ABSTRACT

Title of Thesis: STUDYING THE EFFICACY OF AND DEVELOPING DATA-DRIVEN REAL-TIME CLINICAL DECISION SUPPORT SYSTEMS FOR HYPOTENSION DETECTION

Bryce Yapps, Master of Science, 2016

Thesis Directed By: Assistant Professor Jin-Oh Hahn
Department of Mechanical Engineering

Critically ill patients admitted into intensive care units are prone to reoccurring episodes of sustained hypotension. Prolonged durations of hypotension are correlated to, and potentially cause, permanent body-wide damage to patients if not properly treated, which may result in death. Currently, typical care for the management of hypotension in the critically ill is reactive and delayed, perhaps due to clinical inertia. The purpose of this study is to describe the current problem that is faced in critical care through a retrospective analysis and introduce candidate models that may be used as clinical informatics systems for preemptive hypotension detection to aid clinicians and nurses providing care in the fast-paced clinical environment. The clinical performance of the models is quantified and the efficacy of implementation of these models is discussed.

STUDYING THE EFFICACY OF AND DEVELOPING DATA-DRIVEN REAL-TIME CLINICAL DECISION SUPPORT SYSTEMS FOR HYPOTENSION DETECTION

by

Bryce Yapps

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2016

Advisory Committee:
Professor Jin-Oh Hahn, Chair and Advisor
Professor Balakumar Balachandran
Professor Jeffrey Herrmann
Andrew T. Reisner, M.D.

Acknowledgements

First, I would like to thank my advisor, Dr. Jin-Oh Hahn. He presented me with the opportunity enter the world of research over two years ago. I greatly appreciate all of the guidance, instruction, and support he has provided to me during my time as a research assistant. I would like to wish him the very best with his future endeavors at the University of Maryland.

Additionally, I am very thankful for Dr. Andrew T. Reisner's contributions to this work. Without your clinical insight and helpfulness during the last two years, this research would not be where it is today.

Next, I would like to thank all of the members of the Laboratory for Control and Information Sciences. Being part of such a diverse and friendly group of research colleagues has been a great experience and I have truly enjoyed my time with all of you. Further, I would like to specifically thank Ramin Bighamian who has been a great research mentor to me while simultaneously working diligently on his Ph.D. He has been there every step of the way with my research and his insight and expertise has helped shape my research. I wish you the best of luck.

I am deeply appreciative of my family and friends for their constant love and support. I could not have gotten this far without any of you. Thank you, Mom, for always pushing me to strive for more and encouraging me the entire way. Dad, thank you for demonstrating that hard work and helping others always pays off and for continuously watching over me. I would also like to thank my two sisters, Kaylie and Brianna, for always being so supportive throughout my life. To Jenna, your love and support over the past five years has helped me overcome the many stresses during my

time in academia. And finally, to my friends that I have met at my time at the University of Maryland, thank you for making my time here great and I am excited for what the future holds for us all.

Finally, this research would have not been possible without the generous financial support from the Craig H. Neilsen Foundation (grant number: 297533), as well as the Graduate Research Fellowship (GRFP) awarded to me by the National Science Foundation (grant number: DGE 1322106).

Table of Contents

List of Tables

List of Figures

List of Equations

List of Abbreviations

Abbreviation	Definition
ABP	Arterial blood pressure
ACLS	Advanced Cardiac Life Support
AHE	Acute hypotensive episodes
AIC	Akaike Information Criterion
AR	Autoregressive
ARIMAX	Autoregressive moving average with exogenous input
AUC	Area under curve
CMO	Comfort measures only
CNS	Central nervous system
DAP	Diastolic arterial pressure
HR	Heart rate
ICU	Intensive care unit
MAP	Mean arterial pressure
MGH	Massachusetts General Hospital
MIMIC II	Multi-Parameter Intelligent Monitoring in Intensive Care II
ROC	Receiver operating characteristic
SAP	Systolic arterial pressure
SICU	Surgical intensive care unit
SQL	Structured query language

Chapter 1: Introduction

1.1: Background and Motivation

Circulatory shock is a common condition for critically ill patients; approximately one-third of patients in intensive care units (ICUs) are being treated for one form of shock [1]. Circulatory shock is the result of either severe sepsis, hypovolemia, cardiogenic factors, or obstruction [2]. Most patients treated for circulatory shock experience more than one of these mechanisms. Treating circulatory shock requires suitable hemodynamic (blood flow) support, in which a patient's blood pressure measurements are a typical metric used by clinicians to determine the status of the patient.

For patients being treated for circulatory shock, there are a variety of methods for supporting and healing the patient: ventilator support, fluid resuscitation, and vasoactive agents. Vasopressor therapy is administered when the patient's experiences persistent and severe hypotension (low blood pressure) and the primary therapies (i.e. ventilator support and fluid infusion) to not alleviate the critical state of the patient. Vasopressors are effective in raising the blood pressure of patients experiencing circulatory shock [3].

Current methods of treating severe hypotension in critically ill patients through vasopressor therapy requires clinicians to repetitively monitor the patient's blood pressure, determine if the current state of the patient requires an adjustment in vasopressor dose, and then control the proper adjustment of vasopressor infusion to raise or lower the patient's blood pressure accordingly. The first issue with this type

of treatment is that ICUs, in general, are short-staffed and the clinicians and nurses on duty have many parallel responsibilities, which may result in a constrained bandwidth for continuously monitoring the hemodynamic state of the patient. Secondly, the current treatment method is generally reactive, rather than proactive. This passive response to hypotension may reflect clinical inertia [4], a common phenomenon in which appropriate, and sometimes necessary, clinical responses are delayed. Lastly, with the technological advancements of autonomy and the increasing merger of health and data, it is surprising that treating persistent hypotension in patients experiencing circulatory shock is still primarily a reactive and manual treatment strategy.

More recently, there have been brief studies on data-driven models for clinical informatics systems that may aid in the detection and preventative treatment of severe hypotension [5]–[8], but these studies have been limited with respect to the patient sample size and generalizability. A generalizable approach to detect and assist in the preventative treatment of severe hypotension would be ideal. A clinical informatics system that minimizes the duration of hypotension experienced by the patient, via proactive detection, would vastly improve efficiency in ICUs for clinicians and nurses and improve the health and morbidity of the patient.

1.2: Research Goals

The objectives of this research are listed as the following:

1. Quantify current practices for the treatment of the critically ill patients experiencing prolonged hypotension following circulatory shock by investigating and analyzing the characteristics of episodes of sustained

hypotension of patients in the ICU and the clinical response and treatment to the episodes

2. Show that there is an observable lack of care in ICUs and bright to light the need for data-driven models for decision support in ICUs.
3. Develop and test various models (statistically-driven logistic regression and time series autoregressive forecasting models) that detect and alert for future episodes of sustained hypotension to show the implementation of clinical informatics systems having practical use working with clinicians towards the prevention of prolonged episodes of sustained hypotension
4. Benchmark various detection models against a simple estimate of adherence to existing treatment guidelines to illustrate the general improvement that these "proactive" models instill in patient care.
5. Validate the models on independent and separately sourced datasets to demonstrate the patient-independent generalizability of these models, and ultimately provide proof that population-trained models for hypotension detection would be valuable in the clinical setting.

To achieve these research goals laid out above, an initial retrospective analysis was performed to understand and quantify the existing problem. The next step was to develop and produce the proposed models with the goal of detecting hypotension in an anticipatory manner; with each model then benchmarked against the other model and the threshold detector. Then, a final discussion on the usefulness, limitations, and needed future work for implementing a clinical informatics system for decision support for treating and preventing hypotension.

1.3: Literature Review

1.3.1: Circulatory Shock and Treatment Methods

In a review article by Jean-Louis Vincent and Daniel de Backer, circulatory shock is defined as the failure of the body's circulatory system to provide adequate oxygen to the cells throughout the body [1]. The diagnosis of shock is determined by three main parameters: systemic arterial hypotension, tissue hypoperfusion, and hyperlactatemia. Arterial hypotension is typically when the patient's systolic arterial pressure (SAP) is less than 90 mmHg or their mean arterial pressure (MAP) is less than 70 mmHg. Hypoperfusion is the lack of proper blood flow to the capillaries in the tissues of the patient. Hyperlactatemia is when increased levels of lactate are present in the blood. Treatment goals of circulatory shock are to avoid further decreases in arterial pressure and cardiac output.

If the shock is severe, vasopressor agents are used to elevate arterial pressure in patients. Compared to SAP and diastolic arterial pressure (DAP), MAP is a better reflection of the patient's arterial pressure-head and is commonly monitored when administrating vasopressors. MAP thresholds of 60 – 65 mmHg have been commonly used as goal values for vasopressor infusion [3]. Common vasopressor drugs: dopamine, epinephrine, norepinephrine, and phenylephrine exhibit effective results in raising MAP in patients experiencing shock.

1.3.2: Advancements in Data-Driven Methods for Hypotension Detection

Management of severe hypotension of critically-ill patients experiencing shock is accomplished by manually monitoring a patient's MAP and performing ad

hoc adjustments of vasopressor infusions; clinical informatics systems for detecting the risk of hypotension in critically ill patients has have been sparsely studied.

The PhysioNet/Computers in Cardiology Challenge is an annual research competition where research groups are tasked with creating solutions to a prescribed problem using readily available clinical data gathered from a public research database. For the 10th Annual Challenge in 2009, the problem statement proposed to the involved research teams was the following: develop methods for identifying ICU patients at imminent risk of acute hypotensive episodes (AHE), with the goal of improving care and mortality rates [5] The following paragraph briefly describes various conference papers regarding the Challenge.

One team in this challenge developed a decision tree-based method of classifying patients by level of risk for AHE in the ensuing hour by extracting features of the patient's blood pressure waveform over the previous 12 hours [6]. Specific significant features were then put through a decision tree algorithm to determine the patient's risk. Additionally, a research team investigated the implementation of neural-network multi-models to forecast a patient's arterial blood pressure (ABP) to determine the risk of AHE [7]. Another method proposed a rule-based approach that utilized the prior 20 minutes of a patient's MAP signal to predict AHE according to a series of "yes/no" criteria [8]. All three of these models successfully detected the proper patients at risk for AHE. All of these methods presented in the Challenge demonstrate that data-driven clinical systems for detecting future hypotension are both possible and potentially beneficial in an ICU environment. Though, these results were limited to a very small subset of patients and

hypotensive episodes. Therefore, more investigation must go into developing a reliable system for detecting hypotension for critically ill patient care.

A team from Philips Healthcare (Philips Healthcare, Andover, MA) studied the use of features from a patient's vital signs to predict hemodynamic instability. The team utilized features from various continuous vital sign readings, including heart rate (HR), systolic arterial blood pressure (SAP), mean arterial blood pressure (MAP), and diastolic arterial blood pressure (DAP) to construct an instability index that could determine the increased likelihood of an adverse hemodynamic event [9]. The investigation illustrated how their selected features deviated between stable and unstable patients, as well as showed how the instability index provides early detection prior to a clinical intervention for unstable patients.

1.3.3: Logistic Regression Modeling

Logistic regression provides a simple framework for modeling binary classification given a set of feature parameters. For the case of hypotension detection, we want to determine the likelihood that hypotension will occur in the near future (i.e. Yes, or True) or if hypotension will not occur (i.e. No, or False); this presents itself as a simple binary classification problem. Therefore, logistic regression modeling is an obvious first step for our research purposes.

Regression techniques are a classical method for analyzing data when the goal of the analysis is to develop a relationship between an output variable, or response, and a set of independent, explanatory variables. Different scenarios or data types call for different formats of regression analysis; simple linear regression models are useful when the response and explanatory variables are continuous and linear over

sufficiently small regions [10]. Linear regression follows a simple format where the explanatory variables, x_i, can be linearly parametrized by a set of real coefficients, β_i.

$$\boldsymbol{Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon} \tag{1.3.3.1}$$

The expected value of the response variable is equated to the sum the intercept, β_0, the inner product of vectors $\boldsymbol{\beta}$ and $\boldsymbol{X}$, which consist of the regression coefficients and explanatory variables, respectively, and a residual term ε.

In the case where the response variable is dichotomous (or binary) in nature, a logistic regression model is more suitable [11]. For logistic regression, the response variable is transformed using the logistic function transform. In general, the logistic function transforms a continuous variable, $t \in [-\infty, \infty]$ into a value between 0 and 1, i.e. $\pi(t) \in [0,1]$. Typically, the transformed value is interpreted as a probability. The logistic function is defined as the following equation:

$$\boldsymbol{\pi(t) = \frac{e^t}{e^t + 1}} \tag{1.3.3.2}$$

We can utilize the inverse of the logistic function, sometimes called the logit, to transform a dichotomous response variable (i.e. 0 or 1) into a continuous response variable. As an example, we assume $t = \beta_0 + \beta_1 x_1$:

$$\begin{aligned} &\boldsymbol{t = \beta_0 + \beta_1 x_1} \\ &\boldsymbol{\rightarrow \pi(x) = \frac{e^{\beta_0 + \beta_1 x_1}}{e^{\beta_0 + \beta_1 x_1} + 1}} \\ &\boldsymbol{\rightarrow g\big(\pi(x)\big) = ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1} \end{aligned} \tag{1.3.3.3}$$

The logit, $g(\pi(x))$ is simply the function that transforms a dichotomous variable, $\pi(x)$, into a continuous variable, and then is regressed onto a set of explanatory variables.

1.3.4: Time Series Modeling

Time series modeling is implemented when a set of data either has no suitable set or an excessive set of dynamic equations to describe its behavior. Time series models are also useful for data forecasting, a major tool for the detection of future hypotension. Time series models are attractive for this research, as it can effectively model a system's dynamics and provide a framework for data forecasting.

Time series modeling and forecasting are being used in a wide spectrum of applications, including economic and business planning (i.e. stock price forecasting), production planning, inventory and production, and control and optimization of industrial processes [12]. Time series models take discrete past measurements of a desired output signal, $\{z_{k-1}, z_{k-2}, \dots, z_{k-n},\}$ and (if any) input signals, $\{u_{k-1}, u_{k-2}, \dots, u_{k-n},\}$ to construct a discrete transfer function model that predicts future values of the output signal for any lead time with accompanying probability limits. A common time series model is the autoregressive process model. The most generic form of the autoregressive process model is the Autoregressive Moving Average with Exogenous Input Model (ARIMAX):

$$\begin{aligned} &\boldsymbol{A(q)z(t) = q^{-n_k}B(q)u(t) + C(q)e(t)} \\ &\boldsymbol{\rightarrow z(t) = \frac{q^{-n_k}B(q)}{A(q)}u(t) + \frac{C(q)}{A(q)}e(t)} \end{aligned} \tag{1.3.4.1}$$

Where, q is the forward shift operator (i.e. q^{-1} is the backward shift operator), $A(q)$, $B(q)$, and $C(q)$ are vectors of coefficients, and $e(t)$ is a random white noise shock to the system. For the purposes of our research, we will implement a simpler version of the ARIMAX model, the Autoregressive Model (AR), which is a discrete transfer function of the current and past output signal.

Chapter 2: Acquisition and Pre-Processing of Clinical Data

The first task for this work was to acquire useful data for a retrospective analysis of patients and the clinical care provided to these patients in actual medical environments. Therefore, collecting numerics data (i.e. vital signs) from real patients admitted into ICUs and the corresponding clinician and nursing notes was fundamental for this research. Large amounts of clinical data are difficult to come by, since the processes for measuring, transferring, storing, and extracting are quite laborious. Fortunately, there exist a handful of intensive care unit research databases that contain numerics data and clinical notes that are available for public and academic use [13]–[17].

The Multi-Parameter Intelligent Monitoring in Intensive Care II (MIMIC II) Database was the most comprehensive and useful research database for the goals of this research. Other databases either had missing clinical notes [15], [17] or did not contain high-resolution numerics data [14], [16]. The MIMIC II Database contained both of the desired metrics for this research: minute-by-minute numerics data and clinical notes. The process of downloading the numerics data and clinical notes from MIMIC II is outlined in detail in Section 2.1.

Accompanying the data collected from MIMIC II, additional data was extracted from Surgical ICUs (SICUs) at Massachusetts General Hospital (MGH). Collecting and downloading the data from MGH utilized a different method than the MIMIC II download, and is outlined subsequently in Section 2.2.

Some elementary pre-processing was performed to remove undesired periods or abnormal data from the MIMIC II and MGH datasets. Pre-processing steps are reported in Section 2.3.

2.1: MIMIC II Database

2.1.1: MIMIC II Database Virtual Environment

When navigating the MIMIC II Database, certain tools are utilized to efficiently extract desired data. The goal of this subsection is to define the different tools used for the process of downloading data from MIMICII.

Firstly, there are multiple subsets of the MIMIC II Database that are accessed for data downloading. Within MIMIC II, there is the MIMIC II Clinical Database; the Clinical Database contains the notes and medical records of ICU patients (i.e. clinical notes). A complementary subset is the MIMIC II Waveform Database, which contain data of multiple types of signals for patients in the Clinical Database (i.e. waveforms and numerics). The waveform data are recordings of multiple physiologic signals and numerics data are time series of vital sign signals. For this research, only the numerics data are downloaded.

Secondly, to query the data in the Clinical Database, one makes use of the MIMIC II Explorer/Query Builder. The Explorer is a Structured Query Language (SQL) interface that is used to query and export data from the Clinical Database. The Explorer is restricted to exporting 1,000 rows of data per iteration. Therefore, the use of the MIMIC II Virtual Machine, a pre-made virtual Ubuntu installation, was more effective for our purposes. The Virtual Machine had no limit of data exporting.

Finally, the last tool is the PhysioBank ATM. The PhysioBank ATM organizes and compiles the data found in the Waveform Database [18].

2.1.2: MIMIC II Summary

The MIMIC II Clinical Database contains 25,328 ICU stays from the medical ICU, surgical ICU, cardiac recovery unit, and coronary care unit of the Beth Israel Hospital in Boston, MA [13]. All ICU stays were de-identified in accordance with the Health Insurance Portability and Accountability Act of 1996 [19]. Physiological data, such as waveform and numerics, were collected via patient monitors located at each ICU patient bed. Clinical notes were electronically recorded. With the assistance of the beside monitoring system vendor (Philips Healthcare, Andover, MA), these data were digitized, processed, converted, and uploaded to the MIMIC II Database [13].

2.1.3: Numerics Extraction

We limited our data search to patients who had data in both the Clinical and Waveform Databases; labeled as the matched subset. Patients in this subset had two types of files corresponding to their ICU stay: `.dat` files containing signals and `.hea` header files. The header files were dated in order to match up the timing of raw waveform and numerics data from the Waveform Database and the clinical notes from the Clinical Database. Each header file had the form `SUBJ_ID-YYYY-MM-DD-HH-mm`. An example of the structure of numerics header files is shown below in Figure 2.1.3.1.

```
s00292-3050-10-10-16-50n 14 0.0166666666667/125 14 16:50:47.120 10/10/3050
3478507n.dat 16 10/bpm 16 0 0 0 0 HR
3478507n.dat 16 10/mmHg 16 0 0 0 0 ABPSys
3478507n.dat 16 10/mmHg 16 0 0 0 0 ABPDias
3478507n.dat 16 10/mmHg 16 0 0 0 0 ABPMean
3478507n.dat 16 10/mmHg 16 0 0 0 0 PAPSys
3478507n.dat 16 10/mmHg 16 0 0 0 0 PAPDias
3478507n.dat 16 10/mmHg 16 0 0 0 0 PAPMean
3478507n.dat 16 10/mmHg 16 0 0 0 0 CVP
3478507n.dat 16 10/bpm 16 0 0 0 0 PULSE
3478507n.dat 16 10/pm 16 0 0 0 0 RESP
3478507n.dat 16 10/% 16 0 0 0 0 SpO2
3478507n.dat 16 1/mmHg 16 0 -32768 0 0 NBPSys
3478507n.dat 16 1/mmHg 16 0 -32768 0 0 NBPDias
3478507n.dat 16 1/mmHg 16 0 -32768 0 0 NBPMean
# <age>: 57 <sex>: F
```

Figure 2.1.3.1: Header file from MIMIC II

After gathering a list of all patients in the matched subset with available numerics data in the Waveform Database (stored as a `.txt` file), a list of all patients who received vasopressors (dopamine, epinephrine, norepinephrine, and phenylephrine) was found by running an SQL code in the Virtual machine:

```
SELECT DISTINCT subject_id
FROM mimic2v26.medevents
WHERE (itemid = '43' or itemid = '44' or itemid = '47' or itemid = '127')
```

Figure 2.1.3.2: SQL code for downloading list of patients who received vasopressors

Using both of these lists, a MATLAB routine was ran to download the numerics header files for all patients in the matched subset who also received vasopressors during their ICU stay. In all, 1,332 header files were downloaded for 620 patients (some patients had multiple ICU stays). Within the header files, different numerics signals are stored (e.g. Heart Rate (HR), SAP, DAP, and MAP. Any header files not containing a signal for MAP were removed. Following this first round of data removal, 867 header files from 411 patients remained.

2.1.4: Clinical Notes Extraction

Following the extraction of numerics header files, the next step was to download clinical notes from the Clinical Database corresponding to the 411 patients. There were three types of data related to the clinical notes that were downloaded: the dose, dose volume, and nursing notation. An SQL code was queried in the Virtual Machine to download these three data for each patient. The data were downloaded as `.xlsx` files and combined into single file for each patient using a simple MATLAB routine.

```
SELECT subject_id,itemid,charttime,dose
FROM mimic2v26.medevents
WHERE subject_id in (PASTE ENTIRE COMMA-SEPARATED SUBJECT LIST WITHIN
THESE PARENTHESIS)
AND(itemid = '43' or itemid = '44' or itemid = '47' or itemid = '127')
ORDER BY subject_id, itemid, charttime

SELECT a.poe_id, a.subject_id, start_dt,
stop_dt,drug_name,prod_strength,dose_val_disp,dose_unit_disp
FROM mimic2v26.poe_order a, mimic2v26.poe_med b
WHERE a.subject_id in (PASTE ENTIRE COMMA-SEPARATED SUBJECT LIST WITHIN
THESE PARENTHESIS)
AND a.poe_id = b.poe_id
ORDER BY subject_id, start_dt

SELECT subject_id, charttime, category, title, text
FROM mimic2v26.noteevents
WHERE subject_id in (PASTE ENTIRE COMMA-SEPARATED SUBJECT LIST WITHIN
THESE PARENTHESIS)
AND category in ('DISCHARGE_SUMMARY', 'Nursing/Other')
ORDER BY subject_id, charttime
```

Figure 2.1.4.1: SQL code for downloading clinical notes

Finally, with both numerics header files and clinical notes downloaded for each patient, we determined which patients had numerics data available within the timeframe of the clinical notes. If there were no numerics data within the timeframe, the patient was removed. There were 336 patients with potentially relevant data. Using the PhysioBank ATM, the numerics data for these patients were downloaded as

`.mat` files. Furthermore, the vasopressor dose data from the clinical notes were converted into `.mat` for convenience. Therefore, for each remaining patient, there were now aligned numerics and dose data in `.mat` format. Both types of data have a resolution of 1 measurement/recording per minute.

2.2: MGH Data

2.2.1: MGH Data Summary

The data collected from MGH was from patients admitted to the ICU and administered vasopressors during 2015 – 2016. Waveform and numerics data were collected via patient monitors located at each ICU patient bed. The data from these monitors were then transferred and stored in data servers as `.xml` files using the BedMasterEX system (Excel Medical, Jupiter, FL). Numerics data were sampled at a rate of 0.5 Hz. Clinical notes were hand-written by nurses, and then retroactively converted into electronic format (`.xlsx`). The files were anonymized in accordance with the Health Insurance Portability and Accountability Act of 1996 [19]. The process of extracting usable data form the MGH subset was simpler than the MIMIC II subset, since we preemptively selected patients who fit our inclusion criteria of receiving vasopressor infusion and had matching numerics data and clinical notes.

2.2.2: Numerics Data Extraction

There were 1,507 numerics files for 102 patients from the MGH BedMasterEX servers downloaded to be used for this research. Each numerics `.xml` file was converted to `.mat` format using a Python routine provided by MGH, found in Appendix A. Numerics files associated with the same patient were pieced together

to form a single numerics signal and saved as a single file. The numerics data were median down-sampled in order to have the identical one measurement per minute format as the data from MIMIC II.

2.2.3: Clinical Notes Extraction

Following the extraction of numerics data, the dose data from the clinical notes `.xlsx` files were extracted as `.mat` files using MATLAB. Furthermore, we determined which patients had numerics data available within the timeframe of the clinical notes. If there were no numerics data within the timeframe, the patient was removed. A MATLAB routine was implemented to automate this process. For each remaining patient, there were now aligned numerics and dose data in `.mat` format. Both types of data have a resolution of 1 measurement/recording per minute.

2.3: Data Pre-Processing

With both sets of data extracted from their respective sources, we further limited the scope of patients and data for research purposes.

As stated previously in Section 2.1 and 2.2, we included only patients with valid documentation of infusion of four vasopressor drugs: dopamine, epinephrine, norepinephrine (drug name: Levophed), and phenylephrine (drug name: Neosynephrine). Only patients who survived at least 48 hours after their discharge from the ICU were included. Patients who passed within 48 hours may have been receiving Comfort Measures Only (CMO). CMO refers to the medical treatment of a dying person where the natural dying process is not inhibited and the only goal is to ensure maximum comfort of the patient. We excluded these patients since this type of

care provided is not representative of how typical critically ill patients in an ICU would be treated.

We analyzed each individual patient record and excluded specific intervals of data for one of the six following reasons. First, any interval lacking either MAP or vasopressor dose data was removed. Intuitively, if not all of the desired data is available; there is no need to include it at all. Second, any interval containing non-physiological measurements was removed. For example, MAP measurements exceeding above 300 mmHg, below 5 mmHg, or large jumps in measurements were removed. Third, any interval during the first 30 minutes of a patient's ICU stay was excluded. The first 30 minutes of a patient's ICU stay sometimes involved the clinicians attempting to resuscitate the patient. Initial resuscitation practices may deviate from nominal ICU care to stabilize the patient; additionally behavior of a patient's MAP could be abnormal during the period of initial resuscitation. Therefore, this interval of data is excluded from analysis. Fourth, any interval exceeding maximum vasopressor infusion rates was removed. Per standard hospital protocol, the maximum vasopressor infusion rates are 20 mcg/kg/min, 1 mcg/kg/min, 28 mcg/min, and 290 mcg/min for dopamine, epinephrine, norepinephrine, and phenylephrine, respectively. Periods of super-maximum vasopressor infusion could lead to abnormal behavior of a patient's MAP, and is probable cause that the current treatment is not effective. Fifth, any interval containing infusions of two or more simultaneous vasopressor drugs was removed. For this research, we were only interested in the behavior of MAP during singular vasopressor infusion. Sixth, any interval where the previous recording of vasopressor dose was more than 60 minutes into the past was

excluded. We defined these occurrences as vasopressor dose recording inadequacies. After 60 minutes of no dosage recording in the clinical notes, it is uncertain if the patient is still receiving vasopressor infusion. With this uncertainty, we simply excluded all data associated with these occurrences. Examples of these exclusion criteria are illustrated in Figures 2.3.1 – 2.3.5.

The patient's MAP data was treated with a low-pass filter, in the form of a 5-minute median filter, to remove some of the measurement noise of the signal. Further, intervals of absent MAP data were interpolated if the interval is for less than 15 minutes. Once all of the proper patients were selected, the undesired intervals of data were removed, and the remaining data was filtered and interpolated, we were left with paired MAP and dose data during vasopressor mono-therapy at sub-maximum infusion rates.

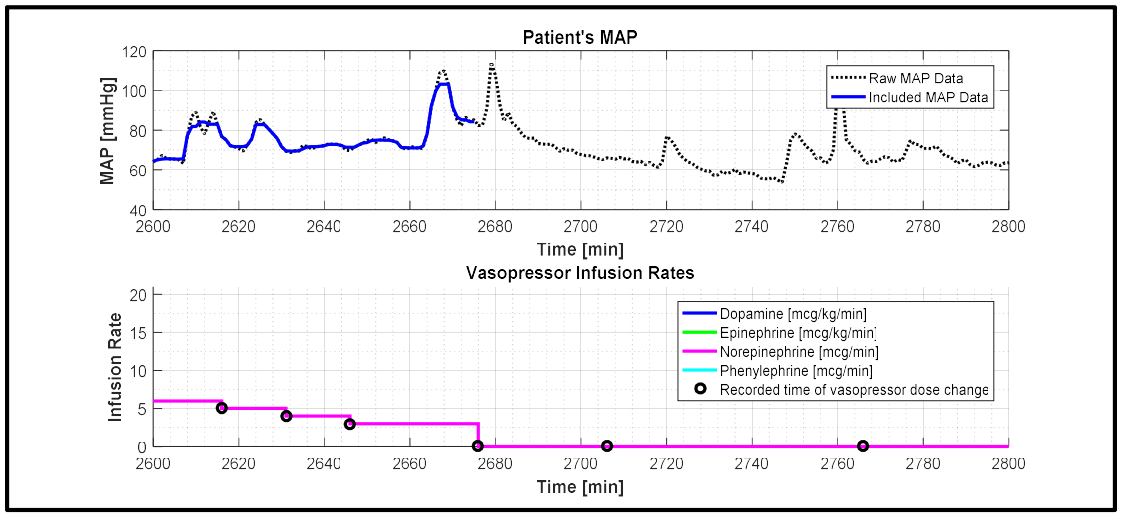

Figure 2.3.1: Zero vasopressor infusion rate example

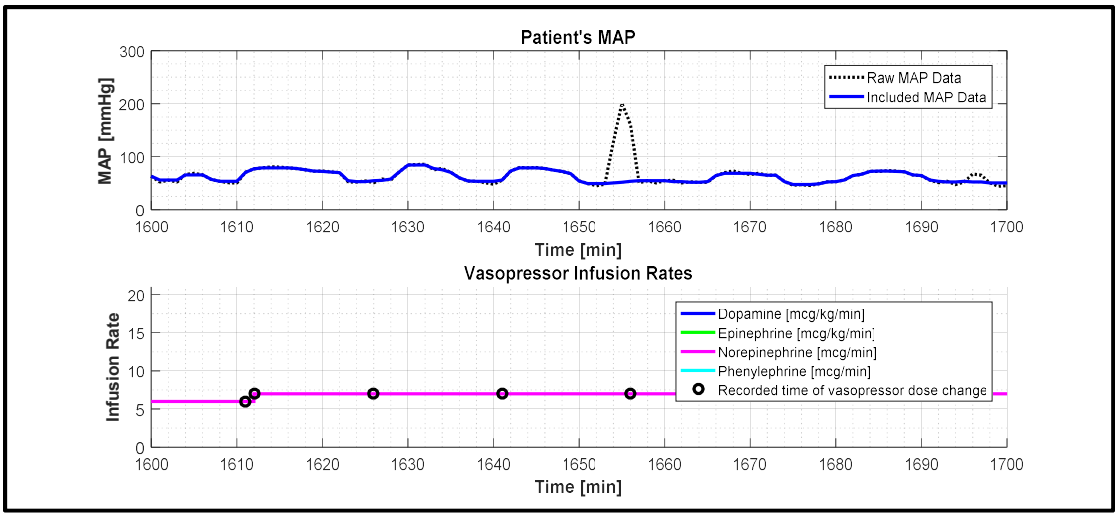

Figure 2.3.2: Non-physiological MAP and interpolation example

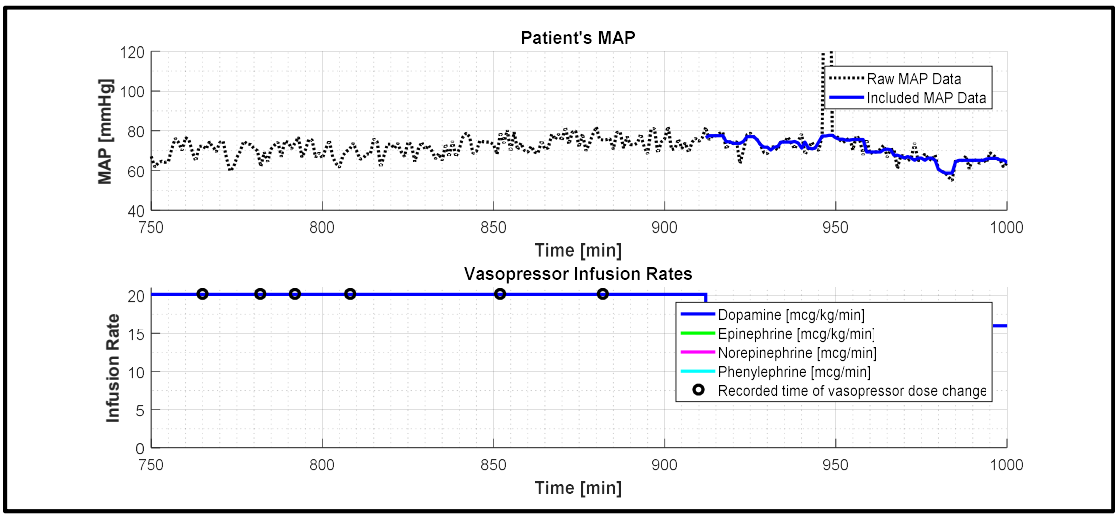

Figure 2.3.3: Maximum vasopressor infusion example

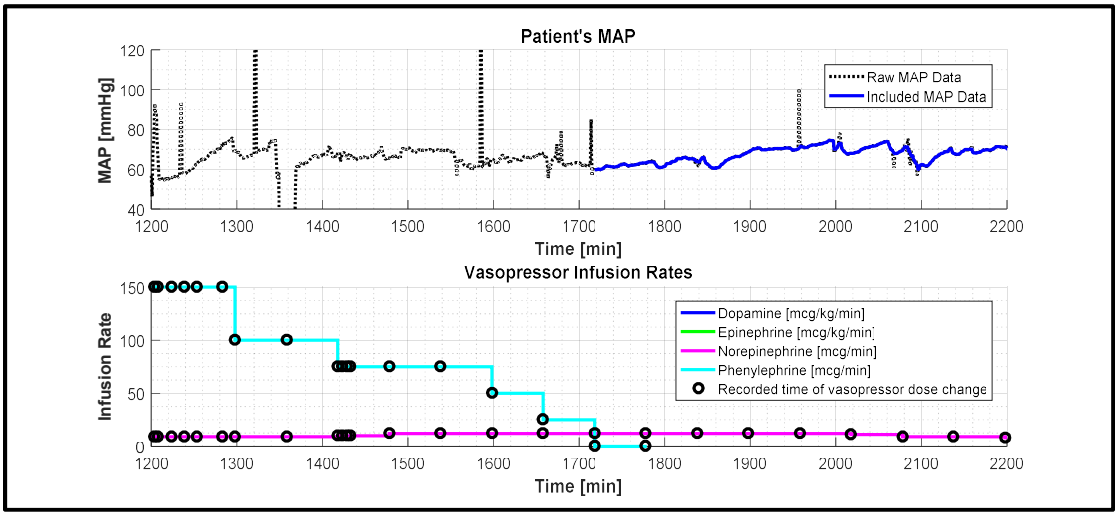

Figure 2.3.4: Multiple vasopressor infusion example

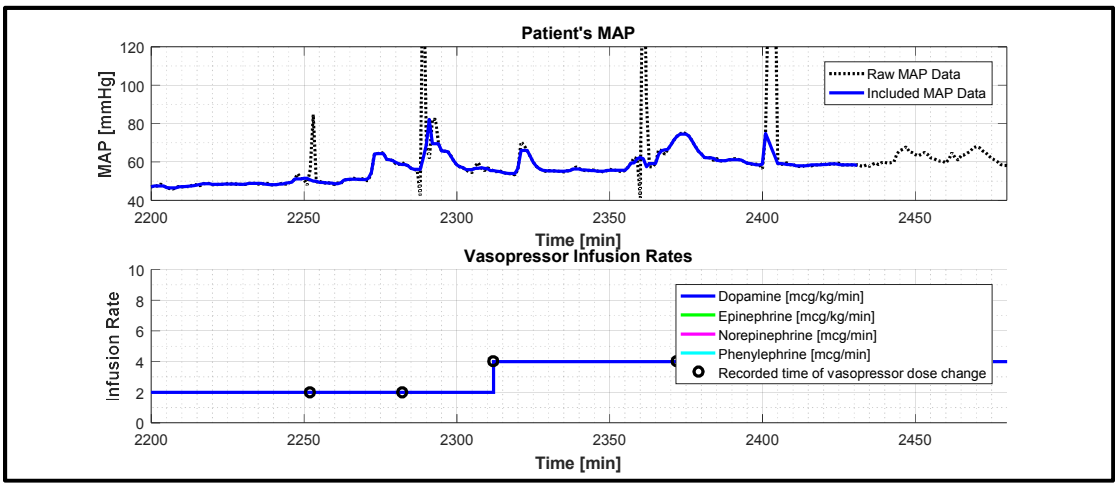

Figure 2.3.5: Vasopressor dose inadequacy example

Chapter 3: Retrospective Analysis of Critically Ill Patients Receiving Vasopressor Infusion

3.1: Background and Motivation

Prior to any engineering or mathematical development, our first research goal was to comprehensively understand the problem we were investigating. A retrospective analysis of the data we acquired from the two medical centers was an ideal way to understand the data we had and craft an eloquent problem statement we found worthy of investigating. The remainder of this chapter details the process of analyzing patient data from ICUs to quantify typical events experienced by patients and the care they received by clinicians and nurses.

Many patients treated for circulatory shock receive hours or days of vasopressor infusion to manage persistent hypotension. This practice is based in part on classical physiology studies showing that central nervous system (CNS) auto-regulation fails in healthy animals for MAP < 65 mmHg [20], [21], and once auto-regulation fails, CNS hypoperfusion and potential ischemia occur. Clinical reports have indeed observed an association between duration of hypotension and measures of end-organ injury in critically-ill [22]–[24] and intraoperative [25] patient populations; it is important to note that the true clinical consequences of hypotension remain unknown, because observational studies do not distinguish between hypotension as a direct cause of end-organ injury versus hypotension as a correlated indicator of disease severity.

It is also reasonable to conjecture that MAP < 65 mmHg could be a causal factor, via CNS ischemia, to the notable cognitive decline reported in some survivors of septic shock [26]. Based on such evidence, the use of vasopressors to prevent MAP < 65 mmHg for sepsis is a Class 1C recommendation according to the consensus Surviving Sepsis guidelines [27], while Advanced Cardiac Life Support (ACLS) advises vasopressor therapy for systolic blood pressure < 70 mmHg [28]. Rates of compliance with such guidelines are not well known. Accordingly, we undertook a pilot analysis studying whether vasopressors were effectively used to prevent hypotension.

3.2: Methods

3.2.1: Defining Regions of MAP

A patient's blood pressure can be within one of three zones: hypertension, in-range (often referred to as "normal"), and hypotension. Each zone has additional subzones to describe the state of a patient, but for our retrospective analysis, we simply defined these three zones.

Hypertension was defined as any interval of time when the patient's MAP was greater than 100 mmHg. Hypotension was defined as any interval of time when the patient's MAP less than 60 mmHg. The remaining intervals of time were defined as in-range (i.e. 60 mmHg ≤ MAP ≤ 100 mmHg). Additionally, we further defined when the patient's MAP was less than 60 mmHg for continuous intervals of 15 minutes or more, as an episode of sustained hypotension. All other intervals of hypotension were considered transient hypotension. Episodes of sustained hypotension began upon the first hypotensive MAP measurement and terminated with any subsequent in-range

MAP measurement. For this retrospective analysis, when the termination of an episode of sustained hypotension was followed within 30 minutes by the onset of another episode, both were combined into a single combined episode of sustained hypotension, called episode spans. Figure 3.2.1.1 illustrates the various regions of blood pressure.

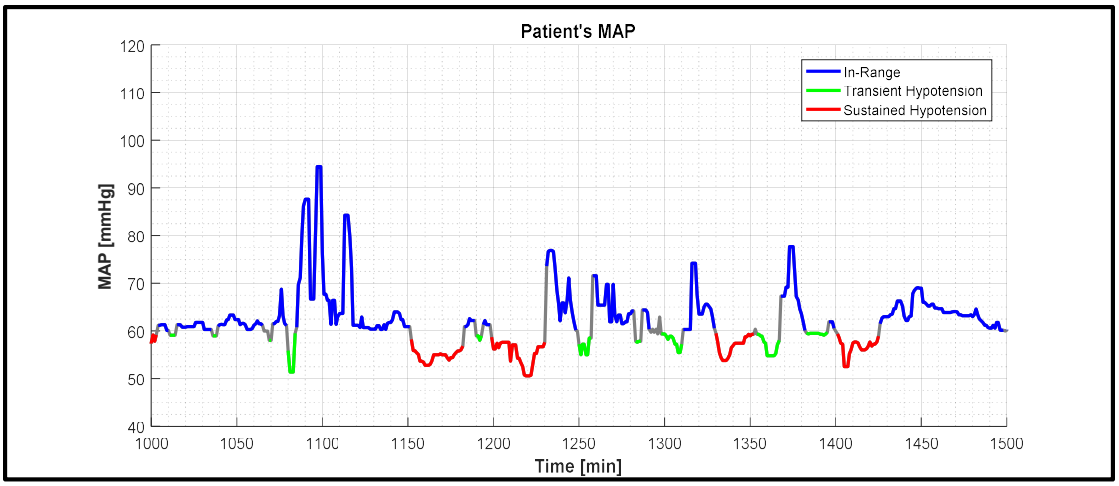

Figure 3.2.1.1: In-range, transient hypotension, and sustained hypotension

3.2.2: Defining Patient Populations

We examined all patients from MIMIC II and MGH with at least 30 minutes of vasopressor infusion. Focusing this retrospective analysis on the patients from MIMIC II, we broke the ICU stays of these patients down into three separate categories. The first category contained all ICU stays where the patient experienced no hypotension (this includes both transient and sustained hypotension). The second category contained all ICU stays where the patient experienced hypotension that was only transient. The final category contained all ICU stays where the patient experienced at least one episode of sustained hypotension. For each ICU stay category, demographic statistics were computed to generally understand any

significant differences between these populations. For each category, we computed the following statistics:

- Number of ICU stays
- Number of unique patients
- Age
- Gender
- Documented indication for vasopressor infusion
- Total duration of vasopressor infusion
- Time between vasopressor dose changes
- Median MAP measurements during vasopressor infusion
- Hourly standard deviation of MAP during vasopressor infusion
- Proportion of MAP measurements within the different MAP regions
- Episodes of sustained hypotension per 24 hours of vasopressor infusion

3.2.3: Analyzing Episodes of Sustained Hypotension

For the ICU stays with at least one episode of sustained hypotension, we further analyzed patterns related to any vasopressor dose changes prior to the onset of the episodes. We wanted to classify all episodes of sustained hypotension based on patterns of care, with respect to vasopressor infusion. We defined four types of episodes. Failed wean episodes were when the vasopressor dosage was decreased within 30 minutes before the onset of the episode. This type of episode was most likely due to a misdiagnosis, and therefore led to the improper weaning of vasopressors, causing the patient's MAP to fall below the 60 mmHg threshold. An example of a failed wean episode is shown in Figure 3.2.3.1. Failed re-dose episodes

were when the vasopressor dosage was increased within 30 minutes before the onset of the episode, and was followed by either no change in vasopressor dose or a decrease in vasopressor dose. Failed re-dose episodes occur when the clinician or nurse attempted to prevent hypotension, but did not sufficiently raise the vasopressor dose to accomplish that goal. An example of a failed re-dose episode is shown in Figure 3.2.3.2. Drift-out episodes occur when there were no vasopressor dose changes within 30 minutes before the onset of the episode. An example of a drift-out episode is shown in Figure 3.2.3.3. Finally, continuation episodes are episodes of sustained hypotension that follow within 30 minutes after the termination of a previous episode (as described in Section 2.2). An example of a continuation episode is shown in Figure 3.2.3.4. If there were missing MAP measurements within 30 minutes before the onset of the episode, the episode was not included in this part of the analysis.

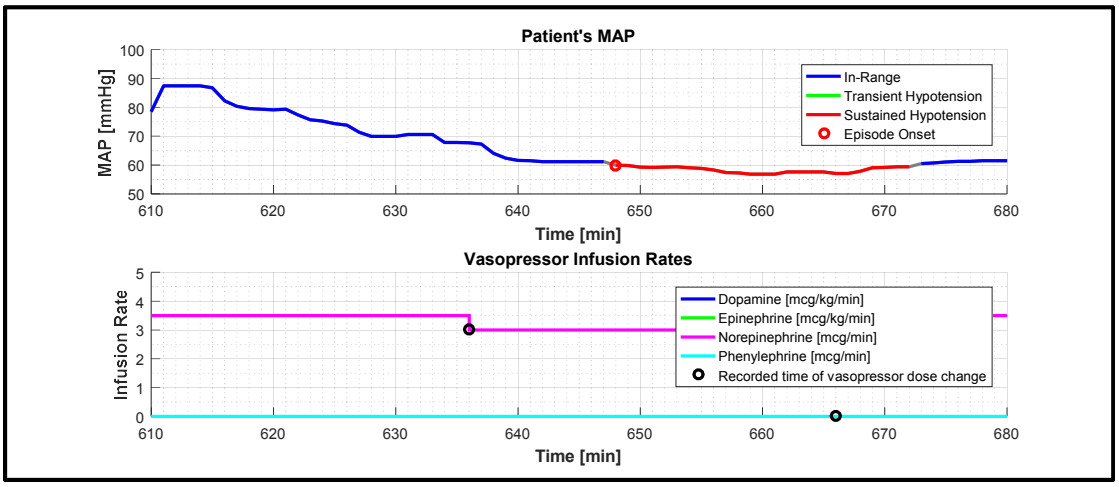

Figure 3.2.3.1: Failed-wean episode of sustained hypotension

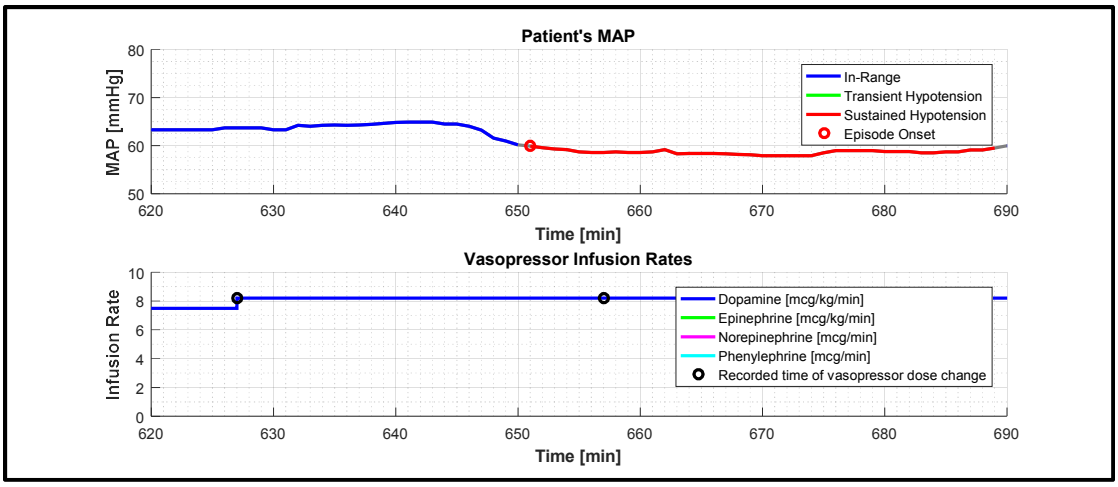

Figure 3.2.3.2: Failed re-dose episode of sustained hypotension

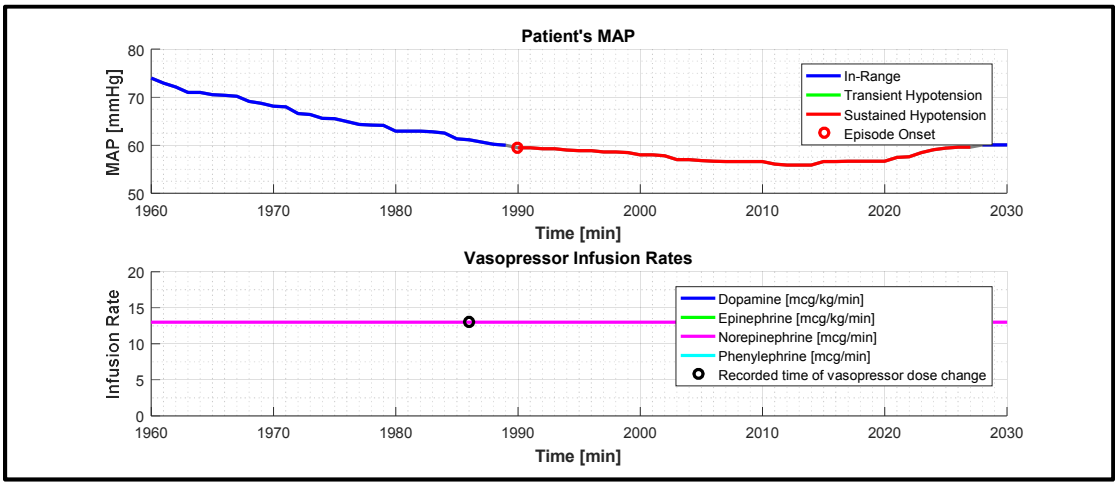

Figure 3.2.3.3: Drift-out episode of sustained hypotension

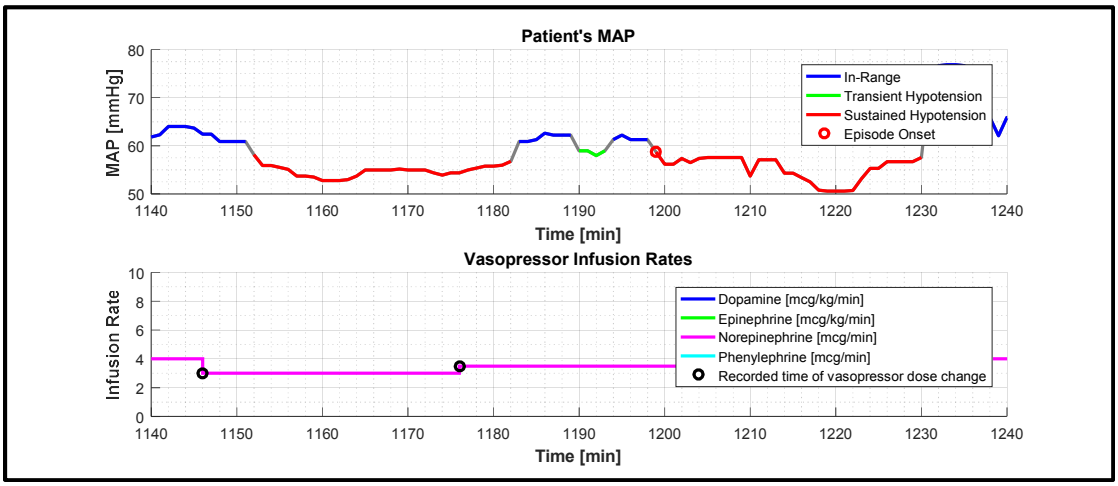

Figure 3.2.3.4: Continuation episode of sustained hypotension

3.2.4: Analyzing Clinical Response to Episodes of Sustained Hypotension

Next, after all episodes of sustained hypotension were categorized, we analyzed the clinical response, in terms of change in vasopressor infusion rates, following the onset of the episode. This analysis consisted of examining if and when there was a change in vasopressor dose following the onset that led to the resolution of the episode. We defined an episode resolution when the patient's MAP passed above the 60 mmHg hypotension threshold and was not followed by another episode of sustained hypotension within 30 minutes. Our definition of an episode resolution was the motive for combining continuation episodes with the previous episode, since the hypotension was not truly resolved.

We derived two categories of episode resolution: Self-resolved and dose-resolved. Intuitively, a self-resolved episode occurred when the episode terminated due to the patient's MAP passing above the 60 mmHg threshold without any increase in vasopressor dose, while dose-resolved episodes occurred when the patient's MAP passed above the 60 mmHg threshold following within 30 minutes of an increase in the vasopressor dose. If there were missing MAP measurements following the resolution of the episode, it was not included in this part of the analysis. Below are examples of both types of resolutions.

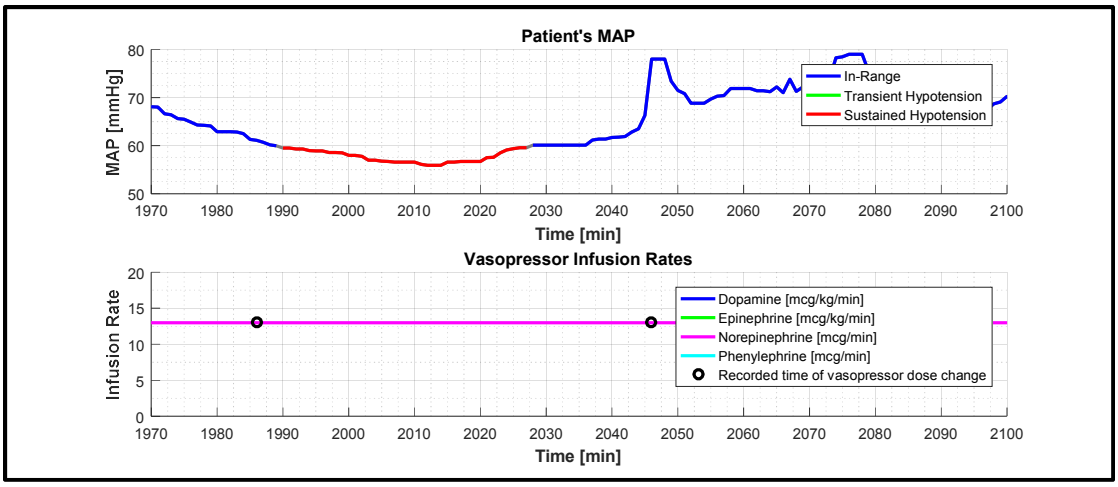

Figure 3.2.4.1: Self-resolved episode of sustained hypotension

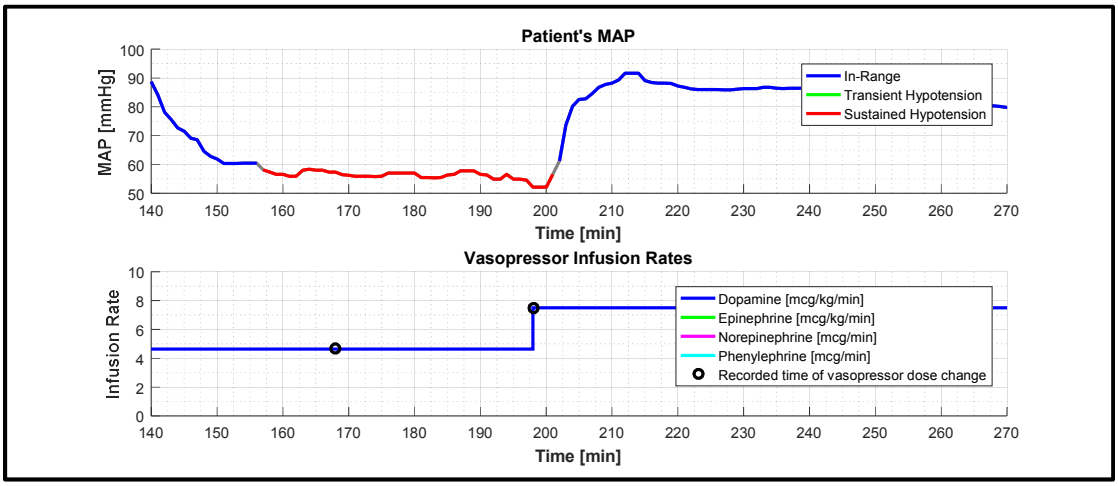

Figure 3.2.4.2: Dose-resolved episode of sustained hypotension

3.3: Results

From MIMIC II, a total of 216 total ICU stays from 176 distinct patients were identified as receiving at least 30 minutes of vasopressor mono-therapy at sub-maximum doses. From MGH, there were a total of 62 patients that were identified as receiving at least 30 minutes of vasopressor mono-therapy at sub-maximum doses. There were 109 ICU stays from MIMIC II, approximately 50%, with at least one episode of sustained hypotension during vasopressor infusion. This population, along

with the ICU stays with transient-only hypotension, ICU stays with no hypotension, and the patients from MGH are characterized in Table 3.3.1 below.

Table 3.3.1: Characteristics of Intensive Care Unit Stays					
	Hospital 1[a]				**Hospital 2**[a]
	Stays w/ sustained hypotension[b]	Stays w/ non-sustained hypotension[b]	Stays w/o hypotension[b]	All Stays	All Stays
Demographics:					
ICU stays, **n** (%)	**109** (51)	**52** (24)	**55** (25)	**216** (100)	**62** (100)
Unique patients[c], **n** (%)	**93** (53)	**49** (28)	**53** (30)	**176** (100)	**62** (100)
Age, **median** (IQR)	**76** (65 – 83)	**74** (64 – 82)	**70** (58 – 78)	**74** (64 – 82)	**69** (61 – 78)
Female, **proportion**, %	**52**	**47**	**51**	**49**	**45**
Male, **proportion**, %	**45**	**49**	**49**	**49**	**55**
Undocumented gender, **proportion**, %	**3**	**4**	**0**	**2**	**0**
Documented indication for vasopressor infusion[d]:					
Sepsis or possible sepsis, **proportion**, %	**56**	**39**	**47**	**48**	**45**
Cardiogenic or possible cardiogenic, **proportion**, %	**54**	**61**	**53**	**54**	**45**
Post-operative care, **proportion**, %	**15**	**16**	**21**	**18**	**68**
Other or unknown, **proportion**, %	**25**	**29**	**43**	**31**	**19**
Characteristics of vasopressor infusion:					
Total duration of vasopressor infusion, **median per stay**, (IQR), hr	**24.2** (10.7 – 48.1)	**13.9** (4.1 – 34.8)	**4.3** (1.6 – 11.5)	**14.0** (4.4 – 35.4)	**38.3** (22.4 – 55.9)
Time between vasopressor dose changes, **median per stay** (IQR), min	**90** (58 – 139)	**80** (35 – 240)	**60** (30 – 98)	**75** (49 – 139)	**60** (42 – 110)

[a] Hospital 1 includes patient data from the MIMIC II database and Hospital 2 includes patient data from a separate medical center

[b] Hypotension is defined as MAP < 60 mmHg and sustained hypotension is defined as at least 15 continuous min of hypotension

[c] Some patients have multiple stays that are in different categories, therefore the number of unique patients for all stays is less than the sum of the first three columns

[d] Patients may have more than one documented indication for vasopressor infusion

Table 3.3.1: Characteristics of intensive care unit stays

Furthermore, for we computed the proportion of time that a patient's MAP was within our pre-defined regions of sustained hypotension, transient hypotension, in-range, and hypertension for each type of patient. In Table 3.3.2, it shows the breakdown of these regions for the patients.

Table 3.3.2: Characteristics of MAP During Mono-Vasopressor Infusion

	Hospital 1[a]				**Hospital 2**[a]
	Stays w/ sustained hypotension[b]	Stays w/ non-sustained hypotension[b]	Stays w/o hypotension[b]	All Stays	All Stays
Statistics:					
MAP during infusion, **median per stay** (IQR), mmHg	**68** (64 – 72)	**75** (71 – 78)	**80** (76 – 88)	**73** (67 – 79)	**75** (71 – 80)
MAP hourly standard deviation during infusion, **median per stay** (IQR), mmHg	**4.2** (2.6 – 5.9)	**3.4** (2.8 – 4.3)	**3.3** (2.4 – 5.8)	**3.8** (2.7 – 5.7)	**3.7** (3.0 – 5.0)
Proportion of 100 ≥ MAP ≥ 60 mmHg, **median per stay** (IQR), %	**80** (68 – 90)	**97** (93 – 99)	**100** (87 – 100)	**90** (77 – 98)	**97** (92 – 99)
Proportion of MAP during transient hypotension, **median per stay** (IQR), %	**3.3** (1.7 – 5.5)	**1.3** (0.7 – 3.5)	**n/a**	**1.5** (0.0 – 4.1)	**0.3** (0.0 – 1.3)
Proportion of MAP during sustained hypotension, **median per stay** (IQR), %	**11** (3.9 – 23)	**n/a**	**n/a**	**0.6** (0.0 – 11)	**0.0** (0.0 – 2.6)
Proportion of MAP during hypertension[c], **median per stay** (IQR), %	**0.2** (0.0 – 3.3)	**0.0** (0.0 – 2.8)	**0.2** (0.0 – 13)	**0.1** (0.0 – 3.8)	**1.0** (0.0 – 3.4)
Episodes of sustained hypotension per 24 hours, **median per 24 hours** (IQR), n	**2.8** (1.2 – 5.0)	**n/a**	**n/a**	**0.0** (0.0 – 3.1)	**0.0** (0.0 – 1.2)

[a] Hospital 1 includes patient data from the MIMIC II database and Hospital 2 includes patient data from a separate medical center

[b] Hypotension is defined as MAP < 60 mmHg and sustained hypotension is defined as ≥ 15 min; see text for details

[c] Hypertension defined as MAP > 100 mmHg

Table 3.3.2: Characteristics of intensive care unit stays

There were 640 episodes of sustained hypotension from 109 ICU stays that we analyzed from MIMIC II. Of the 640, 32 were categorized as failed wean episodes, 39 were categorized as failed re-dose episodes, 338 were categorized as drift-out

episodes, and 231 were categorized as continuation episodes. A majority (53%) of episodes were drift-out episodes, meaning that most episodes did not have a recent change in vasopressor dose leading up to the onset of the episode. The same breakdown of episodes was performed for the MGH ICU stays, and is shown in the right-hand side chart in Figure 3.3.1. Note that there were only 30 episodes from the MGH ICU stays to be analyzed.

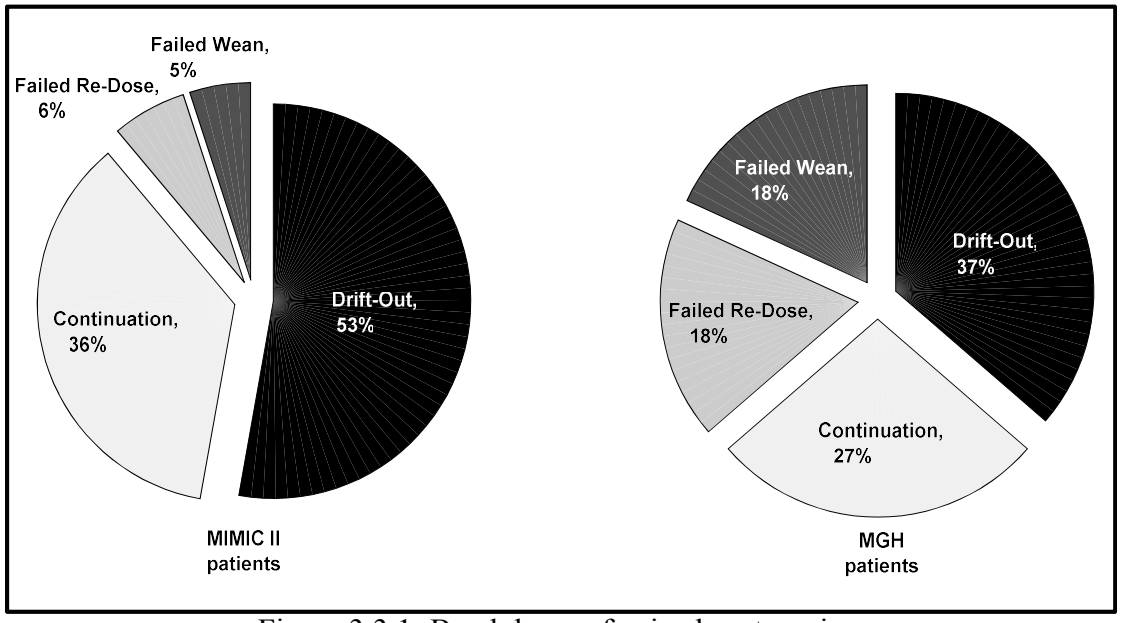

Figure 3.3.1: Breakdown of episode categories

There were 383 episode resolutions analyzed to determine clinical response, in terms of vasopressor infusion. Of the 383, 353 were self-resolved episodes and only 30 were dose-resolved episodes. For the 30 dose-resolved episodes, 25 episodes (83%) were resolved with a single increase in vasopressor dosage; but the median time until the first dose increase following the onset of the episode was 18.5 minutes. The median duration of self-resolved episodes was 38 minutes, and 26 minutes for dose-resolved episodes. We provided the same analysis results for the MGH patients;

note that there were only 22 episode resolutions to be studied from the MGH patients.

Table 3.3.3 provides more detailed statistics on episode resolutions.

Table 3.3.3: Description of Sustained Hypotensive Episodes				
	Hospital 1[a]		**Hospital 2**[a]	
	Episode resolved *w/* increase of vasopressor dose	Episode resolved *w/o* increase of vasopressor dose	Episode resolved *w/* increase of vasopressor dose	Episode resolved *w/o* increase of vasopressor dose
Statistics:				
Episodes, **n** (%)	**48** (8)	**562** (92)	**7** (25)	**21** (75)
Episode spans, **n** (%)	**30** (8)	**353** (92)	**5** (25)	**15** (75)
Duration of episode, **median** (IQR), min	**26** (19 – 45)	**38** (22 – 73)	**27** (21 – 30)	**24** (18 – 70)
Proportion of MAP < 60 mmHg during episode[b], **median** (IQR), %	**100** (87 – 100)	**100** (87 – 100)	**100** (100 – 100)	**100** (88 – 100)
Vasopressor dose changes preceding episode onset:				
Episode onsets with no preceding dose change, **proportion**, %	**70**	**84**	**80**	**40**
Episode onsets with preceding dose decrease, **proportion**, %	**23**	**6.3**	**20**	**27**
Episode onsets with preceding dose increase, **proportion**, %	**7**	**9.7**	**0**	**33**
Vasopressor dose changes during episode:				
Episodes with at least one dose increase, **proportion**, %	**100**	**11**	**100**	**20**
Episodes resolved with a single dose increase, **proportion**, %	**83**	**8.0**	**80**	**13**
Total number of dose increases during episode, **median** (IQR), n	**1** (1 – 2)	**0** (0 – 1)	**1** (1 – 2)	**0** (0 – 1)
Time until first dose increase, **median** (IQR), min	**18.5** (8 – 33)	**n/a**[c] (42 – n/a[c])	**13** (11 – 22)	**n/a**[c] (75 – n/a[c])

[a] Hospital 1 includes patient data from the MIMIC II database and Hospital 2 includes patient data from a separate medical center

[b] When the termination of an episode of sustained hypotension was followed within 30 min by the onset of another period of sustained hypotension, the intervals were combined into a single episode of sustained hypotension, allowing (theoretically) for MAP ≥ 60 mmHg during a portion of the episode

[c] *n/a* signifies that there was *no* increase, at all, in vasopressor dose during the episode

Table 3.3.3: Episode resolution statistics

3.4: Discussion

This investigation found that ICU patients receiving vasopressors commonly experienced episodes of sustained hypotension in which the patient's MAP remained below the limits of CNS autoregulation for 30 minutes or longer. These findings arose from a single institution (via the MIMIC II Database), so the generalizability is of this finding is not known, but supplementary investigations of the MGH ICUs support this finding. This provides a demonstration that patients treated with vasopressors may not be receiving care consistent with treatment guidelines. Our findings are indeed consistent with a recent report about patients with acute spinal cord injury: current guidelines advise maintenance of MAP of 85 – 90 mmHg for the first week after injury, while the investigators reported that MAP was below the guidelines in 42% of the documented values [29].

In terms of clinical context for our findings, most episodes of sustained hypotension occurred not because the vasopressor dose had been decreased in the time preceding the episode (Failed wean episode), nor because there had been an increase in vasopressor dose that was insufficiently large preceding the episode (Failed re-dose episode). Rather, most episodes occurred in the absence of any preceding vasopressor dose change. Typically, the patient's MAP simply drifted out of range.

Moreover, the typical clinical response after episodes of sustained hypotension developed did not involve any vasopressor dose increase. It is notable that, for the minority of episodes in which the vasopressor dose was in fact increased, a reduced duration of sustained hypotension was observed (i.e. 38 minutes to 26

minutes). The generally passive response to hypotension may reflect clinical inertia [4], a common phenomenon in which appropriate clinical responses are delayed. Furthermore, we speculate that one contributing factor might be that low-grade hypotension is clinically indistinct: aside from the low MAP displayed on a monitor, there are often no directly observable sequelae of low-grade hypotension, i.e. no observable cyanosis, posturing, nor abnormal respirations. We did not investigate other clinical responses to episodes of sustained hypotensive episodes, e.g. fluid bolus or reduction in sedation, so it is possible that some ineffective interventions were attempted during some of the episodes of sustained hypotension.

Another potential factor in the scarcity of dose increases before and after episodes of sustained hypotension is that any harm from this practice has not been clearly established, even though clinical guidelines advise vasopressors to avoid hypotension [27]. Yet in animal models it has been shown that, below MAP of 65 mmHg, the CNS cannot effectively autoregulate perfusion, which means that CNS ischemia is likely to occur [20], [21]. Therefore, low-grade sustained hypotension could cause low-grade CNS ischemic injury. Indeed, there is ample evidence of major cognitive injury caused by sepsis [26] and it is reasonable to speculate that sustained hypotension could be directly causing some of the cognitive damage.

Studies of ICU patients have shown a correlation between hypotension and poor clinical outcomes [22]–[24] but of course in retrospective studies, it is impossible to distinguish causation versus correlation. In other words, it is possible that hypotension directly causes bad outcomes, and it is also possible that

hypotension is merely an indicator of disease severity and not the direct cause of the poor outcomes.

The findings from our analysis arise primarily from a single medical center, so it is unknown if this is a truly pervasive issue and warrants additional investigation. As a matter of speculation, we suggest that an academic medical center that is so forward thinking as to freely share ICU records with worldwide researchers to advance clinical care [13] is unlikely to be a totally unique outlier in terms of clinical practice. To some extent, this clinical problem is likely to exist in other medical centers, and these findings are consistent with the findings of Hawryluk, et al. [29]. However, whether these findings are truly applicable to other ICUs cannot be determined definitively without additional investigation.

Chapter 4: Statistics-Based Method for Hypotension Detection

We developed a method for detecting episodes of sustained hypotension via logistic regression. The logistic regression method quantifies the relationship, if any, between a set of independent feature variables and a binary classifier variable. Applying this type of model to hypotension, we defined the expectancy of sustained hypotension at 15 minutes in the future as the binary classifier and defined mathematical trends of a patient's past vasopressor infusion, MAP, and HR as the feature variables. The reason that 15 minutes was chosen as the forecast window was that, based on clinical consensus, it is the approximate time delay for a change in vasopressor to take effect. The following chapter discusses how a logistic regression model was trained to provide reliable detection of sustained hypotension and the subsequent testing, validation, and analysis of the model as a decision support system.

4.1: Methods

The structure of logistic regression is similar to that of linear regression, the output variable, rather than being simply the dependent variable, is the inverse logistic transform of the dependent variable.

$$\boldsymbol{\beta_0 + \beta_1 f_1 + \beta_2 f_2 + \cdots + \beta_p f_p = ln\left(\frac{E}{1 - E}\right)} \qquad (4.1.1)$$

The features, f_i, are certain mathematical trends of past data from a patient's numerics and vasopressor dose information. The output, E, is the expectancy of sustained hypotension occurring 15 minutes in the future (i.e. $E = 1$ if it is certainly occurring, and $E = 0$ if it is certainly not occurring). The regression coefficients, β_i,

are determined by data-driven model fitting techniques that will be described in more detail in this section.

4.1.1: Feature Selection

Choosing the proper features is imperative to any type of model fitting exercise. After our retrospective analysis of episodes of sustained hypotension discussed in the previous chapter, we quantitatively showed that most episodes were preceded by a gradual decline in patient's MAP or by a previous episode. Additionally, vasopressors are medications that act to elevate arterial blood pressure in the critically ill when they suffer body-wide reduction in blood circulation. Therefore, we believed it important to extract features from the trends in a patient's past MAP measurements and the infusion of vasopressor dose, in order to detect the future state of a patient's MAP. Furthermore, for the sake of study, we investigated the usefulness of past HR measurements as a possible avenue for feature selection.

From these three feature sources, we chose to extract the comprehensive behavior of past trends of vasopressor dose, MAP, and HR, via mathematical trends: mean, least-squares linear slope, and standard deviation of various intervals of past data. The intervals of interest were 5, 10, 20, 30, 45, and 60 minutes prior to the present time. We speculated that analyzing up to 60 minutes of past MAP was sufficient enough to make inferences regarding the future 15 minutes of MAP. In the figure below, we show an example of the various features windows for this logistic regression model.

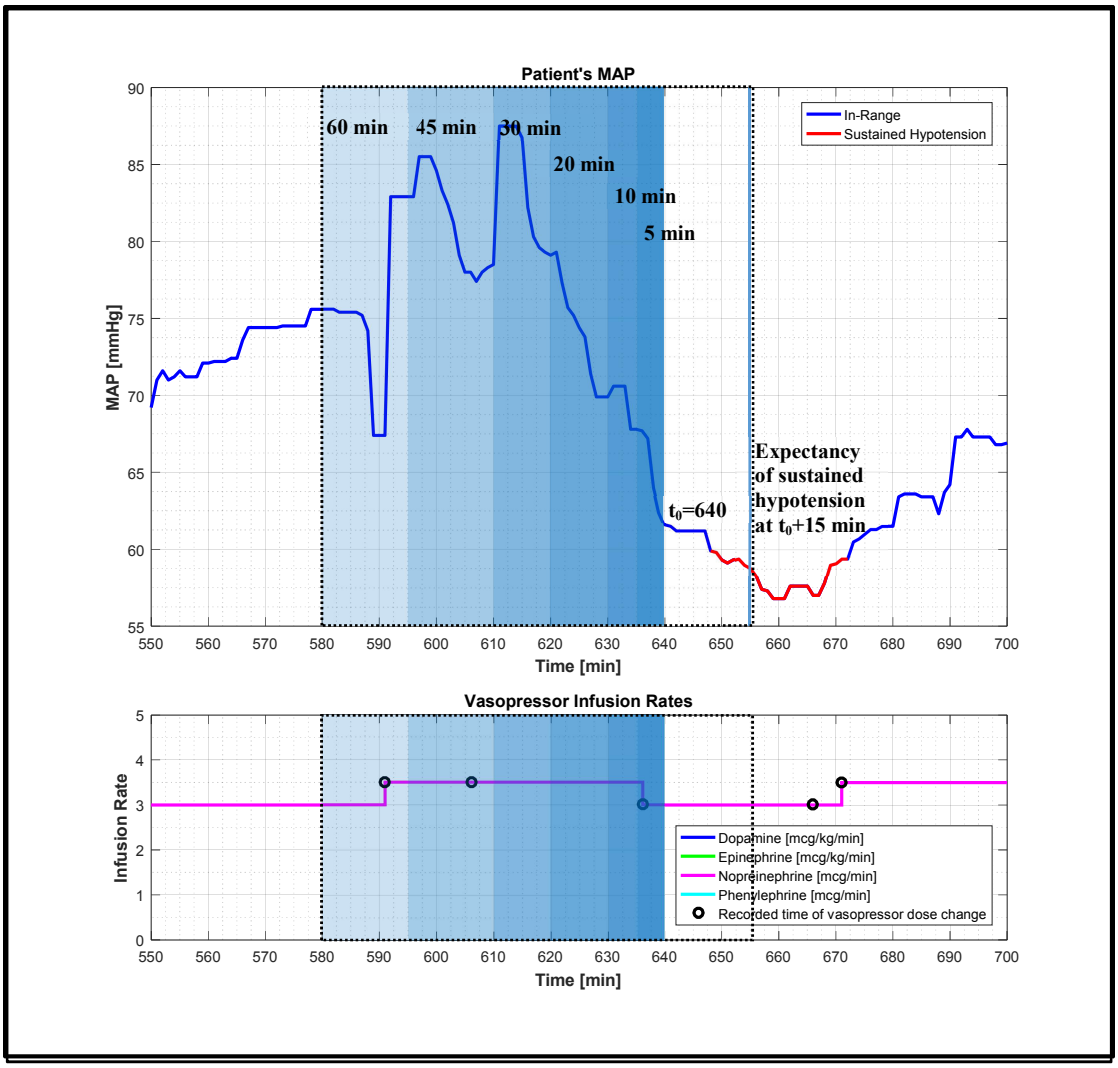

Figure 4.1.1.1: Intervals of past data for feature extraction

For an interval of past data of length, m, the following mathematical trends were computed:

$$\boldsymbol{\mu_m = \frac{1}{m}\sum_{i=1}^{m} MAP(t_0 - i + 1)}$$

$$\boldsymbol{\rightarrow \mu_m = \frac{1}{m}\sum_{i=1}^{m} MAP_i} \tag{4.1.1.1}$$

$$\boldsymbol{\beta_m = \frac{\sum_{i=1}^{m}[MAP(t_0 - i + 1) - \mu_m]\left[(t_0 - i + 1) - \left(\frac{1}{m}\right)\sum_{j=1}^{m}(t_0 - j + 1)\right]}{\sum_{i=1}^{m}(MAP(t_0 - 1 + 1) - \mu_m)^2}} \tag{4.1.1.2}$$

$$\rightarrow \boldsymbol{\beta_m} = \frac{\sum_{i=1}^{m}(\boldsymbol{MAP_i} - \boldsymbol{\mu_m})(\boldsymbol{t_i} - \boldsymbol{\mu_t})}{\sum_{i=1}^{m}(\boldsymbol{MAP_i} - \boldsymbol{\mu_m})^2}$$

$$\boldsymbol{\sigma_m} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{m}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{m}}(\boldsymbol{MAP}(\boldsymbol{t_0} - \boldsymbol{i} + \mathbf{1}) - \boldsymbol{\mu_m})^2} \tag{4.1.1.3}$$

$$\rightarrow \boldsymbol{\sigma_m} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{m}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{m}}(\boldsymbol{MAP_i} - \boldsymbol{\mu_m})^2}$$

Note that the above equations are for extracting mathematical trends from the patient's MAP data, but the same method was used for vasopressor dos and HR.

In total, we developed a list of 54 candidate mathematical features (i.e. 18 features per source) we suspected could be correlated to the expectancy of sustained hypotension at 15 minutes in the future.

4.1.2: Model Training

We developed the logistic regression model using patient data from MIMIC II only. MGH patients would be used for further validation and blind testing of this model. The patients from MIMIC II were separated into training and testing subgroups. This process was completed at random until patients accounting for approximately 75% of the total data were separated into the training subgroup. This randomized process removed any sort of bias that could have stemmed from hand selecting patients for the training and testing subgroups. In the end, patients accounting for 132 ICU stays were grouped as training patients and the remaining patients accounting for 99 ICU stays were put aside to be used as testing patients.

The first step for training the logistic regression model was to obtain segments of data, similar to that shown in Figure 4.1.1.1, containing 75 minutes of continuously available vasopressor dose, MAP, and HR. Each of the 132 ICU stays were scanned for these data segments; each segment was offset by at least 5 minutes from a previous segment to be considered.

For each of these segments, the mathematical features and dependent variable (i.e. expectancy of sustained hypotension 15 minutes in the future) were extracted in order to fit the logistic regression model to this data. Treating each segment like a real-time snapshot of data, the 60^{th} data point in each segment was considered the "current" time. Therefore, the features were extracted from the various intervals by using the data from the 1^{st} to 60^{th} data point. The dependent variable was then extracted by examining the 75^{th} data point from the segment, which in the framework of the real-time snapshot, is 15 minutes in the future. Examining the 75^{th} data point, if that data point corresponded to a MAP value within an episode of sustained hypotension, the dependent variable was given a value of 1 for that segment, and otherwise it was assigned a 0. Note that the only value the dependent variable could take during the model training is 0 or 1, but that was not the necessarily the case during model testing.

After extracting the features and expectancy value from all segments, we organized the features into an nxp matrix and the expectancy value into an $nx1$ vector, where n was the number of segments and p was the number of features extracted from each observation.

$$\boldsymbol{F} = \begin{bmatrix} \overrightarrow{\boldsymbol{f_1}} \\ \overrightarrow{\boldsymbol{f_2}} \\ \vdots \\ \overrightarrow{\boldsymbol{f_n}} \end{bmatrix} \qquad \boldsymbol{Y} = \begin{bmatrix} \boldsymbol{E_1} \\ \boldsymbol{E_2} \\ \vdots \\ \boldsymbol{E_n} \end{bmatrix}$$

In the matrix, F, the vectors $\overrightarrow{f_i}$ contain the p features extracted from all N segments; in the vector Y, the values E_i are the binary variable for the expectancy of sustained hypotension 15 minutes in the future (i.e. 0 or 1). Since the features extracted were measures mean, slope, and standard deviation of pressure (MAP), flow rate (vasopressor dose), and frequency (HR), the features were normalized in order for proper comparisons to be made. The features were normalized using the standard score method: each feature, f, is standardized using its corresponding mean value, μ_f, and standard deviation, σ_f:

$$\bar{\boldsymbol{f}} = \frac{\boldsymbol{f} - \boldsymbol{\mu_f}}{\boldsymbol{\sigma_f}} \tag{4.1.2.1}$$

The model was then fitted utilizing the MATLAB routine, *glmfit*, and a simple greedy backward elimination algorithm. The MATLAB routine, *glmfit*, is a method of generalized linear model regression that estimates the coefficients for a linear regression model based on the input of a matrix of predictors (i.e. features) and a corresponding response vector. Within the options of *glmfit*, the routine can be implemented to fit a logistic regression model. The greedy backward algorithm is a method of eliminating insignificant regressors from a particular model. We implemented this model by iterating the *glmfit* routine in MATLAB, and after iteration, removed the column vector of features that were insignificant, according to its P-value (0.05 significance level).

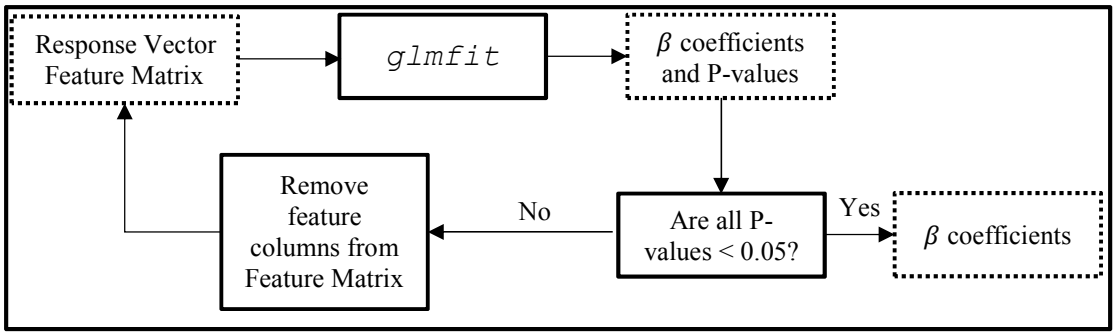

Figure 4.1.2.1: Greedy backward algorithm for logistic regression fitting

Using this algorithm, we derived two logistic regression models. The first model was trained using only features extracted from the patient's MAP data, and the second was trained using all features, from the patient's vasopressor dose, MAP, and HR data. Both of these models, when given the previous 60 minutes of a patient's vasopressor dose, MAP, and HR measurements, would output a value from 0 to 1 that was equated to the model's prediction of the expectancy of sustained hypotension occurring 15 minutes in the future.

4.1.3: Model Testing

We tested the performance of the logistic regression models by simulating the model in real-time with the 99 ICU stays from the testing patients. Starting at the beginning of each stay's data, the model would extract the necessary features from the patient's past data and compute an expectancy value corresponding to that time step. Iterating over the entire stay, a discrete series of expectancy values were created. When the model was unable to produce an output due to lack of available past data, the expectancy value was simply assigned to 0. The figure below shows an example of the discrete series of expectancy values matching up with a patient's MAP data.

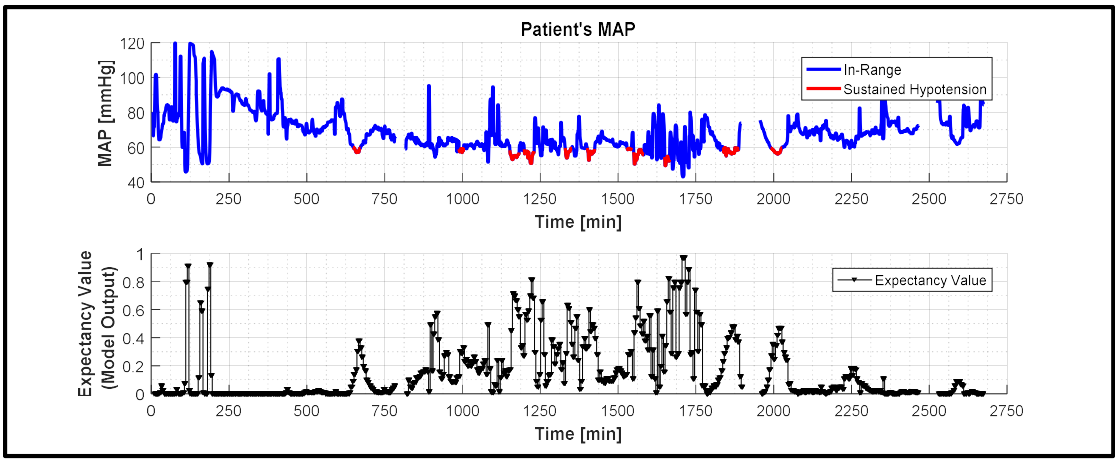

Figure 4.1.3.1: Example of the discrete expectancy values from logistic regression

To utilize this logistic regression model output as a method of detecting sustained hypotension, the discrete series of expectancy values were compared to a value, defined as the alert threshold. Any instance where the expectancy value went above the alert threshold was defined as an alert. These alerts were divided into two categories: true and false alerts.

To categorize alerts properly, we examined the expectancy values from 30 minutes before the onset of episodes of sustained hypotension to 15 minutes after the onset of the episodes. If at any point the expectancy value series passed above the threshold and stayed above the alert threshold until at least 15 minutes after the onset of the episode, it would be categorized as a true alert. Additionally, at any other time the expectancy value series passed above the alert threshold not in proximity to episodes of sustained hypotension, it would be defined as a false alert. The figures below provide examples of true and false alerts. The alert threshold value was determined by finding the threshold that maximized the performance of the model for the training patient subset; model performance is based on three metrics: number of

episodes that have no corresponding alert, the time between the onset of a true alert and the onset of the corresponding episode, and the frequency of false alerts.

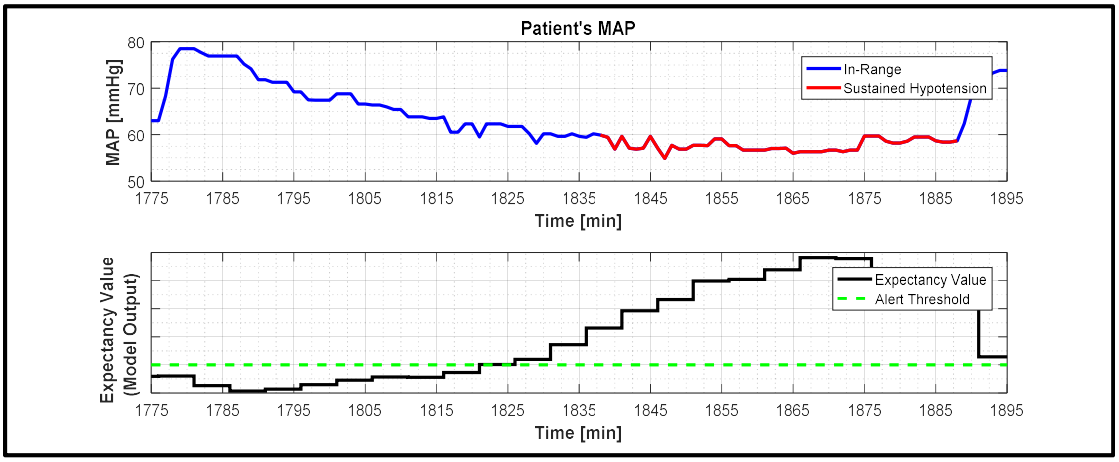

Figure 4.1.3.2: True alert example

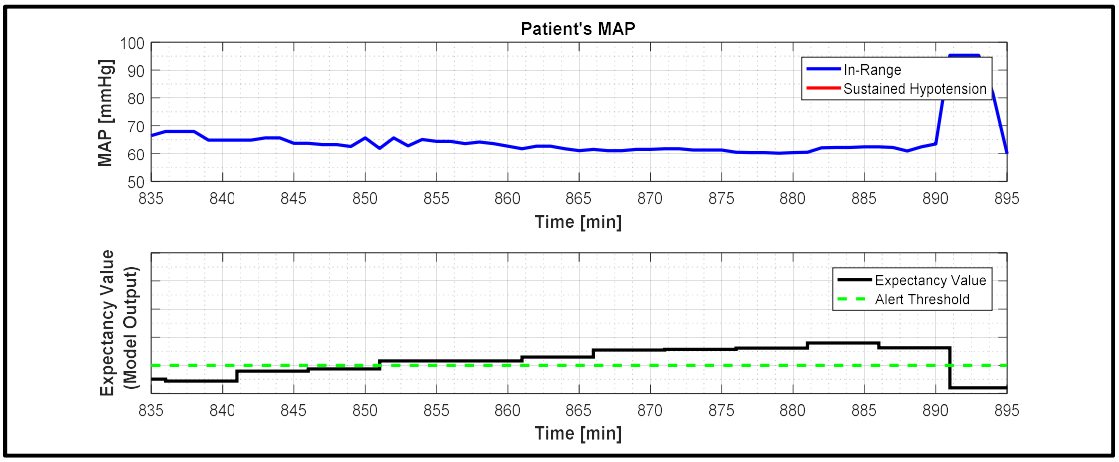

Figure 4.1.3.3: False alert example

The model performance was quantified by three values: Missed Episodes, Advance Warning Time, and False Alert Frequency. A Missed Episode is a false negative, or any instance that the model is unable to detect and alert for an episode of sustained hypotension. The Missed Episodes performance metric was the most important, since the goal was to develop a detection model that has the ability to detect all types of episodes of sustained hypotension. Advance Warning Time is the

duration of time between the onset of a true alert, when the expectancy value passes above the alert threshold, and the onset of the episode of sustained hypotension. False Alert Frequency is the number of false alerts that occur per 24 hours of vasopressor infusion.

To benchmark our logistic regression model against current caregiver practices, we created a separate "model", the threshold detector, which best represents the current practices clinicians should follow, based on consensus guidelines [27], [28]. Simply, If the patient's MAP read from the bedside monitor falls below 60 mmHg (i.e. the threshold for hypotension), the threshold detector will output an expectancy value of 1, otherwise the expectancy value will be 0. The threshold detector is similar to a simplistic system that alerts whenever the MAP signal sourced from the patient monitor falls below 60 mmHg. The same performance metrics were computed for the threshold detector.

4.1.4: Model Blinded Testing

The model was tested using the testing patients from MIMIC II, but both the training and testing patients came from the same source. We found it in our research's best interest to include additional testing that was completely independent from our training and testing patient subgroups and blinded from our model development. We took the logistic regression models developed with our training patients and testing on our testing patients, and implemented that hypotension detection model to the patient data we gathered from MGH. Identical performance metrics were computed for the MGH subset.

4.2: Results

The first configuration of our logistic regression model contained only MAP features, by design. The features that were significant based on the 0.05 significance level following the greedy backwards algorithm were:

- 5-minute MAP slope
- 10-minute MAP mean
- 10-minute MAP slope
- 10-minute MAP standard deviation
- 45-minute MAP slope
- 60-minute MAP standard deviation

There were no significant features from the 20-minute or 30-minute windows of past data. The distribution of feature values corresponding to an expectancy value of 1 versus an expectancy value of 0 from the training segments are compared in the figure below.

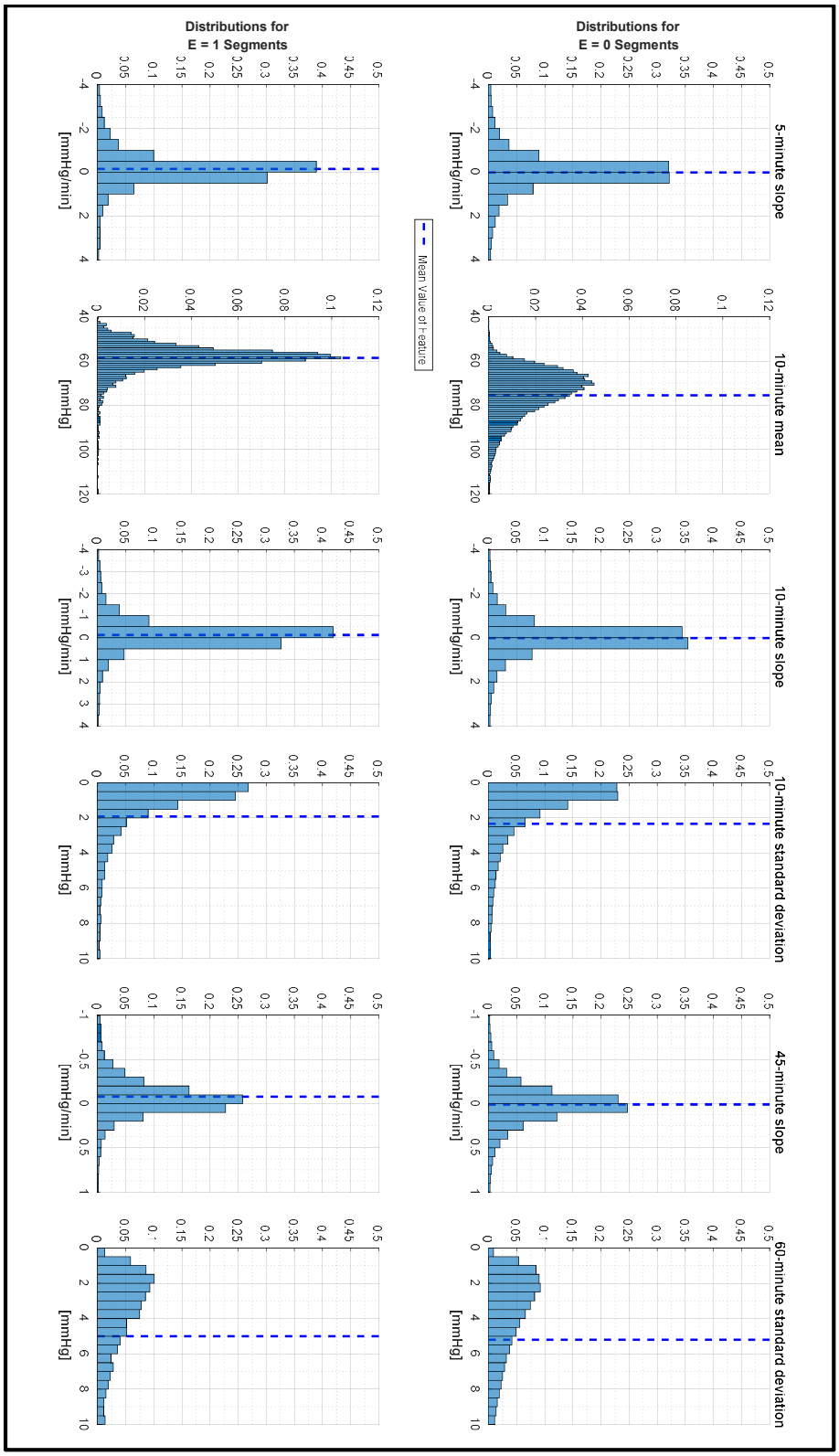

Figure 4.2.1: Training features' distributions

The logistic regression model configuration with vasopressor dose and HR features included in the training (along with MAP features) fitted identical MAP features and only a single additional feature, 5-minute vasopressor dose slope. The table below summarizes the logistic regression coefficients found through *glmfit* and the greedy backwards algorithm. Left out of the table is the constant coefficient, $\beta_0 = -5.58$.

Table 4.2.1: Significant Features[a] for Logistic Regression		
	Model w/ MAP features	Model w/ MAP and dose features
5-minute window features		
MAP mean	--	--
MAP slope	**-0.65**	**-0.66**
MAP std. dev.	--	--
Dose mean	--	--
Dose slope	--	**0.041**
Dose std. dev.	--	--
10-minute window features		
MAP mean	**-7.95**	**-7.95**
MAP slope	**-0.16**	**-0.16**
MAP std. dev.	**0.51**	**0.51**
Dose mean	--	--
Dose slope	--	--
Dose std. dev.	--	--
45-minute window features		
MAP mean	--	--
MAP slope	**0.72**	**0.72**
MAP std. dev.	--	--
Dose mean	--	--
Dose slope	--	--
Dose std. dev.	--	--
60-minute window features		
MAP mean	--	--
MAP slope	--	--
MAP std. dev.	**0.44**	**0.44**
Dose mean	--	--
Dose slope	--	--
Dose std. dev.	--	--

[a] Features are normalized

Table 4.2.1: Significant Features for Logistic Regression

We are not evaluating our model on single detection events, rather we are examining how well our model can detect an episode of sustained hypotension in advance; initially we can examine the Receiver Operating Characteristic (ROC) curve

to study the model's ability to properly detect at individual instants in time. The ROC curve is a tool used to illustrate the performance of a detection system (i.e. binary classifier system) [30].

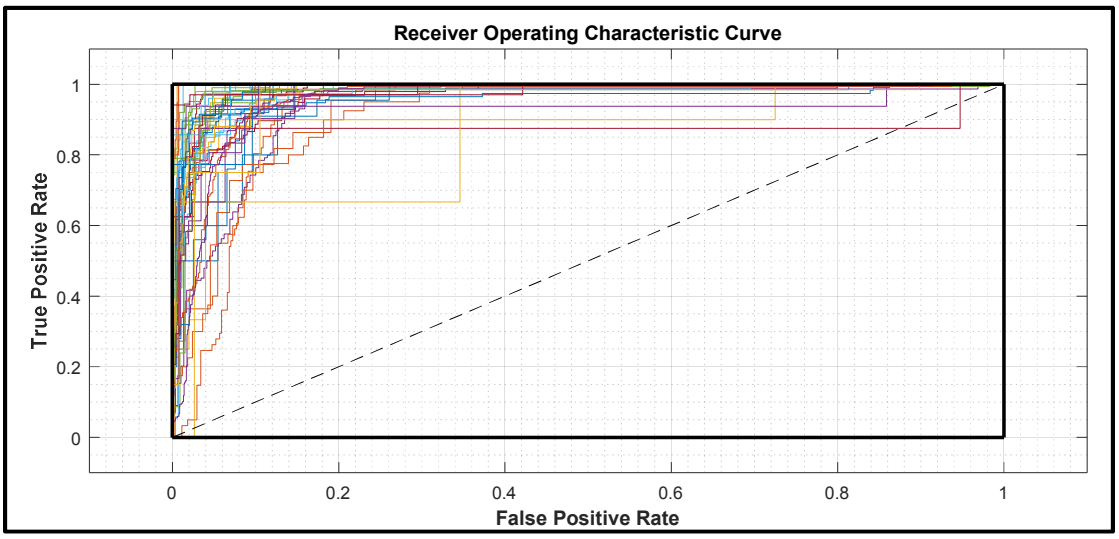

Figure 4.2.2: Receiver operating characteristic curve

Figure 4.2.2 shows the ROC curve for the 99 ICU stays, plotting the true positive (correctly identifying the state of patient's MAP 15 minutes in the future) and false positive rates. The average Area Under Curve (AUC) value for the ICU stays' ROC curve is 0.92; A perfect AUC score is 1.00. But, it is important to note that this AUC value is not indicative of the model's true performance; as it is skewed by the typical patient's ratio of in-range MAP versus hypotensive MAP. Referring to our retrospective analysis, and specifically Table 3.3.2, the median patient's MAP is in-range 90% of the time, and for most cases the model easily detects that there will be no sustained hypotension 15 minutes in the future if the patient's MAP is easily in-range. Therefore, we rely heavily, and solely, on our performance metrics Missed Episodes, Advance Warning Time, and False Alert Rate, rather than the model's ROC curve.

The model's alert threshold was chosen that enabled the best performance metrics for the model when simulated on the training patients; the alert threshold value was $E_t = 0.1$. At this alert threshold for the testing patients, all 290 episodes of sustained hypotension had a corresponding true alert (i.e. 0 Missed Episodes). For the median episode, the Advance warning was 13 minutes. And for the median ICU stay, the model produced 5.3 false alerts per 24 hours of vasopressor infusion. The table below shows complete performance metrics for both logistic regression models and the threshold detector model.

The model's performance was tested also on the MGH patients as a form of blind testing/validation. For this subset of patients, the episode sample size was much smaller so these results need to be expanded with more patients that experienced episodes of sustained hypotension. For the MGH patient subset, all 29 episodes of sustained hypotension were detected by the model; and for the median episode, the advance warning was 22 minutes. The model produced 1.9 false alerts per 24 hours of vasopressor infusion. Figure 4.2.3 shows an example of the model's outputted expectancy value passing above the alert threshold and detecting an episode of sustained hypotension.

Table 4.2.2: Performance of the Logistic Regression Model					
	Hospital 1			**Hospital 2**	
	Model w/ MAP features	Model w/ MAP and dose features	Threshold Detector	Model w/ MAP features	Threshold Detector
Statistics:					
Number of episodes, **n**	**290**	**290**	**290**	**29**	**29**
Proportion of episodes that were undetected, **%**	**0**	**0**	**0**	**0**	**0**
Advance warning time, **median** (IQR)	**13** (0 – 30)	**11** (-1 – 30)	**0** (0 – 0)	**22** (3 – 30)	**0** (0 – 0)
Number of false alerts per 24 hours, **median per stay** (IQR)	**5.3** (1.0 – 10.0)	**7.0** (3.6 – 11.7)	**14** (6.2 – 23)	**1.9** (0.0 – 6.1)	**3.5** (0.3 – 8.3)

[a] Hospital 1 includes patient data from the MIMIC II database and Hospital 2 includes patient data from a separate medical center

Table 4.2.2: Performance of the Logistic Regression Model

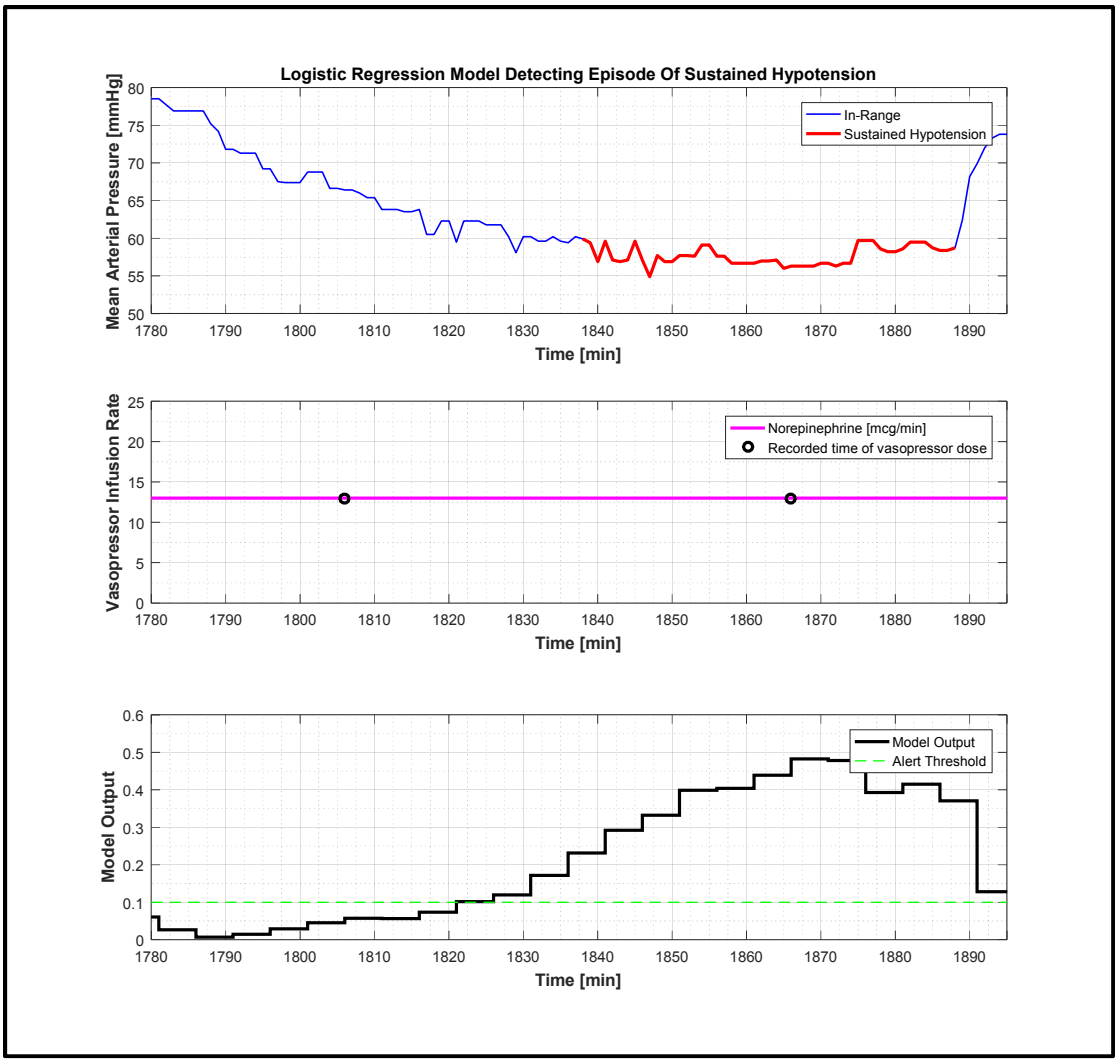

Figure 4.2.3: Logistic regression detection example

4.3: Discussion

We created a logistic regression model that uses a simple collection of mathematical trends of a patient's prior MAP data to infer the expectancy of an imminent episode of sustained hypotension. Proper selection of the features that correlate to this expectancy was imperative to the model's performance. Our model typically alerted for episodes of sustained episodes prior to its onset, and infrequently produced false alerts

We investigated vasopressor dose infusion features and their effectiveness in the logistic regression model for characterizing the dependency of the occurrence of near-term future episodes of sustained hypotension on these features. The only significant vasopressor dose feature was the 5-minute dose infusion slope, with a regression coefficient of $\beta = 0.042$. A hypothesis for the lack of significant dose infusion features is that that MAP features may imbed the effect of past dose infusion trends, therefore eliminating the need for redundant features. The inherent relationship between dose infusion and blood pressure supports this reason [31]. This result-driven hypothesis is not suggesting that dose infusion is not correlated to the future state of a patient's MAP, or that dose infusion information is not potentially relevant for blood pressure prediction.

Many physiological trends were investigated for their correlation to the expectancy of imminent episodes of sustained hypotension; but there was a distinct grouping of near-term (5-10 minute) and far-term (45-60 minute) MAP features that, when combined, provided the best correlation to future episodes in a logistic regression format. Although MAP and HR are related physiologically, when features extracted from a patient's HR (e.g. mean, slope, and standard deviation) were used to train the model, all were insignificantly correlated to the existent of imminent episodes. Further, MAP features extracted from the middle-term (20-30 minute) windows were also insignificantly correlated. Similar to the hypothesis regarding the lack of dose features, the effect of middle-term MAP features and HR features were encapsulated in near- and far-term MAP features.

The significant MAP features derived were also intuitive in nature, making this model transparent to an outside perspective. Mean and slope features have a negative correlation (i.e. lower value leads to higher expectancy) and the standard deviation features have a positive correlation (i.e. higher value leads to higher expectancy). The only exception to this nature was the 45 min MAP slope feature, which is positively correlated. We believe that this occurred due to the fact that many episodes of sustained hypotension are preceded in close proximity by the termination of other episodes.

We compared the logistic regression model against a simple threshold alert. We chose the simple threshold detection scheme as a benchmark as it represents clinicians reacting as soon as the patient's MAP falls below the hypotensive threshold, 60 mmHg. The logistic regression model provides better detection than the simple threshold detector; 13 minutes of advance warning for versus a no predictive warning of 0 minutes for the threshold detector for the testing patients of Hospital 2. The threshold detector has no prediction capability. Further, the logistic regression model produces less false alerts per 24 hours than the threshold detector; 5.3 false alerts versus 14 false alerts for the testing patients of Hospital 1. These results suggest that the logistic regression model could provide valuable insight and alerts for clinicians to improve critically ill patient care.

In addition to testing on MIMIC II patients, we also tested the model further with patients from the MGH dataset. The model produces much less false alerts for the MGH patients than the MIMIC II patients; this is due to the higher level of median MAP for the MGH patient subset. Since the most significant feature of the

logistic regression model is the 10-minute MAP mean, it is the main contributor to false alerts (i.e. patients with low levels of MAP will experience higher rate of false alerts); therefore, the MGH patients will have lower rates of false alerts.

In terms of computational efficiency, the logistic regression model computes an output after every 5 minutes of real time, regardless if features can be extracted. The computational time step was chosen arbitrarily to reduce, and perhaps higher resolution of hypotension expectancies can be achieved by increasing the frequency of model iterations. This change could be implemented when introduced into ICUs for high-resolution outputs of this system.

Chapter 5: Time Series Analysis Forecasting Method for Hypotension Detection

We developed a method for forecasting a patient's MAP future measurements to infer the likelihood of a future episode of sustained hypotension. Autoregressive models exploit only past measurements of a time series signal in order to compute forecasted values of that same time series signal. This model structure fit well with the purposes of MAP forecasting, as exhibited from our logistic regression modeling discussed in the previous chapter, features extracted from only MAP were found to be extremely useful in determining the likelihood of oncoming episodes of sustained hypotension.

5.1: Methods

Discussed in the first chapter, the generalized form of a time series model is the ARMAX model. The ARMAX model is a function of a signal's past measurements, as well as past input and system shock measurements. We simplified the ARMAX model for our purposes to a solely autoregressive (AR) model into the difference equation form:

$$\boldsymbol{A(q)MAP(k) = e(k)}$$

$$\boldsymbol{\rightarrow MAP(k) = \sum_{i=1}^{n} a_i MAP(k-i) + e(k)} \tag{5.1.1}$$

$$\boldsymbol{\rightarrow MAP(k) = a_1 MAP(k-1) + \cdots + a_n MAP(k-n) + e(k)}$$

The current value of MAP, $MAP(k)$ can be expressed as the finite weighted sum of previous values of MAP and a residual term associated with the current value of MAP. We will use this structure of the AR model to develop our MAP forecasting scheme for computing the likelihood of episodes of sustained hypotension.

5.1.1: Model Forecasting and Structure

This subsection will briefly explain how the AR model forecasts future values of MAP using only previous values of MAP. As mentioned previously, the difference equation form of the AR model is:

$$\boldsymbol{MAP(k) = a_1 MAP(k-1) + \cdots + a_n MAP(k-n) + e(k)} \qquad (5.1.1.1)$$

In some cases, an AR model is used to perform one-step ahead prediction; for example, using the measurements from time step $k-1$ to $k-n$ to predict the measurement at time step k. An AR model can be optimized for one-step ahead prediction via least squares analysis since there is only one unknown variable, the measurement at time step k.

This structure can be manipulated into a form that predicts the next step ahead, i.e. $k+1$ measurement of MAP, i.e. $\widehat{MAP}(k+1)$. Further, since the model is predicting a step ahead, the value of $e(k+1)$ has an expected value of 0, as it is a white noise approximation.

$$\boldsymbol{\widehat{MAP}(k+1) = \sum_{i=1}^{n} a_i MAP(k+1-i)} \qquad (5.1.1.2)$$

For our purposes, we wanted to develop an AR model that can forecast MAP measurements up to 15 minutes. If we employed an AR model optimized for one-step ahead prediction, we would have to sample the patient's MAP at 1 measurement

every 15 minutes; this is not ideal. Rather, we look to employ a model for multiple-step ahead prediction, or pure prediction.

As an example, assume we sample the patient's MAP at 1 measurement every s minutes. We then have the following AR model forecasting structure:

$$\text{Base Equation: } \boldsymbol{MAP(k) = a_1 MAP(k-1) + \cdots + a_n MAP(k-n) + e(k)} \tag{5.1.1.1}$$

Forecasting Equations at time step k:

$$\text{k + 1 Forecast} \rightarrow \boldsymbol{\widehat{MAP}(k+1) = a_1 MAP(k) + \cdots + a_n MAP(k+1-n)}$$

$$\begin{aligned}\text{k + 2 Forecast} \rightarrow \boldsymbol{\widehat{MAP}(k+2)} \\ = \boldsymbol{a_1 \widehat{MAP}(k+1) + a_2 MAP(k) + \cdots} \\ \boldsymbol{+ a_n MAP(k+2-n)}\end{aligned} \tag{5.1.1.3}$$

$$\begin{aligned}\text{k + 3 Forecast} \rightarrow \boldsymbol{\widehat{MAP}(k+3)} \\ = \boldsymbol{a_1 \widehat{MAP}(k+2) + a_2 \widehat{MAP}(k+1) + a_3 MAP(k) + \cdots} \\ \boldsymbol{+ a_n MAP(k+3-n)}\end{aligned}$$

And so on ...

The forecast at $k+l$ relies on the forecast from $k+l-1$, and so on. The final forecast, $\widehat{MAP}(k+l)$, corresponds to the forecasted value at 15 minutes in the future.

With this forecasting structure, for example, if a patient's MAP is sampled at 1 measurement every 1 minute (the native sampling rate of the data), the number of forecasts required to forecast 15 minutes is 15. We explore different sampling rates for the AR model to establish a comprehensive analysis: the sample rates investigated are 1 measurement per 1 minute, 3 minutes, and 5 minutes.

Further, the number of previous MAP measurements in the AR difference equation can be adjusted as well; the number of previous measurements, or regressor values, is known as the model order of the system. We vary the model order of the

system from 2 to 15, depending on the sampling rate of the AR model. In the Table 5.1.1, the various model orders and sampling rates are tabulated.

Table 5.1.1: AR Model Structure Variants			
Sampling Rate: / **Model Order:**	**1 pt. / 1 min**	**1 pt. / 3 min**	**1 pt. / 5 min**
2	n = 2, s = 1 min	n = 2, s = 3 min	n = 2, s = 5 min
…	…	…	…
8	n = 8, s = 1 min	n = 8, s = 3 min	n = 8, s = 5 min
9	n = 9, s = 1 min	n = 9, s = 3 min	n = 9, s = 5 min
10	n = 10, s = 1 min	n = 10, s = 3 min	N/A
11	n = 11, s = 1 min	N/A	N/A
…	…	…	…
15	n = 15, s = 1 min	N/A	N/A

Table 5.1.1: AR Model Structure Variants

5.1.2: Model Training Cost Function and Optimization

The goal of the AR model was to forecast a desired amount of patient's MAP. Therefore, it was appropriate to train the AR model to minimize the total error in these forecasts. For our AR model training, we developed cost functions that focus on the minimization of multiple forecast errors. For all models, the goal was to forecast the subsequent 15 minutes of a patient's MAP. Incorporating the variants of the model shown in the previous section, and the fact that the forecast goal is 15 minutes, the following cost function structure was developed.

Batch data segments (at least $N \approx 50{,}000$ segments from the MIMIC II training subpopulation) were used to compute forecasts of a patient's MAP signal to

train the model. The cost function compares the model's forecasted value with actual values of l future time steps of a patient's MAP signal within a segment. The i^{th} step ahead forecast from the j^{th} segment, the forecast residual is calculated and stored in a vector, $\vec{e}_i$, corresponding to the time step ahead of the forecast. These error vectors contain N entries. Finally, the 2-norm of these vectors are computed, summed, and normalized by the total number of time step ahead forecasts, l, to compute the cost function.

$$\boldsymbol{min}\left\{\frac{\mathbf{1}}{\boldsymbol{l}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{l}}\|\vec{\boldsymbol{e}}_{\boldsymbol{i}}\|_{\mathbf{2}}\right\} \tag{5.1.2.1}$$

$$\begin{aligned}
&\boldsymbol{s.t.}\\
&\widehat{\boldsymbol{MAP}}(\boldsymbol{k}+\mathbf{1}) = \boldsymbol{a_1 MAP}(\boldsymbol{k}) + \cdots + \boldsymbol{a_n MAP}(\boldsymbol{k}+\mathbf{1}-\boldsymbol{n})\\
&\widehat{\boldsymbol{MAP}}(\boldsymbol{k}+\mathbf{2}) = \boldsymbol{a_1}\widehat{\boldsymbol{MAP}}(\boldsymbol{k}+\mathbf{1}) + \boldsymbol{a_2 MAP}(\boldsymbol{k}) + \cdots + \boldsymbol{a_n MAP}(\boldsymbol{k}+\mathbf{2}-\boldsymbol{n})\\
&\widehat{\boldsymbol{MAP}}(\boldsymbol{k}+\mathbf{3}) = \boldsymbol{a_1}\widehat{\boldsymbol{MAP}}(\boldsymbol{k}+\mathbf{2}) + \boldsymbol{a_2}\widehat{\boldsymbol{MAP}}(\boldsymbol{k}+\mathbf{1}) + \boldsymbol{a_3 MAP}(\boldsymbol{k}) + \cdots\\
&\qquad\qquad + \boldsymbol{a_n MAP}(\boldsymbol{k}+\mathbf{3}-\boldsymbol{n})\\
&\vdots
\end{aligned} \tag{5.1.1.3}$$

$$\boldsymbol{e_{i,j}} = \widehat{\boldsymbol{MAP}}(\boldsymbol{k}+\boldsymbol{i}) - \boldsymbol{MAP}(\boldsymbol{k}+\boldsymbol{i})\ \boldsymbol{for\ the\ j^{th}\ training\ segment} \tag{5.1.2.2}$$

$$\vec{\boldsymbol{e}}_{\boldsymbol{i}} = \left[\boldsymbol{e_{i,1}}, \boldsymbol{e_{i,2}}, \ldots, \boldsymbol{e_{i,N}}\right]^{\boldsymbol{T}} \tag{5.1.2.3}$$

$$\|\vec{\boldsymbol{e}}_{\boldsymbol{i}}\|_{\mathbf{2}} = \sqrt{\sum_{\boldsymbol{j=1}}^{\boldsymbol{N}} \boldsymbol{e_{i,j}}^{\mathbf{2}}} \tag{5.1.2.4}$$

The AR model was trained using the identical set of training patient data as described in section 4.1.2. First, segments of data were extracted from the patients' data that contained the necessary amount of continuously available MAP measurements; note that the "necessary amount" of measurements varies depending on model order and sample rate, for example, a 4th order model with a sampling rate of 1 measurement every 5 minutes required 15 minutes of the most recent MAP

signal for past MAP values and 15 minutes of the subsequent MAP signal to compare forecasted MAP values with the true values per segment. Figure 5.1.2.1 shows an example of a single segment for this example AR model.

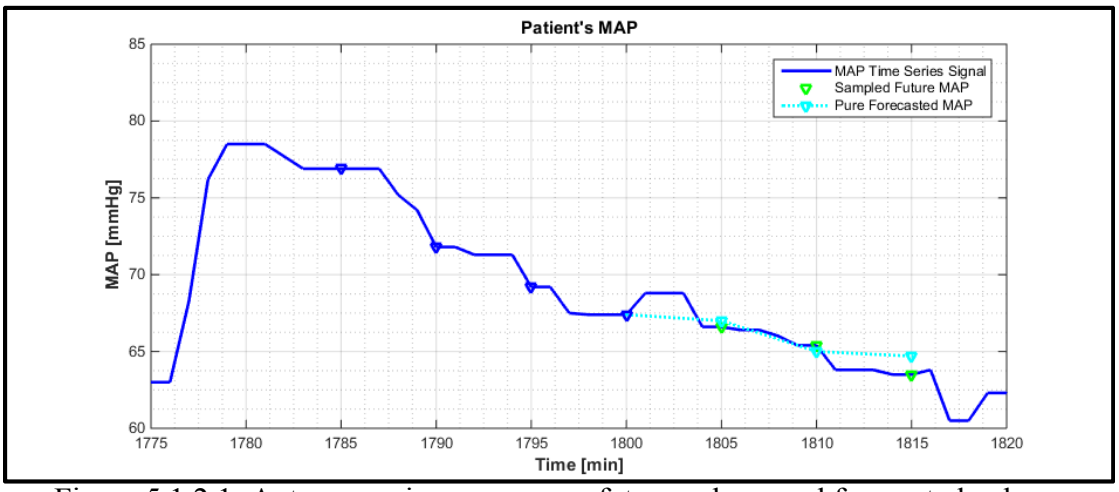

Figure 5.1.2.1: Autoregressive regressors, future values, and forecasted values

The model's autoregressive coefficients, A, were optimized to minimize the cost function described above. The *fmincon* routine in MATLAB finds the optimal coefficients when supplied with a cost function and an initial condition. The initial condition for the coefficients is the initial "guess" for the values of the coefficients. A good starting point was the least squares solution to the simple one-step prediction problem summarized previously. Given the past values of MAP, i.e. $MAP(k), MAP(k-1)$, etc. and the single step ahead future value of MAP, $MAP(k+1)$, the value of the coefficients A could be found that minimize the sum of the square of errors for the single step ahead prediction. Coefficients were computed using the equation below.

$$\boldsymbol{A^{IC} = [X^T X]^{-1} X^T y} \qquad (5.1.2.5)$$

Where X is the matrix of previous values of MAP, each row representing a separate observation (or segment of data) and y is the vector of the corresponding single step ahead future value of MAP for each observation.

We used the least square solution as a starting point for the optimization for *fmincon*. Additionally, we perturbed the coefficients of the initial condition by a factor of 10 to encompass a large proportion of the coefficient vector space; this is to ensure that the true minimum, rather than a local minimum, was found for the optimization. The optimization routine was iterated for each of these perturbations, and the optimal autoregressive coefficients were the coefficients that correspond to the iteration with the minimal cost function value. This optimization was computed for all of the model variants discussed in the previous section.

5.1.3: Model Forecast Envelop

It was not expected that the AR model would be completely accurate in its forecasting, rather, the goal of time series forecasting is to develop probabilistic forecast of the time series trend. Therefore, in addition to simply forecasting future values of a patient's MAP signal, a forecast envelop accompanies the forecasted values to suggest probability limits for the forecasted values. Similar to the forecast envelop that one may see on a weather broadcast for projecting the path of a hurricane, our forecast envelop was put in place to accomplish a similar task: display the highest likelihood projection of a patient's future MAP signal.

To compute probability limits for forecasts computed by AR models, the difference equation derived above needed to be transformed into an infinite series

weighted sum of current and past shock measurements (i.e. model errors) [12]. The form of this time series model looks like the following:

$$\boldsymbol{MAP(k) = \sum_{i=0}^{\infty} \psi_i e(k-i)} \tag{5.1.3.1}$$

To transform the difference equation form into the infinite weighted sum of shocks, the following relationship is used [12]:

$$\boldsymbol{A(q)\psi(q) = \theta(q)} \tag{5.1.3.2}$$

Where $\theta(q)$ is the set of coefficients corresponding to the error terms of the AR difference equation model. For this case, our model contains only the error term $e(k)$, therefore $\theta(q) = 1$. Also, $A(q)$ corresponds to the AR coefficients of the model corresponding to the recent MAP values $MAP(k)$, $MAP(k-1)$, etc. The equation can be expressed as the following.

$$(\mathbf{1} + a_1 q^{-1} + a_2 q^{-2} + \cdots + a_n q^{-n})(\mathbf{1} + \psi_1 q^{-1} + \psi_2 q^{-2} + \psi_3 q^{-3} + \cdots) = \mathbf{1} \tag{5.1.3.3}$$

Solving for the weights, we get the following.

$$\begin{gathered} \boldsymbol{\psi_0 = 1} \\ \boldsymbol{\psi_1 = -a_1} \\ \boldsymbol{\psi_2 = -a_1\psi_1 - a_2} \\ \boldsymbol{\psi_3 = -a_1\psi_2 - a_2\psi_1 - a_3} \\ \vdots \\ \boldsymbol{\psi_j = -a_1\psi_2 - a_2\psi_1 - \cdots - a_j, j \leq n} \end{gathered} \tag{5.1.3.4}$$

Thus, to calculate the expected variance of $\boldsymbol{MAP(k+l)}$, we used infinite sum series with the computed $\boldsymbol{\psi_j}$ values to derive the equation for forecast variance:

$$\boldsymbol{Var[k+l] = \left\{ \sum_{j=0}^{l-1} \psi_j^2 \right\} \sigma_e^2} \qquad (5.1.3.5)$$

Where σ_e^2 is the variance of the shocks $e(t)$ of the model. The value for σ_e^2 was approximated by calculating the variance of the one-step ahead forecast error of previous forecasts. This approximation could be performed in two manners: using the variance from the model training optimization (σ_e^2 is static), and using the variance from one step ahead forecast error of previous forecasts of the last 15 minutes, 30 minutes, 60 minutes, and 120 minutes (σ_e^2 is dynamic). Using the above equations to calculate forecast variance, an accompanying forecast envelop was calculated for each forecast.

5.1.4: Hypotension Detection from Forecasting

The forecasted MAP signal and forecast envelop produced by the model might provide potential insight for clinicians monitoring the patient's state, though it does not directly determine the likelihood of an episode of sustained hypotension. In order to transform the AR model's output into a method for hypotension detection, the forecast and forecast envelop must be further exploited.

Using the forecast envelop and the threshold for hypotension (< 60 mmHg) we constructed a metric that was crudely associated with the likelihood of sustained hypotension. To compute this metric, we simply calculation of the proportion of the amount of the forecast envelop that was below the hypotension threshold (< 60

mmHg) to the total forecast envelop. See Figure 5.1.5 and the equation below for more details.

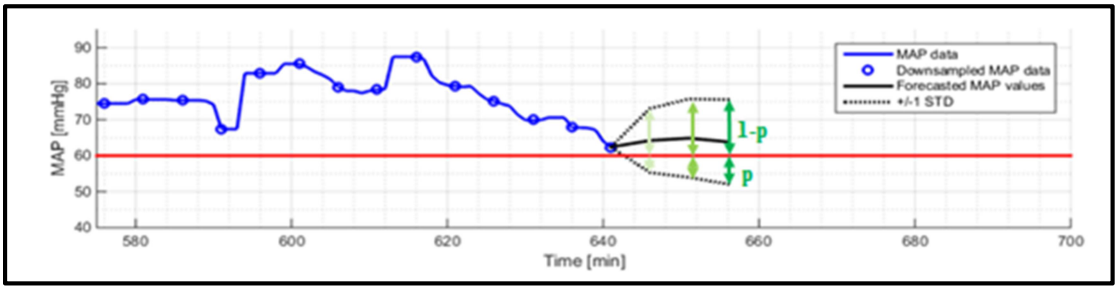

Figure 5.1.5: Forecast envelop and *p*-metric

$$\boldsymbol{p(k+l) = \frac{60\ mmHg - \left(\widehat{MAP}(k+l) - \hat{\sigma}(k+l)\right)}{2\hat{\sigma}(k+l)}} \qquad (5.1.4.1)$$

Similar to the model output of the logistic regression model, E, the *p*-metric could approximate the expectancy of future episodes of sustained hypotension. Additionally, we can determine an optimal threshold value for the *p*-metric that will trigger an alert when the AR model produces a *p*-metric value that passes above it. This exploitation of the AR model and its forecast transforms the remainder of the analysis into the identical process performed for the logistic regression model.

5.1.5: Model Testing and Blind Testing

We tested the performance of the AR models by simulating the model in real-time with the 99 ICU stays from the MIMIC II testing patients. Starting at the beginning of each stay's data, the model would extract the necessary past values of a patient's MAP and iteratively compute an l-step ahead forecast with accompanying forecast envelop at each time step (similarly to the logistic regression model testing, the computation time step was 5 minutes). For each forecast, the *p*-metric was computed by analyzing the last forecasted step and its forecast envelop. This results

in a discrete series of *p*-metric values that were treated as expectancy values for sustained hypotension.

The AR model performance was quantified by the same performance metrics as the logistic regression model. See Subsection 4.1.3 for details. Additionally, blind testing was performed on the entire MGH dataset, see Subsection 4.1.4 for details.

5.2: Results

The model was optimized according to the prescribed cost function for various model orders and sample rates. The one-step prediction forecast residuals were computed for each model as well as the Akaike Information Criterion (AIC). The AIC is a metric that scores a model based on its accuracy and complexity. Table 5.2.1 is a summary of the optimization results.

Table 5.2.1: AR Model Training Results				
Sample Rate	Model Order	Cost Function Value [mmHg]	One-Step Prediction Residual Variance [$mmHg^2$]	Akaike Information Criterion
5 min	2	1671	67	-13
	3	1606	63	-11
	4	1555	61	-9
	5	1511	60	-7
	6	1473	58	-5
	7	1444	56	-3
	8	1419	55	-1
	9	1395	55	1
3 min	2	1646	46	-14
	3	1612	48	-12
	4	1580	46	-10
	5	1528	42	-8
	6	1518	46	-6
	7	1474	41	-4
	8	1463	44	-2
	9	1444	42	0
	10	1420	41	2
1 min	2	14733	1077	-20
	3	1577	11	-14
	4	1549	12	-12
	5	$1*10^{10}$	281	-36
	6	1773	200	-8
	7	2183	38	-7
	8	1848	411	-4
	9	1765	28	-2
	10	1448	10	0
	11	1439	8	2
	12	1444	10	4
	13	1441	10	6
	14	1431	11	8
	15	1399	9	10

Table 5.2.1: AR Model Training Results

Increasing model order led to a lower cost function value for the optimized model, since there were additional parameters to minimize the cost function of the model. Further, for most cases, the one-step prediction residual variance also decreased. Although the model got marginally better in terms of forecast residual

value and variance, the AIC score suggests that higher model orders do not significantly improve the models at the cost of complexity.

The *p*-metric threshold was determined by examining the performance metrics of the model when simulated through the training patient subset. For each model order and sample rate, we calculated the performance metrics across the entire spectrum of *p*-metric values. Below are the results from the models with a 5-minute sampling rate. An identical analysis was performed for each sampling rate variant.

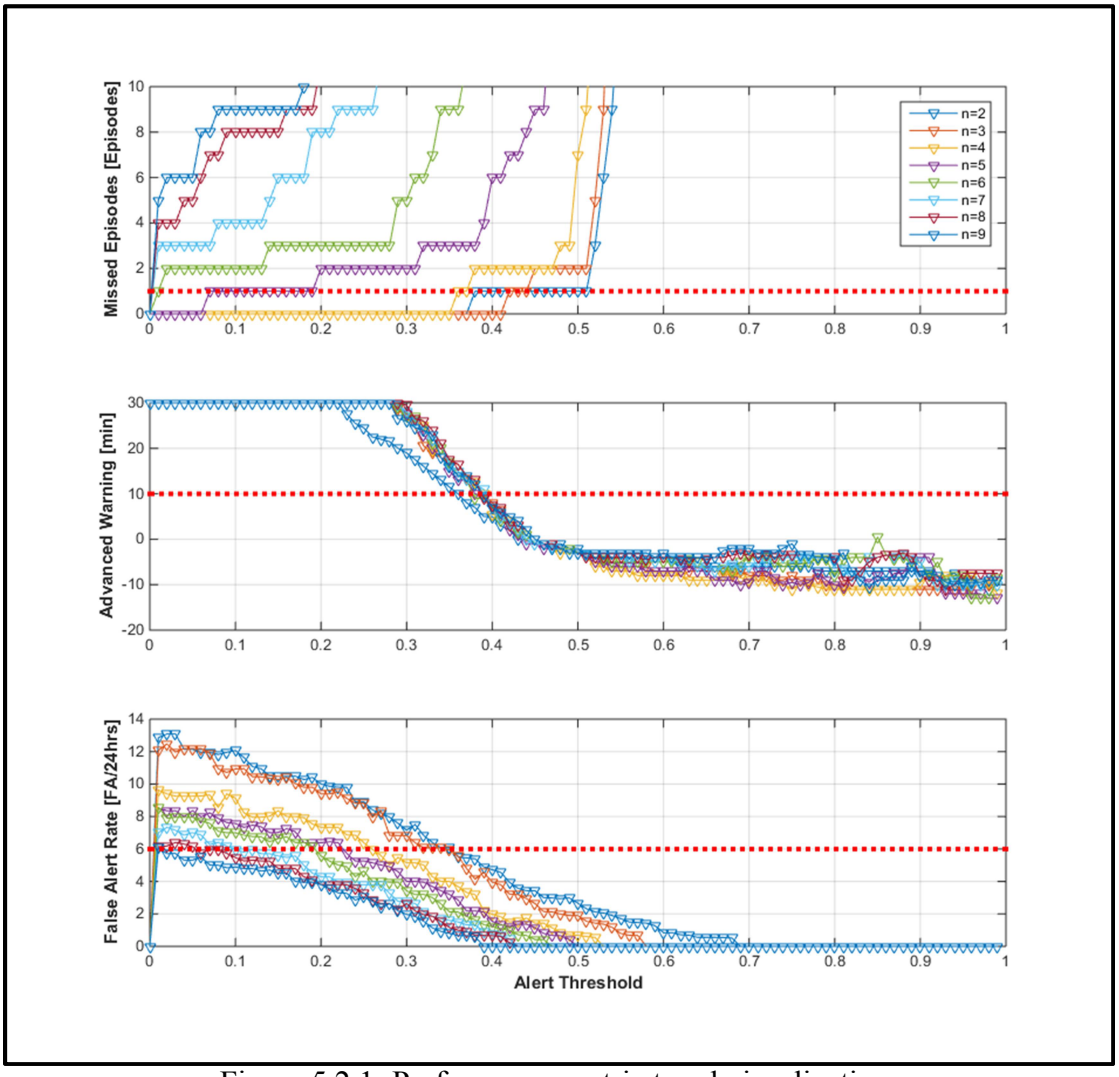

Figure 5.2.1: Performance metric trend visualization

We chose the optimal p-metric threshold after narrowing down the bounds to cases where the model missed zero episodes, the advance warning greater than 10 minutes, and the false alert rate less than 6 false alerts per 24 hours. Ultimately, the best autoregressive model was the fourth order AR model with a p-metric threshold of 0.35. For AR models with a sample rate of 3 minutes, the best model was a 6th order AR model with a p-metric threshold of 0.38. And for AR models with a sample rate of 1 minute, the best configuration was a 14^{th} order model with a p-metric threshold of 0.40.

The figure below shows an example of the model's outputted forecast and the corresponding p-value passing above the alert threshold (0.35 for 5 min sample rate) and detecting an episode of sustained hypotension approximately 18 minutes before its onset.

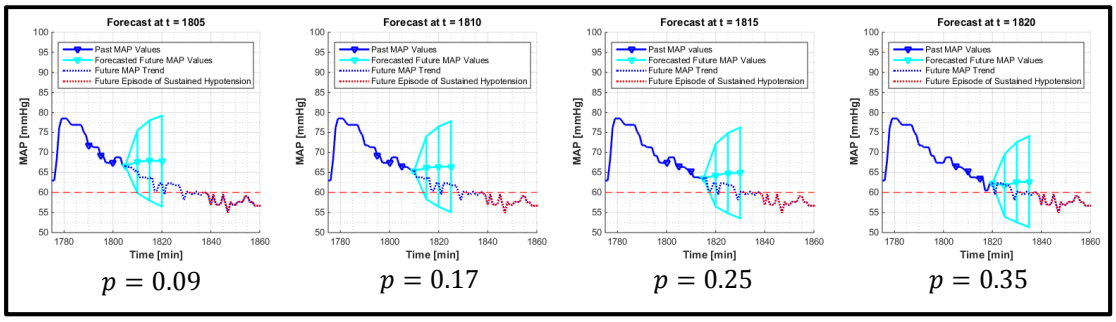

Figure 5.2.2: AR model forecasting and detecting episode of sustained hypotension

Table 5.2.2: Performance of AR Models with Static Forecast Envelop			
	5 min sample rate 4th order AR Model	3 min sample rate 6th order AR Model	1 min sample rate 14h order AR model
Statistics:			
Optimal *p*-metric threshold	**0.35**	**0.38**	**0.40**
Number of episodes, **n**	**290**	**290**	**290**
Proportion of episodes that were undetected, **%**	**0.4** (1 episode)	**0.4** (1 episode)	**0.4** (1 episode)
Advance warning time, **median** (IQR)	**11** (-2 – 30)	**11** (-2 – 30)	**9** (-1 – 30)
Number of false alerts per 24 hours, **median per stay** (IQR)	**4.2** (0 – 8.6)	**4.5** (0 – 8.7)	**4.6** (0 – 9.6)

Table 5.2.2: Performance of AR Models with Static Forecast Envelop

The AR model results above are for the AR model forecasting detection system with a forecast envelop derived from the assumption that the variance of the model shock, σ_e^2, is static. Note that the 1 min sample rate AR model produced unintuitive results due to the order of forecasting needed (15-step prediction is unreasonable). Taking the "best" AR model from the table above (5 min sample rate 4th order model), we then compare how the model performs while varying the derivation of the forecast envelop. We compared the models where the model shock was assumed to be constant, and approximated by the previous 15 minutes, 60 minutes, and 120 minutes of forecasts. The table below summarizes the results we found.

Table 5.2.3: Performance of AR Models with Different Forecast Envelops				
	5 min sample rate 4th order AR Model (constant variance approximation)	5 min sample rate 4th order AR Model (15-minute variance approximation)	5 min sample rate 4th order AR Model (60-minute variance approximation)	5 min sample rate 4th order AR Model (120-minute variance approximation)
Statistics:				
Optimal *p*-metric threshold	**0.35**	**0.20**	**0.23**	**0.28**
Number of episodes, **n**	**290**	**290**	**290**	**290**
Proportion of episodes that were undetected, **%**	**0.4** (1 episode)	**0.4** (1 episode)	**0**	**0**
Advance warning time, **median** (IQR)	**11** (-2 – 30)	**2** (-3 – 24)	**11** (-2 – 30)	**10** (-2 – 30)
Number of false alerts per 24 hours, **median per stay** (IQR)	**4.2** (0 – 8.6)	**7.0** (3.3 – 11.7)	**5.5** (1.6 – 9.4)	**5.1** (0.5 – 7.9)

Table 5.2.3: Performance of AR Models with Different Forecast Envelops

Based on the results from Table 5.2.2 and Table 5.2.3, we chose to further study the 4th order AR model with a static forecast variance approximation and benchmarked this model against the previously studied logistic regression and threshold detection models. The AR model had similar Advance Warning time and a lower False Alert Rate compared to the logistic regression model. Both AR and logistic regression models had lower False Alert Rates compared to the threshold detector when tested on the MIMIC testing patients, but higher False Alert Rates when tested on the MGH patients. The table below summarizes comprehensive benchmarking of the models.

Table 5.2.4: Performance of the AR Model						
	Hospital 1			**Hospital 2**		
	4th Order AR Model	Logistic Regression	Threshold Detector	4th Order AR Model	Logistic Regression	Threshold Detector
Statistics:						
Number of episodes, **n**	**290**	**290**	**290**	**29**	**29**	**29**
Proportion of episodes that were undetected, **%**	**0.4**	**0**	**0**	**0**	**0**	**0**
Advance warning time, **median** (IQR)	**11** (-2 – 30)	**13** (0 – 30)	**0** (0 – 0)	**14** (-1 – 30)	**22** (3 – 30)	**0** (0 – 0)
Number of false alerts per 24 hours, **median per stay** (IQR)	**4.2** (0 – 10.0)	**5.3** (1.0 – 10.0)	**14** (6.2 – 23)	**1.6** (0 – 4.9)	**1.9** (0.0 – 6.1)	**3.5** (0.3 – 8.3)

[a] Hospital 1 includes patient data from the MIMIC II database and Hospital 2 includes patient data from a separate medical center

Table 5.2.4: Performance of the AR, Logistic Regression, and Threshold Detector Models

The fourth order autoregressive model alerted for all but one episode for the testing subset of patients from MIMIC II; note, we examined the missed episode and it appears that the episode is an artifact or a result of non-physiological behavior. For the 289 episodes the model did detect, the median advance warning time was 11 minutes. For the median ICU stay, the model produced 4.2 false alerts per 24 hours of vasopressor infusion.

The model's performance was also tested on the MGH patients as a form of blind testing/validation, identical to the process gone through with the logistic regression model. For the MGH patient subset, the model detected all 29 episodes of sustained hypotension with a median advance warning time of 14 minutes. For the median ICU stay of the MGH patient subset, the model produced 1.6 false alerts per 24 hours of vasopressor infusion.

5.3: Discussion

We developed a model with an underlying autoregressive structure that computes forecasted values of the time series signal of a patient's MAP and determines the expectancy of future episodes of sustained hypotension. Variations of the model were studied to ensure that the best configuration of the model was derived for the goal of hypotension detection. This method of hypotension detection was comparable to the logistic regression model we previously studied, but can allow for much more flexibility.

The AR models we formulated in the previous sections were identified via population-based batch training methods. Similar to the logistic regression modeling, the AR model was trained using a large batch of segments from patient data to fit various AR models to the time series data. Observing the behavior of the population-trained AR model, it is noticeably similar in behavior to a moving average filter. By example, examining the pole-zero mapping of the 4^{th} order AR model extensively studied and a 4^{th} order moving average filter, the pole locations of each filter are in close proximity.

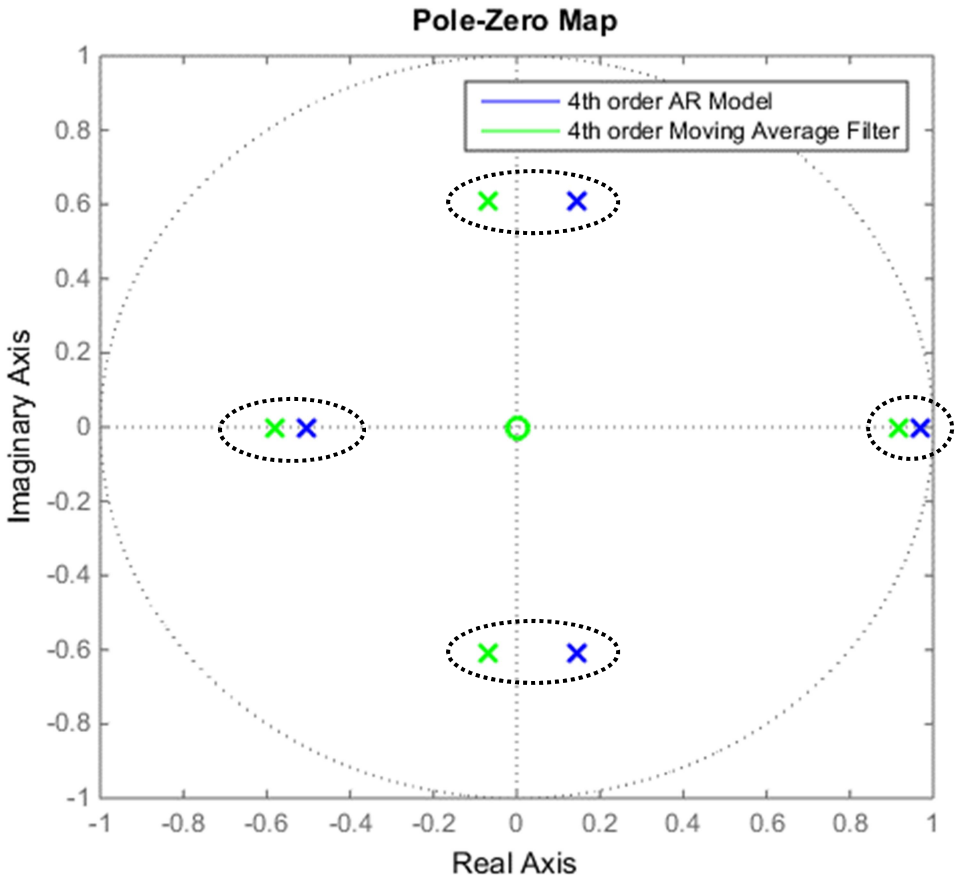

Figure 5.3.1: Pole-Zero map of AR model and moving average filter

This behavior was also noticeable in the model's forecasts. In the example illustrated in Figure 5.2.2, the forecasted values appear to closely resemble the near-term average of the patient's MAP signal. This was an important finding for this autoregressive study, as it corroborated our findings from the logistic regression feature selection: the most significant feature for detecting future episodes was the mean value of the previous 10 minutes of a patient's MAP (intuitively). We have found from both analyses that episodes of sustained hypotension are preceded by very low levels of MAP, rather than sudden drops in the value of MAP.

Additionally, investigating variants of the model, with respect to sampling rate of the data and the calculation of the forecast envelop, yielded the conclusion that a 4^{th} order AR model with a 1 measurement per 5-minute sampling rate and a static forecast envelop approximation was the best formulation for detecting episodes of sustained hypotension. The 5-minute sample interval was able to encompass more information about the past state of the patient's MAP, compared to the 3-minute and 1-minute sampling interval. For the smaller sampling intervals, the model was being fit to the changes in a patient's MAP between a few minutes, or a single minute; Ultimately, the model was attempting to fit to noise, as the changes in the measurements from minute to minute were most likely physiological noise or measurement noise. Therefore, a broader time step between measurements was deemed best for the AR model structure. For the forecast envelop approximation, a static approximation generated relatively better results than the updating approximations. The updating approximations, especially the 15-minute variety, only took into account the very near term past forecasts. In our retrospective study, we observed that most episodes of sustained hypotension were preceded by a gradual drift out of MAP. Therefore, most likely, the variance of the forecast errors preceding most episodes would be small, since the patient's MAP was not highly variable. This results in a very small forecast envelop, and in turn a smaller *p*-metric if the patient's MAP is not in proximity to the 60 mmHg threshold. That is why a static approximation, accounting for the general variability of MAP is a more useful tool for this model.

This AR model was compared to the logistic regression model as well as the simple threshold alert. The AR model provides comparable detection for episodes of sustained hypotension (11 min vs. 13 min Advance Warning Time for MIMIC II patients), The AR model also has comparable rates of false alerts per 24 hours (4.2 vs 5.3 False Alerts per 24 hours). The AR model, like the logistic regression model, provides valuable insight for clinicians that may improve the care of the critically ill.

Further, with the performance of the AR model and logistic regression model being very comparable, it was important to understand the differentiation between these models. The one major improvement that the AR model provided was its flexibility and more comprehensive insight for hypotension detection.

The AR model is trained to minimize forecast residuals of a patient's future MAP. Since the training is not dependent on the threshold of hypotension (i.e. MAP < 60 mmHg for our case), the AR model does not need re-training to be implemented to detect a different definition of hypotension (such as from the Surviving Sepsis Campaign [27]). Since the model's structure is a simple forecasting model, rather than a correlation model to a specific definition of hypotension, the AR model is a much more attractive method for hypotension detection.

The AR model derived in this chapter has proven to be comparable to the logistic regression with respect to our predefined performance metrics, but this new time series forecasting model allows for greater flexibility for real-world application.

Chapter 6: Conclusions

6.1: Conclusions

6.1.1: Retrospective Analysis Conclusions

During vasopressor therapy, episodes of sustained hypotension were common and typically without associated vasopressor dose increases. Since the likelihood is that sustained hypotension causes CNS and other end-organ injury, methods of mitigating episodes of sustained hypotension would be worthy of investigation. This retrospective analysis has shown that most episodes of sustained hypotension that were resolved with a change in vasopressor dosage only required a single adjustment; therefore, the prevalence of these episodes of sustained hypotension may not be due to lack of know-how of the clinician or nurse, but rather the lack of availability of the clinician or nurse to actively monitor the patient's MAP.

The results of this investigation have shown that most episodes of sustained hypotension go undetected by the clinician or nurse. This void in clinical care of the critically ill provides an opportunity to investigate the merit of clinical informatics systems that could support clinicians and nurses with the detection and prompt treatment of sustained hypotension.

6.1.2: Hypotension Detection Methods Conclusions

Both logistic regression and AR models could be useful in real-world application as a clinical informatics system to aid in the care of critically ill patients experiencing severe hypotension. In comparison to current methods of treating chronic hypotension, which are mostly reactive approaches, these models proactively

detect and alert for episodes of sustained hypotension that provide clinicians and nurses ample time to assess and treat the patient with an adjustment in vasopressor dose.

Although the performance and structure of the logistic regression model are enlightening, there are drawbacks that limit its generalizability for practical use. The most obvious limitation is that the model is trained to a specific definition of sustained hypotension (MAP < 60 mmHg for at least 15 consecutive minutes). If this model was to be implemented to aid clinicians, it would need to be retrained if clinicians followed different guidelines, such as Surviving Sepsis (MAP > 65 mmHg) or Advanced Cardia Life Support (systolic blood pressure (SBP) > 70 mmHg)[27], [28].

The AR model attempts to alleviate the constrained nature of the logistic regression model by taking a different approach to episode detection. The AR model is trained to forecast values of a patient's MAP, rather than correlating features prior to an episode's onset; therefore, the AR model does not rely on the definition of an episode. The AR model could be a reasonable system that could be tested for real-time efficacy in a clinical environment. The AR model provides insight regarding the future trend of a patient's blood pressure, a rough estimation for the expectancy of future sustained hypotension, and an alerting system that could aid clinicians in the prevention of hypotension in the critically ill.

6.2: Contributions

The contributions that were made by this research are listed as the following:

- An in-depth analysis of the chronic problem of hypotension management of the critically ill in intensive care units
- Demonstrated the importance of implementing a clinical informatics system to aid in the prevention of reoccurring hypotension by examining adherence to current prevention guidelines
- Developed a statistically-driven model utilizing a simple collection of mathematical trends of a patient's prior MAP signal to detect and alert for future episodes of sustained hypotension
- Developed a time series autoregressive model to forecast a patient's future MAP signal to detect and alert for future episode of sustained hypotension
- Two proof-of-concept solutions that provides over 10 minutes of advance warning for clinicians in intensive care units to provide necessary treatment towards the prevention of sustained hypotension
- Theoretically demonstrated that the developed models can improve proactive treatment and management of reoccurring sustained hypotension while not compromising rates of false alerts.
- Demonstrated how data-driven systems may improve proactivity in managing and treating reoccurring hypotension of the critically ill

6.3: Future Work

This section briefly details ideas for the next steps of the work pertaining to this research. These proposals stem from the discussion of results from this research and the original goals set for the development of an ideal clinical informatics system for the treatment of hypotension.

6.3.1: Testing flexibility of AR models

Developing time series models for hypotension detection arose from the limitation of the logistic regression model: the models fixation to the pre-determined definition of sustained hypotension. The AR model structure was investigated to improve the flexibility of detecting hypotension. In our first investigation we analyzed the AR model to detect episodes of sustained hypotension using our pre-defined definition of sustained hypotension (i.e. MAP < 60 mmHg for at least 15 minutes). We kept this definition initially to compare the AR model with the logistic regression model we previously developed. Though, to truly understand the performance of our AR model, it flexibility must also be studied. Therefore, it would be beneficial to test the AR model against different definitions of sustained hypotension (e.g. MAP < 70 mmHg for at least 30 minutes). With a deeper knowledge of the AR model's flexibility, its advantage would be more concrete.

6.3.2: Individualized time series modeling

In both the logistic regression and AR models, a population-driven training procedure was used to fit these models to the patients' data. It is well known that across patients, the physiological state is different; additionally, across different periods of time of a single patient, the physiological state of that patient will change. Therefore, having dynamic (or updating) models may prove to be favorable. Simply, creating a routine to periodically update parameters from both models that embody various underlying physiological changes to the patient could improve the model's ability to reliably detect future episodes of sustained hypotension. Benchmarking

these updating models against the static population-trained models would ultimately show the best scheme for this type of clinical system.

6.3.3: Clinical application testing

Finally, with any tool or application to be used in the ICU, testing in real-time clinical environments must be done to prove both the model's performance and patient and caregiver safety. A system with the necessary hardware and software to mimic these derived models should be prepared that provide the necessary test bed for studying these systems in a real clinical environment for clinical trials and testing.

Appendix A: MGH Python Code for Converting .xml files to .mat Format

Code.py:

```
'''
Code to iterate through all XML files
Usage: python XML2MAT.py file.xml
'''

import glob

path = 'C:\\Users\\Bryce Yapps\\Desktop\\XML2MAT\\*xml'

for fname in glob.glob(path):

        import subprocess
        proc = subprocess.Popen(['python', 'XML2MAT.py', fname])
        import time
        print('start saving')
        print(fname)
```

XML2MAT.py:

```
'''
Code to convert BedMaster XML files to MAT files
Usage: python XML2MAT.py file.xml

creates a .mat file in the same folder with same name as xml file
'''

from datetime import datetime
from xml.etree.ElementTree import fromstring, ElementTree
import sys
import xml.etree.ElementTree as ET
import scipy.io as sio
from numpy import array
from datetime import datetime, timedelta

startTime = datetime.now()


############################################################################
# function to create datenum format object in matlab
# it takes datetime object as input and provides output as the datenum float
def datetime2matlabdn(dt):
    ord = dt.toordinal()
    mdn = dt + timedelta(days = 366)
    frac = (dt-datetime(dt.year,dt.month,dt.day,0,0,0)).seconds / (24.0 * 60.0 * 60.0)
    return mdn.toordinal() + frac
############################################################################
inFilename=sys.argv[1]
outFilename=inFilename[:-4]+'.mat'

#outFilename='test.mat'

def getmatlabTime(string):
        wv_year_mon=string.split(" ")
        wv_year_mon_spl=wv_year_mon[0].split("/")
        a1=wv_year_mon_spl[0]+" "+wv_year_mon_spl[1]+" "+wv_year_mon_spl[2]+"
"+wv_year_mon[1]+wv_year_mon[2]
        dateobject=datetime.strptime(a1,'%m %d %Y %I:%M:%S%p')
        date_time_num=datetime2matlabdn(dateobject)
        return date_time_num

WV_dict={}
```

```
wv_time_dict={}
VS_dict={}
VS_time_dict={}

def processChunk(buffer1):
        tree = ElementTree(fromstring(buffer1))
        #print buffer1
        root=tree.getroot()
        #print root.tag
        wv_time=(root.attrib['Time'])
        str1=''.join(wv_time)
        date_time_num1=getmatlabTime(str1)
        for subchilds in root:
                wave_tag= subchilds.attrib['Channel']
                wave_info=subchilds.text
                #print wave_info
                for i in wave_info.split(','):
                        i=float(i)
                        if wave_tag!="":
                                try:
                                        WV_dict[wave_tag].append(i)
                                        wv_time_dict[wave_tag].append(date_time_num1)
                                except KeyError:
                                        WV_dict[wave_tag]=[i]
                                        wv_time_dict[wave_tag]=[date_time_num1]

## process the vitalsigns part or the CML chnk

def processChunk1(buffer1):
        tree = ElementTree(fromstring(buffer1))
        #print buffer1
        root=tree.getroot()
        #print root.tag
        wv_time=(root.attrib['Time'])
        str1=''.join(wv_time)
        date_time_num1=getmatlabTime(str1)
        for subchilds in root:
                Par_V,Value_V,AlarmLimitLow_V,AlarmLimitHigh_V= "-","-","-","-"


                #print wave_info
                for sc1 in subchilds:
                        if sc1.tag == "Par":
                                Par_V=sc1.text
                        if sc1.tag == "Value":
                                Value_V = float(sc1.text)
                        if sc1.tag == "AlarmLimitLow_V":
                                AlarmLimitLow_V=sc1.text
                        if sc1.tag == "AlarmLimitHigh_V":
                                AlarmLimitHigh_V=subnodes2.text
                if Value_V != "-":
                        Par_V= Par_V.replace("-","")
                        #print Par_V
                        try:
                                VS_dict[Par_V].append(Value_V)
                                VS_time_dict[Par_V].append(date_time_num1)
                        except KeyError:
                                VS_dict[Par_V]=[Value_V]
                                VS_time_dict[Par_V]=[date_time_num1]

inputbuffer=''
with open(inFilename) as inputFile:
        append=True
        for line in inputFile:
                #print line
                #print '-----'
                if '<Waveforms' in line:
                        inputbuffer=line
                        append=True
                elif '<VitalSigns' in line:
                        inputbuffer=line
```

```
                append=True
            elif '</VitalSigns>' in line:
                inputbuffer +=line
                append=False
                processChunk1(inputbuffer)
                inputbuffer=None
                del inputbuffer
            elif '</Waveforms>' in line:
                inputbuffer +=line
                append=False
                processChunk(inputbuffer)
                inputbuffer=None
                del inputbuffer
            elif append:
                inputbuffer +=line

sio.savemat(outFilename, {'vs': VS_dict, 'vs_time' : VS_time_dict})
print('Saved!')
```

Bibliography

[1] J.-L. Vincent and D. De Backer, "Circulatory Shock," *N. Engl. J. Med.*, vol. 369, no. 18, pp. 1726–1734, Oct. 2013.

[2] M. H. Weil and H. Shubin, "Proposed Reclassification of Shock States with Special Reference to Distributive Defects," in *The Fundamental Mechanisms of Shock*, L. B. Hinshaw and B. G. Cox, Eds. Springer US, 1972, pp. 13–23.

[3] P. Dellinger, "Cardiovascular management of septic shock," *Concise Defin. Rev. Crit. Care Med.*, vol. 31, no. 3, pp. 946–955, 2003.

[4] P. J. O'Connor, J. M. Sperl-Hillen, P. E. Johnson, W. A. Rush, and G. Biltz, "Clinical inertia and outpatient medical errors," in *Advances in Patient Safety: From Research to Implementation (Volume 2: Concepts and Methodology)*, K. Henriksen, J. B. Battles, E. S. Marks, and D. I. Lewin, Eds. Rockville (MD): Agency for Healthcare Research and Quality (US), 2005.

[5] G. Moody and L. Lehman, "Predicting acute hypotensive episodes: The 10th annual PhysioNet/Computers in Cardiology Challenge," in *2009 36th Annual Computers in Cardiology Conference (CinC)*, 2009, pp. 541–544.

[6] F. Chiarugi *et al.*, "Predicting the occurrence of acute hypotensive episodes: The PhysioNet Challenge," in *2009 36th Annual Computers in Cardiology Conference (CinC)*, 2009, pp. 621–624.

[7] J. Henriques and T. Rocha, "Prediction of acute hypotensive episodes using neural network multi-models," in *2009 36th Annual Computers in Cardiology Conference (CinC)*, 2009, pp. 549–552.

[8] M. Mneimneh and R. Povinelli, "A rule-based approach for the prediction of acute hypotensive episodes," in *2009 36th Annual Computers in Cardiology Conference (CinC)*, 2009, pp. 557–560.

[9] H. Cao, L. J. Eshelman, L. Nielsen, B. D. Gross, M. Saeed, and J. J. Frassica, "Hemodynamic Instability Prediction Through Continuous Multiparameter Monitoring in ICU," *J. Healthc. Eng.*, vol. 1, no. 4, pp. 509–534, 2010.

[10] G. A. F. Seber and A. J. Lee, *Linear Regression Analysis*. John Wiley & Sons, 2012.

[11] D. Hosmer and S. Lemeshow, *Applied Logistic Regression*. John Wiley & Sons, 2004.

[12] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 2015.

[13] M. Saeed *et al.*, "Multiparameter intelligent monitoring in intensive care II: a public-access intensive care unit database," *Crit. Care Med.*, vol. 39, no. 5, pp. 952–960, 2011.

[14] J. E. Zimmerman and A. A. Kramer, "Outcome prediction in critical care: the Acute Physiology and Chronic Health Evaluation models:," *Curr. Opin. Crit. Care*, vol. 14, no. 5, pp. 491–497, Oct. 2008.

[15] P. R. Norris and B. M. Dawant, "Closing the loop in ICU decision support: physiologic event detection, alerts, and documentation.," *Proc. AMIA Symp.*, pp. 498–502, 2001.

[16] D. A. Harrison, G. J. Parry, J. R. Carpenter, A. Short, and K. Rowan, "A new risk prediction model for critical care: The Intensive Care National Audit &

Research Centre (ICNARC) model*:," *Crit. Care Med.*, vol. 35, no. 4, pp. 1091–1098, Apr. 2007.
[17] G. B. Moody and R. G. Mark, "A database to support development and evaluation of intelligent intensive care monitoring," in *Computers in Cardiology, 1996*, 1996, pp. 657–660.
[18] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000.
[19] Centers for Disease Control and Prevention, USA, "HIPAA privacy rule and public health. Guidance from CDC and the U.S. Department of Health and Human Services," *MMWR Morb. Mortal. Wkly. Rep.*, vol. 52, no. Suppl. 1, pp. 1–17, 19–20, 2003.
[20] F. A. Finnerty, L. Witkin, and J. F. Fazekas, "Cerebral Hemodynamics during Cerebral Ischemia Induced by Acute Hypotension1," *J. Clin. Invest.*, vol. 33, no. 9, pp. 1227–1232, Sep. 1954.
[21] W. Fitch, E. T. MacKenzie, and A. M. Harper, "Effects of decreasing arterial blood pressure on cerebral blood flow in the baboon. Influence of the sympathetic nervous system," *Circ. Res.*, vol. 37, no. 5, pp. 550–557, 1975.
[22] M. Legrand *et al.*, "Association between systemic hemodynamics and septic acute kidney injury in critically ill patients: a retrospective observational study," *Crit. Care Lond. Engl.*, vol. 17, no. 6, p. R278, 2013.
[23] 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, 2013.
[24] M. Poukkanen *et al.*, "Hemodynamic variables and progression of acute kidney injury in critically ill patients with severe sepsis: data from the prospective observational FINNAKI study," *Crit. Care*, vol. 17, p. R295, 2013.
[25] M. Walsh *et al.*, "Relationship between Intraoperative Mean Arterial Pressure and Clinical Outcomes after Noncardiac Surgery: Toward an Empirical Definition of Hypotension," *Anesthesiology*, vol. 119, no. 3, pp. 507–515, Sep. 2013.
[26] T. J. Iwashyna, E. W. Ely, D. M. Smith, and K. M. Langa, "Long-term cognitive impairment and functional disability among survivors of severe sepsis," *Jama*, vol. 304, no. 16, pp. 1787–1794, 2010.
[27] R. P. Dellinger *et al.*, "Surviving Sepsis Campaign: International Guidelines for Management of Severe Sepsis and Septic Shock, 2012," *Intensive Care Med.*, vol. 39, no. 2, pp. 165–228, Jan. 2013.
[28] E. M. Antman *et al.*, "ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction - Executive summary," *Can. J. Cardiol.*, vol. 20, no. 10, 2004.
[29] G. Hawryluk *et al.*, "Mean Arterial Blood Pressure Correlates with Neurological Recovery after Human Spinal Cord Injury: Analysis of High Frequency Physiologic Data," *J. Neurotrauma*, vol. 32, no. 24, pp. 1958–1967, Feb. 2015.
[30] V. Bewick, L. Cheek, and J. Ball, "Statistics review 13: Receiver operating characteristic curves," *Crit. Care*, vol. 8, p. 508, 2004.

[31] R. Bighamian, A. T. Reisner, and J. O. Hahn, “An Analytic Tool for Prediction of Hemodynamic Responses to Vasopressors,” *IEEE Trans. Biomed. Eng.*, vol. 61, no. 1, pp. 109–118, Jan. 2014.