Vessels on Pulse Oximetry,” *J. Clin. Monit. Comput.*, vol. 18, no. 3, pp. 179–188, Jun. 2004, doi: 10.1023/B:JOCM.0000042924.16052.5e.
[193] K. H. Shelley, D. Tamai, D. Jablonka, M. Gesquiere, R. G. Stout, and D. G. Silverman, “The Effect of Venous Pulsation on the Forehead Pulse Oximeter Wave Form as a Possible Source of Error in Spo2 Calculation,” *Anesth. Analg.*, vol. 100, no. 3, pp. 743–747, Mar. 2005, doi: 10.1213/01.ANE.0000145063.01043.4B.
[194] G. S. Agashe, J. Coakley, and P. D. Mannheimer, “Forehead Pulse Oximetry: Headband Use Helps Alleviate False Low Readings Likely Related to Venous Pulsation Artifact,” *Anesthesiology*, vol. 105, no. 6, pp. 1111–1116, Dec. 2006, doi: 10.1097/00000542-200612000-00010.
[195] A. B. Hertzman and L. W. Roth, “The absence of vasoconstrictor reflexes in the forehead circulation. effects of cold,” *Am. J. Physiol.-Leg. Content*, vol. 136, no. 4, pp. 692–697, Jun. 1942, doi: 10.1152/ajplegacy.1942.136.4.692.
[196] L. Schallom, C. Sona, M. McSweeney, and J. Mazuski, “Comparison of forehead and digit oximetry in surgical/trauma patients at risk for decreased peripheral perfusion,” *Heart Lung*, vol. 36, no. 3, pp. 188–194, May 2007, doi: 10.1016/j.hrtlng.2006.07.007.
[197] J. Allen and A. Murray, “Similarity in bilateral photoplethysmographic peripheral pulse wave characteristics at the ears, thumbs and toes,” *Physiol. Meas.*, vol. 21, no. 3, pp. 369–377, Aug. 2000, doi: 10.1088/0967-3334/21/3/303.
[198] Y. Yao *et al.*, “Unobtrusive Estimation of Cardiovascular Parameters with Limb Ballistocardiography,” *Sensors*, vol. 19, no. 13, Art. no. 13, Jan. 2019, doi: 10.3390/s19132922.
[199] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, “Definitions, methods, and applications in interpretable machine learning,” *Proc. Natl. Acad. Sci.*, vol. 116, no. 44, pp. 22071–22080, Oct. 2019, doi: 10.1073/pnas.1900654116.
[200] Y. Lu, I. Cohen, X. S. Zhou, and Q. Tian, “Feature selection using principal feature analysis,” in *Proceedings of the 15th ACM international conference on Multimedia*, New York, NY, USA, Sep. 2007, pp. 301–304. doi: 10.1145/1291233.1291297.
[201] D. Zhang, R. Wang, and B. Lo, “Surgical Gesture Recognition Based on Bidirectional Multi-Layer Independently RNN with Explainable Spatial Feature Extraction,” *ArXiv210500460 Cs*, May 2021, Accessed: Oct. 19, 2021. [Online]. Available: http://arxiv.org/abs/2105.00460

Appendices

Appendix A[7] - Subjects from MIMIC II Database

Table A.1 details of the NARX model performance on the 15 subjects from MIMIC II database used in Chapter 3.

Table A.1: List of the MIMIC II Clinical Database Public Subset (mimic2cdb) subjects used in the paper. The initial and final time included A, B, and C sections of the data. The absolute error on the waveform estimation is given for both training modes and all inputs.

Subject Number	Absolute Error for Interval Training (mean ± SD mmHg)			Absolute Error for Predictive mode training (mean ± SD mmHg)			Initial Time	Final Time
	ANN_{ECG}	ANN_{PPG}	$ANN_{ECG+PPG}$	ANN_{ECG}	ANN_{PPG}	$ANN_{ECG+PPG}$		
3000397m	2.8±3.2	3.0±3.4	2.5±3.2	5.8±7.4	3.9±4.2	4.5±4.6	20:17:00	1:45:00
3000480m	3.7±3.4	3.3±3.5	2.6±2.5	6.3±5.7	6.9±5.7	5.2±4.6	19:33:42	20:48:22
3000860m	2.6±2.3	3.9±3.3	2.4±2.2	8.0±4.9	9.5±5.0	8.7±5.0	14:50:47	16:01:27
3000989m	1.8±1.7	2.1±2.3	1.5±1.4	2.1±2.0	2.2±2.4	1.6±1.5	22:59:54	23:42:34
3001689m	3.6±4.0	3.7±4.5	2.9±3.2	5.1±5.1	4.1±5.0	3.6±3.9	6:25:03	8:34:23
3002090m	3.0±3.7	4.2±5.3	2.7±3.2	3.9±4.3	5.9±6.0	3.6±3.8	8:02:56	10:12:16
3900975m	3.4±2.9	4.5±4.9	3.1±2.6	4.9±3.6	5.4±5.1	4.8±3.6	12:05:11	14:26:31
3901254m	2.4±2.0	2.8±2.8	2.3±2.0	2.2±2.2	3.1±3.0	2.5±2.4	23:48:30	2:23:10
3901339m	2.8±2.9	3.2±3.6	2.4±2.4	3.3±3.7	4.1±5.7	2.6±2.7	14:06:44	16:24:04
3902929m	4.3±4.5	5.7±6.8	4.1±4.5	10.1±7.3	9.6±8.6	10.2±6.8	19:31:03	2:04:23
3902994m	4.2±3.9	4.3±4.8	3.7±3.8	5.3±4.5	5.4±6.2	5.5±5.3	16:48:44	21:23:24
3903771m	2.0±1.5	2.2±2.3	1.9±1.4	2.7±1.8	3.4±2.4	3.1±1.9	2:20:29	5:08:29
3903773m	3.0±3.9	3.5±3.6	2.5±2.5	4.7±5.3	4.8±4.4	4.1±3.8	1:27:10	3:37:50
3903867m	3.4±2.7	2.5±2.5	2.2±2.5	2.6±3.0	2.3±3.1	2.1±2.2	22:13:07	0:47:47
3904368m	3.9±3.8	3.4±3.8	3.4±3.7	5.9±4.7	5.3±4.1	6.7±4.6	8:23:05	9:39:05

[7] Parts of Appendix A have been adapted from the following publication:

C. Landry, S. D. Peterson, and A. Arami, "Nonlinear Dynamic Modeling of Blood Pressure Waveform: Towards an Accurate Cuffless Monitoring System," *IEEE Sens. J.*, vol. 20, no. 10, pp. 5368–5378, May 2020, doi: 10.1109/JSEN.2020.2967759.

Appendix B[8] - Fusion Results for Diastolic Blood Pressure

Chapter 5 presents the results for SBP. Herein are supplemented the results with analogous findings for diastolic blood pressure (DBP).

The training and testing clusters' error standard deviation (σ_{Err}) for the NARX, ANN_{Dias}, and PAT_{Log} models are shown in Figure B.1 as a function of the clusters' percentage of outliers (lower bound). Note that the training σ_{Err} at 0% of outliers is not defined, as there are no training data in this cluster. In the test data, those are outliers and represent 3.1±1.3%, 4.3±1.7%, and 4.8±3.4% of the BP estimates for the NARX, ANN_{Dias}, and PAT_{Log} models, respectively. The Pearson correlation coefficients between the training and testing σ_{Err} of each cluster are r = 0.80±0.24, 0.72±0.23, and 0.78±0.11 for the NARX, ANN_{Dias}, and PAT_{Log} models, respectively. It can be observed in Figure that the model overestimates on average the σ_{Err} of the test data by 43%, 13%, and 57% for NARX, ANN_{Dias}, and PAT_{Log}, respectively.

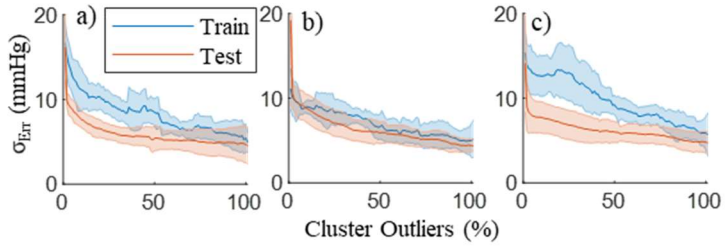

Figure B.1 - The diastolic blood pressure training and testing data standard deviation of error in every cluster for a) NARX, b) ANN_{Dias}, and c) PAT_{Log}. Lines represent the mean of all subjects and the shaded region show ± SD.

The mean and standard deviation (SD) of the MAE, mean error (μ_{Err}), σ_{Err}, r, MeRCI score, and mean of the prediction interval (PI) during activities of daily living computed across the five subjects are shown in Figure B.2.

[8] Parts of Appendix B have been adapted from the following submitted article:

C. Landry, S. D. Peterson, A. Arami, "A Fusion Approach to Improve Cuffless Blood Pressure and Prediction Interval Estimation," submitted to *IEEE Journal of Biomedical and Health Informatics* (August).

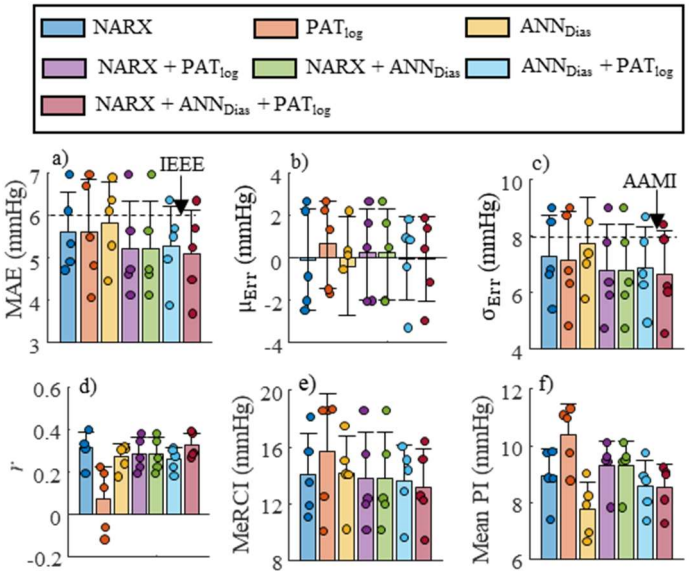

Figure B.2 - Comparison of a) mean absolute error, b) mean error, c) standard deviation of the error, d) Pearson correlation coefficient between different model estimates and the BP measurements, e) MeRCI score, and f) the mean estimated prediction interval. Bars represent the mean of all subjects and the error bars show ± SD. Each data point represents one participant. The IEEE cuffless wearable and the AAMI standard limits are also shown in subplots a), and b-c), respectively.

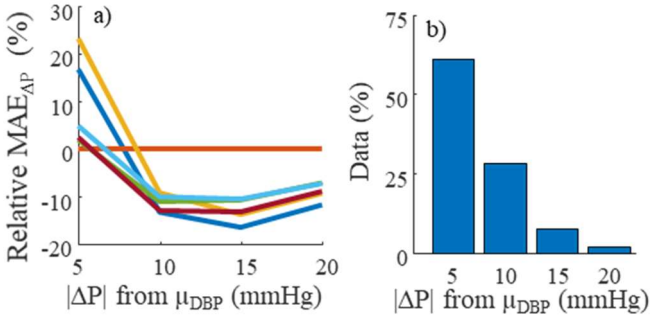

Figure B.3 - a) Mean of the DBP $MAE_{\Delta P}$ relative to $MAE_{\Delta P}$ – PAT_{Log} for all participants, and b) the distribution of data of all participants aggregated against the distribution of $|\Delta P|$ from μ_{SBP} measured throughout the day. The $MAE_{\Delta P}$ are binned in increments of 5 mmHg from -20 mmHg to 20 mmHg then grouped together according to their absolute value.

Figure B.3.a depicts the difference in $MAE_{\Delta P}$ of each bin with analogous bins from PAT_{Log} was computed as a function of $|\Delta P|$; the data distributions when aggregating the data of all participants are

included in Figure B.3.b. The range of DBP shown is the minimum range spanned by all the participants.

The percentage of data removed is plotted against the absolute estimation error in Figure B.4.a for the NARX + ANN_{Dias} + PAT_{Log} model. The tradeoff between amount of data kept and σ_{Err} when varying threshold values σ_T for NARX + ANN_{Dias} + PAT_{Log} is shown in Figure B.4.b.

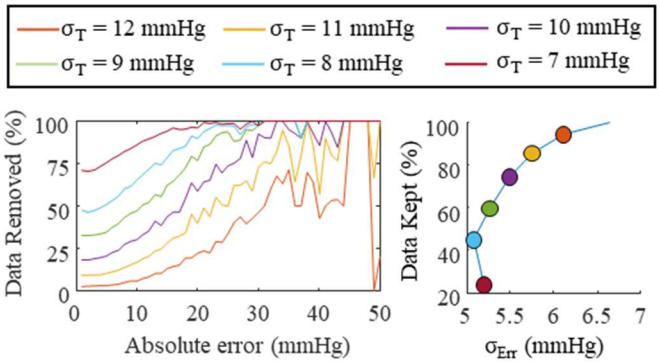

Figure B.4 - a) Percentage of the data removed as a function of the absolute error using different thresholds σ_T. The results are calculated from all the test data from every participant and grouped in increments of 1 mmHg. b) Percentage of data kept as a function of the error SD when varying the threshold on the PI. The line was computed by varying the σ_T, where specific values are marked with circles according to the legend.

The $MAE_{\Delta P}$ as a function of |ΔP| is plotted in Figure B.5.a for the NARX + ANN_{Dias} + PAT_{Log} model using σ_T = 8 and 9 mmHg. The percentage of data removed in each |ΔP| bin is plotted in Figure B.5.b. All the expected $MAE_{\Delta P}$ values are lower when using the OCSVM model PI to remove data with large expected error SD (PI ≥8 or 9 mmHg). In every case, the BP estimations at large |ΔP| by the fused model have a higher expected error than BP estimations closer to the mean BP.

The error SD for the NARX + ANN_{Dias} + PAT_{Log} model is shown in Fig S6.a when estimating BP during sitting, standing, and walking. The σ_{Err} is shown for original beat-by-beat estimation and for the SD threshold σ_T = 8 mmHg case. The percentage of data kept for each activity when using σ_T = 8 mmHg is shown in Figure B.6.b.

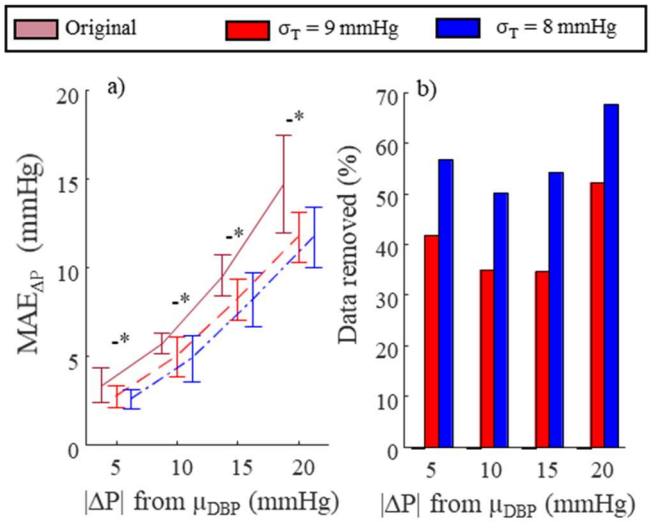

Figure B.5 - a) Mean and SD of $MAE_{\Delta P}$ for all participants, and b) percentage of data removed from each bin, against the $|\Delta P|$ from μ_{DBP} measured throughout the day. In a) bars represent the mean of all subjects and the error bars show ± SD. Legend: (*) Original vs σ_T = 9 mmHg, (-) Original vs σ_T = 8 mmHg, and (+) σ_T = 8 mmHg vs σ_T = 9 mmHg ($p<0.05/3$)

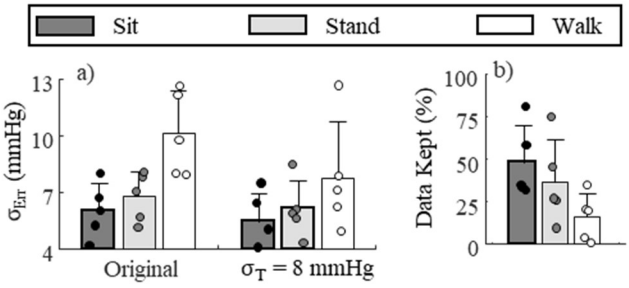

Figure B.6 - a) Comparison of the σ_{Err} for the original NARX + ANN_{Dias} + PAT_{Log} model and the same model using σ_T = 8 mmHg during sitting, standing, and walking, and b) the percentage of data kept (PI < 8 mmHg). Bars represent the mean of all subjects and the error bars show ± SD. Each data point represents one participant.

Appendix C – Results for Low-to-Moderate Cycling Exercise Test

Chapter 6 presents the results for the L-H test. Herein the analogous results for the L-M test are presented. In this section, one participant was removed from the analysis (out of 12) due to a poor PPG signal biasing the results. The PPG signal was found to have small signal to noise ratio, making it difficult to extract the PPG peak.

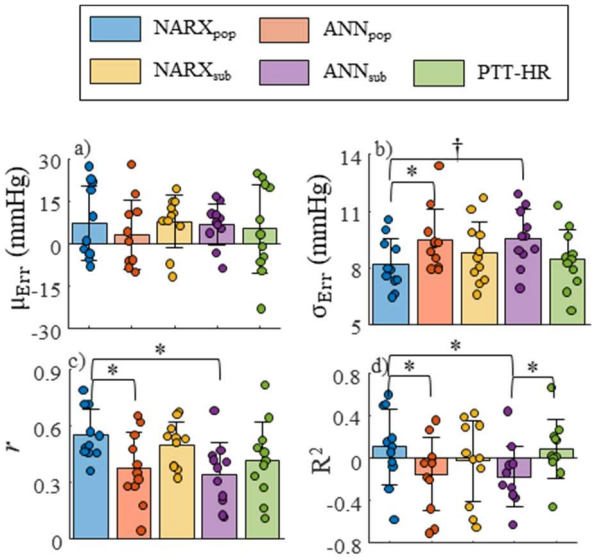

Figure C.1 - Comparison of a) mean error, b) SD of the error, c) Pearson correlation coefficient and d) coefficient of determination between different model estimates and the BP measurements. Bars represent the mean of all subjects and the error bars show ± SD for the L-M test. Each data point represents one participant. Legend: Wilcoxon with (*) p<0.05/10 and (†) p<0.01.

The mean and SD of μ_{Err}, σ_{Err}, r, and R^2 during the L-M test computed across the 11 subjects are shown in Figure C.1 for the different models. The SD of μ_{Err} is relatively large, showing that all BP models can be largely biased for some participants. The results exhibit differences in mean σ_{Err} between the models, ranging from 8.19 mmHg for NARX$_{pop}$ (best) to 9.58 mmHg for ANN$_{sub}$ (worst). As a reference, the MAP SD across the 11 participants is 9.1±1.6 mmHg (mean±SD) during the L-M test. Due to the distribution of results, only the NARX$_{pop}$ σ_{Err} is found to be statistically lower than ANN$_{pop}$ ($p<0.05/10$), and trends toward a statistical difference with ANN$_{sub}$ ($p<0.01$). The average Pearson correlation coefficient between the estimated and measured MAP is higher for NARX$_{pop}$ and statistically different from ANN$_{pop}$(ANN$_{sub}$) ($p<0.05/10$). Similarly, R^2 is the largest for both NARX$_{pop}$

and statistically different from ANN_{pop}(ANN_{sub}) ($p<0.05/10$) and PAT-HR is statistically different from ANN_{sub} ($p<0.01$).

The estimated BP time series from $NARX_{pop}$ and $NARX_{sub}$ are shown for the subjects with the lowest and highest σ_{Err} in Figure 6.3. In Figure 6.3.a-b, the estimated BP waveforms are shown at rest (2 min), during the first PRBS moderate intensity (90%VT) section (9 min), during the low intensity (25 W) section between the two PRBS (14 min), and at the last moderate intensity (90%VT) section (19.5 min). In Figure 6.3.c-d, MAP is shown, as a function of time, for the complete 21 minutes, including the initial resting portion and the complete L-M exercise test.

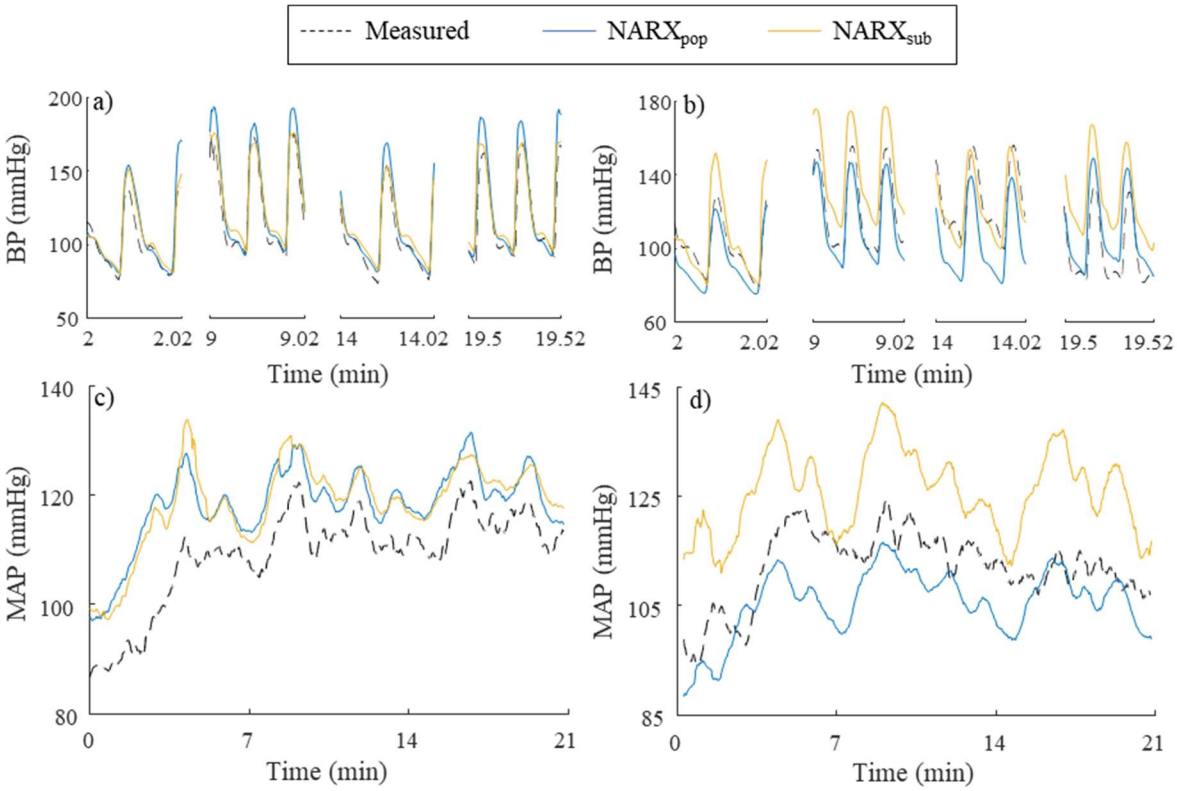

Figure C.2 - Comparison of actual versus estimated BP waveforms and MAP (after applying a moving average). The results are shown for the subject with the lowest, a) and c), and highest, b) and d), SD of the error for the L-M test.

The $NARX_{pop}$ error (with removed bias) evolution over the 21-minute testing period is presented for all 11 participants in Figure 6.4. The red line is the mean of all participants. It can be observed that the model in the majority of subjects underestimates the change in MAP up to the end of the first PRBS

(13 min) and error slowly increases throughout the test, which is consistent with qualitative observations in Figure 6.3.

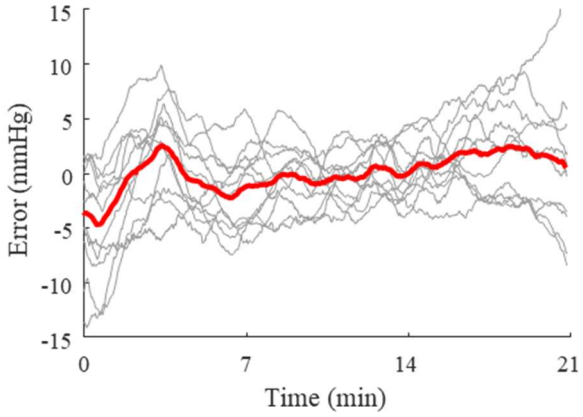

Figure C.3 - Filtered BP error for the 11 subjects for the L-M cycling test. Grey lines represent individual subject data and the red line is the mean of all subjects.

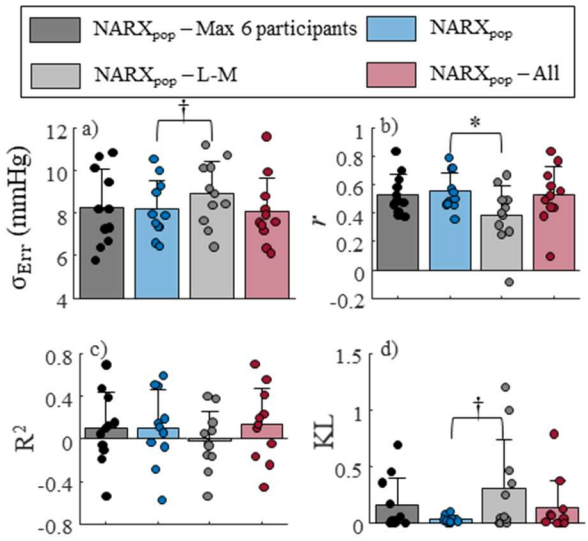

Figure C.4 - Comparison of a) SD of the error, b) Pearson correlation coefficient, c) coefficient of determination, and d) KL divergence between different model estimates and the BP measurements distributions. Bars represent the mean of all subjects and the error bars show ± SD for the L-M test. Each data point represents one participant. Legend: Wilcoxon with (*) p<0.05/6 and (†) p<0.05.

The means and SDs of σ_{Err}, r, R^2, and KL-divergence during the L-M test computed across the 11 subjects are shown in Figure 6.5 for $NARX_{pop}$ trained on different data . In general, training the model on only the L-M test ($NARX_{pop}$ – L-M) leads to poorer outcomes, with statistical difference observed between r of $NARX_{pop}$. Training on the Max tests ($NARX_{pop}$) resulted in higher r, and significantly lower KL-divergence than training the model on L-M tests. On average, training on all tests ($NARX_{pop}$ – All) leads to larger R^2 and smaller σ_{Err} compared to the other models.

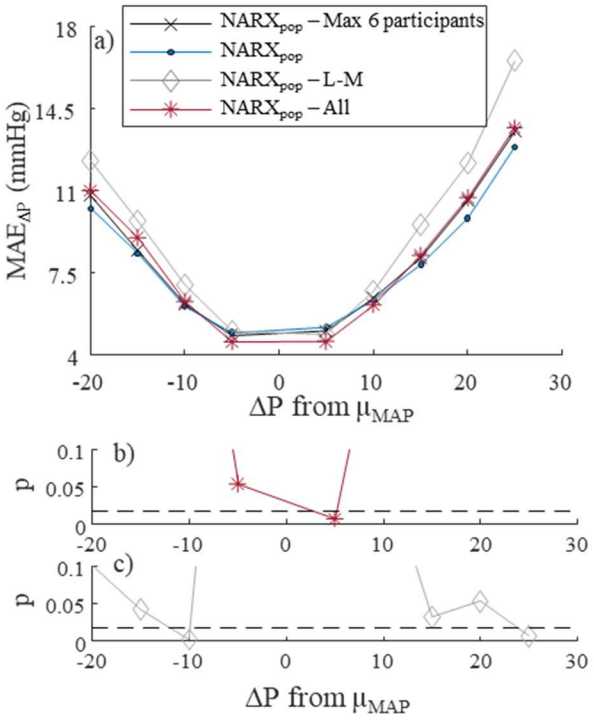

Figure C.5 - a) Mean of $MAE_{\Delta P}$ for all participants against the distribution of ΔMAP from μMAP measured throughout the L-M test. Wilcoxon tests' p-value for b) $NARX_{pop}$ vs $NARX_{pop}$ – All, c) $NARX_{pop}$ vs $NARX_{pop}$ – L-M, and d) $NARX_{pop}$ vs $NARX_{pop}$ – Max 6 participants. Symbols represent the p-values (or $MAE_{\Delta P}$ value) and the location on the abscissa indicates the bin over which the statistical test was performed.

The capability of predicting changes in MAP by $NARX_{pop}$ is shown in Figure 6.6 for the different training datasets. As shown by the p-values obtained from the Wilcoxon tests, using the data of all the tests as training data leads to lower MAE in the small range of ΔMAP without affecting the model's capability to estimate BP at large ΔMAP. The model trained on the Max tests is significantly better at estimating large positive ΔMAP than the training only on L-M. There were no significant differences

between training on six or 11 participants Max tests when comparing every ΔMAP bin, so nothing appears on the figure.

The means and SDs of the σ_{Err}, r, R^2, $r_{err\text{-}PPG}$, and $r_{err\text{-}ECG}$ during the L-M test computed across the 11 subjects are shown in Figure 6.7 for $NARX_{pop}$, $NARX_{sub}$ – L-M (Train), $NARX_{sub}$ – All (Train), and $NARX_{corr}$. The results indicate that when a model can extract more information from the PPG and ECG signals (lower $r_{err\text{-}PPG}$ and $r_{err\text{-}ECG}$), it is better at estimating BP (lower σ_{Err} and higher r). It can also be observed that the employed NARX architectures are not able to extract all the information from the PPG and ECG even when the test data are included in the training set, though this helps the estimation for the moderate exercise case ($NARX_{sub}$ – All (Train)).

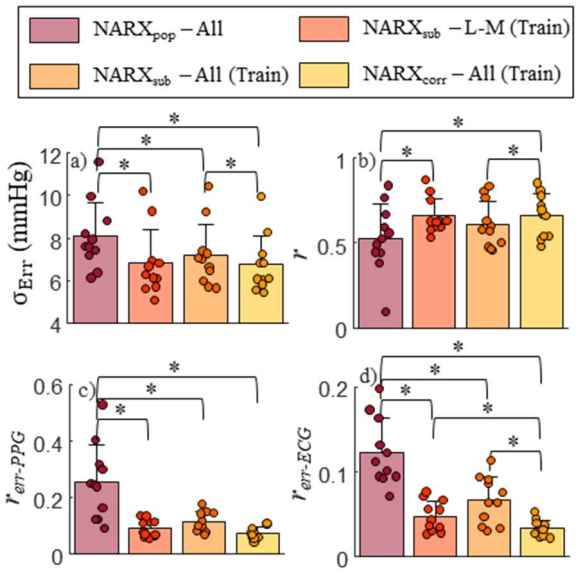

Figure C.6 - Comparison of a) SD of the error, and b) Pearson correlation coefficient between different model estimates and the BP measurements, and the maximum cross-correlation between c) BP and PPG, and d) BP and ECG signals. Bars represent the mean of all subjects and the error bars show ± SD for the L-M test. Each data point represents one participant. Legend: (*) Wilcoxon with p<0.05/6.

The mean of the 11 participants' ΔMAP during the L-M test is shown in Figure 6.8.a. It can be seen that the measured ΔMAP during the second PRBS is lower than the first one, but the difference is much smaller compared to the L-H test. Figure 6.8.b shows HR, PPG_{peak} and 1/PAT, which are all features that keep increasing throughout the test.

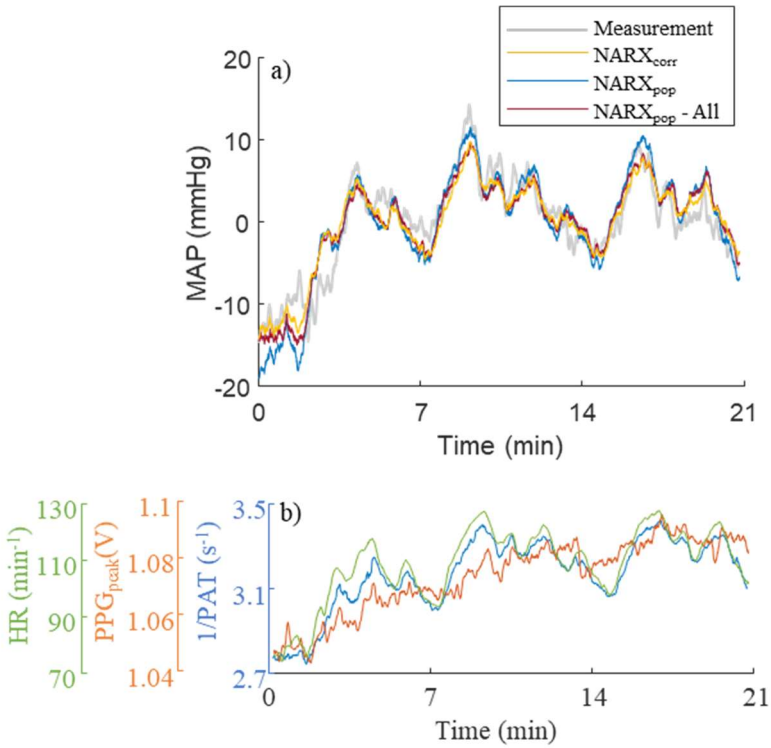

Figure C.7 – Mean of the 11 participants a) measured and estimated ΔMAP and b) common features that correlate with BP for the L-M exercise test.