UNIVERSITY OF CAPE COAST

FLEXIBLE FEATURE-BASED GAUSSIAN PROCESS REGRESSION FOR MULTI-PHYSIOLOGICAL VITAL SIGNS MODELING

EMMANUEL AMEWONOR

2022

UNIVERSITY OF CAPE COAST

FLEXIBLE FEATURE-BASED GAUSSIAN PROCESS REGRESSION FOR MULTI-PHYSIOLOGICAL VITAL SIGNS MODELING

BY

EMMANUEL AMEWONOR

PS/STP/18/0003

Thesis submitted to the Department of Statistics of the School of Physical Sciences, College of Agriculture and Natural Sciences, University of Cape Coast, in partial fulfillment of the requirements for the award of Master of Philosophy in Statistics.

SEPTEMBER, 2022

DECLARATION

Candidate's Declaration

I hereby declare that this thesis is the result of my own original research and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature ..... Date .....

Name: EMMANUEL AMEWONOR

Supervisors' Declaration

We hereby declare that the preparation and presentation of the thesis were supervised in accordance with the guidelines on supervision of thesis laid down by the University of Cape Coast.

Principal Supervisor's Signature ..... Date .....

Name: DR. DAVID KWAMENA MENSAH

Co-Supervisor's Signature ..... Date .....

Name: DR. FRANCIS EYIAH-BEDIAKO

ABSTRACT

This thesis considered a flexible two-stage statistical approach to multi-task modeling of multivariate physiological vital signs using GP regression, in which the joint use of nonparametric and Bayesian GP regression methods are explored. In the first stage, nonparametric schemes based on expected value contribution statistics for fusing multiple physiological vital signs observed over common time-stamps into a composite vital sign are developed. In the second stage, an appropriate Bayesian Gaussian process regression model is developed for the fused vital sign trajectory in relation to the common observation time-stamps. The relationship existing among the multiple vital signs and available non-time-dependent covariates is modeled with the aid of OGK statistics via the covariance function of the assumed GP. Both Variational Bayes and MCMC methods are developed for parameter inference. The coupling of density-based data fusing methods and GP modeling allowed automated extreme value control within both the response and predictor spaces; response dimension reduction; data reduction in the response space and principled modeling of smoothness of the physiological trend. Using both simulation and real data application, the utility of the proposals is illustrated. In terms of fusing of multivariate vital signs, results show that the probability distribution-based features provide a rich source of appealing functional features with the natural ability to ensure that extreme observations are utilized with their effects controlled automatically. For the GP modeling of fused vital signs, the results show that both VB and MCMC algorithms exhibit better fitting performance in terms of MSFE, MAFE, and SMAFE. Thus, the double-stage modeling approach exhibits a great potential for handling multi-task GP regression within the single-task GP framework.

KEY WORDS

Variational Bayes

Kernel density estimation

Gaussian Process Regression

Heart rate

Markov Chain Monte Carlo (MCMC)

Oxygen saturation (SpO_2)

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Dr. David Kwamena Mensah of the Department of Statistics, for his seasoned guidance, advice, encouragement, and the goodwill with which he supervised this work.

I am also grateful to Dr. Francis Eyiah-Bediako, the Head of Statistics Department, and the former Head of Statistics, Prof. Bismark Nkansah of the University of Cape Coast, for their contribution toward the success of this work.

Again, I am thankful to all the lecturers of the Department of Statistics of the University of Cape Coast for their support and contribution towards the successful completion of my programme.

DEDICATION

To my family: Seshie A. Sarah, Aaron, Michael and Miriam.

TABLE OF CONTENTS

Page

DECLARATION i

ABSTRACT ii

KEY WORDS iii

ACKNOWLEDGEMENTS iv

DEDICATION v

LIST OF TABLES ix

LIST OF FIGURES x

LIST OF ABBREVIATION xiii

CHAPTER ONE: INTRODUCTION

Background to the Study 1

Statement of the Problem 7

Objectives of the Study 8

Significance of the Study 9

Delimitation 9

Limitation 9

Organization of the Study 10

Chapter Summary 10

CHAPTER TWO: LITERATURE REVIEW

Introduction 12

Vital Signs 13

Heart Rate 14

Respiratory Rate 15

Oxygen Saturation (SpO_2) 16

Blood Pressure 17

Kernel Density Estimation 19

Gaussian Processes for Vital Sign Modeling 19

Chapter Summary 23

CHAPTER THREE: METHODOLOGY

Introduction 24

Vital Signs Data Fusion 25

Data Fusion with Automated Outlier Control 26

Fusion of Multivariate Data into Univariate Data 29

Gaussian Process Model for Fused Vital Signs Data 31

Bayesian Gaussian Process Approximation 33

Variational Inference for Gaussian Pocess Model 37

MCMC Inference for Gaussian Process Model 39

Measures of Performance for Assessment of Proposed Methods 40

Implementation of Methods 42

Chapter Summary 43

CHAPTER FOUR: RESULTS AND DISCUSSION

Introduction 45

Simulation 45

Health Data Application 57

Data Description 57

Data Pre-processing 58

Preliminary Data Analysis for Health Data 58

Exploratory Analysis Method 59

Preliminary Data Analysis Results 59

Application of Methods 63

Univariate Vital Signs Feature and Data Fusion Application 64

Gaussian Process Regression Application on Fused Vital Signs Data 75

Chapter Summary 79

CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Summary 81

Conclusions 83

Recommendations 84

APPENDICES

List of Tables

1 **Comparison of Variational Bayes (VBSBPM), and its MCMC Counterpart for θ for the Simulated Data Set. Columns 2 and 3 give the Parameter Estimates for VBFBGPR and HGS Respectively. The generated true θ value was 2.91. The last two columns give their corresponding 95% Bayesian credible intervals (BCIs).** 55

2 **Summary Statistics of Physilogical Vital Signs with Sample Size $\mathbf{n = 848 \times 5}$.** 60

3 **Comparison of Vital Sign Specific α and δ for Standardized Dataset. $\mathbf{y_1 = SBP}, \mathbf{y_2 = DBP}, \mathbf{y_3 = MAP}, \mathbf{y_4 = PP}, \mathbf{y_5 = HR}$, $\mathbf{y_c = \sum_{i=1}^{5} y_i}$** 74

4 **Comparison of Vital Sign Specific α and δ for Unstandardized Dataset. $\mathbf{y_1 = SBP}, \mathbf{y_2 = DBP}, \mathbf{y_3 = MAP}, \mathbf{y_4 = PP}, \mathbf{y_5 = HR}$, $\mathbf{y_c = \sum_{i=1}^{5} y_i}$** 74

5 **Fitting Performance of VB and HGS Based on MSEFE for selected $\mathbf{m}$.** 78

6 **Fitting Performance of VB and HGS Based on MAFE for Selected $\mathbf{m}$.** 78

7 **Fitting Performance of VB and HGS Based on SMAFE Selected $\mathbf{m}$.** 79

List of Figures

		Page
1	Probability Density Function of the Random Variable y.	28
2	Nature of the Problem.	28
3	Nature of Simulated Feature, $Z_{(y)}$ and the Gaussian Processes, $f(x)$ and $g(x)$	46
4	Plot of Lower Bound Attained at Convergence for VBFBGPR with $m = 25, 30, 35, 40$.	47
5	Plot of Lower Bound Attained at Convergence for VBFBGPR with $m = 45, 55, 65, 75$.	47
6	Plot of Lower Bound Attained at Convergence for VBFBGPR with $m = 85, 95, 100$.	48
7	Convergence Features of VBFBGPR Based on Synthetic Data. Plot of Lower Bound Attained at Convergence Against Spectral Sample Size (m) (left). Plot of Number of Iterations Against Number of Spectral Sample Size (m) (right).	49
8	Nature of Fitted Trends for VBFBGPR and HGS with $m = 25, 30, 35, 40$.	50
9	Nature of Fitted Trends for VBFBGPR and HGS with $m = 45, 55, 65, 75$.	51
10	Nature of Fitted Trends for VBFBGPR and HGS with $m = 85, 95, 100$.	52
11	Plot of Average Squared Fitted Error for VBFBGPR and HGS over m.	53
12	Plot of Spectral Sample Size m Against Computational Time for VBFBGPR and HGS	54

13 Marginal Posterior Densities of θ Corresponding to m for VBF-BGPR. 56

14 Scatter Plots of Vital Signs. From Left to Right are the Plots of SBP, DBP, MAP, PP, and HR Respectively. 61

15 Boxplot of Vital Signs. From Left to Right are the Plots of SBP, DBP, MAP, PP, and HR Respectively. 62

16 Nature of Vital Signs Data Showing Densities, Inter-relationships and Pearson Correlation Coefficients. 63

17 Nature of Vital Signs Specific Density Plots for Standardized and Unstandardized physiological data. 65

18 Nature of Vital Sgns Specific $T(y)$ Statistics for Standardized and Unstandardized Physiological Data. 65

19 Nature of Vital Signs Specific $V(y)$ Statistics for Standardized and Unstandardized Physiological Data. 66

20 Nature of Vital Signs Specific $S_1(y)$ Statistics for Standardized and Unstandardized Physiological Data. 66

21 Plot of Vital Signs with their Corresponding Order Statistics Trend and $S_1(y)$. 67

22 Plot of Vital Signs with their Corresponding Order Statistics Trend and $S_1(y)$. 69

23 The effect of α_e on $S_1(y)$ Statistics. From Left to Right are the Plots of $S_1(y)$ for 6 Randomly Selected α_e Values Corresponding to $0.15, 0.50, 0.10, 0.80, 0.90$ and 1.00 Values. 70

24 Real data: Dynamics of Vital Sign Specific Features with Vital Sign Specific α_r. 70

25 Nature of Vital Sign Specific Tuning Parameter α_r. 71

26 Dynamics of Vital Signs Specific ρs for Statistics for Standardized and Unstandardized Physiological Data. 71

27 Fused Vital Signs Trend Based on $T(y)$ Statistic. 73

28 Fused Vital Signs Trend Based on $S_1(y)$ Statistic. 73

29 Plot of Llower Bound against Spectral Sample Size m for VBF-BGPR. 76

30 Plot of Iteration against Spectral Sample Size m. for VBFBGPR. 76

31 Fitted Fussed Vital Sign with $m = 40$ (row 1) and $m = 45$ (row 2) 77

List of Abbreviations

GPR	Gaussian Process Regression
KDE	Kernel Density Estimation
MAFE	Mean Absolute Fitted Error
MTGPR	Multi-Task Gaussain Process Regression
OGK	Orthogonalized Gnanadesikan-Ketterning
SMAFE	Standardized Mean Absolute Fitted Error
VBSBPM	Variational Bayes for systolic blood pressure model

CHAPTER ONE

INTRODUCTION

Stochastic modeling of physiological vital signs via Bayesian Gaussian process regression (BGPR) has gained attention within the functional data analysis community recently due to the flexibility with which available information can be utilized in model specification to encode assumed modeling assumptions in BGPR. This has motivated the application of BGPR to physiological vital signs with varying time-stamps datasets within both the single-task and multi-task modeling framework, see for example, Evans, Hodgkinson, and Berry (2001), Gardner-Thorpe, Love, Wrightson, Walsh, and Keeling (2006), Dürichen, Pimentel, Clifton, Schweikard, and Clifton (2014), and Ghassemi et al. (2015).

For multivariate physiological vital sign datasets with common time-stamps and associated with covariates information involving both time-dependent and non-time dependent covariates, the application of BGPR is nontrivial. In particular, how to incorporate the non-time dependent covariate information in the BGPR modeling. The single-task BGPR models a univariate functional dataset with a GP in the Bayesian paradigm whiles the multi-task BGPR extends the BGPR to multiple functional datasets jointly using composite covariance derived from all tasks.

Background to the Study

The modeling of multi-physiological vital signs is gaining attention within the functional data modeling framework based on Gaussian processes regression recently. This is due partly first to the recent spate of pandemics, for example, COVID 19, since pandemics affect human physiology in a way that is not fully understood by scientists. Second, partly due to the availability of physiological vital signs datasets in several scientific fields such as public health, etc. A pandemic that affects human physiology has the potential to render destructive effects on the natural immunity of the human body. When this happens, the hu-

man system becomes prone to several unforeseen issues since immunity can be viewed as a priceless natural building block of the defense system of the human body.

Physiological vital signs, functional readings of the vital organs of the human body, are key elements in health monitoring worldwide (Buist, Burton, Bernard, Waxman, & Anderson, 1999; Evans et al., 2001). These defining features of human physiology can provide insights on early signs of health deterioration if appropriately tracked and monitored. This has led to their wide use in the identification of deterioration in health, in many modeling settings (Gao et al., 2007; Khalid, Clifton, & Tarassenko, 2013). An empirical examination of physiological vital signs using a numerical measure termed Early Warning (EW) scores obtained as a composite score of the associated vital signs have been extensively employed for assessing health deterioration of intensive care unit (ICU) and surgical patients (Evans et al., 2001; Gardner-Thorpe et al., 2006).

Quite a number of vital signs-based health monitoring work in the nonparametric modeling, particularly, kernel density approach has also been registered in the literature. See, for example, Pimentel, Clifton, Clifton, Watkinson, and Tarassenko (2013), who considered modeling multivariate vital signs of upper-gastrointestinal surgery patients in order to identify deterioration. They utilized kernel density estimation (KDE) (Silverman, 1986) approach to building probability density function (pdf)-based monitors for each vital sign characterizing normal and abnormal groups. The grouping was motivated by available patient admission information. Informed decision-making was based on comparison of pdf-based novelty scores with a common threshold computed from the normal group. Velardo et al. (2014) also prioritized chronic obstructive pulmonary patients for clinical review using KDE-based vital sign deterioration detectors in both the univariate and multivariate dimensions.

Nonparametric modeling of clinical data using Gaussian process regression has attracted much attention in the literature. A Gaussian process can simply be viewed as a generalization of the multivariate normal distribution in infi-

nite dimension (Rasmussen & Williams, 2006). Khalid et al. (2013) considered the detection of deterioration in physiological vital signs in the univariate dimension, based on the Gaussian process regression modeling. In comparison with an appropriate empirical approach using the Parzen Window KDE, their proposal reported better performance in terms of lowering the false alarm rate. Gaussian processes provide natural flexibility that allows easy encoding of assumptions about an unknown function that modeling wants to learn via either its mean function or the covariance function (Rasmussen & Williams, 2006).

The above flexibility has resulted in the increasing application of Gaussian processing regression (GPR) and its multi-task extension (MGPR) in physiological vital signs modeling within the modeling functional clinical data analysis field. The multi-task Gaussian process regression extends the usual GPR to modeling multiple vital signs co-currently based on composite covariance obtained from related tasks. It considers the union of all time-stamps of the multivariate vital signs but modeling is fitted for each vital sign separately leading to appropriate joint modeling and estimation for unified inference. With the application of the standard GP formulation and technique, exact inference for MTGPR modeling is achievable. For comprehensive details on the exact inference for MTGPR modeling, see, for example, (Williams, Bonilla, & Chai, 2007). Dürichen et al. (2014) and Ghassemi et al. (2015) adopted the multi-task GP approach to modeling and forecasting clinical multivariate time series data. In particular, Ghassemi et al. (2015) classified intensive care unit (ICU) patients based on the acuity of illness using multi-task GPR. The severity of illness assessment was done using the estimated hyperparameters of the fitted MTGPR modeling with heterogeneous and irregularly sampled clinical multivariate time series from ICU patients.

The authors reported improved performance of their method over an appropriate single-task GPR modeling. Other existing work on the application of MTGPR to clinical time series can be found in Marlin, Kale, Khemani, and

Wetzel (2012), Lasko (2014), Schulam, Wigley, and Saria (2015), Lichman and Smyth (2014), and Clifton, Clifton, Pimentel, Watkinson, and Tarassenko (2013). In each of the above works, it has been observed by Liu and Hauskrecht (2016) that, it seems the dependencies among the different time series were not captured due to the use of forecasting modeling for each time series separately.

To address the above drawback, Liu and Hauskrecht (2016) considered a two-stage modeling perspective using linear dynamical system modeling and multi-task Gaussian process regression modeling for clinical multivariate time series dataset. The authors first applied a linear dynamical system to model the population trend and secondly, employed MTGPR to capture the individual-specific short-term variability. As a result, the interactions among all variables and their dynamics were modeled. With the above two-stage modeling framework, the work of Liu and Hauskrecht (2016) exhibited the potential to learn the population trend from a set of time series from past patients, capture the subject-specific short-term multivariate variability and ensure automatic adaptation of predictions by adjustment using new observations.

The predictive performance of their modeling was better than their appropriate population-based and patient-specific time series counterparts. It is straightforward to observe that the problem of interest in the above works reviewed is centered on joint modeling of multivariate clinical time-series data from varying sources. Despite the attractiveness of the proposals by the authors outlined above, no clinical covariate information was considered or modeled in relation to the multivariate responses. Omitting available clinical covariate information in statistical modeling constitutes a loss of vital information that might aid in the detection and identification of important indicators of the condition of health or otherwise.

Handling of non-time dependent covariates within the Gaussian process regression framework is non-trivial due to the underlying challenges associated with the principled incorporation of such information into the modeling. Although, the GPs have appealing flexibility regarding parameters such as the

mean and covariance functions that can be adapted to encode modeling assumptions, inappropriate modeling of parameters may introduce some intractability into the modeling. Mensah, Nott, Tan, and Marshall (2016) proposed probabilistic modeling of non-time dependent subject-specific covariate information within the GPR framework via the smoothing parameter of covariance of the assumed subject-specific GP modeling. The approach was developed for fitting Bayesian GPR modeling for grouped functional longitudinal data. However, due to the subject-wise modeling of covariate information, the challenges associated with the incorporation of such information were not fully encountered.

Based on the work of Mensah et al. (2016), Ofori (2020) proposed a novel approach for modeling non-time dependent via extraction of robust statistics termed Orthogonalized Gnanadesikan-Kettering (OGK) statistics (Maronna, Martin, Yohai, & Salibián-Barrera, 2019) from the matrix of covariate information. This reduces the non-time dependent covariate designed matrix to a designed vector for easy modeling using either the mean function or the covariance function. This approach is appealing to the best of our knowledge because it has the potential to ensure computational speed-ups in Variational Bayes inference algorithms Mensah, Ofori, and Howard (2021). This approach to covariate modeling is adopted and adapted in this thesis.

Data on physiological vital signs can be univariate or multivariate in nature depending on their source and the sort of generation process. Focusing on multivariate vital sign data type and considering the generating processes within a common medium such as the human system in public health. The generation of multivariate physiological vital signs data usually occurs within the same time-space for some scientific phenomenons. In such situations, it is often of interest to examine the dynamics of these multi-variable vital signs with their relationship with some available health covariates within the common time domain to reduce or circumvent the challenges, for example, modeling and computational burdens, associated with examining them individually using the common time domain.

The above interest calls for a novel modeling perspective in which a two-stage modeling formalism is utilized in a formal way. In such a modeling perspective, first, the multi-vital signs must be fused into a single variate appropriately preserving the intrinsic features as well as borrowing information across the various vital signs for better estimation of free parameters of the corresponding modeling. Secondly, the modeling of composite vital signs data should be flexible enough to be able to incorporate available non-time dependent covariates for informed decision-making. Nevertheless, non-functional (time) dependent covariates generated are generally complex leading to complex design matrices in modeling. Thus, introducing a myriad of challenges in modeling formulation, and, subsequently, fitting and inference.

Interestingly, physiological vital signs are known to exhibit some interrelationships and this information has aided the conventional approach to healthcare provision in public health where medical doctors make health status decisions based on some baseline standard performance measurements of the vital signs observations obtained from vital organs of the human system. This fundamental information allows health monitoring to be executed via simple monitoring of indicators of health with major components being the physiological vital signs. In this thesis, we propose a novel two-stage approach for modeling multivariate vital signs data with common time stamps and endowed with non-time dependent covariates. In particular, in the first stage of the above two-stage modeling framework, nonparametric methods are explored adopting the expected value contribution approach of Mensah, Eyiah., and Assabil (2021) to develop appropriate feature-based schemes for fusing multivariate physiological vitals signs.

The second stage, an exploration of the Gaussian process regression was adopted to develop appropriate single talks Bayesian GPR for the fused data via the features, allowing incorporation of vital sign-specific covariate information. The covariate modeling considered here will be based on the novel approach proposed by Mensah, Ofori, and Howard (2021).

Statement of the Problem

The increasing availability of multivariate clinical datasets in pubic health due partly to the recent outbreak of pandemic and partly, to technological advancements has revived an increasing interest in fast inferential methods for informed decision within the computational statistics community. This has resulted in an extensive application of Bayesian Gaussian process regression (GPR) modeling techniques and its multi-tasks extension (MTGPR) to such datasets, see, for example, Clifton et al. (2013); Dürichen et al. (2014); Ghassemi et al. (2015); Khalid et al. (2013); Lasko (2014); Lichman and Smyth (2014); Liu and Hauskrecht (2016); Marlin et al. (2012); Mensah, Ofori, and Howard (2021); Schulam et al. (2015). GPs generalized multivariate Gaussian random vectors to infinite dimensions (Rasmussen & Williams, 2006). Whiles a GPR is used for a single task (univariate modeling), an MTGPR considers multiple tasks modeling jointly borrowing information across the various tasks.

That is, an MTGPR can be viewed as an extension of the GPR for modeling multiple tasks simultaneously using learned covariance between related tasks. The MTGPR was adopted by Dürichen et al. (2014) for a biomedical application in terms of multiple vital signs monitoring and Ghassemi et al. (2015) for intensive care unit (ICU) patients classification using acuity of illness. Clifton et al. (2013). Liu and Hauskrecht (2016) observed that the dependencies among the different tasks were not fully captured in the above works and they proposed the joint use of the linear dynamical system and multi-task Gaussian process regression modeling framework for modeling clinical multivariate time series dataset.

Also, for a multivariate dataset with p variables, the application of GPR in a multi-task approach without any approximation to the covariance function may increase the above computation and memory scalings. In addition, for multivariate vital signs data with common time stamps, the use of MTGPR may

lead to repeated use of the common timestamp since the MTGPR considers the union of all timestamps associated with different tasks. Another obvious issue is the integration of available covariates especially, the non-time dependent ones. Although the Bayesian approach to inference was considered in the proposals suggested by the authors mentioned above, no clinical covariate information was modelled in relation to the multivariate responses. Omitting available clinical covariate information may lead to loss of information that may affect the performance of fitted statistical modeling.

Though Ofori (2020) considered resolving the issue of incorporation of non-time dependent covariates in the Variational Bayes GPR using the idea of robust statistics (Orthogonalized Gnanadesikan-Kettering (OGK)) (Maronna et al., 2019), it was in the one-dimension direction, just for traumatic systolic blood pressure response. Modeling non-time dependent covariates in the multi-task GPR has not been being addressed in the literature. This thesis aims to develop novel flexible statistical methods for handling multi-tasks GPR problems with common task-specific predictors within the single task GPR modeling framework.

Objectives of the Study

The main objective of this research is to develop inferential methods for reducing multi-task Bayesian Gaussian process regression (MTBGPR) to single-task BGPR. The research is designed based on the following specific objectives.

1. To develop automatic extreme value controlled nonparametric schemes for fusing multivariate physiological vital signs into a single variate vital sign

2. To develop appropriate feature-based Gaussian process regression model for the composite vital sign

3. To develop appropriate inference methods in the context of Variational Bayes and MCMC methods

Significance of the Study

Multivariate physiological vital signs observed over the same time domain require separate models for each vital sign against the common time. The application of multi-task modeling within the Gaussian process framework poses a challenge due to the construction of common covariance functions as well as the computational burden associated with such models. This study offers some significant benefits in the following ways. Firstly, it provides a simple but principled way to transform multi-task GPR problems into single-task GPR problems for physiological vital signs. This transformation leads to significant computational savings allowing easy application to the monitoring of physiological vital signs in practice.

Delimitation

The study was conducted in Ghana but use a data from Biofourmis Private Company Limited, a Health Data Analytic Company in Singapore.
Another delimitation is, the study was conducted in relation to physiolocal vital signs

Limitation

This study was conducted mainly using key variables in the physiological vital signs. The result of the reaserch cannot be generalized in Ghana to reflect the health status of all. Other challenges include time limitation. Since this study has a short duration to be completed.
Also the technical limitation is a major drawback since the researcher do not have high specification computer to run the analysis and have to relied on other departmental labs.

- Kernel Density Estimation: A non-parametric approach for modeling the joint distribution of multiple random variables empirically.

- Gaussian Process: A generalization of the multivariate normal distribution in infinite dimension (Rasmussen & Williams, 2006).
- Gaussian Process Regression: An application of GP in functional regression gives the GPR (single task GPR).
- Multi-Task Gaussain Process Regression: An extension of the GPR for modeling multiple tasks simultaneously using learned covariance between related tasks.

Organization of the Study

This thesis is divided into five chapters under the headings: Introduction, Literature Review, Methodology, Analysis and Summary, Conclusions, and Recommendations. Chapter one is the introduction of the thesis. It presents the background of the study, statement of the problem, objectives, and the outline of the study. Chapter two reviews literature related to the study. It performs an extensive review of existing literature on the Gaussian process and its application in regression in the Bayesian paradigm.

Chapter three deals with methodological development. In particular, the fundamental theoretical aspects of the statistical and computational methods are developed. This entails the modeling framework, inferential methods, technique for data collection, data collection instrument or tool, population and sampling, and data analysis.

Chapter four focuses on the implementation of the developed computational methods in simulation and real data application. Also, exposition on the data analysis and its implication. Finally, Chapter five summarizes the work, presents the conclusions and recommendations.

Chapter Summary

The study aims at developing flexible statistical methods within the Bayesian Gaussian process regression framework for fusing multivariate physiological vi-

tal signs dataset into a composite univariate vital sign and modeling the resultant vital sign trajectory using Gaussian process regression. In other words, transforming multivariate time-space into univariate time-space and applying the Bayesian Gaussian process regression model was the focus of this thesis. This chapter provides the fundamental elements required for gaining insights into the big picture proposed in terms of the major statistical problem of interest intended to find a solution via research. The chapter does it in a structured manner, providing the basic information for easy navigation.

The chapter begins with a general introduction, in brief, followed by a scientific background. The major problem identified is formally stated scientifically next, providing the appropriate perspective for a better understanding of the problem. With the problem formally stated, the objectives of the study become apparently clear and they are formally outlined next.

Existing work in the literature with an appreciable relationship with the current work is outlined and captured into related work. Also, the novelty and contribution to literature are clearly outlined and captured into a state of contribution. Finally, the chapter ends with a road map for the rest of the thesis captured as the organization of the study.

CHAPTER TWO

LITERATURE REVIEW

Introduction

Attention has been given to the fact that, the ill-health of every individual has a direct link to the deterioration in physiological vital signs and that adverse consequences can be avoided by detecting this deterioration early and providing pragmatic treatment (McQuillan et al., 1998). In the early 1990s, it was recognized that the frequent changes in the breathing patterns of individuals, for example, with the aid of other parameters, help to diagnose patients with cardiac arrest which is about 38% accuracy (Bedell, Deitz, Leeman, & Delbanco, 1991; Fieselmann, Hendryx, Helms, & Wakefield, 1993).

The first nationwide survey study (Kause et al., 2004), shows that frequent changes in physiological patterns of individuals have prior consequences on events as cardiac arrest. Kause et al. (2004), are of the view that patients will be admitted to the ICU and the associated death rate which was conducted within three countries such as UK, New Zealand and Australia with 90 hospitals. It was noticed that in three days while these patients were been observed, 60% of the patient's severity was also having this effect in the vital signs just 24 hours of occurrence. This involved 79.4% of the cardiac cases, 54.5% of patients within the ICU and, 11.7% of deaths recorded.

In 2011 - 2012, almost 35% of admissions cases within the UK were grouped as emergency cases (Blunt, Bardsley, & Dixon, 2010). According to a study conducted by (Renton et al., 2011), out of 27 studies on the issues of adverse events before admission to the ICU shows that 37.2% will be admitted to the ICU and 18.3% of these cases will be re-admitted to the ICU. These researchers are of the view that, it should be a priority of medical practitioners to have a system that monitors patients, which will help physicians to make good diagnosis and also aid in the early detection of deterioration of these vital signs in patients. It will also help individual with chronic conditions such as

high blood presure to seek medical care early in case of any change in their vital signs.

Vital Signs

Thran (2018), describes vital signs as the individual's health status within a particular point in time. These vital signs are measured and recorded to aid in the assessment of the general well-being of individuals, with the possibility of detecting any diseases, and taking track of healing progress Lockwood, Conroy-Hiller, and Page (2004). Edmonds, Mower, Lovato, and Lomeli (2002) considers heart rate (HR), respiratory rate (RR), blood pressure (BP), and temperature and oxygen saturation (SpO_2) as the essential health status measuring sign for a person's wellbeing. Even though there are international standards in the interpretation of these vital sign figures, there are a few assumptions as to the best practice of the caring facilities with its protocols involved. Schulman and Staul (2010) gave their opinion on how vital signs are measured and recorded. They are also with the view that accepting the baseline from the international community should not be encouraged since various factors affect an individual's health.

The health experts from these facilities are to spearhead the establishment of their own standard based on the individual needs. The authors also suggested that, the rate at which these vital signs are accessed, some vital signs are to be checked in everyone one hour or even below the hourly rate depending on the patients' stability. In view of this, the framework of assessment can be considered from Hardin and Kaplow (2005), based on stability, complexity, predictability, vulnerability, and resiliency. McGhee, Weaver, Solo, and Hobbs (2016), expressed their opinion on the frequency at which emergency departments are prone to missing records on vital signs for patients. Even though these may serve as a major guide in deteriorating patients.

Heart Rate

Heart rate is measured in beats per minute (beats per min). Its normal range, at rest, varies with age from 110 - 160 beats per min at birth to approximately 60 - 100 beats per min by late adolescence (Malik & Camm, 1990). During adulthood, heart rate varies widely between individuals depending on several factors such as physical activity, body composition, stress levels, or disease. An unusual high heart rate (known as tachycardia) or low rate (known as bradycardia) may indicate an underlying clinical problem.

The electrocardiogram (ECG) is the global standard method for measuring heart rate. It records the electrical activity of the heart with electrodes placed in contact with the patient's skin. Typical configurations for recording the ECG, depending on the clinical information desired, are either 3 electrodes on the chest or 10 to 12 electrodes placed at various locations on the patient's upper body (Clifford, Azuaje, McSharry, et al., 2006).

It is also possible to measure heart rate by using a pulse oximeter, a non-invasive optical sensor usually positioned on the fingers, toes, or ear lobes. Pulse oximeter has made significant contributions to the non-invasive monitoring of physiological signals in a wide variety of clinical situations. Pulse oximeters are regularly used in primary care, the emergency department and at home (Edmonds et al., 2002).

Manual measurement of heart rate is common in clinical practice during the initial assessment, either by using a stethoscope or by palpating the wrist (radial artery) or the neck (carotid artery) (Tarassenko & Fleming, 2010). These manual measurements have been shown to underestimate the true heart rate during a clinical competency assessment in a simulated environment, only 25% of nurses could palpate the heart rate (Suter et al., 2014).

Respiratory Rate

The respiratory system is a complex biological structure composed of multiple organs that enable us to breathe. It filters and transports air into the lungs, where gas exchange occurs through a process called diffusion, allowing the movement of oxygen into the lungs and the release of carbon dioxide. The complete process of gas exchange between the atmosphere and body cells is called respiration. The actions responsible for these air movements are called inspiration and expiration. Respiratory rate is measured in breaths per minute (breaths/min). Similar to heart rate, the normal respiratory rate varies with age, with 30-40 breaths/min at birth, reducing to 15-20 breaths/min by late adolescence.

Inspiration and expiration movements cause changes in thoracic pressure, leading to variations in the air volume within the lungs, which in turn cause expansion and contraction of the thorax that can easily be seen by the eye. This effect is used in hospital settings such as general wards or emergency departments, where the respiratory rate is typically measured by the clinical staff counting chest wall movements. Although this method is simple and does not interfere with the breathing mechanisms, human observation of respiratory rate is time-consuming and the result is not always accurate, with errors reported up to 34% (McQuillan et al., 1998).

To obtain reasonably accurate results, Simoes (2003) discussed the need to count chest wall movements for at least 60 seconds; in clinical practice, however, counting periods of just 15 seconds or at most 30 seconds are more common. The training of clinical staff to adequately measure respiratory rate is also an issue. Pearson and Duncan (2011) reported that only 67% of nurses were able to measure respiratory rate to within 10 breaths/min during an investigation in a simulated environment.

The variation in thorax volume when a patient is breathing can also be measured using a strain-gauge transducer typically held by elastic bands around the chest.

The respiratory movements result in resistance changes of the strain gauge that correspond to the respiratory activity. However, the accuracy of measurements can be affected by body position, variability in the signal recorded among patients, and even in measurements of the same patient at different times (Baird & Neuman, 1991). In addition to measuring the changes in the circumference of the thoracic cavity, there are a number of other methods that estimate respiratory rate from variations in the airflow during expiration.

These methods typically use a face mask, a nasal cannula, or a mouthpiece to collect the air. During expiration, airflow can be detected because the expired air is warmer, has higher humidity, and contains more carbon dioxide than inspired air. However, measuring airflow can affect the respiratory rate. A possible reason for its popularity in a hospital setting is that it can be measured simultaneously with the ECG, using the same electrodes, thereby eliminating the need to attach extra transducers.

Oxygen Saturation (SpO_2)

Oxygen is a chemical substance essential to the functioning of the cells in the human body and, therefore, necessary to sustain life. It is important to monitor if all organs are receiving a sufficient supply of oxygen as it is being delivered to the body. The measurement of blood oxygenation, also known as oxygen saturation, is an important indicator of a patient's health. A prolonged lack of oxygen can rapidly cause permanent damage to cell tissue, leaving patients with devastating neurological handicaps, and has the potential to be life-threatening if cells have a high metabolic rate in organs such as the brain, heart, or the central nervous system is damaged.

Several methods have been developed to analyze oxygen delivery. In principle, oxygen saturation can be measured in vitro such as in haemolysed blood samples in a cuvette or in vivo measurements taken directly from the human body (Van Overmeire, Van de Broek, Van Laer, Weyler, & Vanhaesebrouck,

2001). In-vitro chemical methods were developed during the second half of the twentieth century and are still in use for blood gas analysis as gold standards for oxygen content measurement. Chemically-based methods determine the oxygen content of blood from a sample by using chemical reactions to remove the oxygen from the blood. One of the disadvantages of these methods is that they can be slow and, therefore, measurements are performed infrequently, usually only once or twice a day in the intensive care unit (ICU).

The advent of oximetry as an optical method for measuring the oxygen content in the blood developed rapidly during the 1930s and 1940s. Based on spectrophotometry, oximetry measures the amount of light absorbed by the various molecules that bind to oxygen in the blood (Burris et al., 1997). Early oximeters were calibrated empirically and achieved an acceptable accuracy, but they were too cumbersome and expensive to use in a clinical environment. Pulse oximetry revolutionizes clinical practice, making it possible to measure, non-invasively and continuously, arterial oxygen saturation in the peripheral arteries. Peripheral oxygen saturation gives a measure of the amount of oxygen carried by the arterial blood.

It is given by the percentage of hemoglobin that is bound to oxygen with respect to the total hemoglobin. Unlike heart rate and respiratory rate, SpO_2 does not vary with age and a value between 95% and 100% is considered normal in both children and adults (DeVita et al., 2010). The common measurement sites for pulse oximetry are the finger, toe, and ear lobe. Non-invasive and continuous monitoring of oxygen delivery is an invaluable aid to the physician as an indicator of the patient's health (De Kock & Tarassenko, 1993). Pulse oximeters are small, cheap, portable, non-invasive, and do not require any prior training for their use.

Blood Pressure

Blood pressure is when blood is forced to move into the artery walls during contraction and relaxation in the heart. Any time the heart moves, blood

flows into the arteries, which leads to maximum blood pressure while the moment as heart slim back to its position. The pressure drops any time the heart calms (Magder, 2018).

In determining the blood pressure of individuals, there are two main numbers that are considered. The numerator or the high figure is known as the systolic pressure, which comes by as a result of the artery as at when the heart contract and push blood through the body. The denominator or the lower figure is called the diastolic pressure, which is an indication of the force within when the heart slims back. These two figures have a unit of measurement as "mm Hg" (millimeters of mercury). According to Elliott and Coventry (2012), there are two main ways of checking the blood pressure of patients. That is, the manual system which is called a mercury manometer or sphygmomanometer, and the digital or dial.

There are some key disease conditions that are associated or notably with high blood pressure. Among them are stroke, heart attack, and heart failure. When the pressure reaches its peak, this forces the arteries to exhibit a resistance with the flow of blood which makes the heart to push harder for the flowing of blood in the body. Blood pressure is grouped as normal, elevated, or stage 1 or stage 2 high blood pressure.

Normal: the normal blood pressure is within the range of 120 to 80. Which is, systolic of 120 and diastolic of 80. Elevated: is within the range of 129 to 80. While stage 1: ranges from 139 to 89 and stage 2: is classified as 140 to 90 or higher. (Taylor, Wilt, & Welch, 2011).

When blood pressure is measured for the first time, and it seems to be more than the normal range, it does not make the individual a hypertensive patient or that they should be in a panic situation. A physician may require to get a blood pressure profile, which is a record of blood pressure taken consistently for some days minimum of seven days or a few weeks to be diagnosed with blood pressure and with its treatment (Ramsey III, Medero, & Hood Jr, 1991).

Kernel Density Estimation

The kernel density estimation (KDE) initially proposed by Parzen (1962) provides a non-parametric approach for estimating the underlying distribution of observed datasets either on the univariate or multivariate scale. For patient monitoring, the KDE approach has beens utilized extensively. In particular, the joint distribution of key vital signs, such as heart rate, respiratory rate, SpO2, and blood pressure from a cohort of "healthy" patients are estimated using KDE. Thresholds are set using the estimated probability density function for which predictions are based. Particularly, new thresholds are generated using new patients' vital signs and then compared with the normal threshold and alarm on low-likelihood values reported.

In essence, KDE methods replace an absolute threshold with a probabilistic threshold that can account for the correlation between vital signs. Patient risk, though, is still assessed at a single point in time, thereby losing information from previous measurements, and making the unrealistic assumption of independent and identically distributed (i.i.d.) observations. In addition to deterioration detection, KDE methods have been used to identify artificial anomalies in vital sign data.

Gaussian Processes for Vital Sign Modeling

In this section, we exposit in brief the relationship this project has with existing works in the literature. We point out the similarities and differences in modeling, computational methodology, inferential method, and application. First, we identify existing papers in the same direction as the current work before outlining their features in terms of the above thematic aspects of interest.

GPs are a flexible and principled way to model a variety of functions, including regression, classification, time-series, and Spatio-temporal modeling tasks. It is, therefore, unsurprising that there exists an ever-growing body of

literature describing the use of GPs to model and predict vital signs.

The probabilistic framework of GPs assists in the modeling of vital signs, which are typically recorded using wearable sensors, which induce various noise components, such as sensor noise, quantization, and artifact arising from patient movement or signal processing errors.

The GP's flexibility allows it to handle a variety of modeling tasks. For example, as a regressor, the GP can perform forecasting, interpolation, and missing-value-imputation tasks with measures of uncertainty in its prediction. This allows GPs to serve both as a modeling approach in their own right, as well as serve as a pre-processing step to subsequent analytical steps. As a classifier, GPs can circumvent unnecessary and inaccurate assumptions of linear relationships between predictors and outcomes (Knaus et al., 1991; J. M. Wang, Fleet, & Hertzmann, 2007).

Dürichen et al. (2014) applied multi-task GP regression to assess the inter-correlations existing between physiological vital signs using vital signs data set from a cohort of cancer patients recovering from surgery in the Oxford Hospital NHS Trust. The vital signs spanned heart rate, respiratory rate and systolic blood pressure. The MTGP developed considered convoluted kernels for its covariance function in order to allow different hyperparameters for each vital sign. The utility of their approach was evaluated in comparison with appropriate single-task GP models for each vital sign, for which improved performance was obtained in favor of the MTGP. Clifton, Clifton, Pimentel, Watkinson, and Tarassenko (2012) illustrated the potential of GP regression-based Early Warning Scores (EWS) to improve the efficacy of the well-known manual EWS systems using physiological vital signs data collected manually. They adopted the single-task GP modeling approach.

W. Wang, Feng, Liu, and Chen (2008) used Gaussian processes to impute missing vital sign values as well. By providing a complete set of vital signs, the values could be fed into a patient status index, such as in the KDE-based method from Hann (2008), without resorting to a heuristic imputation at the population

mean. Furthermore, the posterior distribution of the Gaussian process allowed for probabilistic reasoning over both the missing vital sign value and the patient status index which results from those vital sign values.

In Pimentel et al. (2013), multi-task GPs are used to provide an estimate of heart rate and respiratory rate trajectories, which were subsequently clustered via a metric of local likelihood into four template trajectories. These templates were then used to distinguish between deteriorating and non-deteriorating trajectories in a held-out set of test patients. The value of GPR for modeling the trajectories was severalfold, including:

- Imputation of vital sign values at arbitrary and constituent time intervals
- the principled estimation of those values
- a representation of uncertainty at any point (thereby accommodating greater weight for trajectories with greater certainty).

In Lasko (2014), GPs were used to pre-process time-series of uric acid measurements. GPs were used to transform "noisy, irregular, and sparse observations to a longitudinal probability distribution". The timestamps of these posterior GPs were then heuristically time-warped and fed as inputs into an autoencoder with the aim to distinguish between the uric acid measurement features of patients with gout and leukemia. The auto-encoder-learned features were compared to features designed by clinical experts.

Stegle, Fallert, MacKay, and Brage (2008), examined the use of GPs for free-living HR monitoring with 40 adult subjects. modeling-fitting and forecasting were improved by the use of clustering based on "auxiliary" ECG-waveform summary data (for example, the variety of inter-beat intervals, the extrema within these intervals, and the fraction of time these extrema fell outside a credible range). The latter was used to identify noisy periods in the data.

The auxiliary variables were used to generate clusters of variables with different levels of noise. The noise modeling was then a mixture of the different

classes with varying noise. After clustering, the kernel accounted for two additive components of HR variance: a discontinuous short-term variable component and a periodic component for diurnal patterns. modeling parameters were found using expectation maximization to approximate the posterior distribution and then choosing values that maximized log-likelihood. In almost all the above methods, the approach to parameter inference was based on full Bayesian with fitting algorithms constructed via marginal likelihood optimization.

Another related works are the works of Ofori (2020), Mensah, Ofori, and Howard (2021), and Mensah et al. (2016). Mensah et al. (2016) considered functional model for longitudinal data with covariate-dependent smoothness. In their work, functional terms describing subject-specific and group-specific trends were modelled with Gaussian processes.

Allowing the smoothness of specific trends to relate to available subject-specific covariates, a smoothness parameter was assumed and modelled hierarchically through the covariance function of the Gaussian describing the subject-specific trajectory. Variational Bayes inference methods based on sparse spectral approximations were developed for parameter inference. Modeling of covariate information was pretty much simply due to the use of subject-specific models allowing the extension of the Variational Bayes to complex functional models. The utility of the developed methods was illustrated in hydrology with grouped longitudinal streamflow dataset. Mensah, Ofori, and Howard (2021) on the other hand, provided a modification of subject-specific covariate modeling concept to sample-specific covariate modeling within the BGPR framework to assess the relationship trauma-specific covariates have with systolic blood pressure observed during traumatic events.

They proposed the use of OGK statistics (Maronna et al., 2019) for extracting vectorized covariate information for easy modeling via the Gaussian process hyperparameters. In this thesis, the focus is on reducing multiple vital signs data with common time-stamps into a univariate vital sign and modeling the resultant vital sign using Gaussian process regression with sparse spectral

approximation treatment for the GP terms in the Bayesian paradigm. Thus, this project is similar to all the above work in terms of the use of the Bayesian Gaussian process regression modeling. In particular, computational framework is similar to Ofori (2020), Mensah, Ofori, and Howard (2021), and Mensah et al. (2016) in terms of the use of sparse approximation, and the incorporation of non-dependent covariates using OGK statistics. Nevertheless, there exist vast structural differences in models and algorithmic development.

However, the major difference is seen in the modeling. While, single Gaussian process term was employed in modeling the unknown smooth functional observation in Ofori (2020), and Mensah, Ofori, and Howard (2021), two Gaussian process terms are used in this thesis. Also, the response modelled here is a composite functional response obtained via fusing of the multiple vital signs while single functional responses were considered in the papers identified above.

Chapter Summary

This chapter of the thesis provides a brief review of key aspects of the projects in relation to the identified research topic. In particular, fundamental aspects of the research topic that will enable easy understanding of the thesis as well as motivate readers' interest. The chapter gives a general introduction, followed by a brief exposition on the concept of physiological vital signs. In particular, physiological vital signs are clearly defined in relation to the interest of the project with its components briefly introduced. In addition, Gaussian processes are formally defined and their application in functional regression is reviewed in brief. Finally, a brief introduction to the kernel approach to the estimation of the probability density function of a given dataset is provided to end the chapter.

CHAPTER THREE

METHODOLOGY

Introduction

This section of the thesis presents exposition on the appropriate statistical approaches adopted for extracting vital sign specific features from the multivariate vital sign data, a fusion of the extracted multivariate features into a univariate feature, development of functional regression models in the Gaussian process framework, and the development of inferential methods in the Bayesian computational framework. This chapter is structured as follows.

First, an introduction to vital sign data fusion is provided in brief. Second, novel approaches for extracting composite single variate vital signs output from multi-vital signs dataset based on information contained in the probability density function (pdf) and computation of pdf-based statistics is outlined.

Third, the appropriate pdf-based statistics for deriving composite functional vital sign trends and corresponding fusing schemes are outlined. Functional modeling of the derived composite vital sign data via the Gaussian process regression framework is developed with which robust approaches for handling non-functional physiology-specific covariates are introduced next. Appropriate computational methods for parameter inference in the Bayesian formalism are considered next.

In particular, tractable approximations for complex Gaussian process likelihoods in the Variational Bayes perspective are introduced in brief and Variational Bayes inference for unknown parameters is giving a brief exposition and its application made to develop appropriate Variational inference method for the developed Gaussian process model are outlined under this section.

Furthermore, a discussion of an appropriate exact inference method in the advanced Markov Chain Monte Carlo (MCMC) framework is considered. Next, the adopted statistical performance measures are introduced and the outline for implementation of the proposed methods described. Finally, the chapter ends

with a summary.

Vital Signs Data Fusion

Human physiology can be viewed as a complex system consisting of numerous vital signs which work together in a complex manner to define the overall health of an individual. These vital signs are related in some sense to health status in a fashion that is not fully understood by scientists. Nevertheless, medical doctors through convectional ways are able to determine whether a person is sick or otherwise using these physiological vital signs coupled with other factors such as diet, activity, etc.

This suggests that the composite of all vital signs is pivotal in making an informed and precise decision on the health condition of a given individual. However, the question of how to derive an appropriate composite measure of vital signs via fusing the individual vital signs together in an unbiased manner whiles preserving the underlying associations existing among these vital signs becomes clearly apparent. This is because of the challenges associated with the data in terms of its complexity in a unit of measure, auto-correlation, etc.

Particularly, for physiological vital signs, autocorrelations can be informative on the underlying vital sign specific dynamics, thus, can serve as a source of evidence for early signs of health-related issues if detected on time. However, a manifestation of subtle changes in physiological vital signs usually reflects in the deterioration of health condition which when tracked and detected on time could prevent many unforeseen emergencies.

Detection of such subtle changes in vital signs requires the use of appropriate flexible statistical models with the ability to borrow information from the various vital signs to aid better estimation of model parameters. One such novel model framework is fusing of multi-vital signs into a composite single variate vital sign, followed by modeling for a simplistic assessment of the overall effect of the vital signs. In this regard, the essence of fusing multivariate vital signs can easily be underscored.

The fusing of multi-physiological vital signs generated by systems operating under the same conditions into the univariate vital signs is important in many ways. Firstly, it facilitates easy modeling and fitting of complex statistical models in that the un-fused data may require multiple models as well as estimation of unknowing parameters, compounding the computational expenses associated with such datasets.

The combined effect of multiple vital signs can be very informative on some hidden physiological threat or issue that may require urgent attention. Failure or untimely detection of such issues may lead to the overall adverse health issue that may eventually give birth to premature deaths. In the following subsection, methods for extracting composite univariate vital signs from multi-vital signs are considered.

Data Fusion with Automated Outlier Control

Consider multivariate data, $y_t = \Big[y_1, y_2, \ldots, y_p\Big], y_j = \Big[y_{j1}, y_{j2}, \ldots, y_{jn}\Big]^{'}$ observed over common time stamps $x_t = \Big(x_1, x_2, \ldots, x_n\Big)$. Suppose these variables combine in someway to yield trend $Z_{(y)}$ as shown in Figure 3. Further, suppose it is of interest to understand the dynamics of the $Z_{(y)}$ over x_t and its relationship with some set of covariates, $\omega = \Big[\omega_1, \omega_2, \ldots, \omega_p\Big]$. To understand the nature of the univariate process, $Z_{(y)}$, in terms of the underlying covariates, requires knowing the fusion process before modeling its with some probability models. In this regard, we consider fusing y_t into a univariate data, $Z_{(y)}$ and then model it with the covariates ω and x_t. We propose the use of the probability density function for extracting unique features that can be fused easily into a single-variate trend.

The probability distribution of a given data set conveys several vital information, for example, the nature of the data, measures of center, skewness, etc. In addition, the existence of an outlying observation can also be discovered via the distribution of the data. As a result, it can serve as an appealing source of feature extraction for the development of appropriate feature-based models for

learning the important content of the data.

Though, it may be relatively straightforward for univariate feature-based modeling, some complexities exist for multivariate features in that the intrinsic inter-relationships existing among the variables constituting the multivariate data ought to be preserved. In the light of this, extraction of univariate features from multivariate data requires some particular attention and treatment.

We propose the use of probability density function-based data fusion approaches for extracting vital features from multivariate data set. The proposals considered here are based on the contribution of each candidate observation in the common center generated by all the observations (data points). For clarity, consider the empirical estimator of the measure of center, say, the mean, μ for a random, Y, following the probability distribution with probability density function, $f(y)$,

$$\varpi\Big(Y\Big) = \mathbf{E}\Big[Y\Big] = \int yf(y)dy. \tag{3.1}$$

It is easily observed that, the integrand, $yf(y)$, in (3.1), serves as a measure of contribution from the random samples. Based on the likelihood, $f(y)$, as weights, the elements of the random sample contribute differently to the center, with candidates close to the center having more impact than those which are at some appreciable distance from the center.

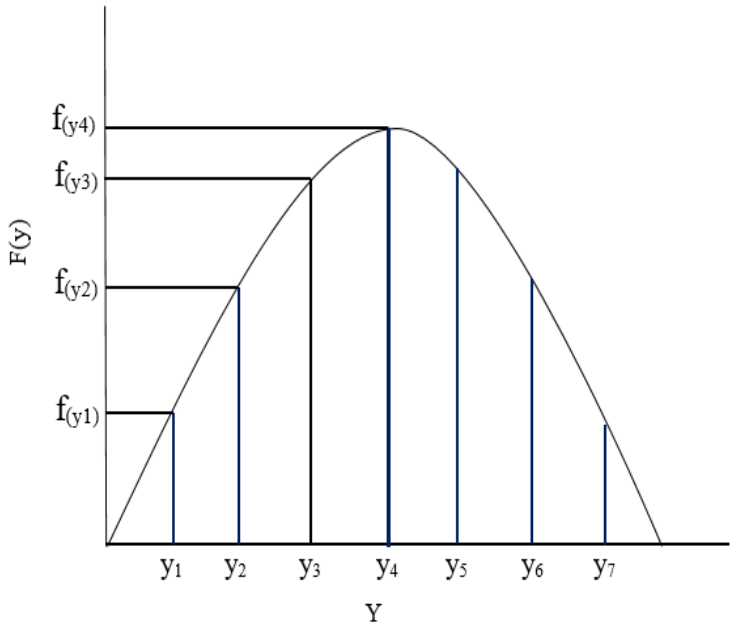

Figure 1: Probability Density Function of the Random Variable y.

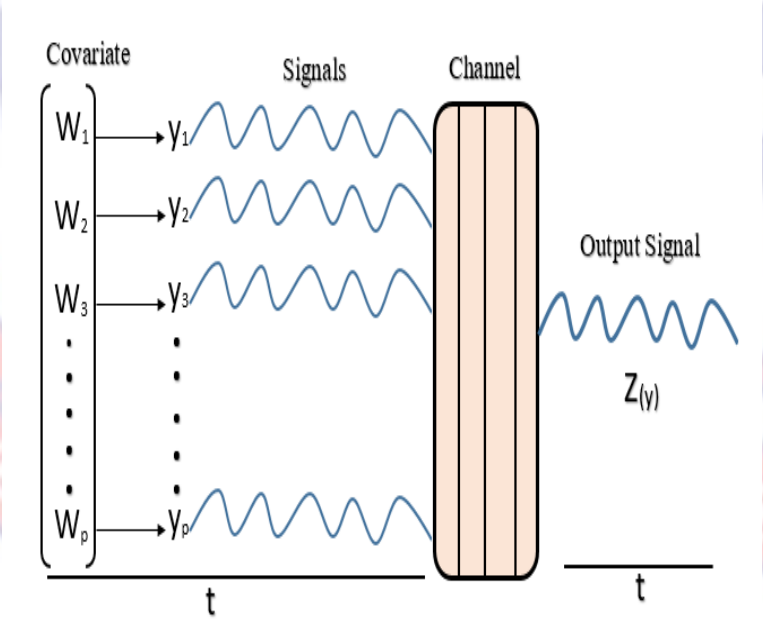

Figure 2: Nature of the Problem.

Probability density-based data fusion using (3.1), is appealing because, it automatically handles outlying observations in an appropriate way through the weight, $f(y)$. Outlying observations are those with a deviant pattern from the common pattern suggested by the data.

They usually appear at the tails of the probability distribution that best defines the data. As a result, they have a low likelihood, $f(y)$, and will yield

relatively lower values, $yf(y)$. If the presence of outlying observation(s) is (are) unknown, then it is clearly seen that the use of features based on (3.1) can offer some level of control over such observations. Furthermore, if those observations are known, it will serve a good purpose not to delete them since deletion constitute loss of information.

Fusion of Multivariate Data into Univariate Data

For multivariate data set in which the likelihood for outlying observations may be high, for example, each defining variable may contribute some of such observations, it will be more beneficial to consider density-based feature extraction for efficient modeling of such data. We consider the following density-based fusion statistics for multivariate physiological vital signs data.

$$T(y_j) = y_j f(y_j) \tag{3.2}$$

$$S(y_j) = \frac{T(y_j) - u_{(y_j)}}{\delta_j}, \delta_j > 0 \tag{3.3}$$

$$V(y_j) = S^2(y_j) \tag{3.4}$$

$$S_1(y_j) = \alpha_j S(y_j), \quad 0 < \alpha_j \leq 1, \tag{3.5}$$

where $u_{(y_j)}$ is a measure of center based on $T(y_j)$. The parameter α_j can be viewed as a data tuning parameter. The choice of it introduces model selection problem which can lead to many interesting learning schemes. However, in this thesis, a data-driven approach was considered.

The probability density function utilized in (3.2), $f(y_j)$, needs to be estimated from the data before it is used. This can be achieved via the nonparamet-

ric density approach. In particular, the kernel density approach to probability density function estimation provides a data-driven approach for understanding the underlying structures in a given dataset through its kernel density. The kernel density estimator of $f(y_j)$ is defined as

$$\hat{f}(y_j) = \frac{1}{n}\sum_{i=1}^{n} K_{\gamma_j}(y_j - y_i), \quad K_{\gamma_j}(y_j - y_i) = \frac{1}{\gamma_j} K\left(\frac{y_j - y_i}{\gamma_j}\right) \tag{3.6}$$

for a symmetric kernel function $K_{\gamma_j}(\cdot)$ and smoothing parameter, γ_j.
For an introduction to kernel density estimation, its implementation, and practical applications see for example (Scott, 2015) and (Silverman, 1986). With the above statistics, more than one possibilities exist for combining multivariate vital sign data into univariate vital sign data for further modeling. The rationale is to extract common features from the vital sign-specific data that constitute the multivariate system and then fuse the features based on an appropriate method. Based on the varied proposals above the following fusion schemes were considered. First, fusing the $T(y_j)$ statistics in (3.2) yields

$$Z_{si} = \rho_i^{'} T(y_i), i = 1, \ldots, n. \tag{3.7}$$

$$Z_{mi} = \bar{T}(y_i) \tag{3.8}$$

$$Z_{oi} = \tilde{T}(y_i) \tag{3.9}$$

Second, fusing the $S_1(y_j)$ statistics in (3.5) yields

$$Z_{si} = \rho_i^{'} S_1(y_i) \tag{3.10}$$

$$Z_{mi} = \bar{S}_1(y_i) \tag{3.11}$$

$$Z_{\mathrm{o}i} = \tilde{S}_1(y_i), \tag{3.12}$$

where ρ_i is a vector of constants, $\bar{S}_{(y_i)}$ denotes vector of row means and $\tilde{S}_{(y_i)}$ is a vector of row OGK statistics (3.19). Alternatively, any of the other statistics derived from $T(y_j)$ can provide a simple way to fuse the multivariate data into univariate data. In particular, applying (3.2), (3.4), (3.3) and (3.5) to y_t, will be appropriate for the fused data, $Z_{(y)}$. A robust consideration for ρ_j based on the moment contribution statistics, $T(y_j)$ is pursued in this thesis and it is of the form

$$\rho_j = \frac{T(y_j)}{\tilde{\eta_j}}, \tag{3.13}$$

where $\tilde{\eta_j}$ denotes the maximum value of $T(y_j)$. That is, $\tilde{n_j} = \max\left[T(y_j)\right]$. In this thesis, the kernel density approach was adopted for the data fusion schemes and coupled with the Gaussian process regression. That is joint use of the non-parametric density estimation and Gaussian process regression principles were explored.

Gaussian Process Model for Fused Vital Signs Data

Given a sample of a single variate vital sign features vector $Z_{(y)} = (Z_1, \ldots, Z_n)$ extracted appropriately from a multivariate vital signs, data observed over time stamps $x = (x_1, \ldots, x_n)$. Suppose further that there exists a set of covariates v computed appropriately from the matrix covariate information that are associated with the $Z_{(y)}$. A detailed exposition on the choice of candidate statistics for v will be provided later. We consider a Bayesian Gaussian process data generative model

$$Z_{(y)} = f(x) + g(x) + \epsilon \tag{3.14}$$

based on Gaussian measurement errors,

$$\epsilon \sim \mathcal{N}\left(0, \sigma_\epsilon^2 \mathrm{I}_n\right)$$

with the following system of prior distribution

$$f(x) \mid \mu_f(x), \kappa_f(x, x^{'}) \sim \mathcal{GP}\left(\mu_f(x), \kappa_f(x, x^{'})\right) \tag{3.15}$$

$$g(x) \mid \mu_g(x), \tau_g(x, x^{'}) \sim \mathcal{GP}\left(\mu_g(x), \tau_g(x, x^{'})\right) \tag{3.16}$$

$$\tau_g\left(x, x^{'}\right) = \sigma_\tau^2 \exp\left(-\beta^2 |x - x^{'}|^2\right) \tag{3.17}$$

$$\kappa_f\left(x, x^{'}\right) = \sigma_\kappa^2 \exp\left(-\theta^2 |x - x^{'}|^2\right) \tag{3.18}$$

$$\theta \mid \lambda, \sigma_\theta^2 \sim \mathcal{N}\left(v^{'}\lambda, \sigma_\theta^2\right),\ \lambda \sim \mathcal{N}(\mu_\lambda^0, \Sigma_\lambda^0)$$

$$\beta \sim \mathcal{N}\left(\mu_{\beta_0}, \sigma_{\beta_0}^2\right),\ \sigma_\theta^2 \sim \mathcal{IG}(\alpha_\theta, \gamma_\theta),\ \sigma_\epsilon^2 \sim \mathcal{IG}\left(\alpha_\epsilon, \gamma_\epsilon\right)$$

$$\sigma_\kappa^2 \sim \mathcal{IG}\left(\alpha_\kappa, \gamma_\kappa\right),\ \sigma_\tau^2 \sim \mathcal{IG}\left(\alpha_\tau, \gamma_\tau\right),$$

where $\mathcal{IG}\left(a, b\right)$ is an inverse gamma probability model with shape and scale parameters a and b. Also, $\mathcal{GP}\left(\mu(x), \delta(x, x')\right)$ denotes a Gaussian process characterized with a mean function μ and process covariance function $\delta(x, x')$ (Scott, 2015). Available covariate information is incorporated into the model via the use

of out-of-sample specific covariate statistic $v = (v_1, \ldots, v_r)$. In what follows, we give a brief exposition on the robust choice for v in terms of Orthogonalized Gnanadesikan-Ketterning (OGK) statistics (Maronna et al., 2019).

Let ω be a single variate with probability density function $h(\omega)$. Then, statistics based on a sample of size n

$$\hat{\mu} = n_1 \sum_{i=1}^{n} m(z_i)\omega_i, \ \hat{\sigma}^2 = n_0 \sum_{i=1}^{n} c\Big(z_i\Big(\omega_i, \hat{\mu}, \tilde{\sigma}_0\Big)\Big), n_0 = \frac{\tilde{\sigma}_0^2}{n}, n_1 = \sum_{i=1}^{n} m(z_i) \tag{3.19}$$

where $z_i\Big(\omega_i, \tilde{\mu}_0, \tilde{\sigma}_0\Big) = \frac{\omega_i - \tilde{\mu}_0}{\tilde{\sigma}_0}$, $m(z_i) = \Big[1 - s^2(\omega, a_1)\Big]^2 \mathbf{I}_{(|\omega| \leq a_1)}, s(\omega, a_1) = \frac{\omega}{a_1}, c(b) = min\left(a_2, b^2\right)$,

$a_1 = 4.5, a_2 = 3$, are termed OGK estimators of the mean and variance of ω.

Bayesian Gaussian Process Approximation

Let $g(x)$ follow a Gaussian process, $g(x) \sim \mathcal{GP}\Big(m(x), \kappa_\xi(x, x')\Big)$. Let the mean $m(x)$ and covariance function $\kappa_\xi(x, x')$ be defined as follows, $m(x) = 0$ and $\kappa_\xi(x, x') = x - x^{'}$, where ξ denotes the hyperparameters of the covariance function. Based on Bochner theorem Bochner et al. (1959), $\kappa_\xi(x, x')$ can be defined in terms of a Fourier transform of the form

$$\kappa_\xi(x, x') = \int f_g(r) \left[\cos\Big(2\pi(x - x^{'})r\Big) + i \sin\Big(2\pi(x - x^{'})r\Big)\right] dr, \tag{3.20}$$

$f_g(r)$ denotes the probability density function of the covariance function if it exists.

There exist a Fourier duality between $\kappa_\xi(x, x')$ and $f_g(r)$ by the the Wiener-Khintchine theorem (Chatfield, 2013)

$$\kappa_\xi(x, x^{'}) = \int f_g(r) \left[\cos\left(2\pi\tau r\right) + i \sin\left(2\pi\tau r\right)\right] dr \tag{3.21}$$

$$f_g(r) = \int \kappa_\xi(\tau) \left[\cos\left(2\pi\tau r\right) - i \sin\left(2\pi\tau r\right)\right] d\tau, \tag{3.22}$$

where $\tau = x - x^{'}$. Model (3.22) defines the power spectrum or the spectral density of the associated covariance function. Since the $g(x)$ is a zero mean Gaussian process, most of its properties will be completely defined by the covariance function, $\kappa_\xi(x, x')$. In that sense, it is straight forward to see that the properties of the stationary Gaussian process will be automatically controlled by the spectral density defined by the covariance function such as (3.22) (Yang, Li, Rana, Gupta, & Venkatesh, 2019). The expectation in (3.21) can be computed using m Monte Carlo simulation methods based on the set of spectral samples , $\{\omega_k, -\omega_k\}_{k=1}^m$, randomly sampled from $f_g(r)$ (Gal & Turner, 2015)

$$\kappa_\xi(x, x^{'}) \approx \frac{1}{m} \sum_{k=1}^{m} \cos\left(2\pi\omega_k \left((x - \nu_k) - (x^{'} - \nu_k)\right)\right) = \tilde{\kappa}_\xi(x, x^{'}). \tag{3.23}$$

where ν denotes an inducing variable given an appropriate deterministic treatment and ω_k, ν_k and x are of the same length. Using the approximated covariance function $\tilde{\kappa}_\xi(x, x^{'})$, the stationary $\mathcal{GP}$ can be written as $g(x) \sim \mathcal{GP}\left(0, \tilde{\kappa}_\xi(x, x^{'})\right)$.

For the covaraince functions defined in (3.17), the following approximation results upon the use of (3.23)

$$\tilde{\tau}_g\left(x, x^{'}, \sigma_\tau, \beta\right) = \frac{\sigma_\tau^2}{m} \sum_{k=1}^{m} \cos\left(2\pi\beta\omega_k \left(\delta_{x,\nu_k} - \delta_{x^{'},\nu_k}\right)\right) \tag{3.24}$$

and

$$\tilde{\kappa}_f\left(x, x^{'}, \sigma_\kappa, \theta\right) = \frac{\sigma_\kappa^2}{m}\sum_{k=1}^{m}\cos\left(2\pi\theta\omega_k\left(\delta_{x,\nu_k} - \delta_{x^{'},\nu_k}\right)\right) \tag{3.25}$$

In what follows, the relationship between the approximate covariance functions in (3.24) and (3.25) and the Bayesian trigonometric models (3.26) and (3.27) is established.

$$f_\tau(x) = \sum_{k=1}^{m}\left[a_{\tau_k}\cos\left(2\pi\beta\omega_k\delta_{x,\nu_k}\right) + b_{\tau_k}\sin\left(2\pi\beta\omega_k\delta_{x,\nu_k}\right)\right] \tag{3.26}$$

and

$$f_\kappa(x) = \sum_{k=1}^{m}\left[a_{\kappa_k}\cos\left(2\pi\theta\omega_k\delta_{x,\nu_k}\right) + b_{\kappa_k}\sin\left(2\pi\theta\omega_k\delta_{x,\nu_k}\right)\right], \tag{3.27}$$

where

$$\delta_{x,\nu} = (x-\nu), a_{\tau_k}, b_{\tau_k} \sim \mathcal{N}\left(0, \frac{\sigma_\tau^2}{m}\right), \; a_{\kappa_k}, b_{\kappa_k} \sim \mathcal{N}\left(0, \frac{\sigma_\kappa^2}{m}\right)$$
$$\omega_k = \mathrm{G}^{-1}\left(\frac{\mu_k}{2}\right), \; \nu_k = F^{-1}\left(\frac{\mu_k}{2}\right), \mu_k = \left[1 + \left(\frac{k}{m+1}\right)\right],$$

$G(z)$ and $F(z)$ are the distribution functions of a uniform random variable on the interval $[0,\ 1]$ Mensah et al. (2016). By the definition of covariance function, the following results are true.

$$\mathrm{E}\left[f_\tau(x)\right] = \sum_{k=1}^{m}\mathrm{E}\left[a_{\tau_k}\cos\left(2\pi\beta\omega_k\delta_{x,\nu_k}\right) + b_{\tau_k}\sin\left(2\pi\beta\omega_k\delta_{x,\nu_k}\right)\right] = 0$$
$$\mathrm{Cov}\left[f_\tau(x)f_\tau(x^{'})\right] = \frac{\sigma_\tau^2}{m}\sum_{i=1}^{m}\cos\left(2\pi\beta\omega_k\left(\delta_{x,\nu_k} - \delta_{x^{'},\nu_k}\right)\right) \tag{3.28}$$

and

$$\mathrm{E}\Big[f_\kappa(x)\Big] = \sum_{k=1}^{m} \mathrm{E}\left[a_{\kappa_k}\cos\left(2\pi\omega_k\delta_{x,\nu_k}\right) + b_{\kappa_k}\sin\left(2\pi\theta\omega_k\delta_{x,\nu_k}\right)\right] = 0$$

$$\mathrm{Cov}\Big[f_\kappa(x)f_\kappa(x')\Big] = \frac{\sigma_\kappa^2}{m}\sum_{i=1}^{m}\cos\left(2\pi\theta\omega_k\left(\delta_{x,\nu_k} - \delta_{x',\nu_k}\right)\right) \tag{3.29}$$

Using equations (3.24), (3.25), (3.28) and (3.29), it is easily observed that $f_\tau(x)$ defined in equation (3.28) and $g(x)$ defined in (3.16) and $f_\kappa(x)$ in (3.29) and $f(x)$ in (3.15) share the same mean and covariance functions respectively. Thus, in terms of tractable alternative forms for easy development of inference methods in the framework of Bayesian spectral approximations, $f_\tau(x)$ and $f_\kappa(x)$ may serve as building blocks.

The spectral approximation framework for Bayesian inference has attracted much attention in the literature recently and myriad of computational friendly methods have been developed from several Gaussian process applications. For details on current applications, readers are referred to Tan, Ong, Nott, and Jasra (2016) and Mensah et al. (2016). Using equations (3.28) and (3.29), writing $S_\theta c$ for $f(x)$ and $W_\beta a$ for $g(x)$, model (3.14), can be redefined as

$$Z_{(y)} = S_\theta c + W_\beta a + \epsilon, \tag{3.30}$$

where

$$c = \left(a_{\kappa_1},\ldots,a_{\kappa_m},b_{\kappa_1},\ldots,b_{\kappa_m}\right) \sim \mathcal{N}\left(0,\ \frac{\sigma_\kappa^2}{m}\mathrm{I}_{2m}\right)$$
$$a = \left(a_{\tau_1},\ldots,a_{\tau_m},b_{\tau_1},\ldots,b_{\tau_m}\right),\ \sim \mathcal{N}\left(0,\ \frac{\sigma_\tau^2}{m}\mathrm{I}_{2m}\right)$$

and S_θ and W_β are $n \times 2m$ design matrices

$$S_\theta = \left[\cos\left(2\pi\omega_k\theta\delta_{x_l,\nu_k}\right), \sin\left(2\pi\omega_k\theta\delta_{x_k,\nu_k}\right)\right]$$
$$W_\beta = \left[\cos\left(2\pi\omega_k\beta\delta_{x_l,\nu_k}\right), \sin\left(2\pi\omega_k\beta\delta_{x_l,\nu_k}\right)\right],$$

for $k = 1, \ldots, m, l = 1, \ldots, n$.

Variational Inference for Gaussian Pocess Model

Given an intractable posterior distribution, say, $p(\xi \mid y)$ defined by sample likelihood, $f(y \mid \xi)$, and an appropriate prior distribution, $g(\xi)$

$$p(\xi \mid y) = \frac{f(y \mid \xi)g(\xi)}{\int f(y|\xi)g(\xi)d\xi}, \tag{3.31}$$

Variational Bayes learning is centered on directly approximating, $p(\xi \mid y)$ by a variational distribution, say, $q(\xi \mid y)$, so that parameter inference is simply made via the q densities depending on an appropriate adopted restriction on the variational posteriors. An iterative algorithm for obtaining parameter estimates is constructed via optimizing a measure of divergence between the two densities termed Kullback-Leibler divergence Attias (2000). The Kullback-Leibler divergence is defined as

$$KL\Big[q(\xi \mid y)|p(\xi \mid y)\Big] = E_q\Big[\log\left\{\frac{q(\xi \mid y)}{p(\xi \mid y)}\right\}\Big]. \tag{3.32}$$

From (3.31), the following decomposition is true

$$\log\left[\int f(y|\xi)g(\xi)\right] = E_q\left[\log\left\{\frac{f(y|\xi)g(\xi)}{q(\xi \mid y)}\right\}\right] + E_q\left(\log\left\{\frac{q(\xi \mid y)}{p(\xi|y)}\right\}\right). \tag{3.33}$$

Dropping the second term, (3.32) can be written in the form

$$\log\left[\int f(y|\xi)g(\xi)\right] \geq E_q\left[\log\left\{\frac{f(y|\xi)g(\xi)}{q(\xi \mid y)}\right\}\right] \tag{3.34}$$

with

$$E_q\left(\log\left\{\frac{q(\xi \mid y)}{p(\xi|y)}\right\}\right) > 0.$$

It is clearly seen that $E_q\left[\log\left\{\frac{f(y|\xi)g(\xi)}{q(\xi|y)}\right\}\right]$ is the lower bound of $\log\left[\int f(y|\xi)g(\xi)\right]$.

Let

$$\mathcal{L}(q) = \log\left[\int f(y|\xi)g(\xi)\right] = E_q\left(\log\left\{\frac{f(y|\xi)g(\xi)}{q(\xi)}\right\}\right) \tag{3.35}$$

Minimizing $\mathcal{L}(q)$ yield an iterative algorithm termed variational Bayes algorithm for obtaining optimal variational densities for posterior parameter inference. See for example, Bishop (2006); Ormerod and Wand (2010), for more detailed background on variatonal Bayes and its adaptations. Applying the variational Bayes technique, we make the following variational distributional assumptions. Let $\xi = (\theta, \beta, \lambda, a, c, \sigma^2_\lambda, \sigma^2_\tau, \sigma^2_\kappa, \sigma^2_\epsilon)$, denotes the set of parameters involved in our model. Then, we consider a fully separable q densities of the form

$$q(\xi) = q(a)q(c)q(\theta)q(\beta)q(\lambda)q(\sigma^2_\theta)q(\sigma^2_\epsilon)q(\sigma^2_\tau)q(\sigma^2_\kappa), \tag{3.36}$$

where

$$q(a) \sim \mathcal{N}\left(\mu^q_a, \Sigma^q_a\right), q(c) \sim \mathcal{N}\left(\mu^q_c, \Sigma^q_c\right), q(\lambda) \sim \mathcal{N}\left(\mu^q_\lambda, \Sigma^q_\lambda\right)$$
$$q(\beta) \sim \mathcal{N}\left(\mu^q_\beta, \Sigma^q_\beta\right), q(\theta) \sim \mathcal{N}\left(\mu^q_\theta, \sigma^{q2}_\theta\right), q(\sigma^2_\epsilon) \sim \mathcal{IG}(\alpha^q_\epsilon, \gamma^q_\epsilon)$$
$$q(\sigma^2_\kappa) \sim \mathcal{IG}\left(\alpha^q_\kappa, \gamma^q_\kappa\right), q(\sigma^2_\theta) \sim \mathcal{IG}\left(\alpha^q_\theta, \gamma^q_\theta\right), q(\sigma^2_\tau) \sim \mathcal{IG}\left(\alpha^q_\tau, \gamma^q_\tau\right)$$

is adopted for the variational approximation of the true joint posterior. The following notations were considered for the spectral matrices involved in the computation of the variational updating equations utilized for the variational

optimization function.

$$S = E_q\left[S_\theta\right], W = E_q\left[W_\beta\right], S^* = E_q\left[S_\theta' S_\theta\right], W^* = E_q\left[W_\beta' W_\beta\right]$$
$$M = \frac{\partial W}{\partial \sigma_\beta^{q2}}, \quad D = \frac{\partial S}{\partial \sigma_\theta^{q2}}, \quad B = \frac{\partial W}{\partial \mu_\beta^q}, \quad A = \frac{\partial S}{\partial \mu_\theta^q}, \quad Z^* = \frac{\partial S^*}{\partial \mu_\theta^q}$$
$$Q = \frac{\partial S^*}{\partial \sigma_\theta^{q2}}, \quad N = \frac{\partial W^*}{\partial \mu_\beta^q}, \quad U = \frac{\partial W^*}{\partial \sigma_\beta^{q2}}$$

Detailed exposition on the derivations and computation of expectations associated with the above notation is outlined in the Appendix. With the above convention for notation, optimal parameter values for the assumed variational distributions are obtained using an iterative algorithm outlined in algorithm 1 presented in Appendix I on page 116.

MCMC Inference for Gaussian Process Model

In this section, a standard approach for making posterior inference for complex Bayesian models via Markov Monte Carlo (MCMC) methods is considered. Bayesian Gaussian process regression models in which some sparse approximation is adopted for computational simplicity usually generate complex joint posterior distributions.

As a result, marginal posterior distributions for some of the parameters cannot be computed in closed form, leading to the blending of two or more MCMC samplers. The GPR model (3.30) considered in the thesis is such that all parameters except β and θ generate marginal posterior condition distributions which have identifiable standard forms.

For such parameters, the usual Gibbs sampler is adopted for sampling from their corresponding full posterior conditionals. However, the nature of the resulting marginal posterior distributions for β and θ inform the use of Metropolis- Hasting Sampler (MHS) for making inference for such parameters (Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953).

Now, joint use or coupling of the above two standard MCMC samplers by

embedding the MHS within the Gibbs sampler yields another standard method for posterior inference. The resulting MCMC method is usually termed the Hybrid MCMC (HMCMC) and is one of the advanced MCMC methods. See for example Givens and Hoeting (2013), Gelman et al. (2013),and Andrieu and Thoms (2008) for detailed exposition on the advanced MCMC methods and their practical implementation guidelines. The version of the hybrid MCMC considered is presented in Algorithm 2 can be found in Appendix *J* on page 115.

Measures of Performance for Assessment of Proposed Methods

This section is focused on the assessment measures considered for analyzing the performance of the developed methods. Several models, as well as parameter estimation performance measures, exist in the literature. These measures can be put into two main accuracy categories namely, fitting and prediction accuracy measures. The prime focus of this thesis was in the direction of the first arm of the above accuracy measures. The rationale is that predictive modeling was not considered, thus not pursued. However, proposals for fitting functional data with non-functional predictors were pursued.

In terms of the modeling principle, a multi-stage modeling framework was adopted. In particular, a two-stage modeling concept was developed in the following direction.

- One-dimensional functional data fusion of a multi-dimensional functional data, with at most one common predictor among several covariates
- Appropriate modeling of the fussed one-dimensional data, incorporating intrinsic relationships with covariates

As a result, the approach for assessment adopted followed the above principle. That is, varied performance assessment perspectives were utilized for the data fusion approaches and the Gaussian process regression modeling. The data

fusion schemes are assessed empirically, based on the underlying pattern, features, and dynamics of the original data both in the individual dimension as well as in the fussed dimension.

The main defining principle for an appropriate statistic as a unidimensional data fusser with input being multi-dimensional data is that it must exhibit an order, pattern, and covariate relationship preserving capabilities in line with the original data. Of importance, is the pattern and covariate preserving features, since the distortion of such features may covey different information which may be so variant from that contained in the original data, eventually leading to model misspecification coupled with inappropriate practical implications. The observations made above are very critical and thus, serve as a benchmark for the assessment undertaken here. Finally, the adopted empirical performance evaluation for the data fusion proposal was based on the following.

- Ability to preserve the original order individually (variable-wise)
- Ability to preserve the original data pattern and its relationship with existing covariates
- Computational savings associated with the scheme.

On the other hand, the assessment for the GP model was based on some specific model fitting measures. However, the focus in the assessment stems from the recovery of underlying trends, in terms of parameter estimation, the choice of spectral approximation scheme, etc. The numerical measures of performance considered are the Mean Squared Fitted Error (MSFE), the Standardized Mean Absolute Fitted Error (MAFE), and the Mean Absolute Fitted Error

(SMAFPE) defined as follows.

$$\text{MSFE} = \frac{1}{n}\sum_{i=1}^{n}\left(Z_{(y)_i} - \hat{Z_{(y)_i}}\right)^2$$

$$\text{MAFE} = \frac{1}{n}\sum_{i=1}^{n}|Z_{(y)_i} - \hat{Z_{(y)_i}}|$$

$$\text{SMAFE} = \frac{1}{n}\sum_{i=1}^{n}|\tilde{Z}_i|,\ \tilde{Z}_i = \frac{Z_{(y)_i} - \hat{Z_{(y)_i}}}{Z_{(y)_i}}$$

where $Z_{(y)_i}$, and $\hat{Z_{(y)_i}}$, are the ith true and fitted fussed observations respectively. It is important to note that the performance assessment conducted for the simulation study followed a different trend. In particular, the mode of assessment followed that of the functional regression modeling via the Gaussian process regression because the simulation was conducted on the developed model using the fussed functional data.

Implementation of Methods

The implementation of the proposed methods and algorithms were conducted using the R statistical software. For the probability density-based fusing methods, the kernel density estimation was done using the kernel smoothing package," ks" in R (Duong et al., 2007). The smoothing parameter γ in (3.6) was set to the smoothed cross-validation estimator implemented in the R package "ks". For comprehensive details on kernel density estimation and its applications in many scientific fields, readers are referred to (Silverman, 1986) and (Scott, 2015).

All the codes were written in R statistical software and run on an Intel (R) Core (TM) $i7$, 6700 processor Windows PC 3.40 GHz workstation. The algorithmic convergence of the variational Bayes method was based on relative convergence criteria. The relative convergence criteria is defined for a generic optimization algorithm based on an optimization function $g\left(\vartheta\right)$ where ϑ is either a parameter or set of parameters of interest as follows.

$$A_{t^*}^{(\Delta\vartheta)} = \frac{|g(\vartheta^{(t+1)}) - g\Big(\vartheta^{(t)})|}{|g(\vartheta^{(t+1)})|}, \tag{3.37}$$

where $g(\vartheta^{(t+1)})$ and $g(\vartheta^{(t)})$ are the current and previous updates. This criterion mandates stopping when $A_t^{(\Delta\vartheta)} < tol$, where *tol* is an experimenter defined tolerance. Applying this to the variational Bayes optimization function, the variational lower bound $\mathcal{L}(q(\xi))$, (3.37) can be expressed as

$$A_{t^*}^{(\Delta\xi)} = \frac{|\mathcal{L}(\xi^{(t+1)}) - \mathcal{L}(\xi^{(t)})|}{|\mathcal{L}(\xi^{(t+1)})|}. \tag{3.38}$$

Using (3.38), the Variational Bayes algorithms were considered converged at any iteration when $A_{t^*}^{(\Delta\xi)} < 10^{-4}$ for both simulation and real data applications. In the case of the hybrid Gibbs (HGS) algorithm, the initialization followed the following setting.

$$\beta_{[0]} = 0.9, \lambda_{[0]} = v, \theta_{[0]} = \sum_i v_i + 0.5\sqrt{\text{var}Z_y}a_{[0]} = \mathbf{0},$$

$$c_{[0]} = \mathbf{0}, \sigma^2_{\epsilon\,[0]} = 1, \sigma^2_{\tau[0]} = 1, \sigma^2_{\theta[0]} = 1, \sigma^2_{\kappa[0]} = 2.$$

Chapter Summary

The chapter focused on the development of the appropriate statistical modeling framework for the project based on the innovation of the existing frameworks. Basically, two modeling and two inferential frameworks were developed. The first framework focused on a nonparametric framework for fusing multivariate or multi-vital signs data measured in the same time domain and thus share a common functional predictor into a single variate vital sign data for flexible modeling in the functional regression context.

Here, the probability density function of a continuous random variable

and its utility in the computation of some statistics were explored to develop novel schemes for deriving composite vital sign feature data for further modeling. In particular, the contributions of candidate data points were explored and employed as fundamental statistics for the development of the fusing schemes.

It was observed that the fusing statistics exhibited the ability of handling data points which show patterns quite different from the underlying pattern exhibited by the majority of the functional data points when the variable is considered at the univariate level. This was realized to have been adopted naturally from the underlying probability density functions of vital sign variables.

In the second framework, the focus was on modeling the fused vital sign in the functional regression domain. For this arm, an appropriate Gaussian process regression framework was developed. The fused vital signs observations were assumed to have been generated by a composite Gaussian process (GP) model (sum of two Gaussian processes) corrupted with some random Gaussian error terms. Non-time dependent (non-functional) vital sign-specific covariates that generate design matrix originally were first transformed to designed vector via the use of robust statistic termed Orthogonalized Gnanadesikan-Kettering (OGK) statistics (Maronna et al., 2019).

The extracted designed covariate vector was then modeled hierarchically through the smoothing parameter of one of the Gaussian process terms leading to flexible GPR model.

Regarding the inferential framework, both approximate and exact inference methods were developed. The approximate inferential framework employed the Variational Bayes method while the exact inference was based on advanced MCMC, in particular, hybrid MCMC. Corresponding model fitting algorithms were developed for fitting and inference for the developed GP models.

CHAPTER FOUR

RESULTS AND DISCUSSION

Introduction

This section of the thesis focuses on the application of the proposed methods to some examples. The first example was tailored towards simulation studies in which appropriate synthetic data was generated from the assumed Gaussian process model (3.14) given some fixed true parameter values. In the second example, real data application was considered in which public health data from collaborators in Singapore was utilized. Details of the application are presented in the subsection below.

Simulation

In this section, empirical assessment of the developed methods through simulation is considered. The simulation is based directly on the model (3.14). Functional features of size $n = 400$ over equally spaced time stamp $x \in [-10, 10]$ are simulated from the feature Gaussian process (3.14) with the following specifications. First, $\beta, \sigma_\tau^2, \sigma_\kappa^2$ and σ_ϵ^2 are set as $0.8, 0.05, 0.025^2, 0.005^2$ respectively. Then, λ is randomly generated from normal,

$$\lambda_j \sim \mathcal{N}\Big(0.85, 0.05\Big), j = 1, \ldots, r.$$

and $r = 4$. The set of covariates, V_i was allowed to be of the form $V_i = [v_{i1}, v_{i2}, v_{i3}, v_{i4}]$, where $v_{i1} \sim \mathcal{N}\Big(0.5, 0.02\Big), v_{i2} \sim \mathcal{N}\Big(0.8, 0.05\Big), v_{i3} \sim$ Unif $[0, 1]$ and $v_{i4} \sim \mathcal{N}\Big(0, 1\Big)$. With which a vector of statistics (median), $V = [v_1, v_2, v_3, v_4]$ were extracted and an intercept term was added for the hierarchical linear model for the parameter λ.

Given λ, θ is generated from normal, $\theta \sim \mathcal{N}\Big(V'\lambda, 0.05\Big)$. The smooth functions, $f(x)$ and $g(x)$ are generated from the corresponding Gaussian process models, $f(x) \sim \mathcal{GP}\Big(\mu_f(x), \kappa_f(x, x')\Big), g(x) \sim \mathcal{GP}\Big(\mu_g(x), \tau_g(x, x')\Big)$. Finally,

the features $Z_{(y)}$ were generated from model (3.14) with the above settings for parameters. For prior setting, the following were considered, $\lambda \sim \mathcal{N}(0.82, 0.10)$ independently for $r = 1, \ldots, 4$ and $\beta \sim \mathcal{N}(0.82, 0.10)$. The remaining prior hyperparameters are set as follows. $\alpha_\theta = 3,\ \gamma_\theta = 2,\ \alpha_\tau = 3, \gamma_\tau = 2, \alpha_\kappa = 2, \gamma_\kappa = 0.1, \alpha_\epsilon = 2, \gamma_\epsilon = 0.01$.

The nature of the simulated Gaussian processes, $f(x)$, $g(x)$ and $Z_{(y)}$ is shown in Figure 3. The varying frequencies underlying the features are clearly evident. We fit model (3.14) based on the developed algorithms with the set of spectral points $m \in \{10, 20, 30, \ldots, 80, 100\}$. Posterior analysis for the advanced MCMC methods was based on chains of length 5000 with a burnin of size 2000. The MCMC algorithms were assessed for convergence by examining representative samples to ensure that the comparison with the corresponding variational Bayes algorithm is fair.

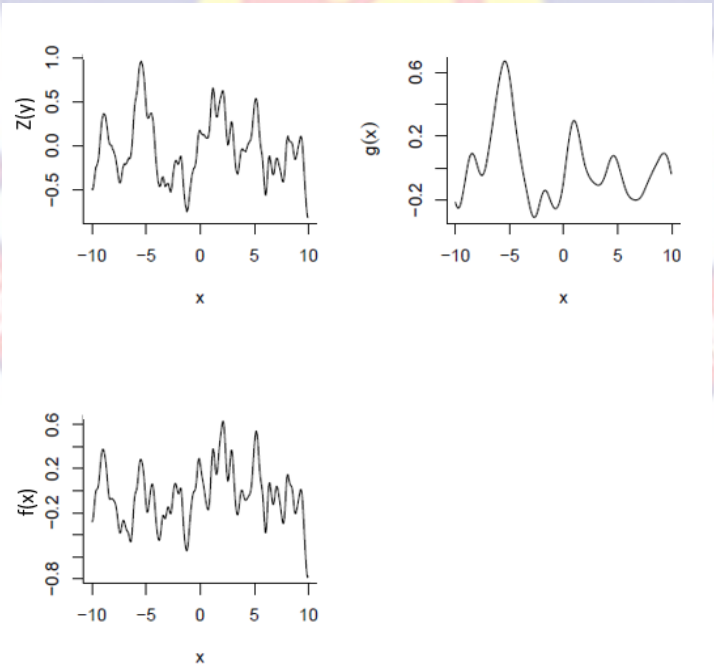

Figure 3: Nature of Simulated Feature, $Z_{(y)}$ and the Gaussian Processes, $f(x)$ and $g(x)$

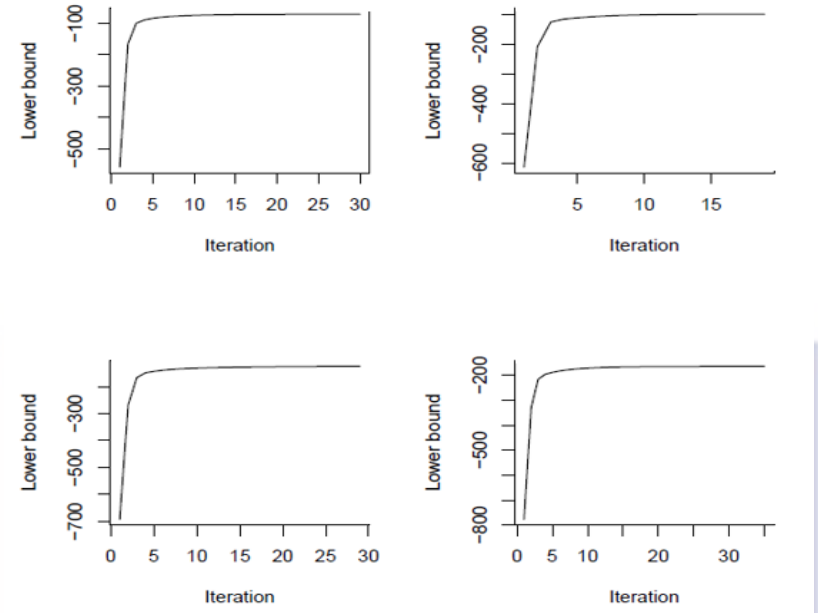

Figure 4: Plot of Lower Bound Attained at Convergence for VBFBGPR with $m = 25, 30, 35, 40$.

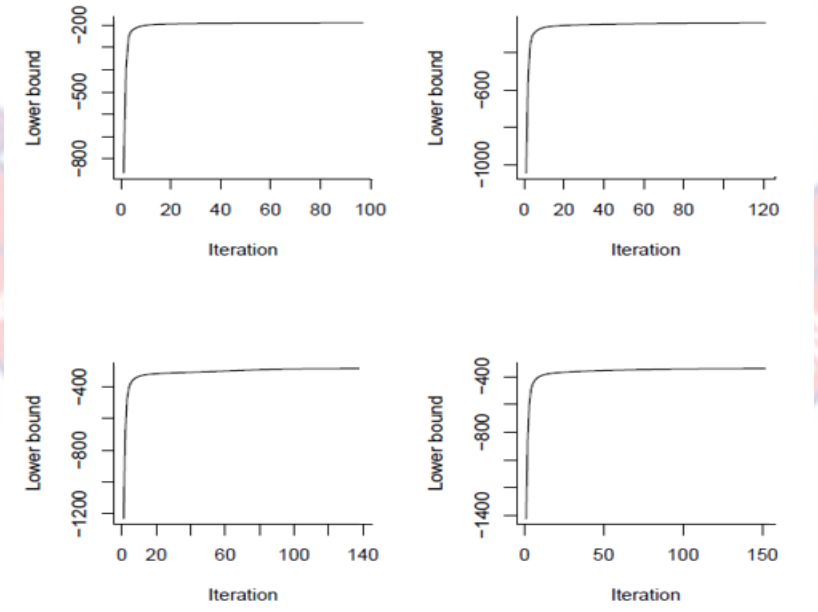

Figure 5: Plot of Lower Bound Attained at Convergence for VBFBGPR with $m = 45, 55, 65, 75$.

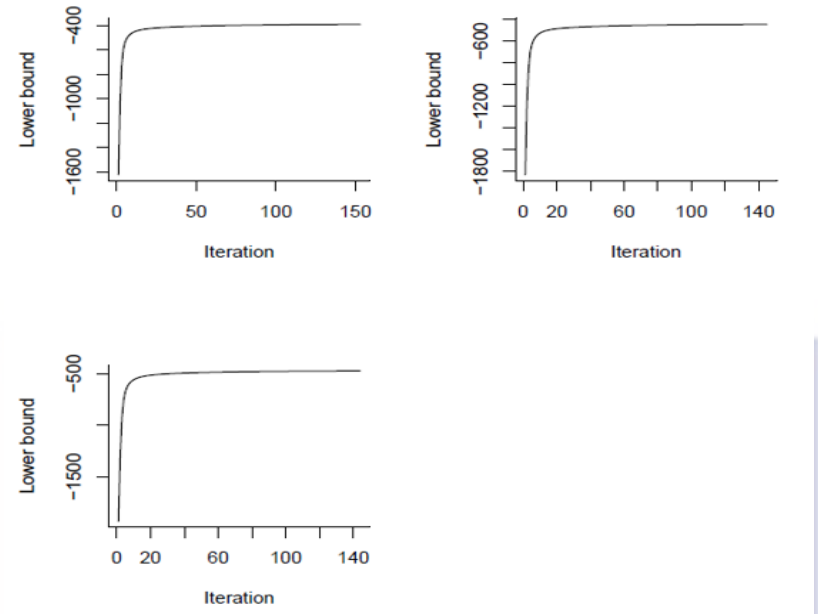

Figure 6: Plot of Lower Bound Attained at Convergence for VBFBGPR with $m = 85, 95, 100$.

First, the nature of the Variational Bayes algorithm is assessed in terms of doing what is expected. A typical Bayesian fitting algorithm constructed in the Variational Bayes framework is expected to exhibit an increasing trend, evidenced by the lower bound as the optimization progresses until convergence. This is because the parameter estimation problem is expressed as a maximization problem via the use of Variational distributions. Figures 4, 5, and 6 show the lower bound pattern exhibited by the VBFBGPR over set of spectral samples $m = \{25, 30, 35, 40\}$,$m = \{45, 55, 65, 75\}$ and $m = \{85, 95, 100\}$ for the simulated dataset. It is clearly obvious that the developed algorithm is conforming to the expected fitting pattern.

Next, the relationship underlying the spectral sample size m and the evidence provided by the model expressed in terms of the lower bound as well as the number of iteration required for convergence are shown in Figure 7. The graph on the left presents the nature of the lower bound attained at convergence against the number of spectral points m. It can be observed that the model evidence decreases with increased spectral sample size , m, for the range of values

considered. The plot on the left illustrates the intrinsic pattern existing among the spectral sample size m and the corresponding number of iterations required for convergence. In general, fewer iterations are required for small spectral samples.

However, a short decreasing and increasing trend is evident between $m = 25$ and 35 whiles a slow change in pattern (gradual decrease) is seen after $m = 80$, suggesting that larger spectral samples ($m > 85$) can yield a variant trend in terms of the number of iterations required for convergence.

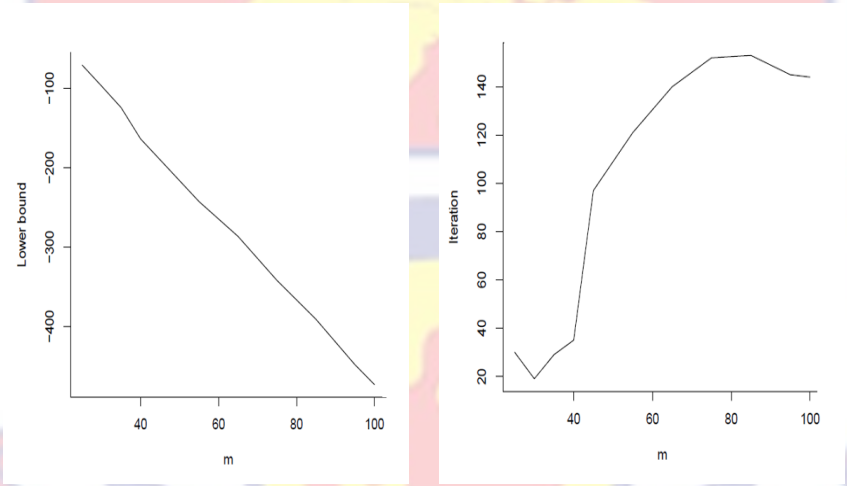

Figure 7: Convergence Features of VBFBGPR Based on Synthetic Data. Plot of Lower Bound Attained at Convergence Against Spectral Sample Size (m) (left). Plot of Number of Iterations Against Number of Spectral Sample Size (m) (right).

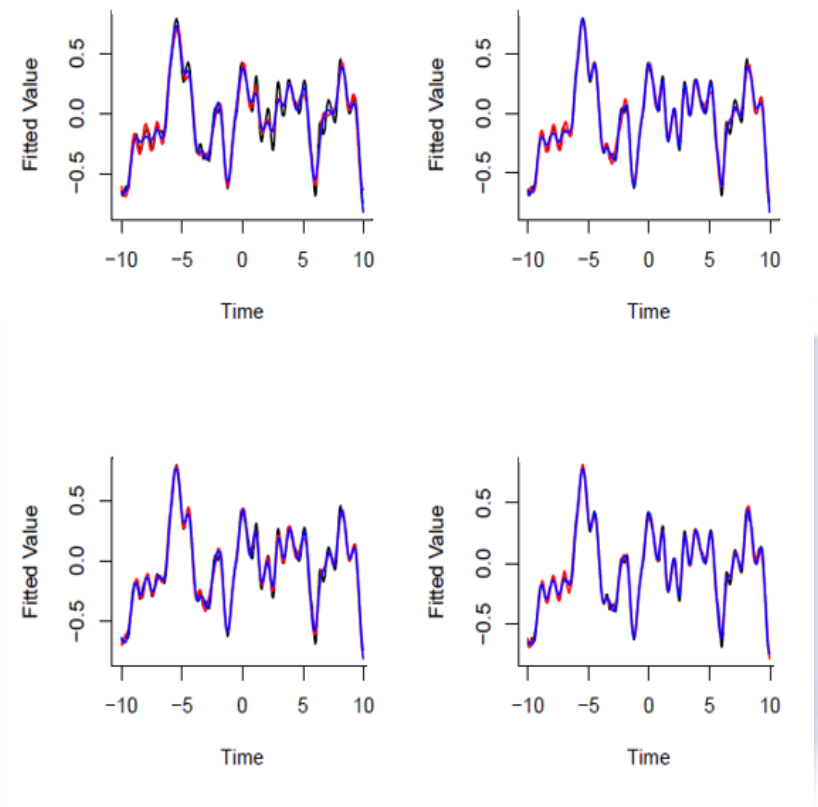

Figure 8: Nature of Fitted Trends for VBFBGPR and HGS with $m = 25, 30, 35, 40$.

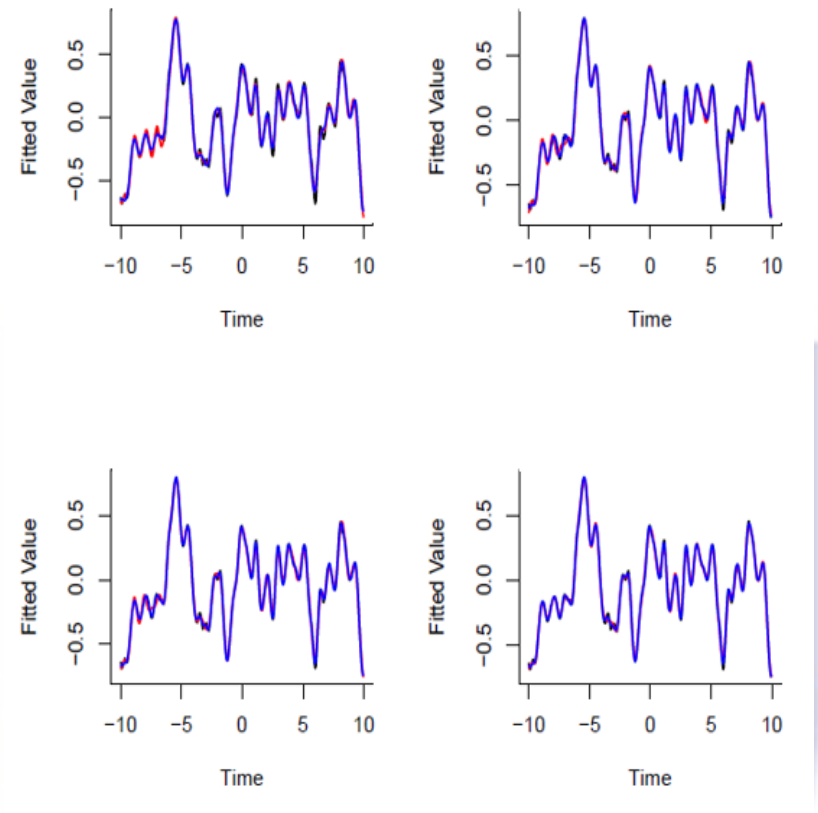

Figure 9: Nature of Fitted Trends for VBFBGPR and HGS with $m = 45, 55, 65, 75$.

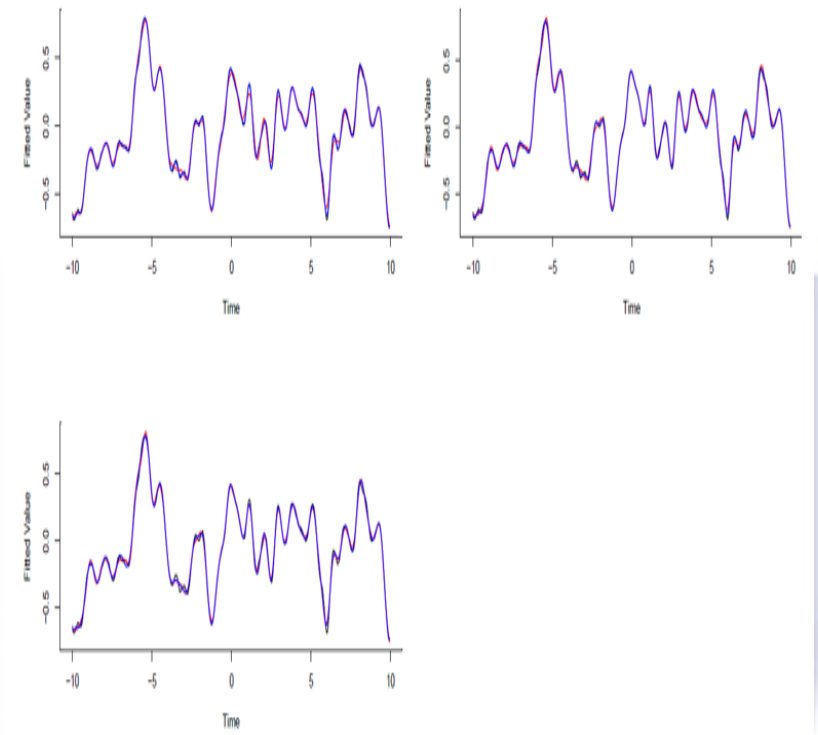

Figure 10: Nature of Fitted Trends for VBFBGPR and HGS with $m = 85, 95, 100$.

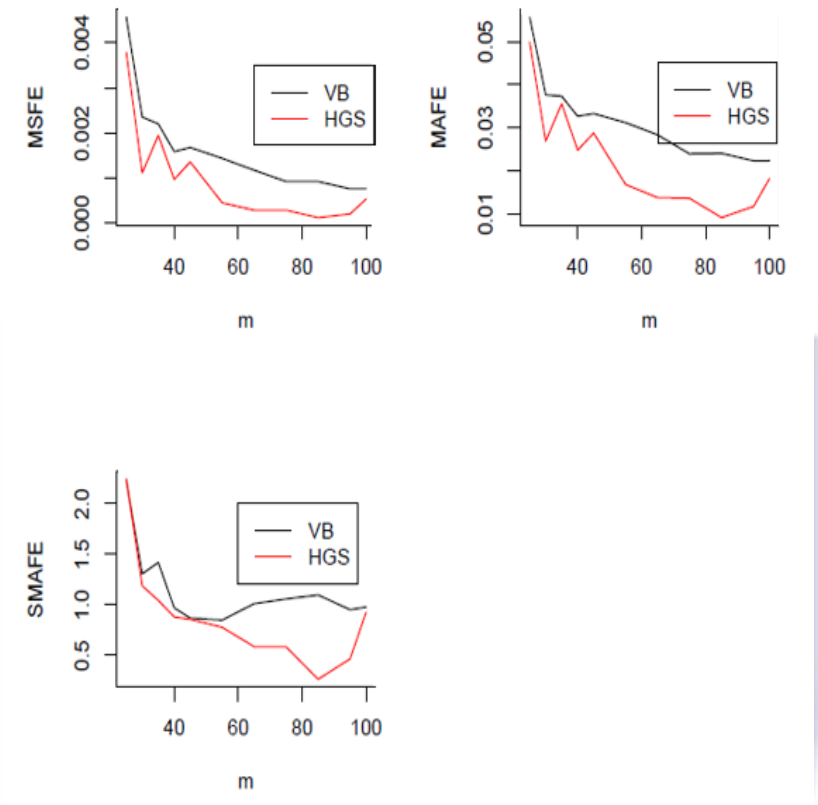

Figure 11: Plot of Average Squared Fitted Error for VBFBGPR and HGS over m.

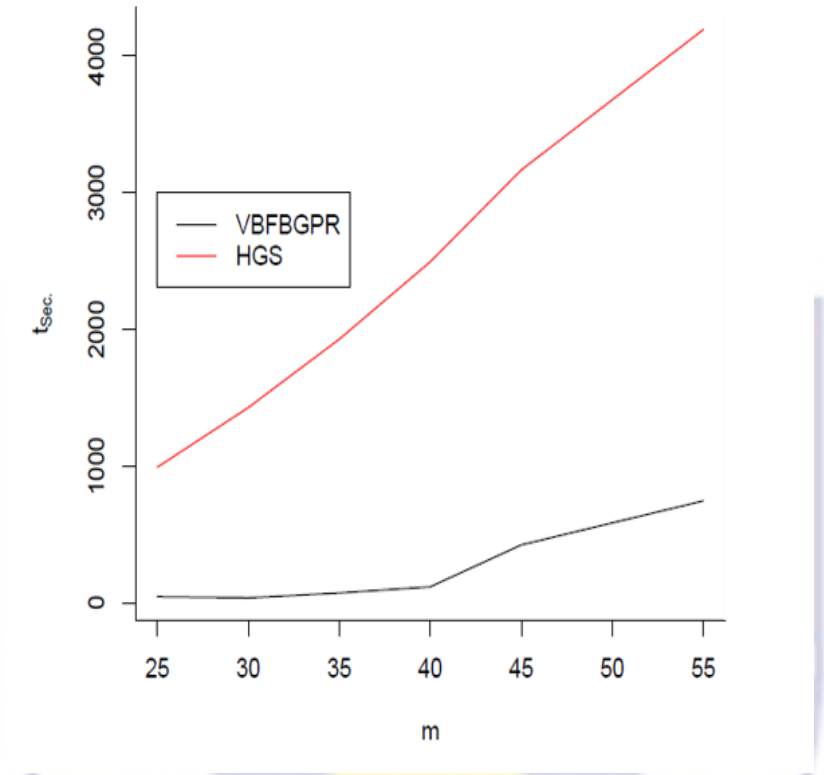

Figure 12: Plot of Spectral Sample Size m Against Computational Time for VBFBGPR and HGS

The fitting performance of the Variational Bayes algorithm and its hybrid MCMC counterparts are shown in Figure 8, Figure 9 and Figure 10, for small, $m = \{25, 30, 35, 40\}$, medium, $m = \{45, 55, 65, 75\}$ and large, $m = \{85, 95, 100\}$ spectral samples respectively. The blue lines are the VBFBGPR fit and the red lines represent the HGS fit. Both algorithms exhibit the ability to recover the underlying functional trend existing in the simulated functional data reasonably well. However, it can be observed that the point-wise fitting performance improve with increasing spectral sample size m.

Figure 11 gives the overall fitting performance of VBFBGPR and HGS, quantified in the average squared error of fit, MSFE, MAFE, and SMAFE for the range of spectral sample sizes considered. The fitting errors are generally small with a similar pattern over m, for both algorithms. This suggests that both algorithms are able to yield better inference for the simulated fused functional data. Although errors decrease with an increasing number of spectral samples,

the decrease is relatively high for HGS than VBFBGPR.

Table 1: ***Comparison of Variational Bayes (VBSBPM), and its MCMC Counterpart for θ for the Simulated Data Set. Columns 2 and 3 give the Parameter Estimates for VBFBGPR and HGS Respectively. The generated true θ value was 2.91. The last two columns give their corresponding 95% Bayesian credible intervals (BCIs).***

	$\theta = 2.91$		VBFBGPR	HGS
m	$\hat{\theta}_{\text{VBFBGPR}}$	$\hat{\theta}_{\text{HGS}}$	95% BCI	95% BCI
25	2.77	2.73	$[2.764, 2.777]$	$[2.477, 2.897]$
30	2.94	4.68	$[2.937, 2.949]$	$[4.436, 4.956]$
35	2.83	2.94	$[2.820, 2.832]$	$[2.540, 3.603]$
40	2.98	3.44	$[2.975, 2.985]$	$[3.077, 3.753]$
45	2.88	2.89	$[2.878, 2.889]$	$[2.637, 3.052]$
55	2.89	3.92	$[2.887, 2.898]$	$[3.669, 4.412]$
65	2.79	4.09	$[2.886, 2.898]$	$[3.679, 4.527]$
75	2.93	3.82	$[2.787, 2.798]$	$[3.336, 4.347]$
85	2.86	4.25	$[2.923, 2.934]$	$[3.997, 4.555]$
90	2.97	3.77	$[2.860, 2.870]$	$[3.355, 4.209]$
100	2.94	2.99	$[2.960, 2.971]$	$[2.838, 3.125]$

Source: Researcher's Computations (2021)

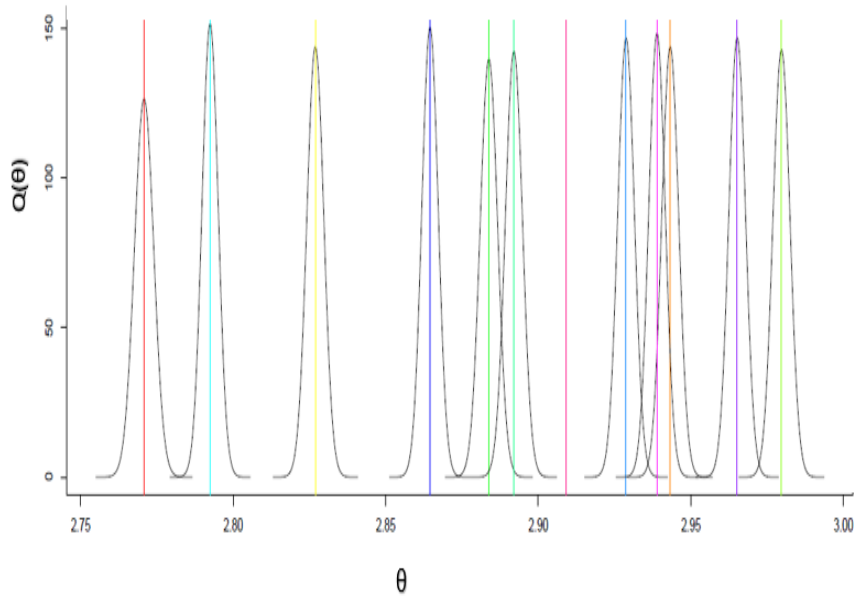

Figure 13: Marginal Posterior Densities of θ Corresponding to m for VBF-BGPR.

The computational expenditure associated with the VB algorithm and its exact inference counterpart (HGS) is shown in Figure 12. It is clearly evident that VBFBGPR is faster than HGS by magnitude.

Table 1 reports the posterior estimates for the hierarchical regression parameter, θ together with their corresponding 95% Bayesian credible intervals (BCIs) for VBFBGPR and HGS over a range of spectral sample sizes m. It can be seen that both algorithms yield posterior estimates that are quite comparable for $m = 25, 35, 45$ and 100. However, for $m = 55, 65, 75, 90$ the estimates exhibit some appreciable differences.

That is the marginal difference are quite significant. Furthermore, it can be observed that for posterior estimates of θ that are similar, 95% BCIs of VBF-BGPR are subsets of BCIs of HGS. On the other hand, for the estimates for which the marginal differences are somewhat large, the 95% BCIs generated by VBFBGPR overlap with those of HGS.

Figure 13 presents the marginal variational posterior densities of θ for the range of spectral sample sizes m considered in the simulation. The vertical lines are

the corresponding posterior estimates reported in Table 1. The differences in the posterior densities in relation to m are evident.

Health Data Application

In this section, a typical application of the developed methods was considered in solving a public health issue. In particular, the issue of health monitoring through monitoring of the indicators of health. In this practical example, the focus was on the physiological factors that define the health status of human beings and are considered as the standard measures for health assessment in the medical field or public health.

The application starts with an examination of the underlying data structure, variable identification and convection for notations. An exploration of the data via preliminary analysis was also considered to allow us to understand in brief what direction does the data present in terms of the analysis pipeline.

Data Description

This section focuses on a brief exposition on the source as well as the nature of the data. Real data made up of observations on physiological vital signs from a group of patients considered in a structured Singapore Heart Foundation (SINGHEART), Singapore project was considered. This dataset was collected for Biofourmis Private Company Limited, a Health Data Analytic Company in Singapore as a result of the partnership they have with the Singapore Heart Foundation. The acquisition of the data was based on the defined international standard by Biofourmis in relation to the data agreement existing between the two companies and the Government of Singapore.

The datasets were released for this research through the main supervisor based on the research relationship existing between the two parties. Documentation on the use of data and publication of output of the research is fully func-

tional. The dataset in context was a complex time series data of size 848×5, on the vital signs, namely Systolic blood pressure (SBP), Diastolic blood pressure (DBP), Mean arterial pressure (MAP), Pulse Rate (PP), and Heart Rate (HR).

In addition, patient-specific covariate information associated with the data was also made available in their complex form. The defining covariate variables were Gender, Age, Height, Weight, and Race. According to the documentation on the data, a group of adult Asians enrolled in a SingHeart study, were continuously monitored over a period based on varied physiological states corresponding to different activities, for example, sleeping, walking, exercising, sitting, etc. It is important to note that all the datasets utilized were de-identified in order to address the issue of sensitivity associated with medical data.

Data Pre-processing

The datasets required for the experimentation were pre-processed by choosing appropriate variable names, re-labeling, and coding the adopted variables. All Categorical variables involved were coded appropriately to suit the model assumptions. Some additional covariates were created where necessary and as guided by the existing ones, to expand the covariate space. Particularly, in the case of continuous covariates, Body Mass Index (BMI) was generated from Age, Height, and Weight using the standard definition of BMI.

The entire dataset was transformed into a common space by subjecting to standardization within variables such that all the random variables generating the dataset follow a common distribution, standard normal. This was to ensure that the variability in a unit of the measure was addressed to provide a unified or common space for the fusion of the multivariate dataset.

Preliminary Data Analysis for Health Data

Exploratory data analysis can be considered as one of the best data analysis procedures in any data analysis problem since it offers the potential to dis-

cover interesting features of the data under consideration; thus, eventually directing the right path for the analysis. This section of the thesis focuses on the preliminary data analysis techniques employed to explore what the data at hand seeks to present in terms of the direction for appropriate methods that could be utilized.

Exploratory Analysis Method

Exploratory analysis in almost all statistical problems is usually based on well-known fundamental statistical methods for performing exploratory analysis of data. These methods span graphing to the application of descriptive statistics. In particular, in terms of graphical methods, scatter plots and boxplots were considered for learning the intrinsic relationships existing among the defining variables of the physiological data. In addition, descriptive measures considered as numerical measures of samples describing the center and dispersion of the underlying distribution of the data were considered.

Preliminary Data Analysis Results

This section presents the analysis results of the application of the exploratory analysis methods. Table 2 shows the summary statistics of the data. The physiological vital signs that characterized the data comprise Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Mean Arterial Pressure (MAP), Pulse Pressure (PP), and Heart Rate (HR). It can be inferred that almost all the values of each of SBP, DBP, MAP, PP, and HR are within $114.20 \pm (3 \times 13.89)$, $72.80 \pm (3 \times 11.38)$, $86.43 \pm (3 \times 10.69)$, $41.40.43 \pm (3 \times 9.99)$ and $71.69 \pm (3 \times 11.05)$ respectively, assuming approximately normal distribution for the vital signs.

Also, considering the standard nature of the physiological vital signs measurements, it is clear that the SBP values exhibit higher dispersion (are more variable) ($sd = 13.89$) than the rest of the vital signs. However, the DBP and

HR values exhibit quite similar dispersion with empirical standard deviations 11.38 and 11.05. The MAP values on the other hand recorded the largest dispersion ($sd = 10.69$) in comparison with the PP. This is typical of physiological vital signs measurements due to the possible complex relationship it has with other defining health indicators such as activity level, sleep pattern, nutrition or diet, environmental factors, etc.

It is well known that the intensity of activity influences the vital physiological organs of the human body, thus affecting the functioning of these organs. This may eventually show up in the readings obtained from vital organs termed physiological vital signs.

Table 2: ***Summary Statistics of Physiological Vital Signs with Sample Size*** $\mathbf{n = 848 \times 5}$.

Statistic	SBP	DBP	MAP	PP	HR
Minimum	78.00	44.00	55.00	20.00	45.00
Maximum	164.00	118.00	131.00	79.00	128.00
Mean	114.20	72.80	86.43	41.40	71.69
Median	114.00	73.00	86.00	41.00	72.00
Standard deviation (sd)	13.89	11.38	10.69	9.99	11.05
Skewness (η_3)	0.154	0.064	0.211	0.390	0.191

Source: Researcher's Computations (2021)

The typical nature of each of the corresponding vital signs characterizing the physiological vital sign data are shown in Figure 14. The functional nature of the vital signs trends is evident with each exhibiting varied frequencies according to the intrinsic features of the sample patients. It is clearly obvious from the graphs that some points are substantially deviating from the common functional pattern evident in each of the vital signs data. This departure observed may to due to many hidden reasons or factors, thus, these points can be viewed as potential outliers.

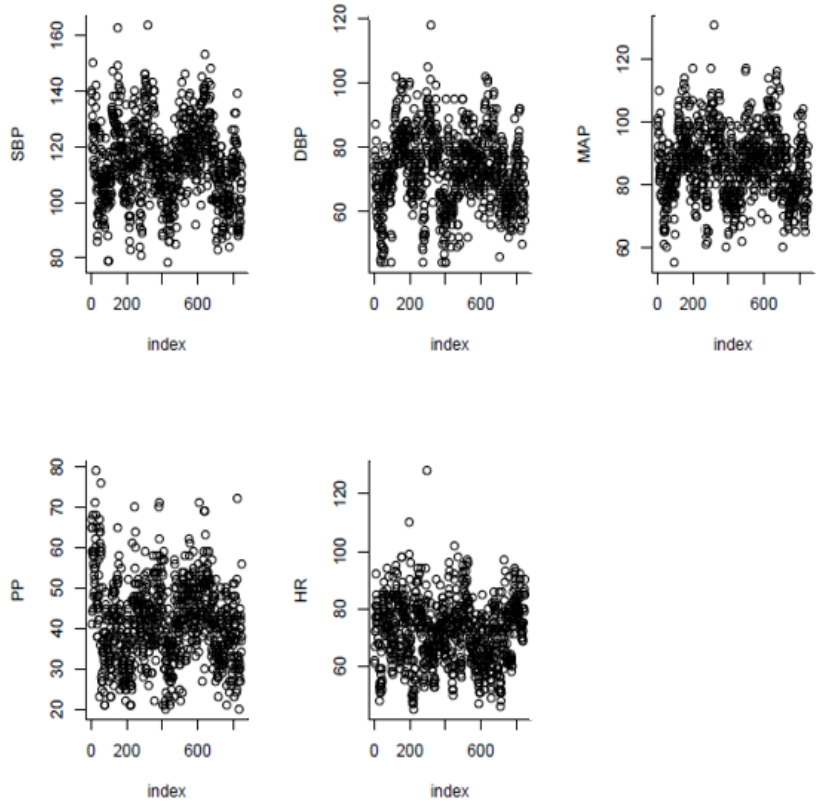

Figure 14: Scatter Plots of Vital Signs. From Left to Right are the Plots of SBP, DBP, MAP, PP, and HR Respectively.

Figure 15 shows the nature of the distribution of the vital signs in terms of the boxplot. Clearly, all the vital signs are skewed to the right with Pulse Pressure (PP) exhibiting the highest skewness and Diastolic Blood Pressure being modest. Though it is hard to quantify the degree of skewness based on Figure 15, Table 2 presents a simple way via a numerical measure termed coefficient of skewness. Thus, arranging the vital signs distributions in terms of the level of skewness associated with each, the obvious order can be expressed as

$$DBP < SBP < HR < MAP < PP.$$

Meaning, the underlying distribution of the Diastolic Blood Pressure observations is less skewed to right than the remaining, while that of the Pulse Pressure shows more shift to the right.

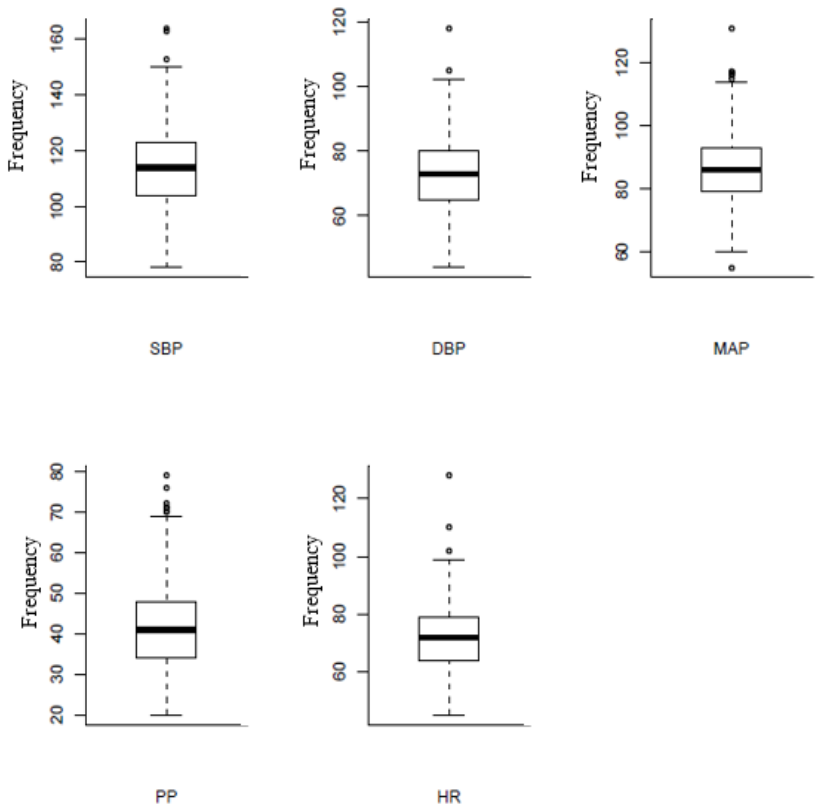

Figure 15: Boxplot of Vital Signs. From Left to Right are the Plots of SBP, DBP, MAP, PP, and HR Respectively.

Figure 16 illustrates the probability distribution function of the vital signs based on Kernel Density Estimation and the underlying inter-relationships existing among the vitals signs in terms of scatter plots and numerical measures termed Pearson Correlation Coefficients. It is apparent that there exist linear relationships among physiological vital signs. In particular, SBP is highly positively related to MAP and DBP with Pearson Correlation Coefficient values of 0.82 and 0.70 respectively.

It also recorded a moderate positive linear relationship with PP, characterized by a Pearson Correlation Coefficient estimate of 0.59. However, an appreciable positive linear relationship exists between SBP and HR. Furthermore, DBP is strongly positively correlated with MAP but negatively correlated with PP on the lower scale with a Pearson Correlation Coefficient value of -0.16. In terms of HR, an appreciable positive linear relationship can be seen with DBP and MAP.

Nevertheless, it does not show any linear pattern with PP. It can be observed further that Figure 16 provides information in support of the right skewness of the distribution of vital signs.

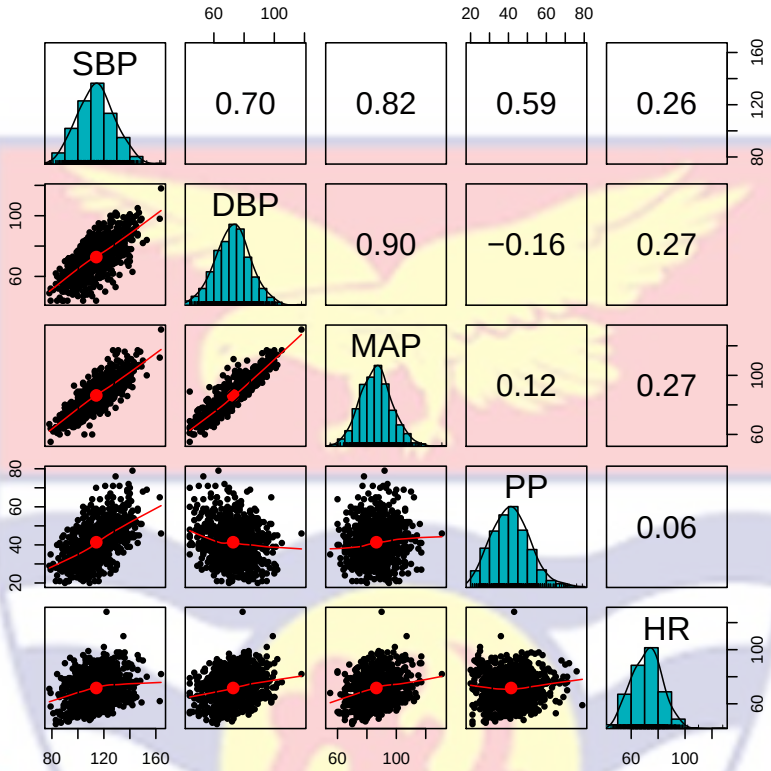

Figure 16: Nature of Vital Signs Data Showing Densities, Inter-relationships and Pearson Correlation Coefficients.

Application of Methods

The proposed schemes were then applied to the pre-processed dataset as follows. First, the data fusion schemes were applied to the dataset variable-wise (i.e each variable was subjected to the data fusion schemes) to extract composite data. That is, one-dimensional datasets were generated from the multi-dimensional dataset. Second, the extracted one-dimensional dataset was then modeled based on the Bayesian Gaussian process regression specified in model (3.14). For the data fusion implementation here, two different ways for selecting the α were considered. The first way is when the experimenter specifies possible values based on the restriction on α. In this direction, the candidate

for α was set as

$$\alpha_e = \{0.1, (0.1+d), (0.1+2d), \ldots, 1\}, \tag{4.1}$$

where $d = 0.05$. The second approach for selecting α considered the statistic based on the ratio of the median of the $T(y)$ and $S(y)$ statistics. Let $\tilde{m}_1$ and $\tilde{m}_2$ denote the medians of $T(y)$ and $S(y)$ respectively. Then, the above statistics is of the form

$$\alpha_r = \frac{\tilde{m}_1}{\tilde{m}_2}, \tag{4.2}$$

The results obtained following the above steps are presented in the next section. In the implementation here, standardized versions of the proposed statistics were considered, particularly, considering response for the Gaussian process regression fitting.

Univariate Vital Signs Feature and Data Fusion Application

This section focuses on the results of the application of data fusion methods on health data. The key principle in fusing multivariate functional data is having similarity or common pattern for defining variables of the dataset. It is well known that near points offer a better prediction of a given point in regression analysis. It can be deduced that similar patterns extracted from each variable on which multivariate dataset is collected can ensure easy fusing than variable patterns. Most especially when the fused pattern is expected to follow or mimic the underlying data structure. For practical purposes, similar features would provide a better background for deriving composite patterns, ensuring an easy basis for comparison.

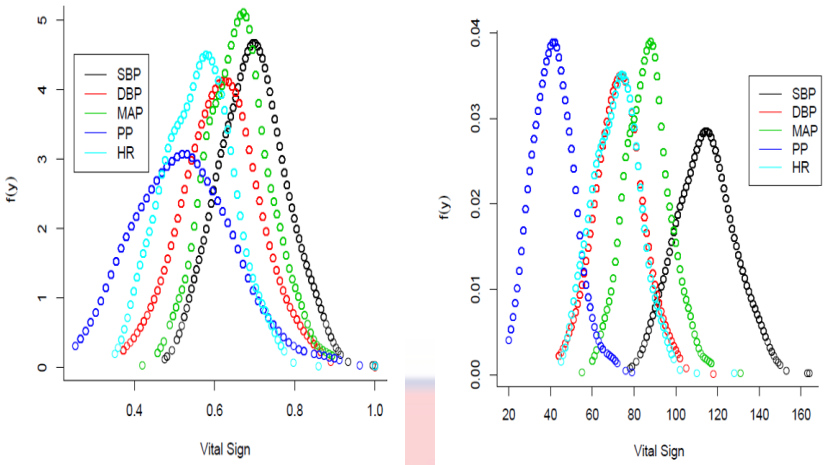

Figure 17: Nature of Vital Signs Specific Density Plots for Standardized and Unstandardized physiological data.

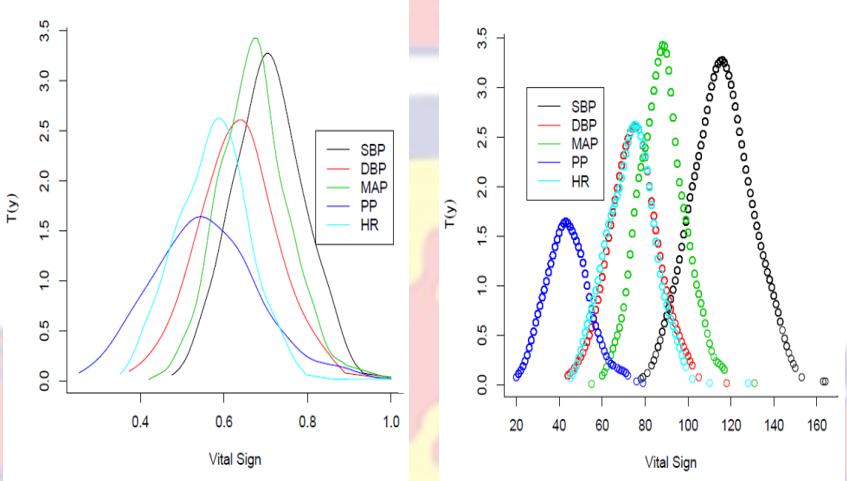

Figure 18: Nature of Vital Sgns Specific $T(y)$ Statistics for Standardized and Unstandardized Physiological Data.

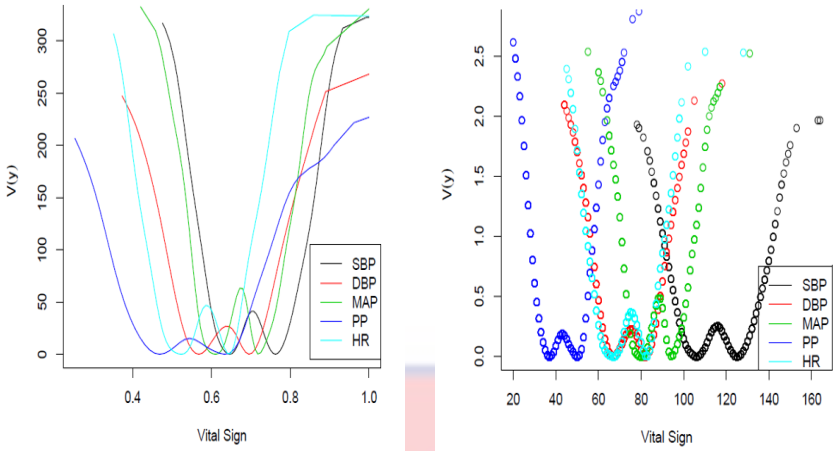

Figure 19: Nature of Vital Signs Specific $V(y)$ Statistics for Standardized and Unstandardized Physiological Data.

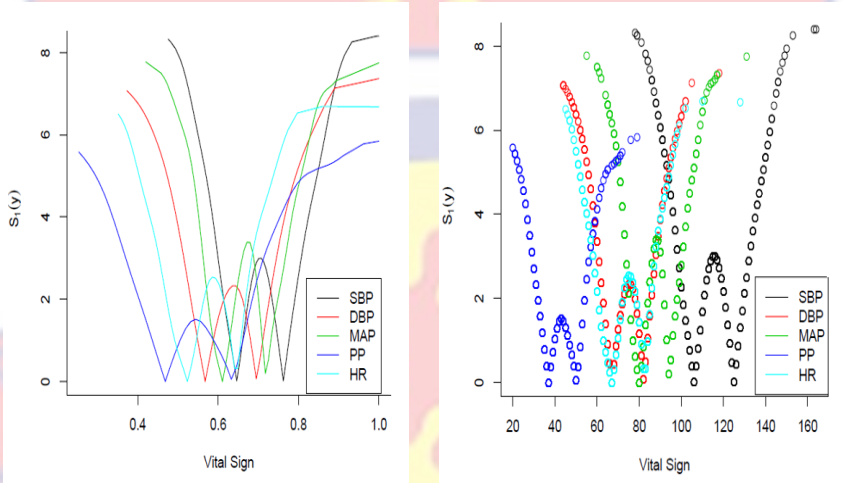

Figure 20: Nature of Vital Signs Specific $S_1(y)$ Statistics for Standardized and Unstandardized Physiological Data.

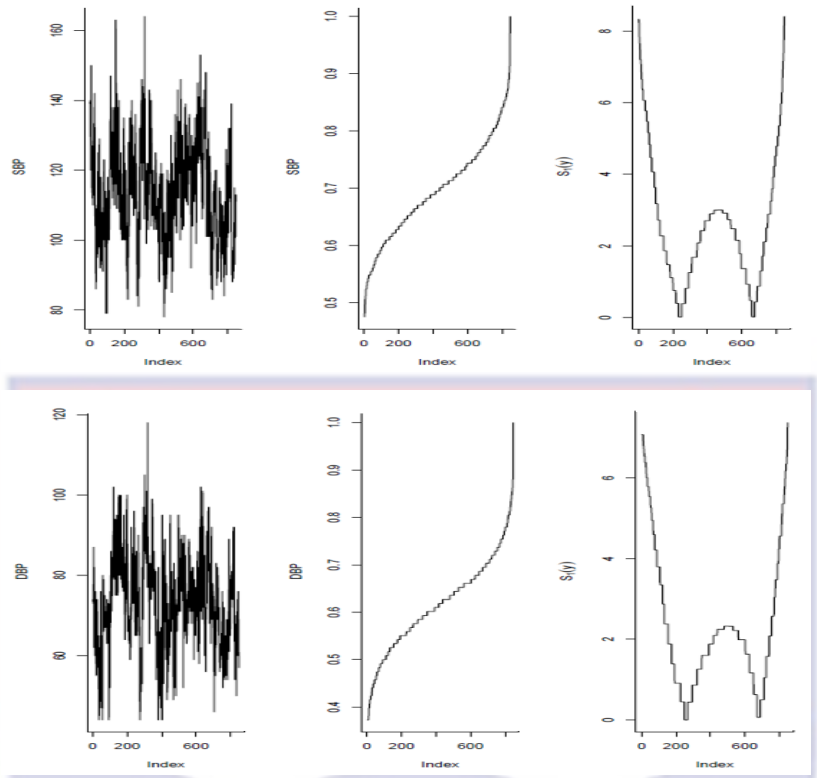

Figure 21: Plot of Vital Signs with their Corresponding Order Statistics Trend and $S_1(y)$.

Figure 17 shows the estimated probability density function for the standardized and unstandardized vital signs data using model (3.6). The plot on the left gives the probability density functions for the standardized vital signs whiles that on the right denotes the probability density functions of the unstandardized counterparts. The differences in vitals signs are clearly characterized by the density functions. The effect of standardization as a way to obtain common distributional characteristics is apparently seen by the level of closeness of the probability density functions observed in both plots. Furthermore, the ability of the density function to handle extreme observations at varying degrees (levels) is well illustrated. This is evidenced by the spread of the density functions.

Figure 18 shows the plot of the statistics $T(y)$ against each of the vital signs observations for both the standardized and unstandardized data. From left to right are the plots for the standardized and unstandardized data respectively. It can be observed that the statistic can inherit the structures of the underlying

probability density functions of the corresponding vital signs observed in Figure 18. For the unstandardized variable, it is difficult to distinguish between some vital signs via the $T(y)$ statistic, for example, that of Heart rate (HR) and Diastolic blood pressure (DBP). However, this challenge is addressed in the case of the standardized variables. In general, the background (foundation) for superimposition is better illustrated in the standardized data than the unstandardized counterpart. This observation presents $T(y)$ as an appealing candidate statistic for fusing multivariate vital signs into a univariate vital sign within the standardized variable context.

The utility of the $V(y)$ statistic in the extraction of common features across vital signs as a basis for fusing multivariate vital signs into a one-dimensional vital sign is shown in Figure 19. The plot on the left is the plot of $V(y)$ against each of the standardized vital signs. The plot on the right shows the plot of $V(y)$ against each of the unstandardized vital signs. It can be observed that the features generated by this statistic exhibit a similar pattern with the underlying difference existing among the vital signs, illustrated high for the unstandardized variables.

Figure 20 shows the nature of the $S_1(y)$ statistics in relation to the corresponding vital signs. Again, the statistic is able to yield common features across vital signs, providing a appealing source for combining the different physiological vital signs into a single physiological vital sign feature. Also, the difference in using the features based on the standardized variables against the unstandardized counterparts becomes apparent.

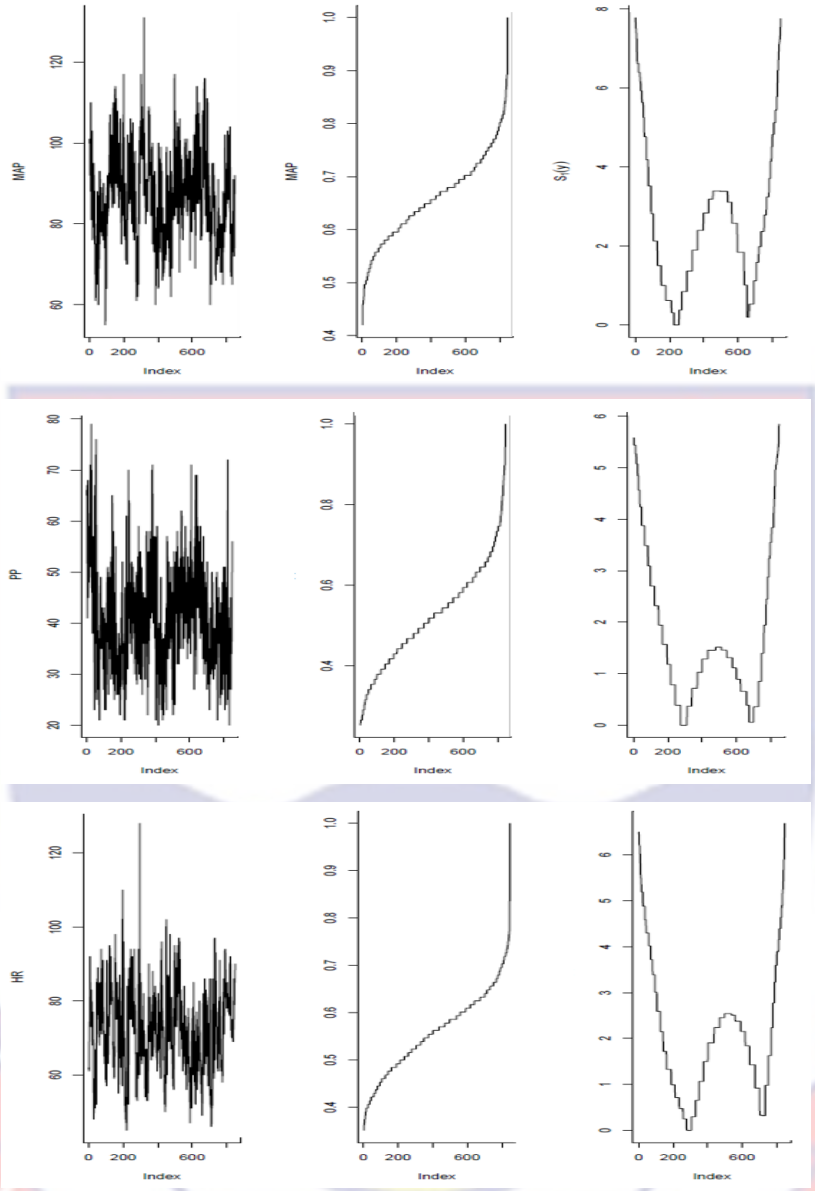

Figure 22: Plot of Vital Signs with their Corresponding Order Statistics Trend and $S_1(y)$.

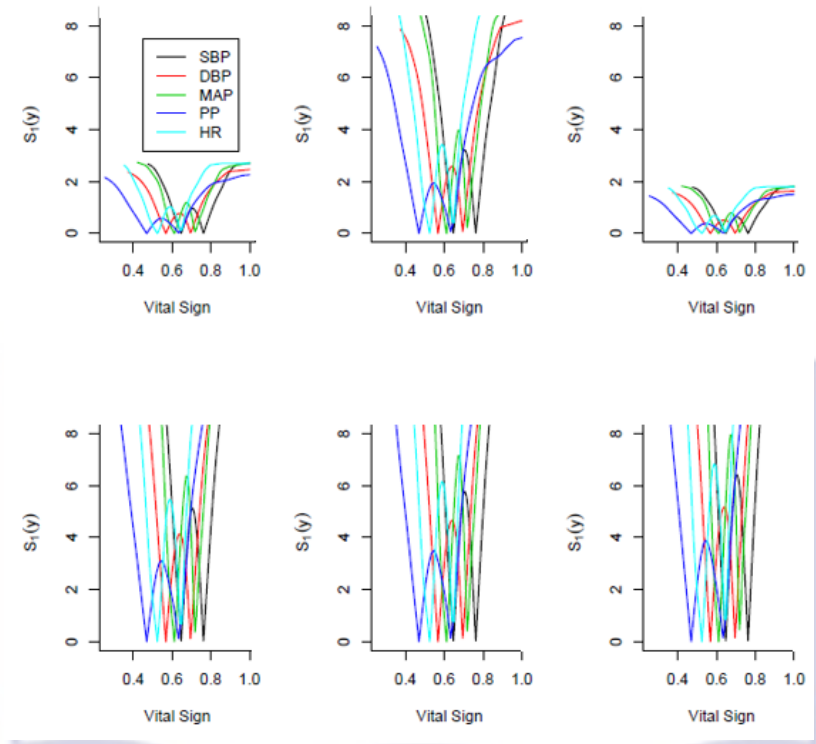

Figure 23: The effect of α_e on $S_1(y)$ Statistics. From Left to Right are the Plots of $S_1(y)$ for 6 Randomly Selected α_e Values Corresponding to $0.15, 0.50, 0.10, 0.80, 0.90$ and 1.00 Values.

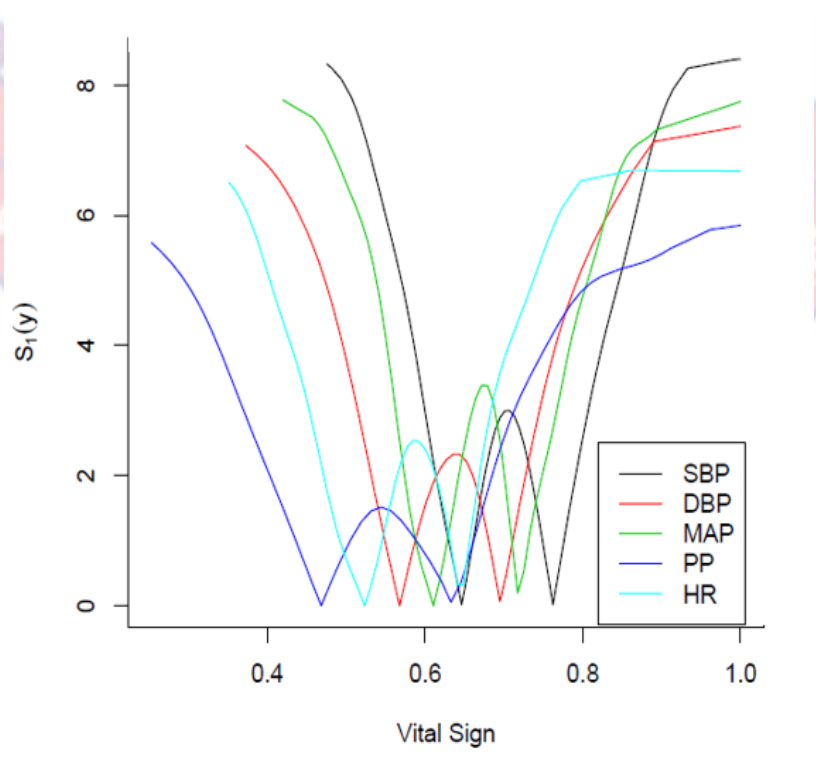

Figure 24: Real data: Dynamics of Vital Sign Specific Features with Vital Sign Specific α_r.

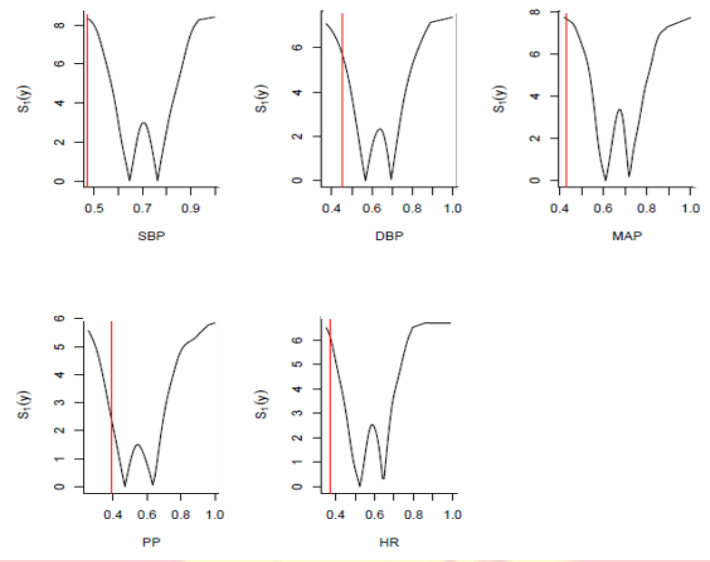

Figure 25: Nature of Vital Sign Specific Tuning Parameter α_r.

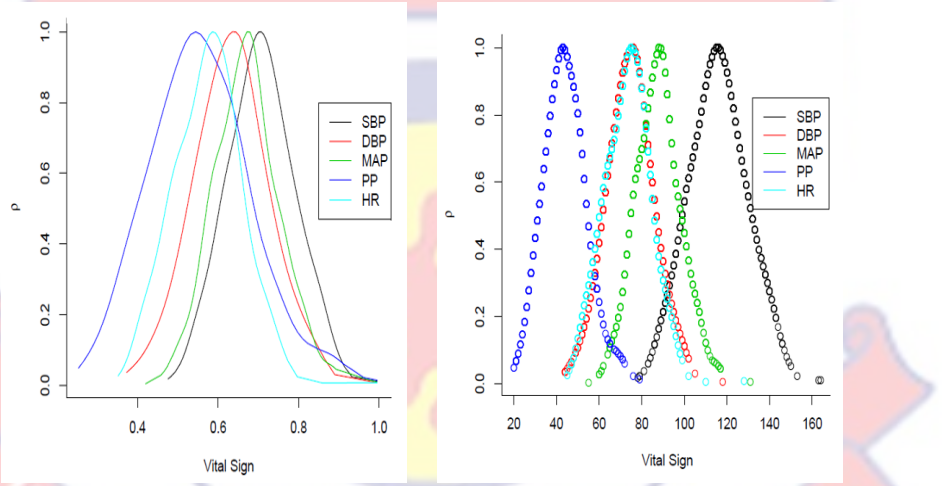

Figure 26: Dynamics of Vital Signs Specific ρs for Statistics for Standardized and Unstandardized Physiological Data.

Figure 21 and Figure 22 show the vital signs and their corresponding features obtained based on the order statistics of the original data. The plots on each row give the scatter plot, the order statistic plot, and $S_1(y)$ statistic plot, for each vital sign. The high-frequency observations at varying degrees, characterizing each vital sign are evident. Clearly, unique features can be extracted from high-frequency observations via density-based features.

The typical effect of the choice of α on the dynamics of the probability density based statistics, $S_1(y)$ in (3.5) for data fusion via features are shown

in Figure 23 and Figure 24. The plots in Figure 23 were obtained from the implementation using α_e specified in (4.1). However, the plot in Figure 24 was based on the implementation when α was set to α_r defined in (4.2). The plots from left to right are the plots of $S_1(y)$ against the standardized vital sign values. The 6 candidates for α_e were randomly selected with numerical value $0.15, 0.50, 0.10, 0.80, 0.90, 1.000$. From left to right are the resulting plots. The αs exhibit scaling effect on the statistic, $S_1(y)$ but does not alter the underlying common structure defined by the features.

Figure 25 presents the nature the vital sign specific tuning parameters, α_r in relation to their corresponding $S_1(y)$ statistics. The red lines are the vital sign-specific estimates of α_r. The variability in the effect of tuning offered by each vital sign is evident.

Figure 26 illustrates the nature of the vital sign specific candidates for ρ defined in (3.13) for both the standardized and unstandardized physiological variables. The graph on the right represents the standardized data whiles the one on the left is that of the unstandardized data. The differences in the ρ values in terms of numerical magnitude are clearly exhibited although, the range is the same.

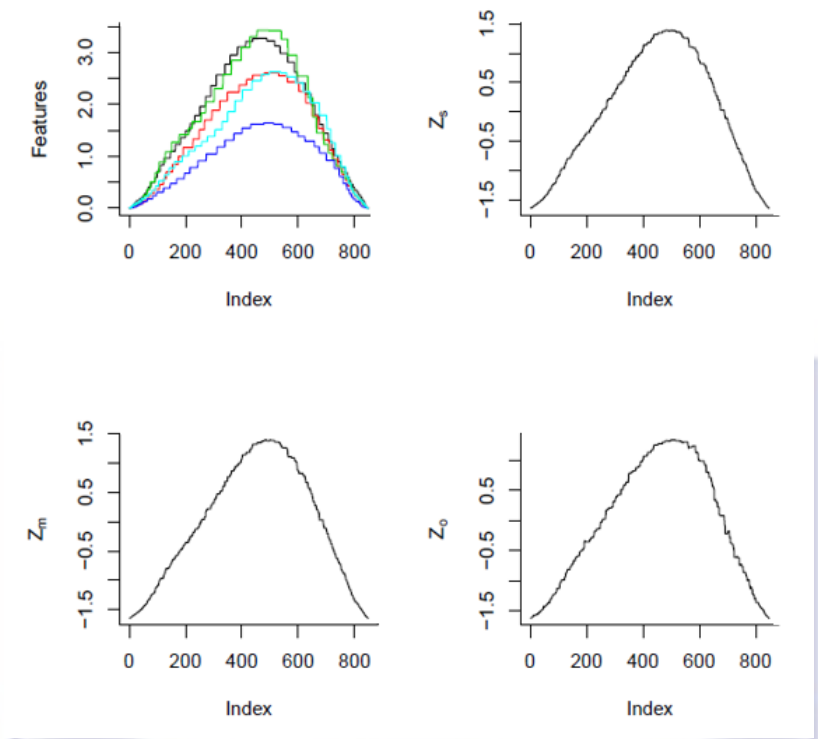

Figure 27: Fused Vital Signs Trend Based on $T(y)$ Statistic.

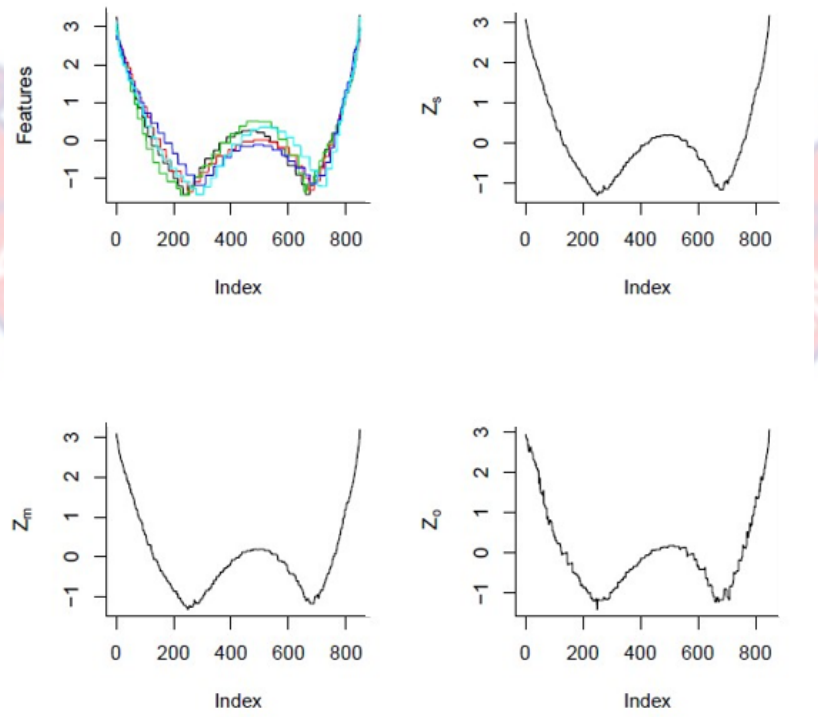

Figure 28: Fused Vital Signs Trend Based on $S_1(y)$ Statistic.

Figure 27 shows the composite vital signs trend obtained from the multivariate vital signs data based on the application of models (3.7),(3.8) and

(3.9). The first graph gives the vital sign specific components, $\rho_i T(y_i)$ for $i = 1, 2, \ldots, 5$. The remaining plots are the fused trends, Z_s, Z_m and Z_o respectively. The nature of the fused vital sign trend obtained from the application of fusion models (3.10),(3.11) and (3.12) are illustrated in Figure 28. From left to right are the vital signs specific components, $\rho_i s_1(y_i), i = 1, \ldots, 5$ plots, and graphs of Z_s,Z_m and Z_o respectively.

Table 3: ***Comparison of Vital Sign Specific α and δ for Standardized Dataset.*** $\mathbf{y_1 = SBP, y_2 = DBP, y_3 = MAP, y_4 = PP, y_5 = HR, y_c = \sum_{i=1}^{5} y_i}$

	y_1	y_2	y_3	y_4	y_5	y_c
α_s	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	$\hat{\alpha}_c$
	0.468	0.450	0.426	0.388	0.371	0.985
δ_s	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_3$	$\hat{\delta}_4$	$\hat{\delta}_5$	$\hat{\delta}_c$
	0.0234	0.0263	0.0195	0.0355	0.0225	0.1333

Source: Researcher's Computations (2021)

Table 4: ***Comparison of Vital Sign Specific α and δ for Unstandardized Dataset.*** $\mathbf{y_1 = SBP, y_2 = DBP, y_3 = MAP, y_4 = PP, y_5 = HR, y_c = \sum_{i=1}^{5} y_i}$

	y_1	y_2	y_3	y_4	y_5
α_u	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$
	5.987	4.886	4.876	3.449	4.199
$\alpha_n = \frac{\hat{\alpha}_u}{\sum \hat{\alpha}_u}$	0.256	0.209	0.208	0.147	0.179
δ_u	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_3$	$\hat{\delta}_4$	$\hat{\delta}_5$
	0.0234	0.0263	0.0195	0.0355	0.0225

Source: Researcher's Computations (2021)

Table 3 and Table 4 report the estimates of α_s and δ for each of the vital signs for the standardized and unstandardized data. $\hat{\alpha}_i$ for $i = 1, \ldots, 5$ are gives the estimates for SBP, DBP, MAP, PP and HR. However, $\hat{\alpha}_c$ denotes the estimate for the combined standardized data $y_c = \sum_{i=1}^{5} y_i$. The α_n in Table 4 are normalized values of α_s for unstandardized data. Again, the underlying

differences existing among the physiological vital signs are clearly seen based on the α values for the two two datasets (standardized and unstandardized). Furthermore, the effect standardization is evidently illustrated.

Gaussian Process Regression Application on Fused Vital Signs Data

This section reports the results obtained from the application of the Gaussian process regression model (3.14) to one of the fused vital sign data proposed in the methodology. The illustration of the Gaussian process regression (GPR) arm of this thesis was given a particular consideration based on the uniqueness of the resulting fused trend such that the GPR's capability for shape functional regression can be assessed. The application of the fusion methods to the physiological vital sign data considered in the previous section provided insightful information on the typical nature of the resulting composite trend obtained from the fusion based on each of the statistics, $T(y)$ and $S(y)$.

It was observed that the $S_1(y)$ based fusion schemes yielded shaped functional trends with relatively complex structures than those of $T(y)$ statistic. Thus, the GPR was conducted on the fused vital signs derived from model (3.10), model (3.11), and model (3.12) respectively.

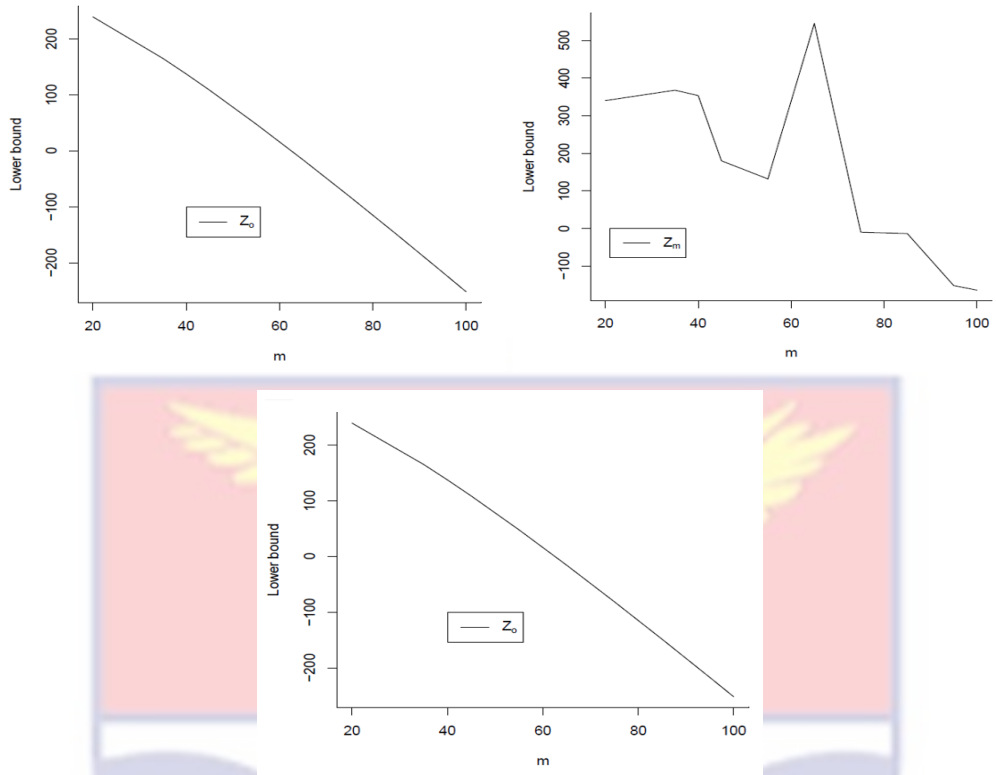

Figure 29: Plot of Llower Bound against Spectral Sample Size m for VBFBGPR.

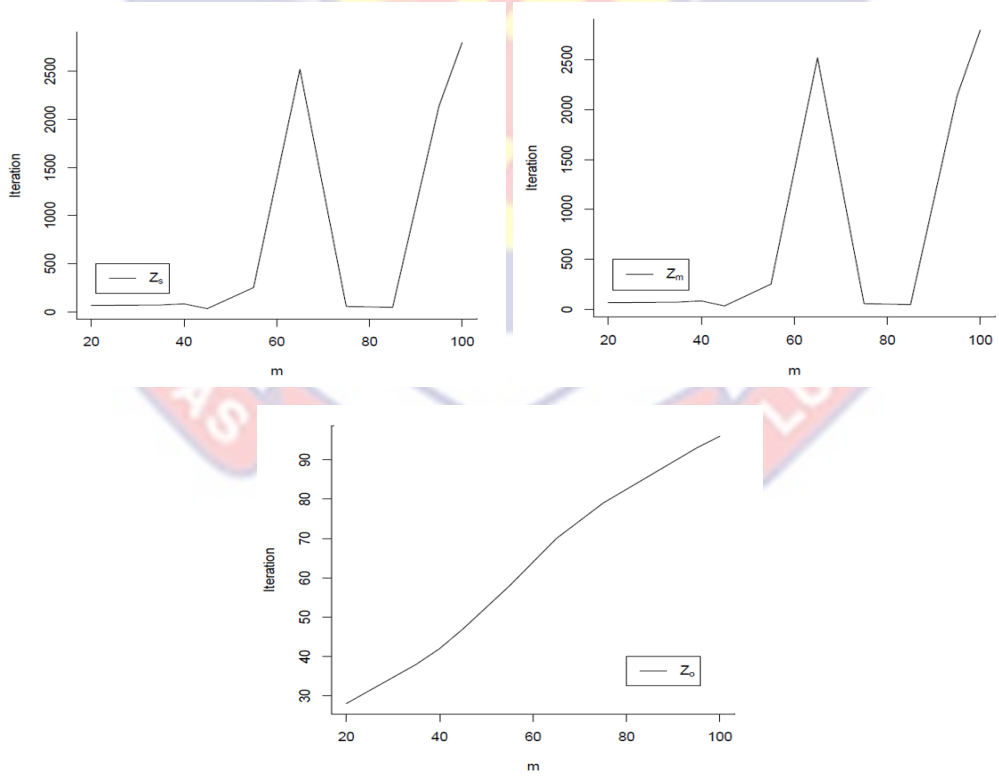

Figure 30: Plot of Iteration against Spectral Sample Size m. for VBFBGPR.

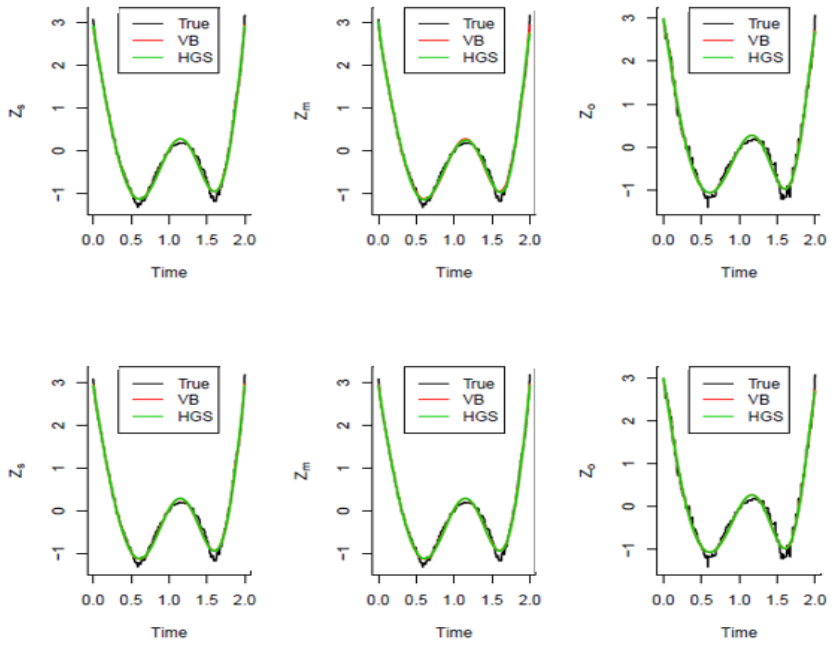

Figure 31: Fitted Fussed Vital Sign with $m = 40$ *(row* 1*) and* $m = 45$ *(row* 2*)*

Figure 29 illustrates the characteristics of the variational lower bound with increasing spectral sample sizes, m, for each of the fused vital sign trends, Z_s ,Z_m, and Z_o. Whiles the lower bound trend for Z_s and Z_m are nonlinear and similar, that of Z_o shows a decreasing linear trend.

Figure 30 shows the plot of the number of iterations to the convergence of the variational algorithm (VBFBGPR) against the spectral samples for fused vital sign trends, Z_s, Z_m and Z_o. Again, a similar nonlinear pattern is exhibited for both Z_s and Z_m. Apparently, an increasing linear pattern is reported in the case of Z_o as m increases.

The performance of the VBFBGPR and HGS in fitting shaped functional fused physiological vital signs is illustrated in Figure 31. The plots in row 1 are the fitted trends for Z_s, Z_m and Z_o respectively using $m = 40$. The row 2 plots give the fitted trends for Z_s, Z_m and Z_o for spectral sample of size, $m = 45$. Again, both methods are able to recover the underlying shaped trend very well.

Table 5: ***Fitting Performance of VB and HGS Based on MSEFE for selected m.***

m	$Z(s)_{VB}$	$Z(s)_{HGS}$	$Z(m)_{VB}$	$Z(m)_{HGS}$	$Z(o)_{VB}$	$Z(o)_{HGS}$
35	0.00788	0.00763	0.00788	0.00765	0.01310	0.01300
40	0.00786	0.00766	0.00786	0.00633	0.01311	0.01303
45	0.00793	0.00767	0.00793	0.00771	0.01312	0.01304
55	0.00765	0.00620	0.00765	0.00621	0.01311	0.01301

Source: Researcher's Computations (2021)

Table 6: ***Fitting Performance of VB and HGS Based on MAFE for Selected m.***

m	$Z(s)_{VB}$	$Z(s)_{HGS}$	$Z(m)_{VB}$	$Z(m)_{HGS}$	$Z(o)_{VB}$	$Z(o)_{HGS}$
35	0.07304	0.07180	0.07304	0.07192	0.09328	0.09279
40	0.07305	0.07203	0.07305	0.06309	0.09333	0.09295
45	0.07275	0.07206	0.07275	0.07225	0.09335	0.09297
55	0.07181	0.06229	0.07181	0.06231	0.09336	0.09285

Source: Researcher's Computations (2021)

Table 7: ***Fitting Performance of VB and HGS Based on SMAFE Selected*** **m.**

m	$Z(s)_{VB}$	$Z(s)_{HGS}$	$Z(m)_{VB}$	$Z(m)_{HGS}$	$Z(o)_{VB}$	$Z(o)_{HGS}$
35	0.33829	0.33620	0.33829	0.33913	0.74441	0.73602
40	0.33361	0.33773	0.33361	0.31638	0.74587	0.73765
45	0.35536	0.33595	0.35536	0.33703	0.74693	0.73895
55	0.33948	0.31362	0.33948	0.31406	0.74805	0.73743

Source: Researcher's Computations (2021)

Table 5, Table 6 and Table 7 show the Gaussian process regression fitted errors for fussed vital signs for the VB and its MCMC counterpart using the health data set for some selected spectral sample sizes namely, $m = 35, 40, 45$ and 55. The selection was based on the fact that the HGS results were not available for those omitted values of m. It can be observed that all the methods exhibit similar fitting performance and that recovery performance increases with increased spectral points. This observation is important in that any one of the methods could be applicable for inference in the future in the case where it is difficult to monitor traumatic events and the Bayesian Gaussian process regression model is adopted as data generating process and trend recovery is of interest.

Chapter Summary

This chapter examined the plausible inferences underlying the results obtained upon application of the developed methods to both the simulated and real datasets. In addition, implications associated with the results in relationship modeling are also established. Beginning with the data fusing methods, the results revealed that the probability density function (pdf)-based features can provide a formal way to extract shape restricted vital sign features that are robust to the presence of extreme measurements.

Also, it can serve as a building block for constructing other interesting functional features with the ability to inherit the automated extreme value controlled property mentioned above. With the Gaussian process application, the results illustrated the ability to fit Gaussian process regression models in the feature space using the Bayesian principle.

In particular, the developed algorithms, namely the VB and its MCMC counterpart (HGS) based on the feature, both have the ability to recover the underlying composite vital sign trajectories in both the simulated and real datasets. However, an increasing pattern exists with the spectral sample size. Furthermore, the overall fitting performance in terms of MSFE, MAFE, and SMAFE are better in both algorithms.

Interestingly, a decreasing fitting error pattern exists with the spectral sample size with the HGS reporting a better (higher) decrease than its VB counterpart. Regarding the computational savings, it turned out that the modeling of the composite vital sign added to the computational savings achieved by the VB algorithm. In general, the possibility to reduce multi-task GPR to composite single-task GPR in the Bayesian framework for a multivariate vital sign with common time-stamps, characterized with both time-dependent and non-time dependent covariates has been illustrated with the use of pdf-based statistics.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Summary

A number of research have been done in the field of Bayesian Gaussian process regression in its sub-domain of single task Gaussian process regression, and multi-task Gaussian process regression. For example, Dürichen et al. (2014) used MTGPR in a biomedical application in relation to vital signs monitoring in intensive care unit (ICU). It was seen from literature that, the inclusion of co-variate in MTGPR was avoided especially the non-time dependents ones. Such information may be vital in explaining the underlying dynamics of the observed physiological trend. Omitting it leads to loss of information that may affect model performance.

Though Ofori (2020) considered resolving the issue of incorporation of non-time dependent covariates in the Variational Bayes GPR using the idea of robust statistics (Orthogonalized Gnanadesikan-Kettering (OGK)) (Maronna et al., 2019), it was in the one-dimension direction, just for traumatic systolic blood pressure response. It is important to note that the observation made by Ofori (2020) regarding the use of non-time dependent covariates was similar to that observed in this thesis.

Ofori (2020) realized that non-time covariates allow easy assessment of the relative importance of both trauma-specific and subject-specific features. Most importantly, these covariate impact the smoothness of the underlying trend existing in the observed data compositely and aid in model fitting. It was also opined that treating the smoothness of trend as a random variable and modeling it allows the uncertainty associated with it to be well quantified and calibrated so that its use for the assessment of how it relates with available non-time co-variates can be achieved properly.

Thus, its utility in public health, particularly, clinical studies was outlined via a comparison with the empirical smoothness index statistic widely used

in the management and treatment of hypertension (Mensah, Ofori, & Howard, 2021). Another vital observation by most of the authors aforementioned (Mensah et al. (2016), Ofori (2020), Mensah, Ofori, and Howard (2021)) who employed the spectral GPR approach is that spectral sample size exhibits some remarkable effect on the computational expenditure of Variational Bayes Algorithms. In particular, larger spectral samples can generate different dynamics for the number of iterations an VB algorithm requires for convergence as realized in this thesis. Regarding the smoothness of trend, this thesis took a different dimension. In that, the composite smoothness of all vital signs was considered so that the joint effect of the vital sign-specific smoothness can be examined.

Another line of uniqueness of the approach adopted in this thesis can be seen in the modeling perspective. In this study, we exploited a novel method for handling multi-task modeling of multivariate physiological vital signs measured over a common time-stamp within the Gaussian process regression framework. Usually, the Gaussian process regression for modeling multivariate physiological vital signs uses the multi-talks approach where each of the multiple vital signs is modeled simultaneously using Gaussian process regression based on a union of time-dependent covariates. This study exploited the joint-use of kernel density estimation and Gaussian process regression within the Bayesian framework to examine the appropriateness of double-stable modeling of multivariate physiological vital signs as an alternative to multi-task Gaussian process regression of such vital signs.

In this regard, the study considered fusing the multivariate vital signs into a univariate vital signs trajectory and modeling the resultant vital sign using Gaussian process regression. The computation of non-central moments for probability density functions was explored to develop appropriate robust data fusion methods for extracting composite vital sign trajectory in the first stage of the dual modeling framework proposed. In the second arm of the dual modeling framework, appropriate Gaussian process regression data generative models were developed for the composite vital sign trajectory.

Non-time-dependent covariates are problematic in Gaussian process regression since their incorporation is non-trial. However, with the use of robust vector covariates based on Orthogonalized Gnanadesikan-Kettering (OGK) statistics (Maronna et al., 2019), non-time-dependent covariates were handled, simply via hierarchical modeling of the Gaussian process covariance function.

Furthermore, both approximate and exact inference methods using Variational Bayes and hybrid MCMC methods were developed for parameter inference. The proposed double-state modeling framework was implemented and experimented using simulated and real datasets. The results illustrated its appropriateness for handling multi-task Gaussian process modeling of multivariate physiological vital signs within the single-task domain.

Conclusions

In this thesis, we have proposed and implemented a novel two-stage modeling framework for handling multi-task Gaussian process regression modeling of multiple physiological vital signs with common time-dependent predictors, within the single-task Gaussian process regression context. The proposal couples nonparametric and Bayesian methods. In particular, nonparametric Kernel density estimation is explored to develop novel probability density function-based schemes for fusing the variant trajectories of the multivariate physiological vital signs into a univariate trajectory for easy modeling, using Gaussian process regression GPR.

The fused (composite) vital sign trajectory is then modeled in relation with the common time-dependent predictor using Gaussian processes with available non-time-dependent covariates incorporated into the GPR via the covariance function, using OGK statistics. Also, we have developed Variational Bayes and hybrid MCMC methods based on sparse spectral approximation using modified Van der Waerden statistics for parameter inference.

The two-stage proposal allowed for automatic control of extreme values in both the response and predictor spaces; ensured dimension reduction in the response space; ensured data reduction in the predictor space and allowed hierarchical modeling of composite smoothness of trend with existing covariates for easy extension of Variational Bayes computational methods. An application to simulated and real physiological vital signs datasets illustrated the potential and the utility of the above proposal for multi-task modeling.

Recommendations

This section of the thesis focuses on some vital recommendations and directions for possible future work based on the results obtained from the implementation of the developed methods. In particular, the following recommendations are considered. We recommend

1. the use of spectral approximation framework for Bayesian inference to avoid computational issues associated with large multivariate physiological vital signs observed over a common time-stamps.
2. the use of probability density function-statistic (features) for extracting single-variate vital sign trajectory from a multivariate vital signs dataset since it can ensure automatic control for extreme values.

In terms of future work, we consider the following.

1. The application of probability density function (pdf) specific features for functional shape restricted regression using Bayesian Gaussian processes. From our experimentation, it was realized that the pdf-based features developed yielded unique shape restricted functional trend dynamics of the observations with the presence or otherwise of extreme observations automatically handled.

2. Extension of the expected value contribution-based features for fusing multivariate datasets to other non-central moments-based methods. The data fusion approach considered in the thesis was centered on the first central moment. It will be interesting to explore the possibility of using other moments for future work.

3. Extension of the developed two-stage modeling framework to multivariate physiological vital sign dataset with uncommon time-stamps and both time-dependent and non-time-dependent covariate information.

REFERENCES

Andrieu, C., & Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and Computing*, *18*(4), 343–373.

Attias, H. (2000). A variational baysian framework for graphical models. In *Advances in neural information processing systems* (pp. 209–215).

Baird, T. M., & Neuman, M. R. (1991). Effect of infant position on breath amplitude measured by transthoracic impedance and strain gauges. *Pediatric pulmonology*, *10*(1), 52–56.

Bedell, S. E., Deitz, D. C., Leeman, D., & Delbanco, T. L. (1991). Incidence and characteristics of preventable latrogenic cardiac arrests. *Jama*, *265*(21), 2815–2820.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

Blunt, I., Bardsley, M., & Dixon, J. (2010). *Trends in emergency admissions in england 2004-2009: is greater efficiency breeding inefficiency?* Nuffield Trust.

Bochner, S., et al. (1959). *Lectures on fourier integrals* (Vol. 42). Princeton University Press.

Buist, M. D., Burton, P. R., Bernard, S. A., Waxman, B. P., & Anderson, J. (1999). Recognising clinical instability in hospital patients before cardiac arrest or unplanned admission to intensive care: A pilot study in a tertiary-care hospital. *Medical Journal of Australia*, *171*(1), 22–25.

Burris, H., Moore, M. J., Andersen, J., Green, M. R., Rothenberg, M. L., Modiano, & Tarassoff, P. (1997). Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial. *Journal of clinical oncology*, *15*(6), 2403–

Chatfield, C. (2013). *The analysis of time series: theory and practice*. Springer.

Clifford, G. D., Azuaje, F., McSharry, P., et al. (2006). *Advanced methods and tools for ecg data analysis* (Vol. 13). McSharry (Ed.). Boston, Artech house.

Clifton, L., Clifton, D. A., Pimentel, M. A., Watkinson, P. J., & Tarassenko, L. (2013). Gaussian processes for personalized e-health monitoring with wearable sensors. *IEEE Transactions on Biomedical Engineering*, *60*(1), 193–197.

Clifton, L., Clifton, D. A., Pimentel, M. A. F., Watkinson, P. J., & Tarassenko, L. (2012). Gaussian process regression in vital-sign early warning systems. In *Engineering in medicine and biology society (embc), 2012 annual international conference of the ieee* (pp. 6161–6164).

De Kock, J., & Tarassenko, L. (1993). Pulse oximetry: Theoretical and experimental models. *Medical and Biological Engineering and Computing*, *31*(3), 291–300.

DeVita, M. A., Smith, G. B., Adam, S. K., Adams-Pizarro, I., Buist, M., Bellomo, R., ... others (2010). "identifying the hospitalised patient in crisis"—a consensus conference on the afferent limb of rapid response systems. *Resuscitation*, *81*(4), 375–382.

Duong, T., et al. (2007). ks: Kernel density estimation and kernel discriminant analysis for multivariate data in r. *Journal of Statistical Software*, *21*(7), 1–16.

Dürichen, R., Pimentel, M. A., Clifton, L., Schweikard, A., & Clifton, D. A. (2014). Multitask gaussian processes for multivariate physiological time-series analysis. *IEEE Transactions on Biomedical Engineering*, *62*(1), 314–322.

Edmonds, Z. V., Mower, W. R., Lovato, L. M., & Lomeli, R. (2002). The reliability of vital sign measurements. *Annals of emergency medicine*, *39*(3), 233–237.

Elliott, M., & Coventry, A. (2012). Critical care: the eight vital signs of patient monitoring. *British Journal of Nursing*, *21*(10), 621–625.

Evans, D., Hodgkinson, B., & Berry, J. (2001). Vital signs in hospital patients: a systematic review. *International journal of nursing studies*, *38*(6), 643–650.

Fieselmann, J. F., Hendryx, M. S., Helms, C. M., & Wakefield, D. S. (1993). Respiratory rate predicts cardiopulmonary arrest for internal medicine in-patients. *Journal of general internal medicine*, *8*(7), 354–360.

Gal, Y., & Turner, R. (2015). Improving the gaussian process sparse spectrum approximation by representing uncertainty in frequency inputs.

Gao, H., McDonnell, A., Harrison, D. A., Moore, T., Adam, S., Daly, K., ... others (2007). Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. *Intensive care medicine*, *33*(4), 667–679.

Gardner-Thorpe, J., Love, N., Wrightson, J., Walsh, S., & Keeling, N. (2006). The value of modified early warning score (mews) in surgical in-patients: a prospective observational study. *The Annals of The Royal College of Surgeons of England*, *88*(6), 571–575.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

Ghassemi, M., Pimentel, M., Naumann, T., Brennan, T., Clifton, D., Szolovits, P., & Feng, M. (2015). A multivariate timeseries modeling approach to severity of illness assessment and forecasting in icu with sparse, heterogeneous clinical data. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 29).

Givens, G. H., & Hoeting, J. A. (2013). *Computational statistics*. Willey.

Hann, A. (2008). *Multi-parameter monitoring for early warning of patient deterioration* (Unpublished doctoral dissertation). University of Oxford.

Hardin, S. R., & Kaplow, R. (2005). Introduction to the aacn synergy model for patient care.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications.

Kause, J., Smith, G., Prytherch, D., Parr, M., Flabouris, A., Hillman, K., et al. (2004). A comparison of antecedents to cardiac arrests, deaths and emergency intensive care admissions in australia and new zealand, and the united kingdom—the academia study. *Resuscitation*, *62*(3), 275–282.

Khalid, S., Clifton, D. A., & Tarassenko, L. (2013). A bayesian patient-based model for detecting deterioration in vital signs using manual observations. In *International symposium on foundations of health informatics engineering and systems* (pp. 146–158).

Knaus, W. A., Wagner, D. P., Draper, E. A., Zimmerman, J. E., Bergner, M., Bastos, P. G., ... others (1991). The apache iii prognostic system: risk prediction of hospital mortality for critically iii hospitalized adults. *Chest*, *100*(6), 1619–1636.

Lasko, T. A. (2014). Efficient inference of gaussian-process-modulated renewal processes with application to medical event data. In *Uncertainty in artificial intelligence: proceedings of the... conference. conference on uncertainty in artificial intelligence* (Vol. 2014, p. 469).

Lichman, M., & Smyth, P. (2014). Modeling human location data with mixtures of kernel densities. In *Proceedings of the 20th acm sigkdd international conference on knowledge discovery and data mining* (pp. 35–44).

Liu, Z., & Hauskrecht, M. (2016). Learning adaptive forecasting models from irregularly sampled multivariate clinical data. In *Proceedings of the... aaai conference on artificial intelligence. aaai conference on artificial intelligence* (Vol. 2016, p. 1273).

Lockwood, C., Conroy-Hiller, T., & Page, T. (2004). Vital signs. *JBI reports*, *2*(6), 207–230.

Magder, S. (2018). The meaning of blood pressure. *Critical Care*, *22*(1), 1–10.

Malik, M., & Camm, A. J. (1990). Heart rate variability. *Clinical cardiology*, *13*(8), 570–576.

Marlin, B. M., Kale, D. C., Khemani, R. G., & Wetzel, R. C. (2012). Unsupervised pattern discovery in electronic health care data using probabilistic clustering models. In *Proceedings of the 2nd acm sighit international health informatics symposium* (pp. 389–398).

Maronna, R. A., Martin, R. D., Yohai, V. J., & Salibián-Barrera, M. (2019). *Robust statistics: theory and methods (with r)*. John Wiley & Sons.

McGhee, T. L., Weaver, P., Solo, S., & Hobbs, M. (2016). Vital signs reassessment frequency recommendation. *Nursing management*, *47*(9), 11–12.

McQuillan, P., Pilkington, S., Allan, A., Taylor, B., Short, A., Morgan, G., ... Smith, G. (1998). Confidential inquiry into quality of care before admission to intensive care. *Bmj*, *316*(7148), 1853–1858.

Mensah, D. K., Eyiah., B., F, & Assabil, S. (2021). *Compendium of tunable data-driven approaches for smoothing parameter learning for kernel density-based normal systolic blood pressure detection.* (unpublished)

Mensah, D. K., Nott, D. J., Tan, L. S. L., & Marshall, L. (2016). Functional models for longitudinal data with covariate dependent smoothness. *Electronic Journal of Statistics*, *10*(1), 527–549.

Mensah, D. K., Ofori, M. A., & Howard, N. (2021). *Traumatic systolic blood pressure modeling: A spectral gaussian process regression approach with robut smaple covariates.* (unpublished)

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, *21*(6), 1087–1092.

Ofori, M. A. (2020). *Gaussian process regression for traumatic systolic blood pressure with robust sample specific covariate dependent smoothness.*

Ormerod, J. T., & Wand, M. P. (2010). Explaining variational approximations. *The American Statistician*, *64*(2), 140–153.

Parzen, E. (1962). Extraction and detection problems and reproducing kernel hilbert spaces. *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*, *1*(1), 35–62.

Pearson, G., & Duncan, H. (2011). Early warning systems for identifying sick children. *Paediatrics and Child Health*, *21*(5), 230–233.

Pimentel, M. A. F., Clifton, D. A., Clifton, L., Watkinson, P. J., & Tarassenko, L. (2013). Modelling physiological deterioration in post-operative patient vital-sign data. *Medical & biological engineering & computing*, *51*(8), 869–877.

Ramsey III, M., Medero, R., & Hood Jr, R. W. (1991, October 1). *Oscillometric blood pressure monitor employing non-uniform pressure decrementing steps.* Google Patents. (US Patent 5,052,397)

Rasmussen, C., & Williams, C. (2006). *Gaussian processes for machine learning*. The MIT Press.

Renton, J., Pilcher, D., Santamaria, J., Stow, P., Bailey, M., Hart, G., & Duke, G. (2011). Factors associated with increased risk of readmission to intensive care in australia. *Intensive care medicine*, *37*(11), 1800.

Schulam, P., Wigley, F., & Saria, S. (2015). Clustering longitudinal clinical marker trajectories from electronic health data: Applications to phenotyping and endotype discovery. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 29).

Schulman, C. S., & Staul, L. (2010). Standards for frequency of measurement and documentation of vital signs and physical assessments. *Critical care nurse*, *30*(3), 74–76.

Scott, D. W. (2015). *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis* (Vol. 26). CRC press.

Simoes, E. A. (2003). Environmental and demographic risk factors for respiratory syncytial virus lower respiratory tract disease. *The Journal of pediatrics*, *143*(5), 118–126.

Stegle, O., Fallert, S. V., MacKay, D. J., & Brage, S. (2008). Gaussian process robust regression for noisy heart rate data. *IEEE Transactions on*

Biomedical Engineering, *55*(9), 2143–2151.

Suter, L. G., Li, S.-X., Grady, J. N., Lin, Z., Wang, Y., Bhat, K. R., ... others (2014). National patterns of risk-standardized mortality and readmission after hospitalization for acute myocardial infarction, heart failure, and pneumonia: update on publicly reported outcomes measures based on the 2013 release. *Journal of general internal medicine*, *29*(10), 1333–1340.

Tan, L. S. L., Ong, V. M. H., Nott, D. J., & Jasra, A. (2016). Variational inference for sparse spectrum gaussian process regression. *Statistics and Computing*, *26*(6), 1243–1261.

Tarassenko, L., & Fleming, S. (2010, November 25). *Method and apparatus for measuring breathing rate.* Google Patents. (US Patent App. 12/452,900)

Taylor, B. C., Wilt, T. J., & Welch, H. G. (2011). Impact of diastolic and systolic blood pressure on mortality: implications for the definition of "normal". *Journal of general internal medicine*, *26*(7), 685–690.

Thran, J. S. (2018). Documentation of vital signs during the post-operative phase.

Van Overmeire, B., Van de Broek, H., Van Laer, P., Weyler, J., & Vanhaesebrouck, P. (2001). Early versus late indomethacin treatment for patent ductus arteriosus in premature infants with respiratory distress syndrome. *The Journal of pediatrics*, *138*(2), 205–211.

Velardo, C., Shah, S. A., Gibson, O., Rutter, H., Farmer, A., & Tarassenko, L. (2014). Automatic generation of personalised alert thresholds for patients with copd. In *Signal processing conference (eusipco), 2014 proceedings of the 22nd european* (pp. 1990–1994).

Wang, J. M., Fleet, D. J., & Hertzmann, A. (2007). Multifactor gaussian process models for style-content separation. In *Proceedings of the 24th international conference on machine learning* (pp. 975–982).

Wang, W., Feng, Q., Liu, L., & Chen, W. (2008). Segmentation of brain mr images through class-adaptive gauss-markov random field model and the

em algorithm. *Journal of Image and Graphics*, *13*(3), 488–493.

Williams, C., Bonilla, E. V., & Chai, K. M. (2007). Multi-task gaussian process prediction. *Advances in neural information processing systems*, 153–160.

Yang, A., Li, C., Rana, S., Gupta, S., & Venkatesh, S. (2019). Sparse spectrum process for bayesian optimisation. *arXiv preprint arXiv:1906.08898*.

APPENDICES

APPENDIX A: REVIEW OF NORMAL AND INVERSE GAMMA RESULTS

This section of the appendix focuses on a brief review of some well known standard results on normal and inverse gamma probability models.

Lemma 1 *(a) Let y denotes a multivariate normal random variate, $y \sim \mathcal{N}_m\Big(\mu, \Sigma\Big)$. For $\nu = (\nu_1, \ldots, \nu_m)$ and a positive definite matrix $\Phi_{m\times m}$, we have*

$$\begin{aligned} E_y\left[\Big(y-\nu\Big)\Phi\Big(y-\nu\Big)\right] &= \Big(\mu-\nu\Big)^{'}\Phi_m\Big(\mu-\nu\Big) + tr\Big(\Phi\Sigma\Big) \\ E_y\left[\log\Big(g(y)\Big)\right] &= -\frac{m}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{m}{2}, \end{aligned}$$

where $g(y)$ denotes probability density function.

(b) Suppose $v \sim IG(\alpha, \beta)$. Then the following results hold.

$$\begin{aligned} E_v\left[\frac{1}{v}\right] &= \frac{\alpha}{\beta} \\ E_v\left[\log(v)\right] &= \log(\alpha) - \psi(\beta) \end{aligned}$$

where $\psi(u)$ is the digamma function.

(c)Let $Y \sim \mathcal{N}\Big(\mu, \sigma^2\Big)$. Let x_1 and x_2 be fixed. Define the following statistics.

$$\begin{aligned} x_{12}^{+} = x_1 + x_2,\ x_{12}^{-} = x_1 - x_2 \\ D(x_{12}^{+}, \sigma^2) = \exp\left\{-\frac{1}{2}\Big(x_{12}^{+}\Big)^2\sigma^2\right\} \\ D(x_{12}^{-}, \sigma^2) = \exp\left\{-\frac{1}{2}\Big(x_{12}^{-}\Big)^2\sigma^2\right\}. \end{aligned}$$

Then the following trigonometric expectations are true.

$$E\left[\cos\left(Yx_1\right)\cos\left(Yx_2\right)\right] = \frac{1}{2}\left[\cos\left(\mu x_{12}^{-}\right)D(x_{12}^{-},\sigma^2) + A^{**}\right] \tag{5.1}$$

$$A^{**} = \cos\left(\mu x_{12}^{+}\right)D(x_{12}^{+},\sigma^2)$$

$$E\left[\sin\left(Yx_1\right)\sin\left(Yx_2\right)\right] = \frac{1}{2}\left[\cos\left(\mu x_{12}^{-}\right)D(x_{12}^{-},\sigma^2) - B^{**}\right] \tag{5.2}$$

$$B^{**} = \cos\left(\mu x_{12}^{+}\right)D(x_{12}^{+},\sigma^2)$$

$$E\left[\sin\left(Yx_1\right)\cos\left(Yx_2\right)\right] = \frac{1}{2}\left[\sin\left(\mu x_{12}^{-}\right)D(x_{12}^{-},\sigma^2) + C^{**}\right] \tag{5.3}$$

$$C^{**} = \sin\left(\mu x_{12}^{+}\right)D(x_{12}^{+},\sigma^2)$$

$$E\left[\cos\left(Yx_1\right)\right] = \cos\left(\mu x_1\right)\exp\left\{-\left(\frac{1}{2}x_1^2\sigma^2\right)\right\} \tag{5.4}$$

and

$$E\left[\sin\left(Yx_1\right)\right] = \sin(\mu x_1)\exp\left\{-\left(\frac{1}{2}x_1^2\sigma^2\right)\right\} \tag{5.5}$$

The proof of the above trigonometric expectations can be found in (Mensah, Ofori, & Howard, 2021).

APPENDIX B : COMPUTATION OF VARIATIONAL OPTIMIZATION FUNCTION

We begin with the unnormalized posterior obtained in the form

$$\begin{aligned} f(\xi \mid Z_{(y)}) &\propto f(Z_{(y)} \mid \xi) g(\xi) \\ &= f(Z_{(y)} \mid \beta, \theta, c, a, \sigma_\epsilon^2) f(\theta \mid \lambda, \sigma_\theta^2) f(a \mid \sigma_\tau^2) \\ &\quad f(c \mid \sigma_\kappa^2) f(\sigma_\epsilon^2 \mid v_\epsilon) g(\sigma_\kappa^2) g(\lambda) g(\sigma_\theta^2) g(v_\epsilon) g(\beta) \end{aligned} \tag{5.6}$$

Using the computational formula obtained for the variational lower bound in (3.35) of Chapter , the following explicit forms of components of the variational lower bounds are obtained.

$$\mathcal{L}(q) = E_q\Big[\log f(Z_{(y)} \mid \xi)\Big] + E_q\Big[\log g(\xi)] - E_q\Big[\log q(\xi)\Big],$$

$$\begin{aligned} E_q\Big[\log f(Z_{(y)} \mid \xi) g(\xi)\Big] &= E_q\Big[\log f(Z_{(y)} \mid \beta, \theta, c, a, \sigma_\epsilon^2)\Big] \\ &\quad + E_q\Big[\log f(\theta \mid \lambda, \sigma_\theta^2)\Big] \\ &\quad + E_q\Big[\log f(a \mid \sigma_\tau^2)\Big] + E_q\Big[\log f(c \mid \sigma_\kappa^2)\Big] \\ &\quad + E_q\Big[\log g(\sigma_\epsilon^2)\Big] \\ &\quad + E_q\Big[\log g(\sigma_\kappa^2)\Big] + E_q\Big[\log g(\beta)\Big] \\ &\quad + E_q\Big[\log g(\lambda)\Big] + E_q\Big[\log g(\sigma_\theta^2)\Big] \end{aligned} \tag{5.7}$$

and

$$\begin{aligned} E_q\Big[\log q(\xi)] &= E_q\Big[\log q(a)\Big] + E_q\Big[\log q(\beta)\Big] + E_q\Big[\log q(\lambda)\Big] \\ &\quad + E_q\Big[\log q(c)\Big] + E_q\Big[\log q(\theta)\Big] + E_q\Big[\log q(\sigma_\tau^2)\Big] \\ &\quad + E_q\Big[\log q(\sigma_\theta^2)\Big] + E_q\Big[\log q(\sigma_\epsilon^2)\Big] + E_q\Big[\log q(\sigma_\kappa^2)\Big] \end{aligned} \tag{5.8}$$

It is important to note that sum of (5.7) and (5.8) yeilds $\mathcal{L}_q$. The corresponding expectations can be computed making using of the results in Appendix A. Let $\mathcal{L}^1(q) = E_q\Big[\log f\Big(Z_{(y)} \mid \beta, \theta, a, c, \sigma_\epsilon^2\Big)\Big]$. Then, the following results are obtainable. $N* = T(S^*) + T(W^*)$

$$
\begin{aligned}
\mathcal{L}^1(q) &= E_q\Big[\log \mathcal{N}\Big(Z_{(y)}; S_\theta c + W_\beta a, \sigma_\epsilon^2 \mathbf{I}_n\Big)\Big] \\
&= -\frac{n}{2\sigma_{\beta_0}^2}\Big[\log(2\pi) - \log(\alpha_\epsilon^q) + \psi(\gamma_\epsilon^q)\Big] \\
&\quad - \frac{a_\epsilon^q}{2b_\epsilon^q}\Big[Z_{(y)}' Z_{(y)} - 2T_{11} - 2T_{12} + 2T_{13} + N*\Big] \\
E_q\Big[\log f(\theta \mid \lambda, \sigma_\theta^2)\Big] &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\Big[\log(\gamma_\theta^q) - \psi(\alpha_\theta^q)\Big] \\
&\quad - \frac{\alpha_\theta^q}{2\gamma_\theta^q}\Big[\Big(\mu_\theta^q - V'\mu_\lambda^q\Big)^2 + \sigma_\theta^{q2} + V'\Sigma_\lambda^q V\Big] \\
E_q\Big[\log f(a \mid \sigma_\tau^2)\Big] &= m\Big[\log m - \log(2\pi)\Big] - m\Big[\log(\gamma_\tau^q) - \psi(\alpha_\tau^q)\Big] \\
&\quad - \frac{m\alpha_\tau^q}{2\gamma_\tau^q}\Big[\mu_a^{q'}\mu_a^q + \mathrm{tr}(\Sigma_a^q)\Big] \\
E_q\Big[\log f(c \mid \sigma_\kappa^2)\Big] &= m\Big[\log m - \log(2\pi)\Big] - m\Big[\log(\gamma_\kappa^q) - \psi(\alpha_\kappa^q)\Big] \\
&\quad - \frac{m\alpha_\kappa^q}{2\gamma_\kappa^q}\Big[\mu_c^{q'}\mu_c^q + \mathrm{tr}(\Sigma_c^q)\Big] \\
E_q\Big[\log g(\sigma_\epsilon^2)\Big] &= -\alpha_\epsilon \log(\gamma_\epsilon) - \log\Gamma(\alpha_\epsilon) - \Big[\log(\gamma_\epsilon^q) - \psi(\alpha_\epsilon^q)\Big] \\
&\quad - \alpha\epsilon\Big(\frac{\alpha_\epsilon^q}{\gamma_\epsilon^q}\Big) \\
E_q\Big[\log g(\sigma_\theta^2)\Big] &= -\alpha_\theta \log(\gamma_\theta) - \log\Gamma(\alpha_\theta) - \Big[\log(\gamma_\theta^q) - \psi(\alpha_\theta^q)\Big] \\
&\quad - \alpha_\theta\Big(\frac{\alpha_\theta^q}{\gamma_\theta^q}\Big) \\
E_q\Big[\log g(\sigma_\tau^2)\Big] &= -\alpha_\tau \log(\gamma_\tau) - \log\Gamma(\alpha_\tau) - \Big[\log(\gamma_\tau^q) - \psi(\alpha_\tau^q)\Big] \\
&\quad - \alpha_\tau\Big(\frac{\alpha_\tau^q}{\gamma_\tau^q}\Big) \\
E_q\Big[\log g(\sigma_\kappa^2)\Big] &= -\alpha_\kappa \log(\gamma_\kappa) - \log\Gamma(\alpha_\kappa) - \Big[\log(\gamma_\kappa^q) - \psi(\alpha_\kappa^q)\Big] \\
&\quad - \alpha_\kappa\Big(\frac{\alpha_\kappa^q}{\gamma_\kappa^q}\Big) \\
E_q\Big[\log g\Big(\beta\Big)\Big] &= -\frac{1}{2}\Big[\log(2\pi) + \log(\sigma_{\beta_0}^2)\Big] \\
&\quad - \frac{1}{2\sigma_{\beta_0}^2}\Big[\big(\mu_\beta^q - \mu_{\beta_0}\big)^2 + \sigma_\beta^{q2}\Big] \\
E_q\Big[\log g\Big(\lambda\Big)\Big] &= -\frac{1}{2}\log|\Sigma_\lambda^0| - T_\mu
\end{aligned}
$$

and

$$
\begin{aligned}
E_q\Big[\log q(a)\Big] &= -m\log\Big(2\pi\Big) - \frac{1}{2}\log|\Sigma_a^q| - m \\
E_q\Big[\log q(c)\Big] &= -m\log\Big(2\pi\Big) - \frac{1}{2}\log|\Sigma_c^q| - m \\
E_q\Big[\log q(\beta)\Big] &= -\frac{1}{2}\log\Big(\sigma_\beta^{q2}\Big) - \frac{1}{2}\log\Big(2\pi\Big) - \frac{1}{2} \\
E_q\Big[\log q(\lambda)\Big] &= -\frac{1}{2}\log|\Sigma_\lambda^q| - \frac{r}{2}\log\Big(2\pi\Big) - \frac{r}{2} \\
E_q\Big[\log q(\theta)\Big] &= -\frac{1}{2}\log\Big(\sigma_\theta^{q2}\Big) - \frac{1}{2}\log\Big(2\pi\Big) - \frac{1}{2} \\
E_q\Big[\log q\Big(\sigma_\epsilon^2\Big)\Big] &= \alpha_\epsilon^q\log\Big(\gamma_\epsilon^q\Big) - \log\Gamma\Big(\gamma_\epsilon^q\Big) \\
&\quad - \Big(\alpha_\epsilon^q + 1\Big)\Big[\log\Big(\gamma_\epsilon^q\Big) - \psi\Big(\alpha_\epsilon^q\Big)\Big] - \alpha_\epsilon^q \\
E_q\Big[\log q\Big(\sigma_\tau^2\Big)\Big] &= \alpha_\tau^q\log\Big(\gamma_\tau^q\Big) - \log\Gamma\Big(\gamma_\tau^q\Big) \\
&\quad - \Big(\alpha_\tau^q + 1\Big)\Big[\log\Big(\gamma_\tau^q\Big) - \psi\Big(\alpha_\tau^q\Big)\Big] - \alpha_\tau^q \\
E_q\Big[\log q\Big(\sigma_\theta^2\Big)\Big] &= \alpha_\theta^q\log\Big(\gamma_\theta^q\Big) - \log\Gamma\Big(\gamma_\theta^q\Big) \\
&\quad - \Big(\alpha_\theta^q + 1\Big)\Big[\log\Big(\gamma_\theta^q\Big) - \psi\Big(\alpha_\theta^q\Big)\Big] - \alpha_\theta^q \\
E_q\Big[\log q\Big(\sigma_\kappa^2\Big)\Big] &= \alpha_\kappa^q\log\Big(\gamma_\kappa^q\Big) - R_c,
\end{aligned}
$$

where

$$
\begin{aligned}
T_\mu &= \frac{1}{2}\Big[\big(\mu_\lambda^q - \mu_\lambda^0\big)'\,\Sigma_\lambda^{0\,-1}\big(\mu_\lambda^q - \mu_\lambda^0\big) + \mathrm{tr}\Big(\Sigma_\lambda^{0\,-1}\Sigma_\lambda^{0\,q}\Big)\Big] \\
T(W^*) &= \mathrm{tr}\Big(\Big(\mu_a^q\mu_a^{q'} + \Sigma_a^q\Big)W^*\Big), \quad T(Z^*) = \mathrm{tr}\Big(\Big(\mu_c^q\mu_c^{q'} + \Sigma_c^q\Big)Z^*\Big) \\
T_{11} &= Z_{(y)}'S\mu_c^q, T_{12} = Z_{(y)}'W\mu_a^q, T_{13} = \mu_c^{q'}S'W\mu_a^q \\
R_c &= \log\Gamma\Big(\gamma_\kappa^q\Big) - \Big(\alpha_\kappa^q + 1\Big)\Big[\log\Big(\gamma_\kappa^q\Big) - \psi\Big(\alpha_\kappa^q\Big)\Big] - \alpha_\kappa^q
\end{aligned}
$$

APPENDIX C: COMPUTATION OF VARIATIONAL PARAMETER-LEVEL UPDATES

This section of the Appendix considers the parameter-level updates of the variational algorithm. Based on the defined joint posterior distribution (5.6) and the assumed variational approximation, the respective parameter-wise variational posteriors can be derived as follows.

$$\begin{aligned} q(\sigma_\epsilon^2) &\propto \exp\left[\mathrm{E}_{-q(\sigma_\epsilon^2)} \log\Big(f(Z_{(y)} \mid \beta, \theta, c, a, \sigma_\epsilon^2)\Big) g(\sigma_\epsilon^2 \mid v_\epsilon)\right] \\ &\propto \exp\left[\mathrm{E}_{-q(\sigma_\epsilon^2)} \left[\log\left\{\mathcal{N}\Big(Z_{(y)}; S_\theta c + W_\beta a,\ \sigma_\epsilon^2 \mathrm{I}_n\Big)\right\} + \log \alpha^{**}\right]\right] \\ &= \mathrm{IG}\Big(\alpha^*,\ \gamma^*\Big), \end{aligned} \tag{5.9}$$

where

$$\begin{aligned} \alpha^{**} &= \left\{\mathrm{IG}\Big(\sigma_\epsilon^2;\ \alpha_\epsilon,\ \gamma_\epsilon\Big)\right\} \\ \alpha^* &= \frac{n}{2} + \alpha_\epsilon \\ \gamma^* &= \frac{1}{2}\left[Z_{(y)}' Z_{(y)} - 2Z_{(y)}' S\mu_c^q - 2Z_{(y)}' W\mu_a^q + 2(S\mu_c^q)'(W\mu_a^q) + A_1 + A_2\right] + \gamma_\epsilon \\ A_1 &= \mathrm{tr}\Big(\Big(\mu_c^q \mu_c^{q'} + \Sigma_c^q\Big) S^*\Big),\ A_2 = \mathrm{tr}\Big(\Big(\mu_a^q \mu_a^{q'} + \Sigma_a^q\Big) W^*\Big) \end{aligned}$$

Comparing the variational distribution $q(\sigma_\epsilon^2) \sim \mathrm{IG}\Big(a_\epsilon^q,\ b_\epsilon^q\Big)$ with (5.9), the variational hyperparameter updates are as follows.

$$a_\epsilon^q \leftarrow \alpha^*, \quad b_\epsilon^q \leftarrow \gamma^*$$

$$\begin{aligned} q\Big(\sigma_\kappa^2\Big) &\propto \exp\left[\mathrm{E}_{-q(\sigma_\kappa^2)} \log f(c \mid \sigma_\kappa^2) g(\sigma_\kappa^2)\right] \\ &\propto \exp\left[\mathrm{E}_{-q(\sigma_\kappa^2)} \left[\log\left\{\mathcal{N}\Big(c; 0,\ \frac{\sigma_\kappa^2}{m}\mathrm{I}_{2m}\Big)\right\} + \log\left\{\mathrm{IG}\Big(\sigma_\kappa^2;\ \alpha_\kappa, \gamma_\kappa\Big)\right\}\right]\right] \\ &\propto \exp\left[\log\Big(\sigma_\kappa^2\Big)^{-(m+\alpha_\kappa)-1} - \frac{1}{\sigma_\kappa^2}\left\{\frac{m}{2}\left[\mathrm{tr}\Big(\mu_c^q \mu_c^{q'} + \Sigma_c^q\Big)\right] + \gamma_\kappa\right\}\right] \\ &\sim \mathrm{IG}\Big(m + \alpha_\kappa,\ \frac{m}{2}\left[\mathrm{tr}\Big(\mu_c^q \mu_c^{q'} + \Sigma_c^q\Big)\right] + \gamma_\kappa\Big) \end{aligned} \tag{5.10}$$

Upon comparing $q(\sigma_\kappa^2) \sim \text{IG}\left(\alpha_\kappa^q,\ \gamma_\kappa^q\right)$ with (5.10), we obtain the updates of the form

$$\alpha_\kappa^q \leftarrow m + \alpha_\kappa, \quad b_\kappa^q \leftarrow \frac{m}{2}\left[\text{tr}\left(\mu_a^q \mu_a^{q'} + \Sigma_a^q\right)\right] + \gamma_\kappa$$

$$\begin{aligned} q\left(\sigma_\tau^2\right) &\propto \exp\left[\text{E}_{-q(\sigma_\tau^2)} \log f(a \mid \sigma_\tau^2) g(\sigma_\tau^2)\right] \\ &\propto \exp\left[\text{E}_{-q(\sigma_\tau^2)}\left[\log\left\{\mathcal{N}\left(a; 0,\ \frac{\sigma_\tau^2}{m}\text{I}_{2m}\right)\right\} + \log\left\{\text{IG}\left(\sigma_\tau^2;\ \alpha_\tau, \gamma_\tau\right)\right\}\right]\right] \\ &\propto \exp\left[\log\left(\sigma_\tau^2\right)^{-(m+\alpha_\tau)-1} - \frac{1}{\sigma_\tau^2}\left\{\frac{m}{2}\left[\text{tr}\left(\mu_a^q \mu_a^{q'} + \Sigma_a^q\right)\right] + \gamma_\tau\right\}\right] \\ &\sim \text{IG}\left(m + \alpha_\tau,\ \frac{m}{2}\left[\text{tr}\left(\mu_a^q \mu_a^{q'} + \Sigma_a^q\right)\right] + \gamma_\tau\right) \end{aligned} \tag{5.11}$$

Upon comparing $q(\sigma_\tau^2) \sim \text{IG}\left(\alpha_\tau^q,\ \gamma_\tau^q\right)$ with (5.11), we obtain the updates of the form

$$\alpha_\tau^q \leftarrow m + \alpha_\tau, \quad b_\tau^q \leftarrow \frac{m}{2}\left[\text{tr}\left(\mu_a^q \mu_a^{q'} + \Sigma_a^q\right)\right] + \gamma_\tau$$

$$\begin{aligned} q(\sigma_\theta^2) &\propto \exp\left[\text{E}_{-q(\sigma_\theta^2)} \log p\left(\theta \mid \lambda, \sigma_\theta^2\right) p\left(\sigma_\theta^2\right)\right] \\ &\propto \exp\left[\text{E}_{-q(\sigma_\theta^2)}\left[\log\left\{\mathcal{N}\left(\theta;\ V'\lambda, \sigma_\theta^2\right)\right\} + \log\left\{\text{IG}\left(\sigma_\theta^2;\ \alpha_\theta,\ \gamma_\theta\right)\right\}\right]\right] \\ &\propto \exp\left[\log\left(\sigma_\theta^2\right)^{-\left(\frac{1}{2}+\alpha_\theta\right)-1} - \frac{1}{\sigma_\theta^2}\left\{\frac{1}{2}\left[FF^*\right] + \gamma_\theta\right\}\right] \\ &\sim \text{IG}\left(\frac{1}{2} + \alpha_\theta,\ \frac{1}{2}\left(\left(\mu_\theta^q - V'\mu_\lambda^q\right)^2 + \sigma_\theta^{q2} + V'\Sigma_\lambda^q v\right) + \gamma_\theta\right) \end{aligned} \tag{5.12}$$

$$FF^* == \left(\mu_\theta^q - v'\mu_\lambda^q\right)^2 + \sigma_\theta^{q2} + V'\Sigma_\lambda^q V$$

Matching the hyperparmeters of $q(\sigma_\theta^2) \sim \text{IG}\left(\alpha_\theta^q,\ \gamma_\theta^q\right)$ with (5.12), it is clear that

$$\alpha_\theta^q \leftarrow \frac{1}{2} + \alpha_\theta, \quad \gamma_\theta^q \leftarrow \frac{1}{2}\left(\left(\mu_\theta^q - V'\mu_\lambda^q\right)^2 + \sigma_\theta^{q2} + V'\Sigma_\lambda^q V\right) + \gamma_\theta\right)$$

APPENDIX D: VARIATIONAL UPDATES FOR GAUSSIAN PARAMETERS

Recall the assumed variational models for a, c and λ $a \sim \mathcal{N}\left(\mu_a^q, \Sigma_a^q\right)$, $c \sim \mathcal{N}\left(\mu_c^q, \Sigma_c^q\right)$, and $\lambda \sim \mathcal{N}\left(\mu_\lambda^q, \Sigma_\lambda^q\right)$. The corresponding updating equations can be are obtained follows. First, we considering $q(\lambda)$, we can write

$$\begin{aligned} q(\lambda) &\propto \exp\left[\mathrm{E}_{-q(\lambda)} \log f\left(\theta \mid \lambda, \sigma_\theta^2\right) g\left(\lambda\right)\right] \\ &\propto \exp\left[\mathrm{E}_{-q(\lambda)} \log\left\{\mathcal{N}\left(\lambda; V'\lambda, \sigma_\theta^2\right)\mathcal{N}\left(\lambda;\, \mu_\lambda^0, \Sigma_\lambda^0\right)\right\}\right] \end{aligned}$$

Note that

$$\begin{aligned} &\mathcal{N}\left(\lambda; V'\lambda, \sigma_\theta^2\right)\mathcal{N}\left(\lambda;\, \mu_\lambda^0, \Sigma_\lambda^0\right) \\ &\propto \exp\left\{-\frac{1}{2\sigma_\theta^2\Sigma_\lambda^0}\left[\Sigma_\lambda^0\left(\theta - V'\lambda\right)^2 + H_{(\theta,\lambda)}\right]\right\} \end{aligned} \tag{5.13}$$

where $H_{(\theta,\lambda)} = \sigma_\theta^2\left(\lambda - \mu^0\right)'\left(\lambda - \mu^0\right)$. We complete squares in λ of the exponent to obtain

$$\mathcal{N}\left(\lambda; V'\lambda, \sigma_\theta^2\right)\mathcal{N}\left(\lambda;\, \mu_\lambda^0, \Sigma_\lambda^0\right) \propto \exp\left\{-\frac{1}{2\Sigma_\lambda^*}\left[\left(\lambda - \mu_\lambda^*\right)'\left(\lambda - \mu_\lambda^*\right)\right]\right\},$$

where

$$\Sigma_\lambda^* = \frac{\sigma_\theta^2\Sigma_\lambda^0}{\Sigma^0 VV' + \sigma_\theta^2\mathbf{I}}, \quad \mu_\beta = \frac{\theta\Sigma_\lambda^0 V + \sigma_\theta^2\mu_\lambda^0}{\Sigma^0 VV' + \sigma_\theta^2\mathbf{I}}.$$

Taking natural logs followed by expectations with respect to the variational distributions of all parameters except λ, it can be observed that $q(\lambda)$ has a standard distributional form

$$q(\lambda) \sim \mathcal{N}\left(\mu_\lambda^{q*},\ \Sigma_\lambda^{q*}\right), \tag{5.14}$$

where

$$\Sigma_\lambda^{q*} = \left[\frac{\alpha_\theta^q}{\gamma_\theta^q} VV' + \Sigma_\lambda^{0\,-1}\right]^{-1}, \quad \mu_\lambda^{q*} = \Sigma_\lambda^{q*}\left[\frac{\alpha_\theta^q}{\gamma_\theta^q}\mu_\theta^q V + \Sigma_\lambda^{0\,-1}\mu_\theta^0\right]$$

The updating rules are obtained by comparing $q\Big(\lambda\Big) \sim \mathcal{N}\Big(\mu_\beta^q, \Sigma_\beta^q\Big)$ and with (5.14)

$$\Sigma_\lambda^q \leftarrow \Sigma_\lambda^{q*}; \quad \mu_\lambda^q \leftarrow \mu_\lambda^{q*}$$

Next we consider the variational updating rules for $q(a)$. Using the general updating rule, we have

$$\propto \exp\left[\mathrm{E}_{-q(a)} \log\left\{\mathcal{N}\Big(Z_{(y)}; S_\theta c + W_\beta a, \sigma_\epsilon^2 \mathrm{I}_n\Big)\mathcal{N}\Big(a;\, 0, \frac{\sigma_\tau^2}{m}\mathrm{I}_{2m}\Big)\right\}\right]$$

We first concentrate on the product $\mathcal{N}\Big(Z_{(y)}; S_\theta c + W_\beta a, \sigma_\epsilon^2 \mathrm{I}_n\Big)\mathcal{N}\Big(a;\, 0, \frac{\sigma_\tau^2}{m}\mathrm{I}_{2m}\Big)$. Following similar steps to those utilized in the derivations for $q(\lambda)$, we can write

$$\mathcal{N}\Big(Z_{(y)}; S_\theta c + W_\beta a, \sigma_\epsilon^2 \mathrm{I}_n\Big)\mathcal{N}\Big(a;\, 0, \frac{\sigma_\tau^2}{m}\mathrm{I}_{2m}\Big) \propto \exp\left\{-\frac{1}{2\Sigma_a^*}\left[a'a - 2aD_*\right]\right\} \tag{5.15}$$

where

$$\Sigma_a^* = \frac{\sigma_\epsilon^2\sigma_\tau^2}{\sigma_\tau^2 W_\beta' W_\beta + \sigma_\epsilon^2 m \mathrm{I}_{2m}}, \quad D_* = \frac{Z_{(y)}' W_\beta - (S_\theta c)' W_\beta}{\sigma_\tau^2 W_\beta' W_\beta + \sigma_\epsilon^2 m \mathrm{I}_{2m}}$$

We complete squares in a of the exponent of (5.15) to obtain

$$\begin{aligned}&\mathcal{N}\Big(Z_{(y)}; S_\theta c + W_\beta a, \sigma_\epsilon^2 \mathrm{I}_n\Big)\mathcal{N}\Big(a;\, 0, \frac{\sigma_\tau^2}{m}\mathrm{I}_{2m}\Big) \\ &\quad \propto \exp\left\{-\frac{1}{2}\left[(a-\mu_a)'\Sigma_a^{-1}(a-\mu_a)\right]\right\},\end{aligned} \tag{5.16}$$

where

$$\Sigma_a = \Sigma_a^*, \ \mu_a = D_*$$

Taking natural logs followed by expectations with respect to the variational distributions of all parameters except a, $q(a)$ has a multivariate normal form.

$$q(a) \sim \mathcal{N}\Big(\mu_a^{q*},\ \Sigma_a^{q*}\Big), \tag{5.17}$$

where

$$\Sigma_a^{q*} = \left[\frac{\alpha_\epsilon^q}{\gamma_\epsilon^q} W^* + m\frac{\alpha_\tau^q}{\gamma_\tau^q}\mathbf{I}_{2m}\right]^{-1}, \quad \mu_a^{q*} = \frac{\alpha_\epsilon^q}{\gamma_\epsilon^q}\Sigma_a^{q*}\left[\Big(Z_{(y)} - S\mu_c^q\Big)' W\right]$$

Comparing $q(c) \sim \mathcal{N}\Big(\mu_c^q, \Sigma_c^q\Big)$ and (5.20), the following updates are obtained

$$\Sigma_a^q \leftarrow \Sigma_a^{q*} \quad \mu_a^q \leftarrow \mu_a^{q*}$$

Finally we consider the variational updating rules for $q(c)$ following similar steps as that for $q(a)$.

$$\propto \exp\left[\mathrm{E}_{-q(c)} \log\left\{\mathcal{N}\Big(Z_{(y)}; S_\theta c + W_\beta a, \sigma_\epsilon^2\mathbf{I}_n\Big)\mathcal{N}\Big(c;\ 0, \frac{\sigma_\kappa^2}{m}\mathbf{I}_{2m}\Big)\right\}\right]$$

We first concentrate on the product $\mathcal{N}\Big(Z_{(y)}; S_\theta c + W_\beta a, \sigma_\epsilon^2\mathbf{I}_n\Big)\mathcal{N}\Big(c;\ 0, \frac{\sigma_\kappa^2}{m}\mathbf{I}_{2m}\Big)$. We have

$$\begin{aligned}\mathcal{N}\Big(Z_{(y)}; S_\theta c + W_\beta a, \sigma_\epsilon^2\mathbf{I}_n\Big)\mathcal{N}\Big(c;\ 0, \frac{\sigma_\kappa^2}{m}\mathbf{I}_{2m}\Big) \\ \propto \exp\left\{-\frac{1}{2\Sigma_c^*}\left[c'c - 2cK_*\right]\right\}\end{aligned} \tag{5.18}$$

where

$$\Sigma_c^* = \frac{\sigma_\epsilon^2\sigma_\kappa^2}{\sigma_\kappa^2 S_\theta' S_\theta + \sigma_\epsilon^2 m\mathbf{I}_{2m}}, \quad D_* = \frac{\sigma_\tau^2\Big(Z_{(y)}' + W_\beta a\Big)' S_\theta}{\sigma_\kappa^2 S_\theta' S_\theta + \sigma_\epsilon^2 m\mathbf{I}_{2m}}$$

Completing squares in c of the exponent of (5.18), we obtain

$$\begin{aligned}\mathcal{N}\Big(Z_{(y)}; S_\theta c + W_\beta a, \sigma_\epsilon^2 \mathrm{I}_n\Big)\mathcal{N}\Big(c;\, 0, \frac{\sigma_\kappa^2}{m}\mathrm{I}_{2m}\Big) \\ \propto \exp\left\{-\frac{1}{2}\left[(c-\mu_c)'\Sigma_c^{-1}(c-\mu_c)\right]\right\},\end{aligned} \tag{5.19}$$

where

$$\Sigma_c = \Sigma_c^*,\ \mu_c = K_*$$

Taking natural logs followed by expectations with respect to the variational distributions of all parameters except c, $q(c)$ has a multivariate normal form.

$$q(c) \sim \mathcal{N}\Big(\mu_c^{q*},\ \Sigma_c^{q*}\Big), \tag{5.20}$$

where

$$\Sigma_c^{q*} = \left[\frac{\alpha_\epsilon^q}{\gamma_\epsilon^q} S^* + m\frac{\alpha_\kappa^q}{\gamma_\kappa^q}\mathrm{I}_{2m}\right]^{-1}, \quad \mu_a^{q*} = \frac{\alpha_\epsilon^q}{\gamma_\epsilon^q}\Sigma_c^{q*}\left[\Big(Z_{(y)} - W\mu_a^q\Big)' S\right]$$

Comparing $q(a) \sim \mathcal{N}\Big(\mu_a^q, \Sigma_a^q\Big)$ and (5.20), the following updates are obtained

$$\Sigma_c^q \leftarrow \Sigma_c^{q*} \quad \mu_c^q \leftarrow \mu_c^{q*}$$

APPENDIX E: VARIATIONAL UPDATES FOR β AND θ

The variational updating rules for finding the optimal parameters for $q(\beta)$ and $q(\theta)$ can be derived as follows.

$$\sigma_{\beta}^{q\,2} \leftarrow -\frac{1}{2}\left[\frac{\partial}{\partial \sigma_{\beta}^{q\,2}} f_1(\mu_{\theta}^{q}, \mu_{\beta}^{q}, \sigma_{\theta}^{q\,2}, \sigma_{\beta}^{q\,2}) + \frac{\partial}{\partial \sigma_{\beta}^{q\,2}} f_2(\mu_{\beta}^{q}, \sigma_{\beta}^{q\,2})\right]^{-1} \tag{5.21}$$

$$\mu_{\beta}^{q} \leftarrow \mu_{\beta}^{q} + \sigma_{\beta}^{q\,2}\left[\frac{\partial}{\partial \mu_{\beta}^{q}} f_1(\mu_{\theta}^{q}, \mu_{\beta}^{q}, \sigma_{\theta}^{q\,2}, \sigma_{\beta}^{q\,2},) + \frac{\partial}{\partial \mu_{\beta}^{q}} f_2(\mu_{\beta}^{q}, \sigma_{\beta}^{q\,2})\right] \tag{5.22}$$

$$\sigma_{\theta}^{q\,2} \leftarrow -\frac{1}{2}\left[\frac{\partial}{\partial \sigma_{\theta}^{q\,2}} f_1(\mu_{\theta}^{q}, \mu_{\beta}^{q}, \sigma_{\theta}^{q\,2}, \sigma_{\beta}^{q\,2}) + \frac{\partial}{\partial \sigma_{\theta}^{q\,2}} f_3(\mu_{\theta}^{q}, \sigma_{\theta}^{q\,2})\right]^{-1} \tag{5.23}$$

$$\mu_{\theta}^{q} \leftarrow \mu_{\theta}^{q} + \sigma_{\theta}^{q\,2}\left[\frac{\partial}{\partial \mu_{\theta}^{q}} f_1(\mu_{\theta}^{q}, \mu_{\beta}^{q}, \sigma_{\theta}^{q\,2}, \sigma_{\beta}^{q\,2},) + \frac{\partial}{\partial \mu_{\theta}^{q}} f_3(\mu_{\theta}^{q}, \sigma_{\theta}^{q\,2})\right], \tag{5.24}$$

where

$$\begin{aligned} f_1(\mu_{\theta}^{q}, \mu_{\beta}^{q}, \sigma_{\theta}^{q\,2}, \sigma_{\beta}^{q\,2}) &= -\frac{n}{2\sigma_{\beta_0}^{2}}\Big[\log(2\pi) - \log(\alpha_{\epsilon}^{q}) + \psi(\gamma_{\epsilon}^{q})\Big] \\ &\quad - \frac{\alpha_{\epsilon}^{q}}{2\gamma_{\epsilon}^{q}}\Big[Z_{(y)}^{'} Z_{(y)} - 2T_{11} - EE^{*}\Big] \\ f_2(\mu_{\beta}^{q}, \sigma_{\beta}^{q\,2}) &= -\frac{1}{2}\Big[\log(2\pi) + \log(\sigma_{\beta_0}^{2})\Big] - \frac{1}{2\sigma_{\beta_0}^{2}}[FF^{*}] \qquad (5.25) \\ f_3(\mu_{\theta}^{q}, \sigma_{\theta}^{q\,2}) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\Big[\log(\gamma_{\theta}^{q}) - \psi(\alpha_{\theta}^{q})\Big] \qquad (5.26) \\ &\quad - \frac{\alpha_{\theta}^{q}}{2\gamma_{\theta}^{q}}\left[\left(\mu_{\theta}^{q} - V^{'}\mu_{\lambda}^{q}\right)^{2} + \sigma_{\theta}^{q\,2} + V^{'}\Sigma_{\lambda}^{q}V\right] \qquad (5.27) \end{aligned}$$

$EE^{*} = 2T_{12} + 2T_{13} + T(S^{*}) + T(W^{*})$, *and* $FF^{*} = \left(\mu_{\beta}^{q} - \mu_{\beta_0}\right)^{2} + \sigma_{\beta}^{q\,2}$. The partial derivatives involved the updating equations (5.21), (5.22),(5.23) and (5.24) are computed as followed.

$$\frac{\partial f_1(\mu_\theta^q, \mu_\beta^q, \sigma_\theta^{q2}, \sigma_\beta^{q2})}{\partial \sigma_\theta^{q2}} = -\frac{\alpha_\epsilon^q}{2\gamma_\epsilon^q}\left[2\Big(W\mu_a^q - Z_{(y)}\Big)^{'} D\mu_c^q +\right] \tag{5.28}$$

$$\frac{\partial f_1(\mu_\theta^q, \mu_\beta^q, \sigma_\theta^{q2}, \sigma_\beta^{q2})}{\partial \mu_\theta^q} = \frac{\alpha_\epsilon^q}{\gamma_\epsilon^q}\left[2\Big(Z_{(y)} - W\mu_a^q\Big)^{'} A\mu_c^q - \frac{1}{2}CC^*\right] \tag{5.29}$$

$$CC^* = \text{tr}\Big(\Big(\mu_c^q u_c^{q'} + \Sigma_c^q\Big)Z^*\Big), and\ DD^* = \text{tr}\Big(\Big(\mu_c^q u_c^{q'} + \Sigma_c^q\Big)Q\Big)$$

$$\frac{\partial f_1(\mu_\theta^q, \mu_\beta^q, \sigma_\theta^{q2}, \sigma_\beta^{q2})}{\partial \sigma_\beta^{q2}} = -\frac{\alpha_\epsilon^q}{2\gamma_\epsilon^q}\left[2\Big(S\mu_c^q - Z_{(y)}\Big)M\mu_a^q + BB^*\right] \tag{5.30}$$

$$\frac{\partial f_1(\mu_\theta^q, \mu_\beta^q, \sigma_\theta^{q2}, \sigma_\beta^{q2})}{\partial \mu_\beta^q} = \frac{\alpha_\epsilon^q}{\gamma_\epsilon^q}\left[2\Big(Z_{(y)-S\mu_c^q}\Big)B\mu_a^q - \frac{1}{2}AA^*\right] \tag{5.31}$$

$$AA^* = \text{tr}\Big(\Big(\mu_a^q\mu_a^{q'} + \Sigma_a^q\Big)N\Big),\ and\ BB^* = \text{tr}\Big(\Big(\mu_a^q\mu_a^{q'} + \Sigma_a^q\Big)U\Big)$$

$$\frac{\partial f_2(\mu_\beta^q, \sigma_\beta^{q2})}{\partial \sigma_\beta^{q2}} = -\frac{1}{2\sigma_{\beta_0}^2} \tag{5.32}$$

$$\frac{\partial f_2(\mu_\beta^q, \sigma_\beta^{q2})}{\partial \mu_\beta^q} = -\frac{\Big(\mu_\beta^q - \mu_{\beta_0}\Big)}{\sigma_{\beta_0}^2} \tag{5.33}$$

$$\frac{\partial f_3(\mu_\theta^q, \sigma_\theta^{q2})}{\partial \sigma_\theta^{q2}} = -\frac{\alpha_\theta^q}{\gamma_\theta^q} \tag{5.34}$$

$$\frac{\partial f_3(\mu_\theta^q, \mu_\theta^{q2})}{\partial \sigma_\theta^{q2}} = -\frac{\alpha_\theta^q\Big(\mu_\theta^q - V^{'}\mu_\lambda^q\Big)}{\gamma_\theta^q} \tag{5.35}$$

The computational expressions for the spectral matrices associated with the derivatives (5.28), (5.30), (5.32) and (5.34), namely, $S, W, S^*, W^*, D, A, M, B, U, N, Q$, and Z^* can be obtained following similar argument used by Mensah, Ofori, and Howard (2021).

APPENDIX F: MCMC POSTERIOR FULL CONDITIONALS

In this section of the appendix, the full conditionals required for the development of approximate MCMC algorithms for posterior inference are considered. In particular, the derivation here begins with the definition or expression for the joint posterior distribution involved in the model considered. The various components of the conditionals distributions needed are extracted from the joint according to the dependencies a given parameter has with the remaining accordingly. The joint distribution of all parameters an variables can be written in the form

$$\begin{aligned} p(\xi \mid Z_{(y)}) &= p(Z_{(y)} \mid \xi)p(\xi) \\ &= p\Big(Z_{(y)} \mid \beta, \theta, a, c, \sigma_\epsilon^2\Big)p\Big(\theta \mid \lambda, \sigma_\theta^2\Big)p\Big(a \mid \sigma_\tau^2\Big)p\Big(c \mid \sigma_\kappa^2\Big) \\ &\times p\Big(\sigma_\epsilon^2\Big)p\Big(\beta\Big)p\Big(\lambda\Big)p\Big(\sigma_\theta^2\Big)p\Big(\sigma_\tau^2\Big)p\Big(\sigma_\kappa^2\Big). \end{aligned}$$

APPENDIX G: POSTERIOR FULL CONDITIONALS FOR INVERSE GAMMA PARAMETERS

The variance parameters involved the feature based GPR are $\sigma_\epsilon^2, \sigma_\tau^2, \sigma_\kappa^2$ and σ_θ^2. Their corresponding model can be expressed as follows.

$$\sigma_\epsilon^2 \sim \text{IG}\Big(\alpha_\epsilon, \gamma_\epsilon\Big),\ \sigma_\theta^2 \sim \text{IG}\Big(\alpha_\theta, \gamma_\theta\Big),\ \sigma_\tau^2 \sim \text{IG}\Big(\alpha_\tau, \gamma_\tau\Big),\ \sigma_\kappa^2 \sim \text{IG}\Big(\alpha_\kappa, \gamma_\kappa\Big).$$

The full posterior conditional distributions associated with the above parameters are derived as follows.

$$\begin{aligned}
p\Big(\sigma_\epsilon^2|\text{rest}\Big) &= f\Big(Z_{(y)} \mid \beta, \theta, a, c, \sigma_\epsilon^2\Big) f\Big(\sigma_\epsilon^2 \mid \alpha_\epsilon, \gamma_\epsilon\Big) \\
&= \mathcal{N}\Big(Z_{(y)};\ S_\theta c + W_\beta a, \sigma_\epsilon^2 \text{I}_n\Big) \text{IG}\Big(\sigma_\epsilon^2; \alpha_\epsilon, \gamma_\epsilon\Big) \\
&\propto (\sigma_\epsilon^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left[\Big(Z_{(y)} - S_\theta c - W_\beta a\Big)' B^*\right]\right\} \\
&* (\sigma_\epsilon^2)^{-\alpha_\epsilon - 1} \exp\left\{-\frac{\gamma_\epsilon}{\sigma_\epsilon^2}\right\} \\
&= (\sigma_\epsilon^2)^{-\Big(\frac{n}{2} + \alpha_\epsilon\Big) - 1} g_1,\ and\ B^* = \Big(Z_{(y)} - S_\theta c - W_\beta a\Big)
\end{aligned}$$

where $g_1 = \exp\left\{-\frac{1}{\sigma_\epsilon^2}\left[\frac{1}{2}\Big(Z_{(y)} - S_\theta c - W_\beta a\Big)'\Big(Z_{(y)} - S_\theta c - W_\beta a\Big) + \frac{\alpha_\epsilon}{\gamma_\epsilon}\right]\right\}$

Using the standard properties of inverse gamma distribution, $f(\sigma_\epsilon^2|\text{rest})$ can be identified as an inverse gamma,

$$f\Big(\sigma_\epsilon^2|\text{rest}\Big) = \text{IG}\Big(\frac{n}{2} + \alpha_\epsilon, b_\epsilon\Big),$$

where $b_\epsilon = \left[\frac{1}{2}\Big(Z_{(y)} - S_\theta c - W_\beta a\Big)'\Big(Z_{(y)} - S_\theta c - W_\beta a\Big) + \frac{\alpha_\epsilon}{\gamma_\epsilon}\right]$

$$
\begin{aligned}
f\left(\sigma_{\kappa}^{2} \mid \text{rest}\right) &\propto f\left(c|\sigma_{\kappa}^{2}\right) g\left(\sigma_{\kappa}^{2}\right) \\
&= \mathcal{N}\left(c; 0, \frac{\sigma_{\kappa}^{2}}{m}\mathrm{I}_{2m}\right) \mathrm{IG}\left(\alpha_{\kappa}, \gamma_{\kappa}\right) \\
&\propto \left(\sigma_{\kappa}^{2}\right)^{-\left(m+\alpha_{\kappa}\right)-1} \exp\left\{-\frac{1}{\sigma_{\kappa}^{2}}\left(\frac{mc^{'}c}{2}+\gamma_{\kappa}\right)\right\} \\
&= \mathrm{IG}\left(m+\alpha_{\kappa}, \frac{mc^{'}c}{2}+\gamma_{\kappa}\right)
\end{aligned}
$$

$$
\begin{aligned}
f\left(\sigma_{\tau}^{2} \mid \text{rest}\right) &\propto f\left(a|\sigma_{\tau}^{2}\right) g\left(\sigma_{\tau}^{2}\right) \\
&= \mathcal{N}\left(a; 0, \frac{\sigma_{\tau}^{2}}{m}\mathrm{I}_{2m}\right) \mathrm{IG}\left(\alpha_{\tau}, \gamma_{\tau}\right) \\
&\propto \left(\sigma_{\tau}^{2}\right)^{-\left(m+\alpha_{\tau}\right)-1} \exp\left\{-\frac{1}{\sigma_{\tau}^{2}}\left(\frac{ma^{'}a}{2}+\gamma_{\tau}\right)\right\} \\
&= \mathrm{IG}\left(m+\alpha_{\tau}, \frac{ma^{'}a}{2}+\gamma_{\tau}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
f\left(\sigma_{\theta}^{2}|\text{rest}\right) &= f\left(\theta|\lambda, \sigma_{\theta}^{2}\right) g\left(\sigma_{\theta}^{2}\right) \\
&= \mathcal{N}\left(\theta; V^{'}\lambda,\ \sigma_{\theta}^{2}\right) \mathrm{IG}\left(\sigma_{\theta}^{2}; \alpha_{\theta}, \gamma_{\theta}\right) \\
&= \left(\sigma_{\lambda}^{2}\right)^{-\left(\frac{1}{2}+\alpha_{\lambda}\right)-1} \exp\left\{-\frac{1}{\sigma_{\lambda}^{2}}\left(\frac{1}{2}\left(\lambda - V^{'}\beta\right)^{2}+b_{\lambda}^{0}\right)\right\} \\
&= \mathrm{IG}\left(\left(\frac{2\alpha_{\theta}+1}{2}\right), \frac{1}{2}\left(\theta - V^{'}\lambda\right)^{2}+\gamma_{\theta}\right)
\end{aligned}
$$

APPENDIX H: POSTERIOR CONDITIONALS FOR PARAMETERS WITH NORMAL MODELS

The full posterior conditionals for parameters modeled with normal probability models are considered in this section of the appendix. Recall the normal parameters among the set of parameters.

$$a \sim \mathcal{N}\left(0, \frac{\sigma_\tau^2}{m}\mathrm{I}_{2m}\right), c \sim \mathcal{N}\left(0, \frac{\sigma_\kappa^2}{m}\mathrm{I}_{2m}\right)$$
$$\theta \sim \mathcal{N}\left(V'\lambda, \sigma_\theta^2\right), \lambda \sim \mathcal{N}\left(\mu_\lambda^0, \Sigma_\lambda^0\right), \beta \sim \mathcal{N}\left(\mu_\beta^0, \sigma_{\beta_0}^2\right)$$

$$\begin{aligned} f\left(c|\text{rest}\right) &\propto f\left(Z_{(y)} \mid \beta, \theta, a, c, \sigma_\epsilon^2\right) f\left(c \mid \sigma_\kappa^2\right) \\ &= \mathcal{N}\left(Z_{(y)}; S_\theta c + W_\beta a,\ \sigma_\epsilon^2 \mathrm{I}_n\right) \mathcal{N}\left(c; 0, \frac{\sigma_\kappa^2}{m}\mathrm{I}_{2m}\right) \\ &\propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(Z_{(y)} - S_\theta c - W_\beta c\right)'\left(Z_{(y)} - S_\theta c - W_\beta a\right)\right\} g_2 \\ &\propto \exp\left\{-\frac{\left(\sigma_\kappa^2 S_\theta' S_\theta + \sigma_\epsilon^2 m \mathrm{I}_{2m}\right)}{2\sigma_\kappa^2\sigma_\epsilon^2}\left[c'c - 2\ \ A^*\right]\right\} \\ &where\ \ A^* = \frac{\sigma_\kappa^2\left(Z'_{(y)} - W_\beta a\right)' S_\theta}{\left(\sigma_\kappa^2 S_\theta' S_\theta + \sigma_\epsilon^2 m \mathrm{I}_{2m}\right)} \end{aligned}$$

where $g_2 = \exp\left\{-\frac{mc'c}{2\sigma_\kappa^2}\right\}$. Completing squares in c, yields

$$f\left(c|\text{rest}\right) \propto \exp\left\{-\frac{1}{2\Sigma_c}\left[(c-\mu_c^*)'(c-\mu_c^*)\right]\right\}.$$

$f\left(c|\text{rest}\right)$ can easily be recognized as multivariate normal, $\mathcal{N}_{2m}\left(\mu_c^*, \Sigma_c^*\right)$ with the following specifications for the associated parameters.

$$\Sigma_c^* = \left[\frac{1}{\sigma_\epsilon^2}S_\theta^* + \frac{1}{\sigma_\kappa^2}m\mathrm{I}_{2m}\right]^{-1}, \quad \mu_c^* = \Sigma_c^*\left[\frac{1}{\sigma_\epsilon^2}\left(Z_{(y)} - W_\beta a\right)' S_\theta\right]$$

$$
\begin{aligned}
f\Big(a|\text{rest}\Big) &\propto f\Big(Z_{(y)} \mid \beta, \theta, a, c, \sigma_\epsilon^2\Big) f\Big(a \mid \sigma_\tau^2\Big) \\
&= \mathcal{N}\Big(Z_{(y)}; S_\theta c + W_\beta a,\ \sigma_\epsilon^2 \mathrm{I}_n\Big) \mathcal{N}\Big(a; 0, \frac{\sigma_\tau^2}{m}\mathrm{I}_{2m}\Big) \\
&\propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\Big(Z_{(y)} - S_\theta c - W_\beta c\Big)^{'}\Big(Z_{(y)} - S_\theta c - W_\beta a\Big)\right\} g_3 \\
&\propto \exp\left\{-\frac{\Big(\sigma_\tau^2 W_\beta^{'} S_\beta + \sigma_\epsilon^2 m \mathrm{I}_{2m}\Big)}{2\sigma_\tau^2\sigma_\epsilon^2}\Big[c^{'}c - 2\ \ C^{**}\Big]\right\} \\
& where\ \ C^{**} = \frac{\sigma_\tau^2\Big(Z_{(y)}^{'} - S_\theta c\Big)^{'} W_\beta}{\Big(\sigma_\tau^2 W_\beta^{'} W_\beta + \sigma_\epsilon^2 m \mathrm{I}_{2m}\Big)}
\end{aligned}
$$

where $g_3 = \exp\left\{-\frac{ma^{'}a}{2\sigma_\tau^2}\right\}$. Completing squares in a, we obtain

$$
f\Big(a|\text{rest}\Big) \propto \exp\left\{-\frac{1}{2\Sigma_a}\Big[(a-\mu_a^*)^{'}(a-\mu_a^*)\Big]\right\}.
$$

We can identify $f\Big(a|\text{rest}\Big)$ as multivariate normal, $\mathcal{N}_{2m}\Big(\mu_a^*, \Sigma_a^*\Big)$ with the following expressions for parameters,

$$
\Sigma_a^* = \left[\frac{1}{\sigma_\epsilon^2} W_\beta^* + \frac{1}{\sigma_\tau^2} m \mathrm{I}_{2m}\right]^{-1}, \quad \mu_a^* = \Sigma_a^*\left[\frac{1}{\sigma_\epsilon^2}\Big(Z_{(y)} - S_\theta c\Big)^{'} W_\beta\right]
$$

$$
\begin{aligned}
f\Big(\lambda|\text{rest}\Big) &\propto f\Big(\theta \mid \lambda, \sigma_\theta^2\Big) g\Big(\lambda\Big) \\
&\mathcal{N}\Big(\theta, ; V^{'}\lambda, \sigma_\theta^2 \mathrm{I}_n\Big) \mathcal{N}\Big(\lambda; \mu_\lambda^0, \Sigma^0{}_\lambda\Big) \\
&\propto \exp\left\{-\frac{\Big(\theta V^{'}\Sigma_\lambda^0 + \sigma_\theta^2 \mu_\lambda^0\Big)}{2\sigma_\theta^2\Sigma_\lambda^0}\Big[\lambda^{'}\lambda - 2\lambda B_*\Big]\right\},
\end{aligned}
$$

where $B_* = \frac{\theta V^{'}\Sigma_\lambda^0 + \sigma_\theta^2\mu_\lambda^0}{\Sigma_\lambda^0 V^{'}V + \sigma_0^2 \mathrm{I}}$.

Completing squares in λ gives

$$f\Big(\lambda|\text{rest}\Big) \propto \exp\left\{-\frac{1}{\Sigma^*_\lambda}\left[\Big(\lambda-\mu^*_\lambda\Big)'\Big(\lambda-\mu^*_\lambda\Big)\right]\right\}$$

$f\Big(\lambda|\text{rest}\Big)$ can be identified as a multivariate normal distribution $\mathcal{N}_{2m}\Big(\mu^*_\lambda, \Sigma^*_\lambda\Big)$, where

$$\Sigma^*_\lambda = \left[\frac{1}{\sigma^2_\theta}V'V + \Sigma^{0\,-1}_\lambda\right]^{-1}$$
$$\mu^*_\lambda = \Sigma^*_\lambda\left[\frac{1}{\sigma^2_\theta}\theta V + \Sigma^{0\,-1}_\lambda\mu^0_\lambda\right]$$

For β and θ there exists no closed for expressions for the full conditionals. That is the full conditionals can be identified as following some known standard probability distributions. As result, we consider the full conditionals up to some proportionality constants.

$$\begin{aligned} f\Big(\beta|\text{rest}\Big) &\propto f\Big(Z_{(y)} \mid \beta,\theta,c,a,\sigma^2_\epsilon\Big)g\Big(\beta\Big) \\ &= \mathcal{N}\Big(Z_{(y)}; S_\theta c + W_\beta a,\ \sigma^2_\epsilon \mathbf{I}_n\Big)\mathcal{N}\Big(\beta; \mu_{\beta_0}\ \sigma^2_{\beta_0}\Big) \\ &\propto \exp\left\{-\frac{1}{2\sigma^2_\epsilon}\Big(Z_{(y)} - S_\theta c - W_\beta c\Big)'\Big(Z_{(y)} - S_\theta c - W_\beta a\Big)\right\} g_5 \end{aligned}$$

where $g_4 = exp\left\{-\frac{1}{2\sigma^2_\beta}\Big(\beta-\mu_{\beta_0}\Big)^2\right\}$.

Finally, for θ, we have

$$\begin{aligned} f\Big(\theta|\text{rest}\Big) &\propto f\Big(Z_{(y)} \mid \beta,\theta,c,a,\sigma^2_\epsilon\Big)f\Big(\theta \mid \lambda,\sigma^2_0\Big) \\ &= \mathcal{N}\Big(Z_{(y)}; S_\theta c + W_\beta a,\ \sigma^2_\epsilon \mathbf{I}_n\Big)\mathcal{N}\Big(\theta; V'\lambda\ \sigma^2_{\beta_0}\Big) \\ &\propto \exp\left\{-\frac{1}{2\sigma^2_\epsilon}\Big(Z_{(y)} - S_\theta c - W_\beta c\Big)'\Big(Z_{(y)} - S_\theta c - W_\beta a\Big)\right\} g_4 \end{aligned}$$

where $g_5 = exp\left\{-\frac{1}{2\sigma^2_\theta}\Big(\theta - V'\lambda\Big)^2\right\}$.

APPENDIX I: ALGORITHM ONE

Algorithm 1 Variational Bayes Algorithm for model (3.14) (VBFBGPR)

Initialize: $\gamma_\epsilon^q = 0.5Z'_{(y)}Z_{(y)}$, $\gamma_\theta^q = 0.5\gamma_\theta$, $\gamma_\tau^q = \gamma_\tau$, $\gamma_\kappa^q = \gamma_\kappa$, $\mu_\lambda^q = \mathbf{0}$, $\Sigma_\lambda^q = 0.5\Sigma_\lambda^0$, $\sigma_\beta^{q\,2} = \sigma_{\beta_0}^2, \mu_\beta^q = \mu_{\beta_0}$, $\sigma_\theta^{q\,2} = \sigma_{\theta_0}^2, \mu_\theta^q = \mu_{\theta_0}$.

Set $\alpha_\epsilon^q = \frac{n}{2} + \alpha_\epsilon$, $\alpha_\theta^q \leftarrow 0.5 + \alpha_\theta$, $\alpha_\kappa^q = m + \alpha_\kappa^q$, $\alpha_\tau^q = m + \alpha_\tau^q$.

Do until the change in the lower bound is less than a specified tolerance:

- $\Sigma_c^q \leftarrow \left[\frac{\alpha_\epsilon^q}{\gamma_\epsilon^q} S^* + m\frac{\alpha_\kappa^q}{\gamma_\kappa^q}\mathrm{I}_{2m}\right]^{-1}, \quad \mu_c^q \leftarrow \frac{\alpha_\epsilon^q}{\gamma_\epsilon^q}\Sigma_c^q\left[\left(Z_{(y)} - W\mu_a^q\right)' S\right]$
- $\Sigma_a^q \leftarrow \left[\frac{\alpha_\epsilon^q}{\gamma_\epsilon^q} W^* + m\frac{\alpha_\tau^q}{\gamma_\tau^q}\mathrm{I}_{2m}\right]^{-1}, \quad \mu_a^q \leftarrow \frac{\alpha_\epsilon^q}{\gamma_\epsilon^q}\Sigma_a^q\left[\left(Z_{(y)} - S\mu_c^q\right)' W\right]$
- $\Sigma_\lambda^q \leftarrow \left[\frac{\alpha_\theta^q}{\alpha_\theta^q} VV' + {\Sigma_{\lambda_0}^0}^{-1}\right]^{-1}, \quad \mu_\lambda^q \leftarrow \Sigma_\lambda^q\left[\left(\frac{\alpha_\theta^q}{\alpha_\theta^q}\right)\mu_\theta^q V + \mu_\lambda^0{\Sigma_\lambda^0}^{-1}\right]$
- $\gamma_\epsilon^q \leftarrow \frac{1}{2}\left\{Z'_{(y)}Z_{(y)} - 2Z'_y Z\left(S\mu_c^q + W\mu_a^q\right) + 2\mu_a^{q\prime} S' W\mu_a^q + C^*\right\} + \frac{1}{2}\left\{+\mathrm{tr}\left((\mu_a^q\mu_a^{q\prime} + \Sigma_a^q)W^*\right)\right\} + \gamma_\epsilon$
- where $C^* = \mathrm{tr}\left((\mu_c^q\mu_c^{q\prime} + \Sigma_c^q)S^*\right)$
- $\gamma_\theta^q \leftarrow \frac{1}{2}\left[\left(\mu_\theta^q - V'\mu_\lambda^q\right)^2 + \sigma_\theta^{q\,2} + V'\Sigma_\lambda^q V\right] + \gamma_\theta$
- $\gamma_k^q \leftarrow \frac{m}{2}\mathrm{tr}\left(\left(\mu_c^q\mu_c^{q\prime} + \Sigma_c^q\right)\right) + \gamma_k$
- $\gamma_\tau^q \leftarrow \frac{m}{2}\mathrm{tr}\left(\left(\mu_a^q\mu_a^{q\prime} + \Sigma_a^q\right)\right) + \gamma_\tau$
- $\sigma_\theta^{q\,2} \leftarrow \left[\frac{\alpha_\epsilon^q}{\gamma_\epsilon^q}\left\{2\left(W\mu_a^q - Z_{(y)}\right)' D\mu_c^q + \mathrm{tr}\left(\left(\mu_a^q\mu_a^{q\prime} + \Sigma_a^q\right)Q\right)\right\} + \frac{\alpha_\theta^q}{\gamma_\theta^q}\right]^{-1}$
- $\mu_\theta^q \leftarrow \mu_\theta^q + \sigma_\theta^{q\,2}\left[\frac{\alpha_\epsilon^q}{\gamma_\epsilon^q}\left\{\left(Z_{(y)} - W\mu_a^q\right)' A\mu_c^q - \frac{1}{2}\mathrm{tr}\left(\left(\mu_c^q\mu_c^{q\prime} + \Sigma_c^q\right)Z^*\right)\right\} - B^*\right]$
- where $B^* = \frac{\alpha_\theta^q}{\gamma_\theta^q}(\mu_\theta^q - V'\mu_\lambda^q)$
- $\sigma_\beta^{q\,2} \leftarrow \left[\frac{\alpha_\epsilon^q}{\gamma_\epsilon^q}\left\{2\left(S'\mu_c^q - Z_{(y)}\right)' M\mu_a^q + \mathrm{tr}\left(\left(\mu_a^q\mu_a^{q\prime} + \Sigma_a^q\right)U\right)\right\} + \frac{1}{\sigma_{\beta_0}^2}\right]^{-1}$
- $\mu_\beta^q \leftarrow \mu_\beta^q + \sigma_\beta^{q\,2}\left[\frac{\alpha_\epsilon^q}{\gamma_\epsilon^q}\left\{\left(Z_{(y)} - S'\mu_c^q\right)' B\mu_a^q - \frac{1}{2}\mathrm{tr}\left(\left(\mu_a^q\mu_a^{q\prime} + \Sigma_a^q\right)N\right)\right\} - A*\right]$
- where $A* = \frac{1}{\sigma_{\beta_0}^2}\left(\mu_\beta^q - \mu_{\beta_0}\right)$

APPENDIX J: ALGORITHM TWO

Algorithm 2 Hybrid Gibbs Sampler (HGS)

1: Initialization: Set MCMC sample size, $\mathcal{N}$,

2: Set $b = 0$ and select starting values $\beta_{[0]}$, $\lambda_{[0]}$,$\theta_{[0]}$, $a_{[0]}$, $c_{[0]}$, $\sigma^2_{\epsilon\,[0]}$,$\sigma^2_{\tau\,[0]}$,$\sigma^2_{\theta\,[0]}$,$\sigma^2_{\kappa\,[0]}$.

3: Sample $\beta_{[b+1]}$ via Metropolis step

1. Sample β_* from $\mathcal{N}\left(\beta_{[b-1]}, \sigma^2_\beta\right)$
2. Choose $\beta_{[b+1]}$ based on:

$$\beta_{[b+1]} = \begin{cases} \beta_* & \text{with probability min } \left\{\mathcal{L}\left(\beta_{[b]}, \beta_*\right), 0\right\} \\ \beta_{[b]} & \text{otherwise,} \end{cases}$$

where

$$\mathcal{L}\left(\beta_{[b]}, \beta_*\right) = log\left[\frac{f\left(\beta^*|\text{rest}\right)}{f\left(\beta_{[b]}|\text{rest}\right)}\right]$$

$$f\left(\beta^*|\text{rest}\right) = f\left(\beta_*|\beta_{[b]}, a_{[b]}, \theta_{[b]}, S_{[b]}, W_{[b]}, \sigma^2_{\epsilon\,[b]}, Z_{(y)}\right)$$

$$f\left(\beta_{[b]}|\text{rest}\right) = f\left(\beta_{[b]}|\beta^*, a_{[b]}, \theta_{[b]}, S_{[b]}, W_{[b]}, \sigma^2_{\epsilon\,[b]}, Z_{(y)}\right)$$

3. Update spectral matrices $W_{[b+1]}, W^*_{[b+1]}$

APPENDIX K: ALGORITHM TWO CONTINUED

Algorithm 2 Hybrid Gibbs Sampler (HGS)(continued)

4: Sample $\theta_{[b+1]}$ via Metropolis step.

1. Sample θ_* from $\mathcal{N}\left(\theta_{[b-1]}, \sigma_*^2\right)$
2. Choose $\theta_{[b+1]}$ based on:

$$\theta_{[b+1]} = \begin{cases} \theta_* & \text{with probability min } \left\{\mathcal{L}\left(\theta_{[b]}, \theta_*\right), 0\right\} \\ \theta_{[b]} & \text{otherwise,} \end{cases}$$

where

$$\mathcal{L}\left(\theta_{[b]}, \theta_*\right) = log\left[\frac{f\left(\theta^*|\text{rest}\right)}{f\left(\theta_{[b]}|\text{rest}\right)}\right]$$

$$f\left(\theta^*|\text{rest}\right) = f\left(\theta_*|\theta_{[b]}, a_{[b]}, \theta_{[b]}, S_{[b]}, W_{[b]}, \sigma^2_{\epsilon\,[b]}, Zy\right)$$

$$f\left(\theta_{[b]}|\text{rest}\right) = f\left(\theta_{[b]}|\theta_*, a_{[b]}, \theta_{[b]}, S_{[b]}, W_{[b]}, \sigma^2_{\epsilon\,[b]}, Z_{(y)}\right)$$

3. Update spectral matrices $S_{[b+1]}, S^*_{[b+1]}, W_{[b+1]}$, and $W^*_{[b+1]}$

5: $c_{[b+1]} \sim \mathcal{N}\left(\sigma^{2\,-1}_{\epsilon\,[b]}\Sigma_c^*\left[\left(Z_{(y)} - W_{[b+1]}a_{[b]}\right)' S_{[b+1]}\right]\Sigma_c^*\right)$,

$$\Sigma_c^* = \left(\sigma^{2\,-1}_{\epsilon\,[b]}S^*_{[b+1]} + m\sigma^{2\,-1}_{\kappa\,[b]}\mathbf{I}_{2m}\right)^{-1}$$

6: $\sigma^2_{\tau\,[b+1]} \sim \text{IG}\left(m + \alpha_\tau,\ \frac{m}{2}a'_{[b]}a_{[b]} + \gamma_\tau\right)$

7: $\sigma^2_{\theta\,[b+1]} \sim \text{IG}\left(\frac{1}{2} + \alpha_\theta,\ \frac{1}{2}\left(\theta_{[b+1]} - v'\lambda_{[b]}\right)^2 + \gamma_\theta\right)$

8: $\sigma^2_{\kappa\,[b+1]} \sim \text{IG}\left(m + \alpha_\kappa,\ \frac{m}{2}c'_{[b+1]}c_{[b]} + \gamma_\kappa\right)$

9: $\sigma^{2\,[b+1]}_{\epsilon} \sim \text{IG}\left(\frac{n}{2} + \alpha_\epsilon,\ \gamma^*_\epsilon\right)$,

$$\gamma^*_\epsilon = \frac{1}{2}\left[C'_{(y)}C_{(y)}\right] + \gamma_\epsilon,\ C_{(y)} = \left(Z_{(y)} - S_{[b+1]}c_{[b+1]} - W_{[b]}a_{[b+1]}\right)$$

$$\Sigma_a^* = \left(\sigma^{2\,-1}_{\epsilon\,[b+1]}W^*_{[b+1]} + m\sigma^{2\,-1}_{\tau\,[b+1]}\mathbf{I}_{2m}\right)^{-1}$$

10: Set $b = b + 1$

11: If $b < \mathcal{N}_0$ repeat steps $4 - 7$. Otherwise stop.