Regularized Identification of Dynamic Models for the Personalization of a Physical Activity Intervention

by

Tarunima Mandal

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved July 2020 by the
Graduate Supervisory Committee:

Daniel E. Rivera, Chair
Jennie Si
Konstantinos Tsakalis

ARIZONA STATE UNIVERSITY

August 2020

ABSTRACT

Physical activity helps in reducing the risk of many chronic diseases, and plays a key role in maintaining good health of an individual. *Just Walk* is an intensively adaptive physical activity intervention, which has been designed based on system identification and control engineering principles. The goal of *Just Walk* is to design interventions that are responsive to an individual's changing needs, and thus encourage the individual to increase the number of steps walked.

Regularization is widely used in the field of machine learning. The goal of this thesis is to see how classical system identification principles in combination with machine learning methods like regularization help towards getting improved model estimates for complex systems. Estimating individual behavioral models using traditional prediction error methods can be done using an order selection. However, this method is can be computationally expensive due to the extensive search performed on a large set of order combination. If order selection is not done properly, it can cause bias (low order) and variance (high order) issues. In such cases regularization plays an important role in addressing the bias-variance trade-off.

One of the most important applications of identifying individual behavioral models is to understand what factors impact most the behavior of the person. Here "factors" can be considered as inputs (designed or environmental) to the participant over the course of the study, and the "behavior" is the step count of the participant under study. This is done by estimating models with different input combinations and then seeing which combinations of inputs (influence behavior most) give the best model estimate (best describe behavior of the person). As a part of this thesis, it is studied how regularized models can give a better estimation of personalized behavioral models, for the *Just Walk study*, which can further help in designing personalized interventions.

Upon study and examination of regularization approaches for four individual participants in *Just Walk*, it was observed that the regularized estimates using the DI (Digonal) kernel gave satisfactory results in terms of the model fit for these participants, while the using the (Stable Spline) SS kernel was better for personalization.

DEDICATION

Dedicated to my family and friends...

ACKNOWLEDGMENTS

I would like to start my acknowledgement by expressing the deepest appreciation to my thesis advisor, Dr. Daniel E. Rivera, who has been a pillar of support, and has encouraged me through this entire period. I would like to thank him for providing me with the opportunity to work on the thesis, as well as the NLM grant project, under his keen guidance and supervision, which has been a great learning experience for me. Being his student, and working with him, I have learnt many new things, which has contributed significantly to my knowledge and overall development. His dedication, patience, and continuous feedback on improvement, has helped me to a great extent in my thesis work.

I would also like to thank, Dr. Si and Dr. Tsakalis, for taking out time from their busy schedules, and being a part of my thesis defense committee. I also thank my professors at Arizona State University.

Finally, I would like to thank my parents for supporting me throughout my Master's program and always encouraging me. I would also like to thank my friends and colleagues for constantly motivating me, and always being there to help me out.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

1.1 Motivation

An emerging application of system identification and control theory is the design of optimized interventions in health behavior [17, 35]. *Just Walk* is a behavioral intervention which was designed using open-loop system identification methods to develop dynamical models, where the goal was to try to understand the outcome, i.e. the step count per day of an individual using the identified models. Here, system identification methods help in estimating dynamical models for physical activity (PA) in sedentary adults using an idiographic approach. The idiographic approach focuses on individual behavior as opposed to a nomothetic approach, that forms a general conclusion for all participants. Hence, these idiographic models can be used to design personalized interventions, which can in turn result in an increase in the number of steps (measure of PA) of that individual [28].

System identification is basically a method that is used for estimating mathematical models of dynamical systems, using the system's input-output data [1]. These models give an idea about the system's behavior. One of the major issues faced while building a model using system identification techniques is to achieve a good balance between the fit to the data and model complexity. For example, a high order system may give a good fit to the estimation data, but may cause overfitting and result in a model that does not generalize well to new data. On the contrary, a low order system might fail to capture important patterns and features in the estimation data set, and cause underfitting. This is an issue which is commonly seen for all kinds of system identification problems.

In classical linear time-invariant (LTI) system identification problem, prediction error methods (PEM) have typically been used. In PEM methods, model complexity depends on the model orders. To find the most appropriate model order, models are evaluated over a range of orders using cross validation techniques, information criteria such as AIC and BIC, and other model validation techniques. However, these classical methods have limitations. Firstly, they might give less desirable results for data that is short and noisy. Also, these methods can be computationally expensive, since a search has to be run over a range of orders. Many different structures have to be tested before the most appropriate one can be decided on. This complexity increases with an increase in the number of inputs and outputs, and can hence further worsen for MIMO (multi-input-multi-output) systems. From the plots in Fig. 1.1, obtained

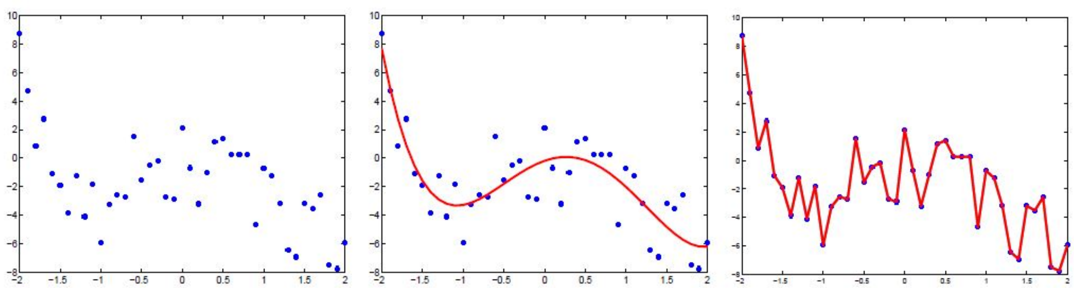


Figure 1.1: Fitting a model to given data [20]

from the work of Ljung and Chen [20], the first plot represents the raw data, the third plot shows the case of overfitting, and the second case, which is the best case, where the model fits adequately to the data under consideration. To achieve this "best case", an alternative to the classical system identification methods can be used, which is regularization of a high order model. It is a statistical technique that can help in dealing with ill-conditioned parameter estimation problems. Using a high order model allows the use of more parameters in the model structure, which ensures that the estimated model captures all features, and underfitting is avoided. Using regularization on this high order model can provide an “automatic” way of finding the most important parameters. A good parameterization is obtained by giving more weight to the important parameters, and less weight to the ones with least importance. This ensures that they do not contribute much to the variance error [14, 38].

Hence, to observe the advantages of regularized models over those obtained by classical system identification methods, both were implemented on the data set obtained from the *Just Walk* study, and then studied and compared. Two types of regularization methods are discussed as a part of the thesis [14, 23]:

1. Ridge Regression, which is basically an example of *Tikhonov regularization* for ill-posed problems.

2. Kernel Based Methods, which uses the finite impulse response model to determine the underlying model structure. This is done using kernel structures, which encode prior knowledge on the system to be identified. The hyper-parameter used for parameterization of the kernel structure decides the model complexity, which is tuned using the empirical Bayes method, where the marginal likelihood is maximized.

More background on these approaches have been provided in detail in Chapter 2.

1.2 *Just Walk* Intervention

It is understood that physical activity (PA) aids in reducing the risk of chronic diseases, and contributes to maintaining a good health of the individual [16]. However in society, a significant section of the population fails to meet the minimum PA guidelines. Thus the concern is to find a way to encourage these people to maintain favorable and sustained PA levels, especially for middle aged adults who lead a sedentary lifestyle. Since each person is different in terms of behavior and lifestyle (which might change on a daily basis based on certain factors - internal or external), the strategies for fitting regular PA into one's daily schedule should also be idiosyncratic and dynamic, and so should the time and place where individuals fit PA into their daily routine. This can be accomplished by the use of mobile health (mHealth) technologies that deliver personalized and adaptive PA interventions which are designed as per an individual's changing needs. The main advantages of the mHealth technology lie in its cost-effectiveness and convenience [16].

1.2.1 Overview of the Study

Just Walk is an intensively adaptive physical activity intervention designed using system identification and control engineering principles. The goal was to develop behavioral models for individuals under study using the ARX estimation-validation procedure. The intervention components, "goals" (daily step goals) and "expected points" (reward targets) for a participant were designed using pseudo random multi-sine signals. Analysis of the estimated dynamic models were used for understanding the PA behavior of the participant. This gave insights into what factors influence participant activity the most. These are critical for building semi-physical models, for which the basis is a dynamic extension of the *Social Cognitive Theory*, described in section 1.2.2 [16].

The participants recruited for this study were healthy, middle aged individuals between the age of 40-65 years, leading a sedentary lifestyle. These people were were overweight, with a body mass index (BMI) in the range of 25-45 kg/m^2. The requirement was that the participants own an Android (v 2.3 or higher) phone, which would be capable of connecting to a Fitbit Zip (activity tracker) via Bluetooth 4.0. Also, the participants were residents of the United States, and were willing to actively participate in the walking intervention for a period of 14 weeks. The intervention system consisted of the following:

- *Android App*: this front-end app as seen in Fig. 1.2, served as medium through which participants received a daily step goal, and were notified of the available reward points on achieving the set goal. The participants were also required to complete daily morning and evening ecological momentary assessment (EMA) measures, via this app.
- *Fitbit Zip*: it was used to measure and keep a track of the PA of the participant
- *Back-end server*: which basically stored and processed all participant data that was collected during the study, and sync the Fitbit data with the smartphone application.

Figure 1.2: Screenshot of the *Just Walk* App [16]

The 14 week study period included an initial baseline period of 2 weeks, during which no step goals were delivered to the participant. In this period, the median daily step value was calculated for each participant. Based on this data, for the rest of the study period, the participant received individualized step goals, which were in an ambitious, but doable step range. The intervention phase started from week 3 and continued though week 14, during which the participants received interventions, and were required to complete EMA via the app, which covered a range of psychometric measures like confidence in achieving goal, predicted business for that day, previous night's sleep quality, etc. The points granted, based on whether or not the goal was achieved were converted into Amazon gift cards [16, 28].

Some of the input variables considered while estimating behavioral models in the *Just Walk* study have been illustrated using time series plots in Fig. 1.3, which has been obtained from [16]. Here, the manipulated inputs are *Goals*, *Expected Points* and *Granted Points*, while *Predicted Busyness*, *Predicted Stress*, *Predicted Typical*, and *Weekday-Weekend* are the measured disturbances. Additional information on the intervention has been given in Chapter 3.

1.2.2 Behavioral Medicine and Social Cognitive Theory

Adaptive interventions are those which are personalized depending on the person's needs, which can vary based on various factors, like time, location, etc. These interventions give much better results when compared to static ones [17]. For designing adaptive interventions, it is important to be able to identify the dynamics which vary depending on the individual under consideration, as well as those that are universal, and do not depend on individual characteristics. To achieve this parsimonious modeling based on some a priori knowledge is essential. This can be accomplished by using behavior change theory, along with control systems engineering principles to develop behavioral models that can help in the creation of decision frameworks for interventions, where the ultimate goal is to increase PA among individuals who are not very active. This is what the *Just Walk* intervention aims to achieve [16].

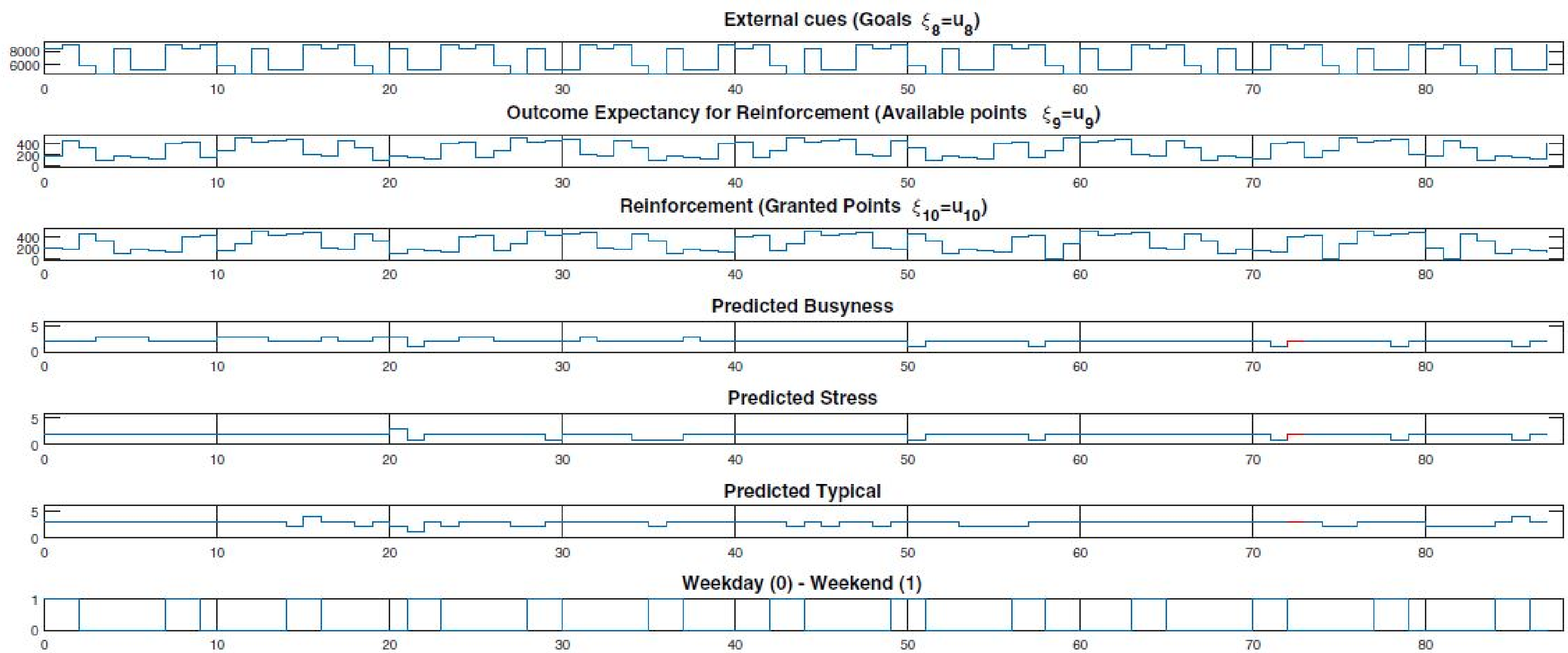


Figure 1.3: Time series plot for seven selected input sequences (manipulated inputs and measured disturbances) [16]

One of the behavior change theories that play an important role in this study is the *Social Cognitive theory* (SCT), that was proposed by Bandura [9]. The concepts for this theory have been represented in the form of a simple fluid analogy representation, as seen in Fig. 1.4, which has been referred from the work of Martín *et al.* in [27, 11]. The SCT gives an idea about the potential constructs which have a significant effect

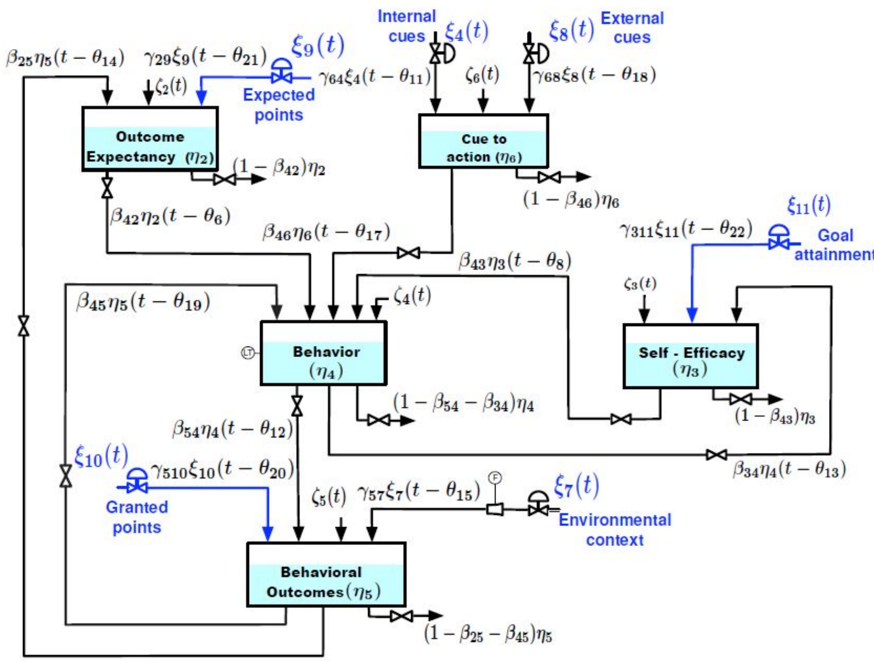


Figure 1.4: Dynamical fluid analogy model of Social Cognitive Theory [27, 11]

on participant behavior, and should be taken into consideration for the prediction of behavioral models. One such example is the Environmental Construct mentioned in Fig. 1.4, which can have variables like weather, busyness, stress, weekday, mood, and many more [16]. This model shows how different factors can interact to influence the behavior of a person over time, and this is interpreted on a daily time-scale. The primary constructs, which are self-efficacy, outcome expectancy, behavioral outcomes, cue to action, and behavior, are the 'inventories'. The other variables which can increase or decrease the behavior of an individual are the inputs, which can be external or internal.

Tailoring variables comprise of information that forms the basis for the operation of decision rules to select the intervention option for adaptive interventions. The *Just Walk* study was executed considering those variables that showed great potential for being tailoring variables (e.g. stress, mood, weather, and self-efficacy), that could be used for informing the decision made at each daily decision point. The variables under consideration were interpersonal states, and environmental context, which include the following: predicted stress, predicted typicality, predicted busyness, and whether it is a weekday/weekend [28].

1.3 Identification of Dynamical Systems and Regularization

The structure $\mathcal{M}$ of a model can be defined as a parameterized collection of models, which describe the relationship between the input and output of the system. These parameters are given by θ. $\mathcal{M}(\theta)$ is a particular model which predicts one-step ahead output at time t, based on the previous input-output data for up to time $t-1$ (given by Z^{t-1}), to give [20]

$$\hat{y}(t|\theta) = g(t, \theta, Z^{t-1}) \tag{1.1}$$

Model parameters can be obtained by fitting measured data to the predicted response of the model. This is done by comparing values in (1.1) predicted by the model, with the actual output of the system, which gives the prediction errors

$$\varepsilon(t, \theta) = y(t) - \hat{y}(t|\theta) \tag{1.2}$$

A criterion of fit is given by

$$F_N(\theta) = \sum_{t=1}^{N} \left[\varepsilon(t, \theta)^T \Lambda^{-1} \varepsilon(t, \theta)\right] \tag{1.3}$$

where Λ is a positive semi-definite (psd) matrix, weighting together the different output components (channels) [20].

In the absence of regularization, minimizing the weighted quadratic norm of prediction errors $\varepsilon(t, \theta)$, as shown in (1.3), give the parameter estimates. Since the flexibility of the model is determined by its order, an increase in the order will give a model with a better fit to the observed data. However, this might lead to higher uncertainty in the estimates because of increased random or "variance" error. Lowering the order too much has its disadvantages as well. It can lead to large systematic errors, also known as "bias" error. To get a good model estimate, which is close to the ideal situation, the model parameters should be chosen such that it minimizes the mean square error (MSE), since both, the systematic error (bias) and the random error (variance) contribute to the MSE as seen below [8]

$$\text{MSE} = |\text{Bias}|^2 + \text{Variance} \tag{1.4}$$

A reasonable estimation criterion can be

$$\text{Model} = \underset{\text{Model Class}}{\arg\min} \; [\text{Fit to Data} + \text{Penalty on Flexibility}] \tag{1.5}$$

Regularization introduces an additional penalty term, which gives more control over the bias-variance trade-off, by penalizing model flexibility. This reduces the variance, while slightly increasing the bias. The following equation shows the addition of a quadratic norm to $F_N(\theta)$, as per (1.5), which penalizes the model flexibility:

$$V_N(\theta) = F_N(\theta) + \lambda(\theta - \theta^*)^T R(\theta - \theta^*) \tag{1.6}$$

(λ is a scaling and R is a psd matrix). The parameter estimates can then be determined by

$$\hat{\theta}_N = \underset{\theta \in \mathcal{D}_{\mathcal{M}}}{\arg\min} V_N(\theta) \tag{1.7}$$

The amount of penalty added depends on λ and R, which decides the criterion for balance between model fit and penalty on the model parameter size [8, 21].

1.3.1 Regularization Example: Bias-variance Trade-off in FIR Modeling

This example replicates the work of Ljung and Chen [20, 22], for which the reference code and data have been obtained from MATLAB website [2]. Here the problem is to estimate an FIR model for the impulse response of a linear system:

$$y(t) = \sum_{k=0}^{n_b} g(k)u(t-k) \tag{1.8}$$

The choice of n_b, is critical as this decides the trade-off between bias and variance. n_b should be large enough, so that it can capture the slowly decaying impulse responses without too much error. However, if n_b is too large, it will call for the estimation of many parameters, which in turn will lead to increased variance [2].

This can be illustrated using a simple second order Butterworth filter (1.9), whose impulse response is shown in Fig. 1.5.

$$G(z) = \frac{0.02008 + 0.04017z^{-1} + 0.02008z^{-2}}{1 - 1.561z^{-1} + 0.6414z^{-2}} \tag{1.9}$$

Consider that the input to the system is a low-pass filtered white noise, and a small output disturbance is added, which is a white noise, with a variance of 0.0025. The input-output data (regularizationExampleData.mat) file, shown in Fig. 1.6 has been obtained from the MATLAB example in [2]. Over a range of $n_b = 1, \ldots, 50$, all models were examined, and it was observed that for $n_b = 13$, the best possible error norm to the impulse response, which is mse = 0.2522, gave a model with the best fit to the impulse response. Fig. 1.7 shows the comparison of the estimated impulse response with the true response. It is seen that even though the data set has sufficient length (1000 data points), and has a good signal to noise ratio, the estimated model is

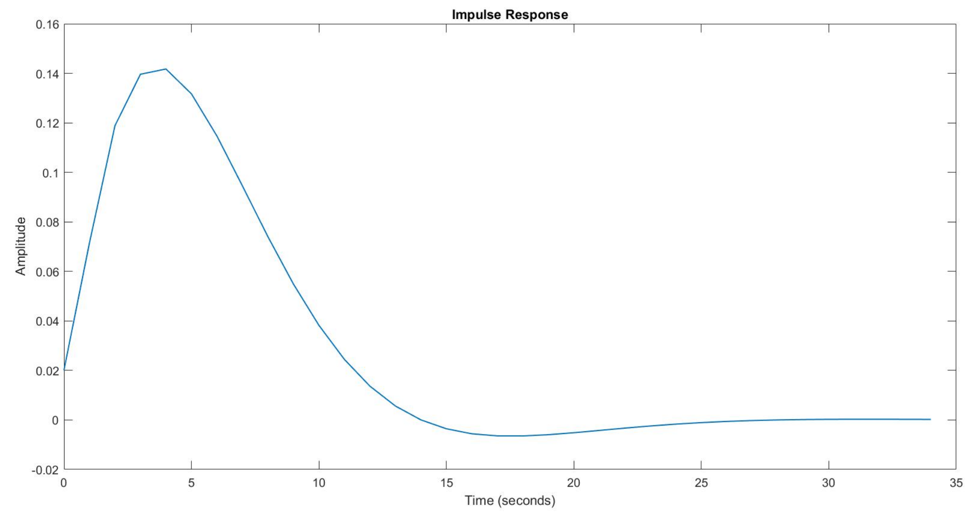


Figure 1.5: The true impulse response for the system (example) [2, 22]

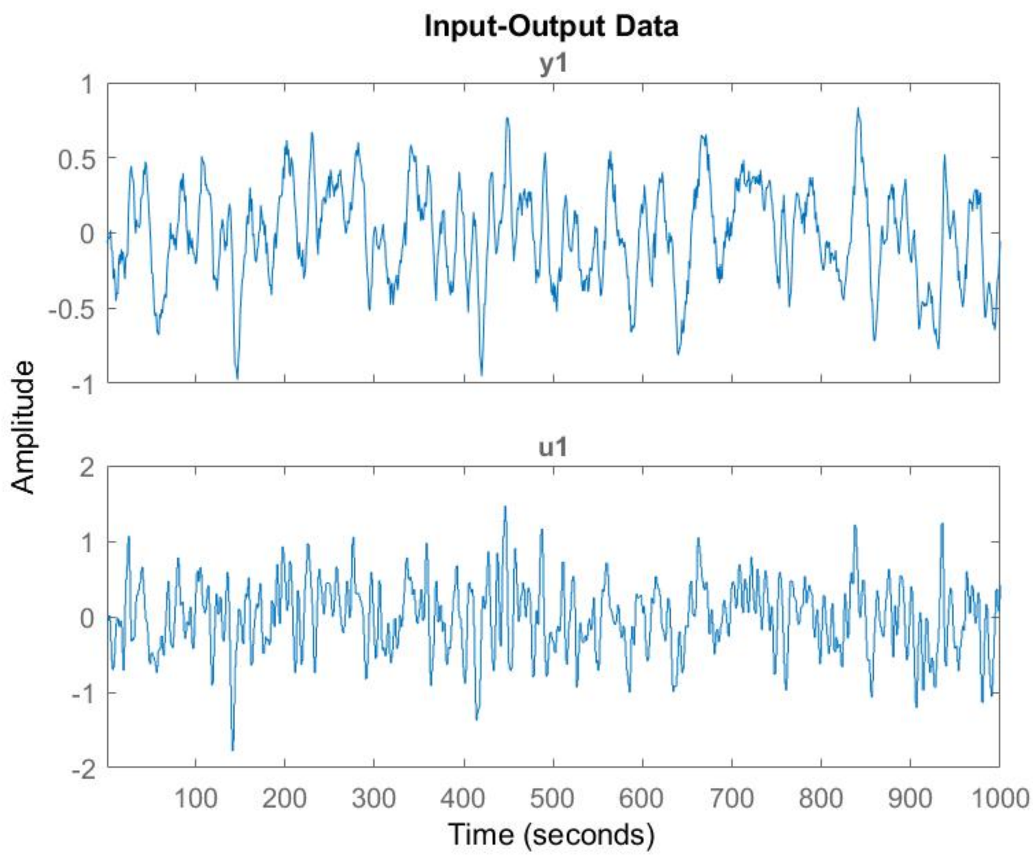


Figure 1.6: The data used for estimation (example) [2, 22]

not good enough. Also, there is significant variance. In order to get a good trade-off between bias and variance, ridge regression is tried, where a simple penalty of $\|\theta\|^2$ is applied to an FIR model of order 50. The estimate obtained using this approach has a lower error norm of 0.1171. The response is shown in Fig. 1.8.

From the plot in Fig. 1.8, it is seen that even a simple form of regularization can improve results significantly by giving a better bias-variance trade-off, than selecting an FIR model with optimum order in the absence of regularization [2].

Next, the kernel based methods, which have been further been described in Chapter 2, are also implemented to see if there is any improvement in the results. From the plot in Fig. 1.9, it can be seen that all regularization methods give better results than the FIR model with optimum order. Comparing the regularization methods, it is seen that ridge regression gives reasonable results. The kernel-based methods TC, SE, SS, and DC give very good results, and the impulse response of model estimate fits almost perfectly to the true impulse response. The HF and DI kernels perform worse, but only slightly than the ridge regression, with HF being the worst among the regularized estimates. The fit percentage values for all estimation methods have been listed in Table 1.1.

1.4 Contribution of the Thesis

In this thesis, a study is made to determine if using regularization in system identification for determining the model structure has advantages over traditional PEM methods, for data from *Just Walk*, a behavioral intervention. Behavioral dynamical models are estimated for four individual participants, which will help in designing personalized interventions. For the design of personalized interventions, it is important to be able to determine which factors impact individual behavior the most. While using the traditional order selection approach, many a times it is observed that there

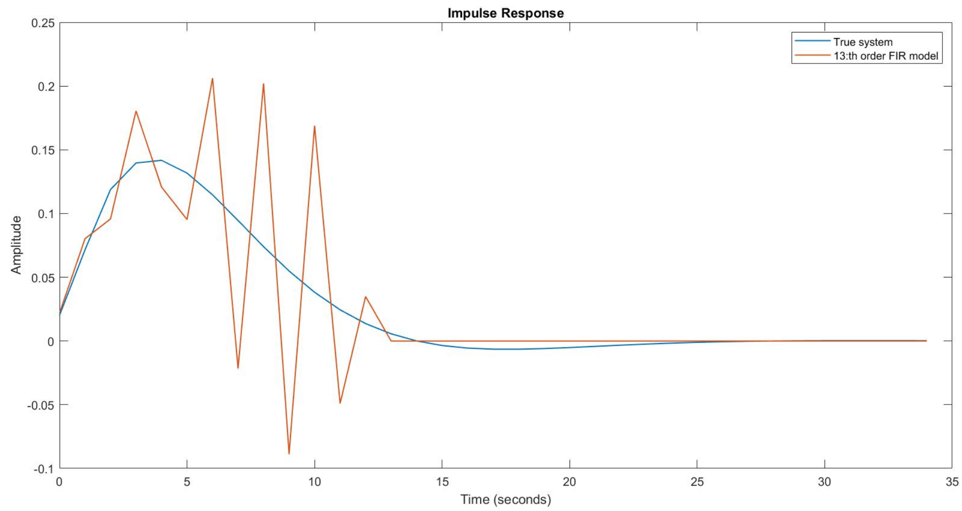


Figure 1.7: The true impulse response together with the estimate for order $n_b = 13$ (example) [2, 22]

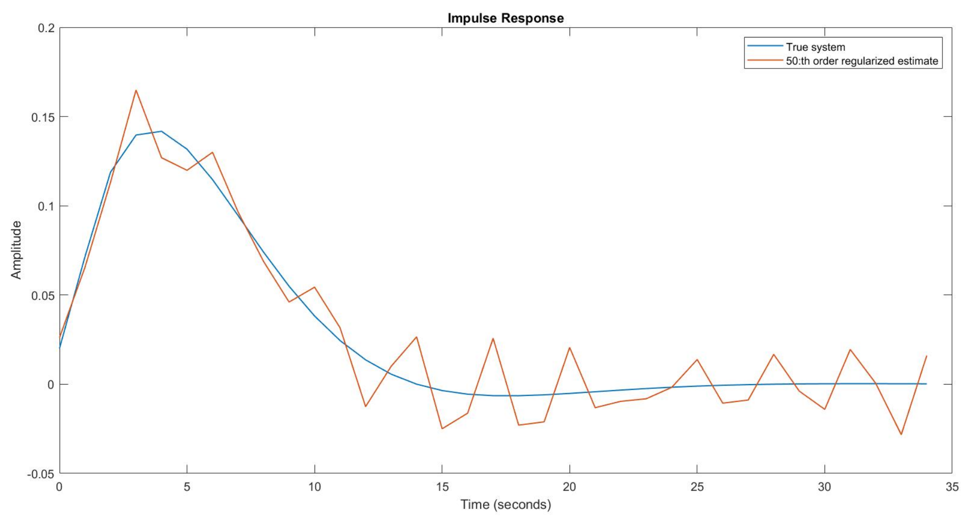


Figure 1.8: The True Impulse Response Together with the Ridge-regularized Estimate for Order $n_b = 50$ (example) [2, 22]

Table 1.1: Fit percentages for all model estimation methods (example)

Method	**% Fit**
FIR ($n_b = 13$)	83.20
Ridge Regression	84.41
Kernel TC	84.17
Kernel SE	84.26
Kernel SS	84.16
Kernel HF	84.24
Kernel DI	84.21
Kernel DC	84.17

is very little or almost no change in the % fit measure, when different combination of inputs are used for model estimation. Thus it is difficult to decide what factors influence the person's activity the most. In contrast, models estimated using some of the regularization methods show a peak in % fit when a certain combination of inputs is used for model estimation. This shows that regularization can be used for the personalization of a physical activity intervention, like the *Just Walk* intervention. Even though traditional methods like order selection give good results, from the results of this thesis, it can be seen that some regularization methods can provide comparable, or even better results than the order selection, when judged on the basis of % fit. However, one of the most important advantages of using regularized models is that estimation of these models are computationally less complex, especially when it comes to systems with multiple inputs and outputs, and are also more robust.

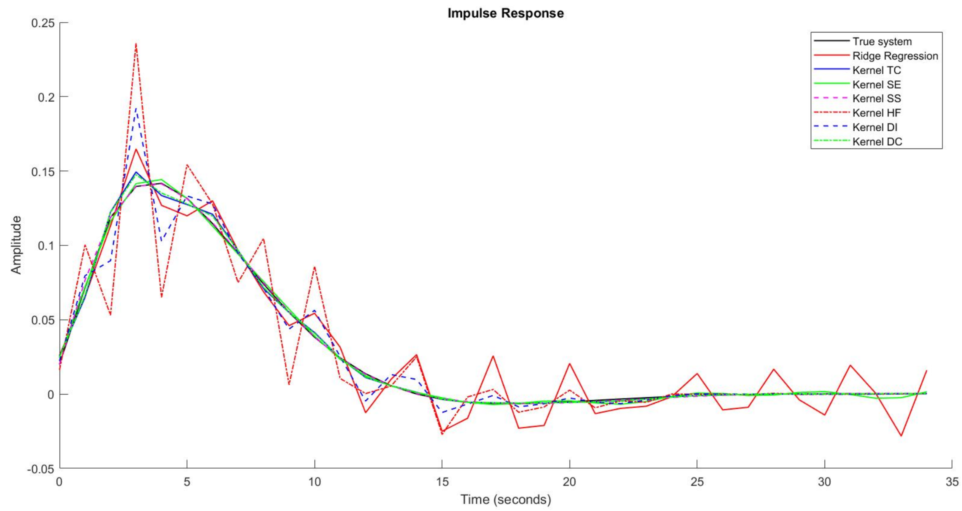


Figure 1.9: The true impulse response together with the regularized estimates using ridge regression and various kernel-based methods for order $n_b = 50$ (example)

1.5 Thesis Outline

The thesis has been organized into five chapters. *Chapter 1* is the introductory chapter, which gives an overview of the work being done as a part of this thesis. *Chapter 2* describes the various methods used (classical PEM estimates, as well as the regularized approach) for model estimation and validation. It also points at the issues related to model structure, faced while using classical system identification methods, and the reason why regularization can be considered as an option to overcome these problems. *Chapter 3* discusses the procedure for obtaining model estimates using the methods described in Chapter 2, as well as the results and conclusions obtained for a set of four participants (164, 180, 222, 230) that were studied. *Chapter 4* focuses on how addition of inputs to the standard 3-input model can have an effect on the % fit, which in turn determines whether addition of a particular input, or a combination of certain inputs can have a positive effect on the model estimate quality. This can help in determining what factors influence the behavior of a particular participant the most, and contribute to the design of personalized interventions. *Chapter 5* summarizes all results and conclusions, and gives an insight into future work that can be done based on the results obtained from this thesis.

Chapter 2

MODEL ESTIMATION AND VALIDATION: METHODS OVERVIEW

2.1 Overview

This chapter gives a briefing about various model estimation methods used in this thesis, which will help get a better understanding of the work and results described in the following chapters. The content mentioned here is an overview of established literature, and contains no new results.

2.2 Classical System Identification

System identification principles can be used for building mathematical models for dynamic systems using the input-output data that is measured. The system output is a convolution between the input to the system, and the system's impulse response. One of the ways in which system identification can be interpreted as a deconvolution problem [30]. This mainly involves determining the impulse response of the system from the collected data. Estimation paradigms from mathematical statistics, and classical methods like Maximum Likelihood (ML), are some some of the important elements from which most of the system identification methods have originated [30, 20].

2.2.1 Parametric Model Structures

For single input–single output (SISO) linear time-invariant, stable and causal systems, the general model structure from input to output can be described by the transfer function G, and the transfer function from a white noise source e to output additive disturbances is given by H [30, 20]:

$$y(t) = G(q, \theta)u(t) + H(q, \theta)e(t) \tag{2.1}$$

$$E\left[e^2(t)\right] = \lambda; \quad E\left[e(t)e(k)\right] = 0 \text{ if } k \neq t \tag{2.2}$$

where E denotes mathematical expectation. This model is in discrete time, where q denotes the shift operator $qy(t) = y(t+1)$. The sampling interval is assumed to be one time unit. The expansion of $G(q, \theta)$ and $H(q, \theta)$ in the inverse (backwards) shift operator gives the impulse responses of the two systems [30]:

$$G(q, \theta) = \sum_{k=1}^{\infty} g_k(\theta) q^{-k} \tag{2.3}$$

$$H(q, \theta) = h_0(\theta) + \sum_{k=1}^{\infty} h_k(\theta) q^{-k} \tag{2.4}$$

For normalization purposes, the function H is assumed to be monic, i.e. $h_0(\theta) = 1$.

Assuming $H(q, \theta)$ is inversely stable, the natural one-step-ahead predictor for (2.1) is given by

$$\hat{y}(t|\theta) = \frac{H(q, \theta) - 1}{H(q, \theta)} y(t) + \frac{G(q, \theta)}{H(q, \theta)} u(t) \tag{2.5}$$

Since the expansion of H starts with a "1" ($h_0(\theta) = 1$), the numerator in the terms starts with $h_1(\theta)q^{-1}$ and $g_1(\theta)q^{-1}$, respectively. This shows that there is a delay in both, y and u. The most important step which comes next, is the parameterization G and H [20, 30].

2.2.1.1 Black-box System Identification: ARX Models

Black-box models are those for which there is no prior insight available. Considering G and H to be rational in the shift operator q, the *ARX-model* (AutoRegressive with eXternal input model), which is one of the common forms of black-box models, is given by [20, 19, 30]:

$$y(t) = G(q)u(t) + H(q)e(t) \tag{2.6a}$$

$$= \frac{B(q)}{A(q)}u(t) + \frac{1}{A(q)}e(t), \quad \text{or} \tag{2.6b}$$

$$A(q)y(t) = B(q)u(t) + e(t) \tag{2.6c}$$

$$\text{where,} \quad A(q) = 1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a}$$

$$B(q) = 1 + b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}$$

Substituting $A(q)$ and $B(q)$ in (2.6c) gives:

$$y(t) + a_1 y(t-1) + \cdots + a_{n_a} y(t - n_a) \tag{2.6d}$$

$$= b_1 u(t-1) + \cdots + b_{n_b} u(t - n_b) + e(t) \tag{2.6e}$$

where, positive integers, n_a and n_b are the orders of the *ARX-model*. Other common *black-box* structures of this type are *FIR-model* (Finite Impulse Response model), *OE model* (Output Error model), *ARMAX-model* (AutoRegressive–Moving-Average model with eXogenous inputs), and *BJ-model* (Box–Jenkins) [19].

2.2.2 Approximating Linear Systems by ARX Models: Fitting Time-domain Data

For a single-input–single-output (SISO) linear stable system given by

$$y(t) = G_0(q)u(t) + H_0(q)e(t), \quad \text{or} \tag{2.7a}$$

$$y(t) = G_0(q)u(t) + \nu(t) \tag{2.7b}$$

where, $H_0(q)e(t) = \nu(t)$, is the additive noise, which is independent of the input $u(t)$. The transfer function is given by

$$G_0(q) = \sum_{k=1}^{\infty} g_k^0 q^{-k} \tag{2.8}$$

The coefficients g_k^0, $k = 1, \ldots, \infty$, form the impulse response of the system. The corresponding frequency response is defined as

$$G_0(e^{j\omega}) = \sum_{k=1}^{\infty} g_k^0 e^{-i\omega k} \tag{2.9}$$

Consider that the only information available about the input-output time domain data record for the system is that is has been collected from a linear stable system with additive noise.

$$Z^N = \{u(1), y(1), \ldots, u(N), y(N)\} \tag{2.10}$$

To find the best possible estimate $\hat{G}_N(e^{j\omega})$ of $G_0(e^{j\omega})$, a model of a particular order having an impulse response close to that of the system under consideration needs to be estimated. One of the approaches to accomplish this is to try models over a range of different orders, and use order selection, where the error in the estimate is quantified and assessed, to get the PEM estimate with the best model order [15].

Consider, an ARX model (2.6b) is estimated for orders n_a and n_b, where the estimates are calculated by linear least squares (LS) techniques. To form the criterion of fit, the values predicted by the estimated model (2.5), are compared with the actual outputs

$$\nu_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} |y(t) - \hat{y}(t|\theta)|^2 \tag{2.11}$$

which gives the parameter estimate

$$\hat{\theta}_N^{LS} = \arg\min_{\theta} \nu_N(\theta) \tag{2.12}$$

A finite-dimensional parameterization $G(q, \theta)$ in terms of θ can be formed, for which θ is then estimated in order to get the estimate $\hat{G}_N(e^{j\omega}) = G(e^{j\omega}, \hat{\theta}_N)$. One of the issues faced while doing this is determining the size of the parameter vector θ, and computing the error that is contributed by $G_0(e^{j\omega})$ from being outside the set of functions covered within the parameterization [15]. If multiple solutions to the optimization problem exist in (2.12), the equality has to be interpreted as a set inclusion. This is known as the Prediction Error Method (PEM), which coincides with the Maximum Likelihood, ML, method when the noise source e is Gaussian [20, 30].

For high order *ARX-models*, i.e. as n_a and n_b tend to infinity, and the number of data N increases, the ARX-model estimate $\hat{A}_{n_a}(q)$ and $\hat{B}_{n_b}(q)$ is given by [24, 30]:

$$\frac{\hat{B}_{n_b}(q)}{\hat{A}_{n_a}(q)} \to G_0(q), \qquad \frac{1}{\hat{A}_{n_a}(q)} \to H_0(q) \quad \text{as } n_a,\, n_b \to \infty \tag{2.13}$$

One of the disadvantages of high order ARX models is high variance. A solution to this could be estimation of a high order ARX model, followed by model reduction, which can be a solution to the numerically demanding PEM criterion minimization. This has been further discussed in the following section 2.2.3.

2.2.3 Bias and Variance Issues

The quality of the model estimate $G_0(e^{j\omega})$ is measured by the distance between the estimate and the true value, for which one of the approach is to determine the mean square error (MSE), given by [15, 30]

$$M_N(\omega) = E|\hat{G}_N(e^{j\omega}) - G_0(e^{j\omega})|^2 \tag{2.14}$$

Here, the expectation E is with respect to the output noise process $\nu(t)$. The MSE $M_N(\omega)$ is split into a bias part, and a variance part. The bias is given by the difference between the mean and a true description of the system

$$B_N(\omega) = E\hat{G}_N(e^{j\omega}) - G_)(e^{j\omega}) \tag{2.15}$$

the variance is given by

$$V_N(\omega) = E|\hat{G}_N(e^{j\omega}) - E\hat{G}_N(e^{j\omega})|^2 \tag{2.16}$$

such that

$$M_N(\omega) = V_N(\omega) + |B_N(\omega)|^2 \tag{2.17}$$

Hence, it can be said that the total error in a system is as a result of the bias and variance component.

The quality of estimated model depends on two factors. First is the quality of the measured data which is used for estimation of the model. The second is the flexibility of the chosen model structure (2.1). A more flexible model will have a smaller bias, i.e. it may be closer to the true system. But the disadvantage is that increase in the flexibility contributes to a higher variance. Thus to achieve a small total error, it is important to balance the trade-off between bias and variance by a proper choice of the model flexibility, which in turn depends on the order of the estimated model [30].

2.2.4 Selection of Model Flexibility: Order Selection

For black-box estimation, there is almost no prior information available about the system. In such cases, a reasonable approach for determining the best order for ARX model estimation, would be to evaluate over a range of orders, and select the one that seems to work the best. The aim should be to use the most appropriate model order, which should not be higher than necessary. This can be determined by analyzing how model % fit improves as a function of the model order. A validation data-set should be used to avoid overfitting [3].

2.2.4.1 Cross Validation

In this thesis, cross validation (CV) was used for order selection, which is widely used for model selection. In this method, for different choices of θ, an estimate of the predictive capability of the model is obtained, which is then optimized for parameter selection [30].

In *Holdout validation*, is the simplest form of CV in which the available data is split in two parts

- *Estimation Set*: Used for estimating the model
- *Validation Set*: Used for assessing the predictive capability of the model

The model order with the best fit to the validation data is then selected [4].

2.3 Regularization of Linear Regression Models

The trade-off between bias and variance errors can be examined by performing cross-validation tests on a set of models by varying its flexibility. However, this approach may not give very good results always because of the factors described below [8]:

- There is no a-priori information about the system, which can be used for improving the model quality.
- The quality of estimated model is governed by the quality of the data used for estimation, i.e. it should be rich enough to capture the full range of dynamic behavior of the system
- Varying the model order does not allow for explicitly shaping the variance of the underlying parameters.

In such cases, regularization gives a better control over the bias versus variance trade-off by introducing an additional term in the minimization criterion that penalizes the model flexibility.

2.3.1 Linear Regression Models: FIR Model

The simplest approach to estimate $G(q, \theta)$ is to truncate the expansion (2.8) at a finite number of impulse response coefficients.[15]

$$G(q, \theta) = \sum_{k=1}^{n} g_k q^{-k}, \quad \theta = [g_1 g_2 \dots g_n] \tag{2.18}$$

where n is the order of the FIR model. The vector θ can then be estimated by the least squares method. This model is written as a linear regression model, which has the form [15, 30]

$$y(t) = \varphi^T(t)\theta + \nu(t), \quad \varphi(t) = [u(t-1) \dots u(t-n)]^T \tag{2.19a}$$

Here, the observed variables are the output y, and the regression vector φ. ν is a noise disturbance, and θ is the unknown parameter vector. $\nu(t)$ is assumed to be independent of $\varphi(t)$. For convenience, (2.19a) is rewritten in vector form, by stacking all the row elements in $y(t)$ and $\varphi^T(t)$ to form the vectors Y_N and ϕ_N^T, which gives

$$\text{or} \quad Y_N = \phi_N^T\theta + \Lambda_N \tag{2.19b}$$

$$\text{where} \quad Y_N = [y(n+1)\ y(n+2)\ \dots\ y(N)]^T \tag{2.19c}$$

$$\phi_N = [\varphi(n+1)\ \varphi(n+2)\ \dots\ \varphi(N)] \tag{2.19d}$$

$$\Lambda_N = [\nu(n+1)\ \nu(n+2)\ \dots\ \nu(N)]^T \tag{2.19e}$$

The least squares (LS) estimate of the parameter θ is

$$\hat{\theta}_N^{LS} = [\hat{g}_1^{LS}\ \hat{g}_2^{LS}\ \dots\ \hat{g}_n^{LS}]^T = \arg\min \nu_N(\theta) \tag{2.20a}$$

$$\nu_N(\theta) = \|Y_N - \phi_N^T\theta\|^2 = \sum_{t=n+1}^{N} (y(t) - \varphi^T(t)\theta)^2 \tag{2.20b}$$

$$\hat{\theta}_N^{LS} = (\phi_N\phi_N^T)^{-1}\phi_N Y_N = R_N^{-1}F_N \tag{2.20c}$$

$$R_N = \phi_N\phi_N^T = \sum_{t=n+1}^{N} \varphi(t)\varphi(t)^T \tag{2.20d}$$

where $\|\cdot\|$ represents the Euclidean norm. $\phi_N\phi_N^T$ is assumed to be nonsingular. In cases where $\phi_N\phi_N^T$ is singular or ill-conditioned, a regularization term is introduced in (2.20a) by means of a regularization matrix D and the regularized least squares is considered instead [15, 30].

Since $u(-n+1), \dots, u(0)$ are not known, the summation in (2.20b) starts at $n+1$ to allow $\varphi(t)$ to be formed. From (2.19b) it can be seen that the first n outputs,

$y(1), y(2), \ldots, y(n)$ in the data set $Z^N = \{u(t), y(t), t = 1, \ldots, N\}$ are not used. [15] Assume that,

$$E\nu(t) = 0, \qquad E\nu(t)\nu(s) = \sigma^2 \delta_{t,s} \tag{2.21}$$

where, $\delta_{t,s}$ denotes the Kronecker delta function. i.e. if $t = s$, $\delta_{t,s} = 1$, otherwise $\delta_{t,s} = 0$. The input $u(t)$, and thus $\varphi(t)$, are seen as deterministic variables. For simplicity of the conceptual analysis, it is assumed that there exists $\mu > 0$ such that

$$\frac{1}{N-n} R_N \to \mu I_n \qquad \text{as } N \to \infty \tag{2.22}$$

where, I_n denotes the $n \times n$ identity matrix. This means that for reasonably large N,

$$\frac{1}{N-n} R_N \approx \mu I_n \tag{2.23}$$

Then it can be shown that

$$E\hat{\theta}_N^{LS} = \theta_0 = [g_1^0 \; g_2^0 \; \ldots \; g_n^0]^T \tag{2.24}$$

$$E(\hat{\theta}_N^{LS} - \theta_0)(\hat{\theta}_N^{LS} - \theta_0)^T = \sigma^2 R_N^{-1} \approx \frac{\sigma^2}{(N-n)\mu} I_n \tag{2.25}$$

which gives the bias, variance, and MSE, corresponding to (2.15) to (2.17), as follows

$$B_N(\omega) = \sum_{k=n+1}^{\infty} g_k^0 e^{i\omega k} \tag{2.26a}$$

$$V_N(\omega) \approx \frac{n\sigma^2}{(N-n)\mu} \tag{2.26b}$$

$$M_N(\omega) \approx \frac{n\sigma^2}{(N-n)\mu} + \left| \sum_{k=n+1}^{\infty} g_k^0 e^{i\omega k} \right|^2 \tag{2.26c}$$

When the order n increases to infinity with the number of data N, sufficiently slowly, the model (2.18) converges to the true transfer function (2.8). For minimizing the MSE $M_N(\omega)$ with respect to the order n for a given data size N, it is necessary to have an idea about the size of $B_N(\omega)$ as a function of n. Assuming the system has all poles inside a circle with radius $\bar{\lambda}$, there exists a $\bar{c} > 0$ such that [15]

$$|g_K^0| < \bar{c}\bar{\lambda}^k \tag{2.27a}$$

$$|B_N(\omega)| < \frac{\bar{c}\bar{\lambda}^{n+1}}{1-\bar{\lambda}} \tag{2.27b}$$

Since the squared bias decreases like $\bar{\lambda}^{2n}$ as a function of n, and the variance increases like n (for large N), an upper bound on the MSE $M_N(\omega)$ is minimized by an order n that increases with N like

$$n_{opt} \sim \log N \tag{2.28}$$

Thus, the upper bound on the MSE $M_N(\omega)$ is minimized at relatively low orders compared to the data size [15].

2.3.2 Regularized Least Squares

In order to regularize the estimate, a regularization term $\theta^T D\theta$ is added in (2.20a) to obtain the *regularized least squares* (ReLS) [30]. This is done by replacing the criterion $\nu_N(\theta)$ in (2.20) by

$$\nu_N^R(\theta, D) = \sum_{t=n+1}^{N} (y(t) - \varphi(t)\theta)^2 + \theta^T D\theta \tag{2.29a}$$

where the regularization matrix D, is positive semi-definite, and it changes the estimate to be

$$\hat{\theta}_N^R = [\hat{g}_1^R \; \hat{g}_2^R \; \dots \; \hat{g}_n^R]^T \tag{2.29b}$$

$$= (R_N + D)^{-1}F_N = (R_N + D)^{-1}R_N\hat{\theta}_N^{LS} \tag{2.29c}$$

Disregarding the tail of the impulse response g_k^0, $k > n$ and assuming that the true system is given by a FIR model of order n, the main task at hand now is the determination of the regularization matrix D. The optimal regularization matrix should be such that it should minimize the mean square error matrix and give a theoretical upper bound on the accuracy that can be achieved. Considering all expectations are with respect to $\nu(t)$ [15]

$$E\hat{\theta}_N^R = (R_N + D)^{-1} R_N \theta_0 \tag{2.30a}$$

$$\theta_{bias}^R = E\hat{\theta}_N^R - \theta_0 = -(R_N + D)^{-1} D\theta_0 \tag{2.30b}$$

$$\tilde{\theta} = \hat{\theta}_N^R - E\hat{\theta}_N^R = (R_N + D)^{-1} R_N (\hat{\theta}_N^{LS} - \theta_0) \tag{2.30c}$$

$$E\tilde{\theta}\tilde{\theta}^T = (R_N + D)^{-1} \sigma^2 R_N (R_N + D)^{-1} \tag{2.30d}$$

$$\begin{aligned} MSE(\hat{\theta}_N^R) &= E(\hat{\theta}_N^R - \theta_0)(\hat{\theta}_N^R - \theta_0)^T \\ &= E\tilde{\theta}\tilde{\theta}^T + \theta_{bias}^R(\theta_{bias}^R)^T \\ &= (R_N + D)^{-1}(\sigma^2 R_N + D\theta_0\theta_0^T D_T)(R_N + D)^{-1} \end{aligned} \tag{2.30e}$$

where $MSE(\hat{\theta}_N^R)$ is the MSE matrix of $\hat{\theta}_N^R$ with respect to the true impulse response coefficients vector θ_0 in (2.25). Suppose, $D = diag(d_1, d_2, \ldots, d_n)$, and (2.23) is used for R_N, the $(k, k)^{th}$ element of $MSE(\hat{\theta}_N^R)$ will have the form

$$MSE(\hat{g}_k^R) \approx \frac{\sigma^2 \mu(N-n) + d_k^2 (g_k^0)^2}{(\mu(N-n) + d_k)^2} \tag{2.31}$$

which is minimized with respect to d_k by $d_k = \sigma^2/(g_k^0)^2$. This gives an idea about how to choose the regularization matrix D, i.e. if the system is stable as in (2.27a), the diagonal of D should increase exponentially [15, 30]:

$$d_k = \frac{\sigma^2}{c\lambda^k}, \quad k = 1, \ldots, n \tag{2.32}$$

where $\lambda = \bar{\lambda}^2$ and c = $\bar{c}^2$

2.4 Ridge Regression

In Ridge Regression, regularization modifies the cost function by adding a term proportional to the square of the norm of the parameter vector θ, so that the parameters θ are obtained by:

$$\hat{\theta}_N^R = \arg\min_{\theta} \|Y_N - \phi_N^T \theta\|_2^2 + \gamma \|\theta\|_2^2 \tag{2.33}$$

where γ is a positive constant that has the effect of trading variance error in $\nu_N(\theta)$ for bias error, i.e. the larger the value of γ, the higher the bias and lower the variance of θ. The added term penalizes the parameter values with the effect of keeping their values small during estimation [8, 29].

Cross validation is used for tuning the regularization, where the data set is split into two parts: estimation and validation data. Regularized models are then estimated using the estimation data for various values of the regularization variables, and these models are then evaluated to check how well they can reproduce the validation data. The regularization variables that give the model the best fit to validation data, are then picked [20].

2.5 Kernel-based Regularization

In the kernel-based regularization methods, the penalty term is made more effective by using a positive definite matrix, which allows weighting and/or rotation of the parameter vector. This square matrix gives additional freedom for:

- Shaping the penalty term to meet the required constraints, such as keeping the model stable
- Adding known information about the model parameters, such as reliability of the individual parameters in the θ vector

Interpretation in the Bayesian setting is that θ has a prior distribution that is Gaussian. Prior information is introduced in the identification process by modeling the impulse response as a zero-mean Gaussian process, and introducing prior information in the identification process by assigning a covariance, known as kernel in the machine learning literature. These kernels include information on impulse response exponential stability and depend on some hyperparameters which can be estimated from the data e.g. using marginal likelihood maximization. This procedure is much more robust than the PEM methods [8, 30].

2.5.1 A Bayesian Interpretation

As per the Bayesian interpretation, the the parameter to be estimated is a random variable, whose posterior distribution is estimated from the given observations. Following this, a jointly Gaussian random variable is conditioned, which is a key element in Bayesian calculations [15]. Assuming $x|y \sim \mathcal{N}(m, P)$ denotes that conditioned on y, x is a multivariate Gaussian random variable with mean vector m and covariance matrix P, let

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m_1 \\ m_2 \end{bmatrix}, \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \right) \tag{2.34a}$$

Then, the conditional distribution of x_1, given x_2 is

$$x_1|x_2 \sim \mathcal{N}(m, P) \tag{2.34b}$$

$$m = m_1 + P_{12}P_{22}^{-1}(x_2 - m_2) \tag{2.34c}$$

$$P = P_{11} - P_{12}P_{22}^{-1}P_{21} \tag{2.34d}$$

Recalling the following simple matrix equality:

$$A(I_j + BA)^{-1} = (I_k + AB)^{-1}A \tag{2.35}$$

where A is an $k \times j$ matrix and B is a $j \times k$ matrix. For symmetric matrices A and B, let $A \geq B$ denote that $A - B$ is a positive semi-definite matrix [15].

The parameter of the n^{th} order FIR model (2.18), i.e., the impulse response coefficients vector θ as a random variable, is considered to be of Gaussian distribution with zero mean and covariance matrix P_n:

$$\theta \sim \mathcal{N}(\theta^{ap}, P_n), \quad \theta^{ap} = 0 \tag{2.36}$$

If the input $u(t)$ (and $\varphi(t)$, refer (2.19a)) is known, and the noise $\nu(t)$ is independently Gaussian distributed with

$$\nu(t) \sim \mathcal{N}(0, \sigma^2) \tag{2.37}$$

then with

$$Y_N = \phi_N^T + \Lambda_N \tag{2.38}$$

Y_N and θ will be jointly Gaussian variables:

$$\begin{bmatrix} \theta \\ Y_N \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} P_n & P_n\phi_N \\ \phi_N^T P_n & \phi_N^T P_n \phi_N + \sigma^2 I_{N-n} \end{bmatrix} \right) \tag{2.39}$$

The posterior distribution of θ given Y_N follows from (2.34)

$$\theta | Y_N \sim \mathcal{N}(\hat{\theta}_N^{apost}, P_N^{apost}) \tag{2.40a}$$

$$\hat{\theta}_N^{apost} = P_n\phi_N(\phi_N^T P_n \phi_N + \sigma^2 I_{N-n})^{-1} Y_N \tag{2.40b}$$

$$= (P_n\phi_N\phi_N^T + \sigma^2 I_n)^{-1} P_n \phi_N Y_N \tag{2.40c}$$

$$= (R_N + \sigma^2 P_n^{-1})^{-1} F_N \tag{2.40d}$$

$$= ((\sigma^2 R_N^{-1})^{-1} + P_n^{-1})^{-1} (\sigma^2 R_N^{-1})^{-1} \hat{\theta}_N^{LS} \tag{2.40e}$$

$$P_N^{apost} = P_n - P_n\phi_N(\phi_N^T P_n \phi_N + \sigma^2 I_{N-n})^{-1} \phi_N^T P_n \tag{2.40f}$$

$$= P_n - (P_n\phi_N\phi_N^T + \sigma^2 I_n)^{-1} P_n \phi_N \phi_N^T P_n$$

$$= ((\sigma^2 R_N^{-1})^{-1} + P_n^{-1})^{-1} \tag{2.40g}$$

where F_N, R_N, $\hat{\theta}_N^{LS}$ are defined in (2.20) [15].

It is observed that this *aposteriori* estimate $\hat{\theta}_N^{apost}$ is the same as the regularized estimate $\hat{\theta}_N^{R}$ if the regularization matrix D is chosen as

$$D = \sigma^2 P_n^{-1} \tag{2.41}$$

which indicates that regularization is closely related to prior estimates, and gives an insight into how the regularization matrix can be chosen such that it reflects the size and correlations of the impulse response coefficients. The size depends on how the diagonal elements are chosen (2.32). In case of a smooth impulse response (for example a fast sampled continuous system), the diagonals of P_n close to the main diagonal can show high correlation in order to reflect this [15]. A simple choice is to let the correlation coefficient between g_k and g_j in (2.18) be $\rho^{|k-j|}$. With diagonal elements of P_n being $c\lambda^k$ as in (2.32), a covariance matrix P_n is obtained, whose $(k,j)^{th}$ element is

$$c\rho^{|k-j|}\lambda^{(k+j)/2} \tag{2.42}$$

where $c \geq 0$, $0 \leq \lambda \leq 1$ and $|\rho| \leq 1$. Hence, the Bayesian perspective has given additional insights in the choice of D [15].

2.5.1.1 Estimating Hyper-parameters

If prior knowledge does not give a definite choice of P_n, it is natural to let it depend on unknown hyper-parameters β, $P_n(\beta)$ (like $\beta = [c\ \lambda]$ in (2.32)). From (2.39) it is seen that [15]

$$Y_N \sim \mathcal{N}(0, \sigma^2 I_{N-n} + \phi_N^T P_n(\beta)\phi_N) \tag{2.43a}$$

hence, with a slight twist in the Bayesian framework, the likelihood function of the observation Y_N can be formed, given β, and β can be estimated by the maximum likelihood (ML) method:

$$\hat{\beta} = \arg\min_{\beta} Y_N^T \Sigma(\beta)^{-1} Y_N + \log\det\Sigma(\beta) \tag{2.43b}$$

where, $\Sigma(\beta) = \sigma^2 I_{N-n} + \phi_N P_n(\beta)\phi_N^T$. This method of estimating hyper-parameters in the prior distribution is known as the *Empirical Bayes* method [10, 15].

2.5.1.2 'DI', 'DC' and 'TC' Kernels

Since the estimation of FIR models is a linear regression problem, regularization and estimation of hyper-parameters can thus be applied to estimation of FIR models. Suitable choices of P should reflect a reasonable assumption about the system's impulse response. For example, if the system is exponentially stable, the impulse response coefficients g_k should decay exponentially, and if the impulse response is smooth, neighboring values should have a positive correlation. Based on the Bayesian method (2.40) and (2.43), with the following prior covariances: the diagonal (2.32) and the correlation (2.42), a suitable regularization matrix P for θ could be a matrix whose k, j element is

$$P_{DI}(k,j) = \begin{cases} c\lambda^k & \text{if } k = j \\ 0 & \text{else} \end{cases} \qquad \text{('Diagonal')} \tag{2.44}$$

$$P_{DC}(k,j) = c\rho^{|k-j|}\lambda^{(k+j)/2} \qquad \text{('Diagonal/correlated')} \tag{2.45}$$

where the hyper-parameters are $c \geq 0$, $0 \leq \lambda \leq 1$ and $|\rho| \leq 1$. Here λ accounts for the exponential decay along the diagonal, while ρ describes the correlation across the diagonal (the correlation between neighboring impulse response coefficients) [30, 15].

The complexity of the prior (2.45) can be reduced by linking $\rho = f(\lambda)$, where $f(\cdot)$ could be, for example, either a non-decreasing function that satisfies $f(0) = 0$ and $f(1) = 1$ or a non-increasing function that satisfies $f(0) = 0$ and $f(1) = -1$. A special case of (2.45) by linking $\rho = \lambda^{1/2}$ gives:

$$P_{TC}(k,j) = c\min(\lambda^j, \lambda^k) \qquad \text{('Tuned/correlated')} \tag{2.46}$$

which is known as TC for *Tuned/Correlated* kernel (2.46), which corresponds to the *First-order Stable Spline* [15, 30].

2.5.2 *Gaussian Process Regression for Transfer Function Estimation*

Gaussian process regression (GPR), is a method for inferring an unknown function $f(x)$ from measurements y_k, $k = 1, 2, \ldots, N$ that bear some information about $f(x)$. The argument x can either be a continuous or a discrete variable. The prior information about the function is that it is a Gaussian process, with a certain mean and covariance function. This means that the vector $[f(x_1), f(x_2), \ldots, f(x_n)]$, for a collection of points x_i is a jointly Gaussian random vector, with mean $m(x) = Ef(x)$ and covariances [15]

$$Cov(f(x_i), f(x_j)) = P(x_i, x_j) \tag{2.47}$$

where $P(x_i, x_j)$ is the kernel. Often $m(x) \equiv 0$. The observation y_k is a linear functional of $f(x_i)$, measured in additive Gaussian noise. This causes $[f(x), y_1, \ldots, y_N]$ to be a jointly Gaussian vector, which means that the posterior distributions,

$$p(f(x_1), \ldots, f(x_n)|y_1, \ldots, y_N) \tag{2.48}$$

can be calculated by the rules for conditioning jointly Gaussian random variables, (2.36).

For a sampled model, the impulse response function is given by g_0^k, $k = 1, \ldots, \infty$ in (2.8). The observation y_k is the measured output in (2.7b) at time $t = k$. Modeling the impulse response function as a Gaussian process means that, for any n,

$$[g_1, \ldots, g_n] \sim \mathcal{N}(0, P_n) \tag{2.49}$$

where P_n is the $n \times n$ upper left block matrix of the semi-infinite matrix P defined in (2.47). This is similar to the situation in the Bayesian perspective (2.36)–(2.40). The Gaussian process estimate of any collections of impulse response coefficients is thus given by (2.40) [15].

The considerations for choosing P_n in (2.49) and in (2.36) must be the same, and the relation to the regularization matrix D in (2.41) still holds. Several standard choices for (2.47) exist in GPR. The the kernels/covariance functions mentioned in the following section 2.5.2.1 , have further been discussed in [31]

2.5.2.1 'SE', 'SS' and 'HF' Kernels

The *Squared Exponential* kernel, which is probably the most widely-used kernel within the kernel machines field [34], is defined as

$$P_{SE}(k,j) = ce^{-\frac{(k-j)^2}{2\lambda^2}} \quad \text{('Squared Exponential')} \tag{2.50a}$$

For a stable optimal controller, the variance of θ is expected to tend to zero exponentially, with a certain decay rate λ. This is because the impulse response of a stable system annihilates for large time t. Also, if the impulse response is supposed to be smooth, the neighbour coefficients have positive correlation. From these two hypotheses, *Stable Spline Kernel* can be built as [32, 15]

$$P_{SS}(k,j) = \begin{cases} c\dfrac{\lambda^{2k}}{2}\left(\lambda^j - \dfrac{\lambda^k}{3}\right), & k \geq j \\ c\dfrac{\lambda^{2j}}{2}\left(\lambda^k - \dfrac{\lambda^j}{3}\right), & k < j \end{cases} \quad \text{('Stable Spline')} \tag{2.50b}$$

where the hyper-parameters $c \geq 0$, $0 \leq \lambda \leq 1$. If the impulse response of the controller is non-smooth and rapidly varying, it is necessary to include the case of negative correlation between impulse response coefficients. In this case, a suitable choice of the kernel can be the *High Frequency Stable Spline* defined as [32]

$$P_{HF}(k,j) = c\begin{cases} \max(\lambda k, \lambda j) & \text{if } k+j \text{ is even} \\ -\max(\lambda k, \lambda j) & \text{if } k+j \text{ is odd} \end{cases} \tag{2.50c}$$

2.6 Summary

This chapter was basically to give an insight into the background literature, based on which the work in this thesis was done. A detailed explanation has been given in this chapter regarding the bias-variance trade-off issues faced while using traditional system identification methods, and how regularization can help overcome this. The ridge regression, as well as the kernel-based regularization methods have been further examined in the subsequent chapters.

Chapter 3

MODEL ESTIMATION AND VALIDATION OF *JUST WALK* DATA: EXPERIMENTAL SETUP AND RESULTS

3.1 Experimental Setup

This section describes how control system methodologies, especially system identification, can be used for the development of dynamic models of physical activity, based on an idiographic, which is a person specific approach. These models estimated can then be used to personalize interventions in a systematic and scalable way. In this chapter, the goal is to apply system identification methods to develop dynamical models of PA for individuals, which can help in getting insights on potential tailoring variables. Tailoring variables are predictors which can be expected to influence step count for the person the most, and these are determined using the idiographic models. This is done to ensure that, individual differences, contextual differences, and changes to both the person and context over time can all be taken into account while providing the optimal intervention for a participant when they need it, which helps in providing more personalized adjustments of the intervention dosages over time [28].

3.1.1 Input Variables

The input variables for the *Just Walk* study consists of two manipulated input variables, and a set of seven measured disturbance variables, which have been described in Table 3.1. A time series plot for a selected participant that shows the behavior as well as the eight inputs, some of which are shown in Fig. 1.3, which has been obtained from the work of Freigoun *et al.* in [16].

Table 3.1: Variables used in the *Just Walk* analyses

Variable	Design/Measurement	Scale/Unit
Step goals	Multisine input	Steps/day
Expected points	Multisine input	100-500
Granted points	If goals are met	0-500
Weekday/weekend		0/1
Predicted stress	How stressful do you expect your day to be?	1-5 scale (*not at all to very*)
Predicted busyness	How busy today is going to be for you?	1-4 scale (*not at all to extremely*)
Predicted typicality	How typical do you expect today to be for a [*day of the week*]?	1-4 scale (*not at all to completely*)
Temperature		$^\circ F$

3.1.1.1 Manipulated Input Signal Design

The study used an orthogonal multisine experimental design [26] to deliver excitation to the intervention inputs. A "multisine" pseudo-random signal is one that appears to be random but is in fact deterministic and periodic, which can help produce sufficient excitation in the behavior, which is important for generating dynamic models. The pseudo-randomization strategies used in these experiments are carried out at the within-person level, allowing inferences to be drawn at the individual level [28].

In the *Just Walk* study, there are two manipulated input signals u_n, generated from a multisine signal:

- *Step Goals*: These establish the desired behavior in a quantitative form

- *Expected Points*: These are the daily available points announced each morning, granted upon goal achievement. Once these are sufficiently accumulated, they can be traded for Amazon gift cards.

The multisine signal is defined as

$$u_n(k) = \lambda_n \sum_{j=1}^{N_s/2} \sqrt{2\alpha_{[n,j]}} \cos\left(\omega_j k T_s + \phi_{[n,j]}\right) \tag{3.1}$$
$$\omega_j = \tfrac{2\pi j}{N_s T_s}, \quad k = 1, \ldots, N_s$$

where λ_n is the scaling factor, N_s is the number of samples per period, T_s is the sampling time. For the j^{th} harmonic of the signal, $\alpha_{[n,j]}$ is a factor used to specify the relative power of the harmonic, ω_j is the frequency, and $\phi_{[n,j]}$ is the phase. To obtain independent transfer function and uncertainty estimates, factors $\alpha_{[n,j]}$ are chosen to excite input signals orthogonally in frequency. Two signals are orthogonal if a nonzero Fourier coefficient at a specific frequency in one signal implies a zero-valued Fourier coefficient at the same frequency for the other, which is known as a "zippered" spectra design [16].

The conceptual representation of the "zippered" design [11] is shown in Fig. 3.1 Freigoun *et al.* [16]. There is another work by Rivera *et al.* [36], that describes this design. For n_u design inputs and n_s independently excited sinusoids, the Fourier coefficients are specified as

$$\alpha_{[n,j]} = \begin{cases} 1, & \text{if } j = n_u(i-1) + (n-7) \\ & \text{for } i = 1, 2, \ldots, n_s \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

Using the ω_j frequencies defined in (3.1), and the Nyquist- Shannon sampling theorem, the following bound for N_s is defined:

$$N_s \geq 2n_s \tag{3.3}$$

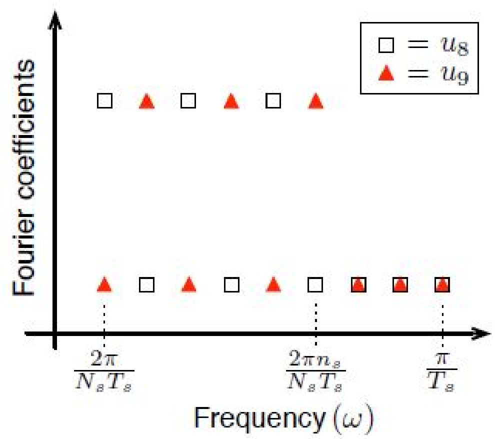


Figure 3.1: Conceptual representation of a "zippered" spectra design for $n_u = 2$ design inputs, and $n_s = 6$ harmonic rrequencies [11, 16]

If $n_s = 6$ excited sinusoids are selected for the $n_u = 2$ design inputs, then from applying (3.3), $N_s = 16$ days is selected as a feasible option. Phases $\phi_{[n,j]}$ are selected so as to minimize the crest factor of the signal [16]. The amplitudes for input signals (Goals: u_8 and Expected Points: u_9) were chosen based on the initial baseline level of PA for the participant, which was varied if the actual baseline step level of individuals was too high or low.

3.1.2 Data Pre-processing and Characteristics

The pre-processing of the collected data included interpolation (to account for missing data), and shifting Actual Steps and Granted Points by one sample to reflect temporal precedence. The absence of drifts in the data leads to the assumption that the noise characteristics of the problem remain essentially unchanged during the course of the intervention, and hence the data is considered stationary over the the intervention period [16].

For cross-validation, which is one of the conventional approaches in system identification (where the model fit is evaluated over data that is not used for estimation), a certain percentage of data is assigned for estimation, followed by validation. Following this, the participant data was split into two parts:

- *Estimation Data*: In our analysis, this corresponds to the first (70%) of the time-series data
- *Validation Data*: The remaining (30%) of the time-series data

This was chosen keeping in mind that sufficient data points should be available for estimation to get a good model estimate, and the validation data should contain at least one experimental cycle. Even though the data was assumed to be stationary, it is reasonable to expect that noise and disturbance characteristics will vary over long-duration interventions such as *Just Walk* [16].

3.2 Estimation of Behavioral Models

For the estimation of behavioral models, initially the traditional black-box system identification was used, to estimate ARX models. The ARX model structure was determined using order selection. Since, this is a computationally expensive method, and may have bias and variance errors, as described in previous chapters, a more robust approach, i.e. the regularized identification of dynamic models is used.

For this thesis, four participants (164, 180, 220, 230) were studied. The approach was to start with a basic 3-input model consisting of Goals (u_8), Expected Points (u_9), and Granted Points (u_{10}), and then add additional measured inputs (Predicted Busyness, Predicted Stress, Predicted Typical, Weekday-Weekend, and Temperature) to this basic model [16]. All possible 3, 4 and 5 input combinations of these inputs were estimated, using different model estimation methods (ARX and regularized identification methods), which are described in the following sections .

Model validation following estimation, plays an important role in determining which of these inputs are most important in describing individual behavior [16], which in turn contributes towards the design of personalized interventions for an individual. This has been discussed in the following chapter. Model validation involves comparing the identified model output with the measured output to get a measure of the goodness of the fit between the simulated response and the measurement data [5]. To quantify model fits, the normalized root mean square error (NRMSE) fit index is used

$$\text{model fit } (\%) = 100 \times \left(1 - \frac{\|y(k) - \hat{y}(k)\|_2}{\|y(k) - \bar{y}\|_2}\right) \tag{3.4}$$

$y(k)$ is the measured output, $\hat{y}(k)$ is the simulated output, $\bar{y}$ is the mean of all measured $y(k)$ values, and $\|\cdot\|_2$ indicates a vector l_2-norm ($\|x\|_2 \overset{\text{def}}{=} \sqrt{x^T x}$) [16, 28].

3.2.1 Black-box System Identification for PA Interventions

In this method, pre-processed data are fitted to an ARX model structure, ARX-$[n_a, n_{b_1}, \ldots, n_{b_{n_u}}, n_{k_1}, \ldots, n_{k_{n_u}}]$ which can be expressed as:

$$y(t) + \sum_{l=1}^{n_a} a_t y(t-l) = \sum_{j=1}^{n_u} \sum_{i=0}^{n_{b_j}-1} b_{(i+1)(j)} u_j(t - n_{k_j} - i) + e(t) \tag{3.5}$$

where $y(t)$ is the measured output (e.g., steps/day), $u_j(t)$ is the measured input j, $e(t)$ is the prediction error, all measured/estimated on day t. The ARX model in (3.5) is estimated by using linear least-squares regression, which has attractive statistical properties such as consistency [16].

Table 3.2: Order selection results for all combination of 3, 4 and 5 inputs (all combinations have the basic 3 input combination: G-EP-GP) for participants 164, 180, 222 and 230

Input	Participant 164			Participant 180			Participant 222			Participant 230		
Combinations	n_a	n_b	n_k	n_a	n_b	n_k	n_a	n_b	n_k	n_a	n_b	n_k
G-EP-GP	3	[3, 2, 8]	[1, 1, 0]	4	[2, 3, 4]	[1, 1, 0]	1	[1, 1, 6]	[1, 0, 0]	5	[2, 4, 6]	[0, 1, 0]
G-EP-GP-PB	2	[1, 6, 1, 7]	[1, 1, 0, 1]	4	[3, 3, 3, 7]	[0, 1, 0, 0]	1	[1, 1, 1, 1]	[1, 0, 0, 0]	3	[2, 4, 6, 7]	[0, 1, 0, 0]
G-EP-GP-PS	1	[2, 2, 7, 1]	[1, 0, 0, 1]	4	[2, 1, 1, 1]	[1, 1, 0, 0]	6	[2, 2, 1, 7]	[1, 1, 0, 1]	6	[1, 6, 4, 2]	[1, 0, 0, 1]
G-EP-GP-PT	5	[1, 1, 5, 8]	[1, 0, 0, 0]	5	[2, 1, 1, 5]	[1, 1, 0, 0]	1	[1, 1, 1, 5]	[1, 0, 0, 0]	5	[1, 6, 1, 3]	[1, 0, 0, 0]
G-EP-GP-W	4	[1, 1, 7, 1]	[1, 0, 0, 1]	4	[2, 3, 4, 2]	[1, 1, 0, 0]	1	[1, 1, 6, 1]	[1, 0, 0, 0]	5	[1, 4, 6, 4]	[1, 1, 0, 1]
G-EP-GP-T	3	[2, 1, 8, 3]	[1, 1, 0, 0]	5	[7, 1, 2, 3]	[1, 1, 0, 0]	1	[2, 1, 1, 1]	[0, 0, 0, 0]	4	[6, 4, 5, 4]	[0, 0, 0, 0]
G-EP-GP-PB-PS	4	[5, 2, 4, 3, 1]	[1, 1, 0, 0, 0]	5	[7, 1, 3, 6, 1]	[0, 1, 0, 0, 1]	7	[2, 2, 1, 2, 7]	[1, 1, 0, 1, 1]	3	[2, 7, 1, 7, 1]	[0, 1, 0, 0, 0]
G-EP-GP-PB-PT	1	[2, 6, 1, 7, 7]	[1, 1, 0, 1, 1]	7	[4, 1, 4, 6, 3]	[0, 1, 0, 1, 0]	1	[2, 1, 1, 1, 4]	[0, 0, 0, 0, 0]	3	[1, 7, 1, 7, 3]	[1, 0, 0, 0, 0]
G-EP-GP-PB-W	1	[1, 6, 1, 6, 1]	[1, 1, 0, 1, 1]	7	[2, 2, 1, 6, 1]	[1, 1, 0, 0, 0]	3	[1, 1, 6, 1, 4]	[1, 0, 0, 0, 0]	5	[6, 2, 6, 7, 1]	[1, 0, 0, 0, 1]
G-EP-GP-PB-T	3	[2, 1, 5, 1, 3]	[1, 1, 0, 1, 0]	5	[7, 2, 2, 5, 2]	[1, 1, 0, 1, 0]	1	[2, 1, 1, 1, 3]	[0, 0, 0, 1, 0]	4	[5, 3, 5, 1, 4]	[1, 0, 0, 1, 0]
G-EP-GP-PS-PT	4	[5, 1, 4, 2, 1]	[1, 1, 0, 1, 0]	1	[6, 3, 1, 1, 6]	[1, 1, 0, 0, 0]	1	[1, 1, 1, 1, 6]	[1, 0, 0, 0, 0]	5	[1, 7, 1, 1, 3]	[1, 0, 0, 1, 0]
G-EP-GP-PS-W	2	[5, 1, 5, 2, 2]	[1, 1, 0, 1, 0]	4	[2, 3, 4, 2, 3]	[1, 1, 0, 0, 1]	1	[1, 1, 5, 1, 5]	[1, 0, 0, 0, 1]	6	[7, 7, 1, 4, 3]	[0, 0, 0, 0, 1]
G-EP-GP-PS-T	2	[2, 2, 1, 1, 5]	[1, 1, 0, 1, 0]	5	[7, 1, 2, 1, 3]	[1, 1, 0, 0, 0]	1	[2, 1, 1, 1, 1]	[0, 0, 0, 0, 0]	5	[5, 3, 5, 3, 3]	[1, 0, 0, 1, 0]
G-EP-GP-PT-W	3	[3, 1, 1, 6, 2]	[1, 0, 0, 1, 1]	5	[2, 4, 4, 4, 3]	[1, 1, 0, 0, 1]	3	[1, 1, 5, 5, 3]	[1, 0, 0, 1, 0]	5	[1, 7, 1, 3, 1]	[1, 0, 0, 0, 0]
G-EP-GP-PT-T	3	[2, 1, 3, 2, 3]	[1, 1, 0, 0, 0]	3	[7, 1, 1, 5, 2]	[0, 1, 0, 0, 1]	1	[2, 1, 1, 4, 1]	[0, 0, 0, 1, 0]	4	[1, 6, 7, 3, 4]	[1, 1, 0, 0, 1]
G-EP-GP-W-T	1	[3, 1, 1, 2, 3]	[1, 1, 0, 1, 0]	1	[5, 3, 4, 2, 7]	[1, 1, 0, 1, 0]	1	[1, 1, 5, 5, 1]	[1, 0, 0, 1, 0]	5	[6, 4, 6, 6, 5]	[1, 0, 0, 0, 0]

3.2.1.1 Order Selection

The ARX model orders n_a, n_b and n_k are determined using **Order Selection** functionality in the MATLAB System Identification toolbox. This is done by exhaustively examining a range of model orders, and using model validation procedures to determine the most suitable structure, a stable ARX model with highest predictive ability. This step, though computationally expensive, provides a safeguard against overparameterization. The order selection results for all four participants in three inputs and all combination of four and five inputs have been listed down in Table 3.2.

3.2.2 Regularized Estimates of Model Parameters

Model parameters are obtained by fitting measured data to the predicted model response, where the the measure of the model flexibility is directly proportional to its order. Even though increasing the model order causes the model to fit the observed data with increasing accuracy, it comes with the price of higher uncertainty in the estimates, measured by a higher value of random or variance error. On the other hand, choosing a model with too low an order leads to larger systematic errors, known as bias error. Hence, to avoid such errors that may arise as a result of improper choice of model order, regularization is used, which gives a better control over the bias versus variance trade-off by introducing an additional term in the minimization criterion that penalizes the model flexibility [8].

In this thesis, regularized models have been estimated, where regularization has been implemented on models with $n_a = 0$ (which makes it FIR), $n_b = 10$ (for 3 and 4 input cases), $n_b = 9$ (for 5 input models), and $n_k = 0$. For this thesis two types of regularization methods have been used, ridge regression and kernel-based methods, which have been described in previous chapters.

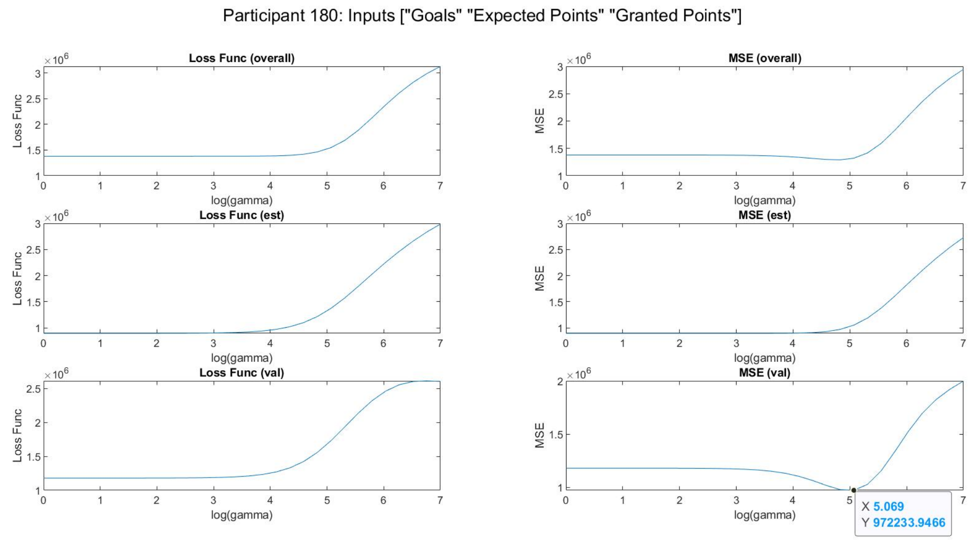
3.2.2.1 Determine Regularization Constant γ for Ridge Regression

For ridge regression, the regularization parameter γ, as seen in (2.33), is usually determined by following the steps, as described below [8]:

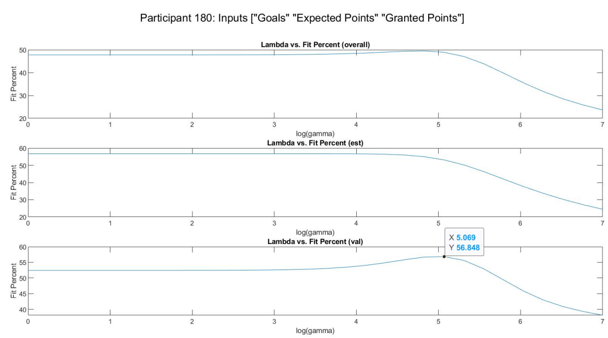
- Create a list of γ value that need to be tested, which is done by generating 30 logarithmically spaced values of γ in the range of 10 to 10^7.
- Estimate a regularized model for each value of γ in the set that minimizes the objective function in (2.33).
- Compute MSE and Loss Function, parameter 2-norm, and % fit for each estimated model with respect to the validation, overall, and estimation data. These are then plotted against $\log \gamma$ in order to observe how these parameters vary with change in γ

Here, the loss function, or cost function, is a positive function of prediction errors (weighted sum of squares of the errors) [6].

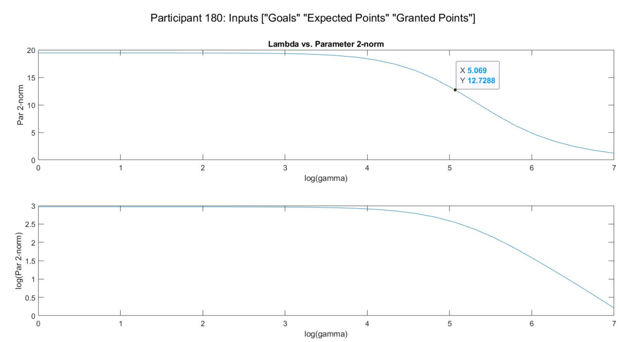
The procedure for choosing the regularization constant γ has been demonstrated for participant 180 (for the basic 3-input case) in Fig 3.2. From the plots, the value of γ for which the MSE and loss function (Fig. 3.2a) are seen to have a minimum, is the one that is selected. For this value of the regularization constant, the estimated regularized model will have a good fit to the measured data, which can be observed from the plot of % Fit against $\log \gamma$ (Fig. 3.2b). Also, the parameter 2-norm (Fig. 3.2c) should be low.



(a) Plot of MSE and Loss Function versus $\log \gamma$

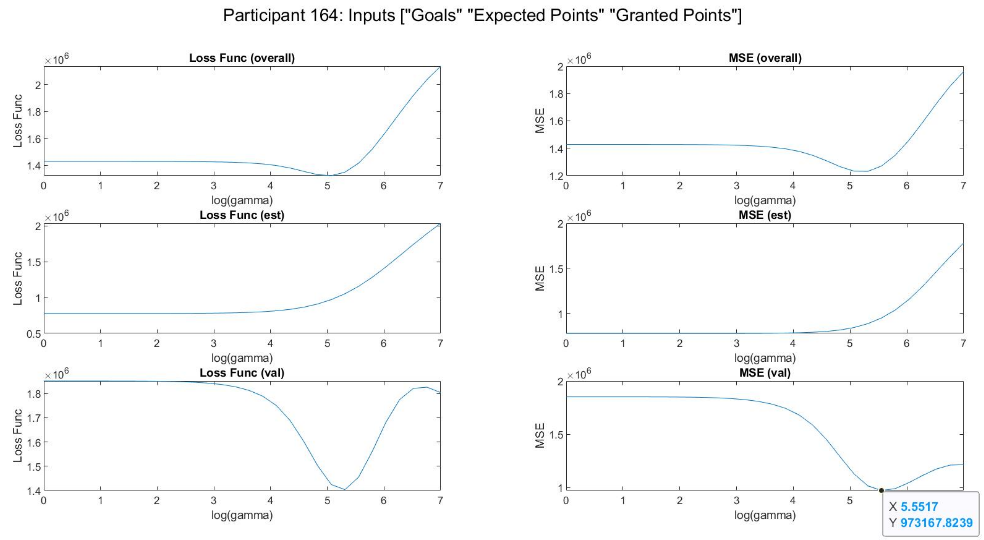


(b) Plot of % Fit versus $\log \gamma$

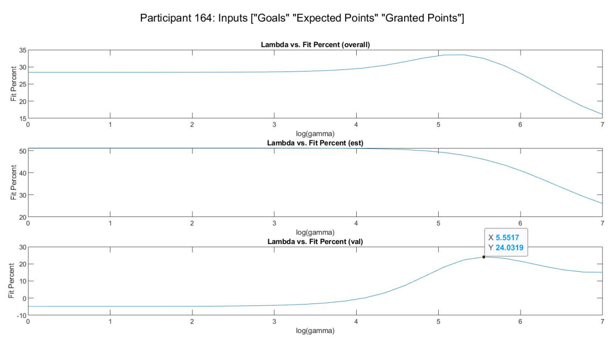


(c) Plot of parameter 2-norm and its log value versus $\log \gamma$

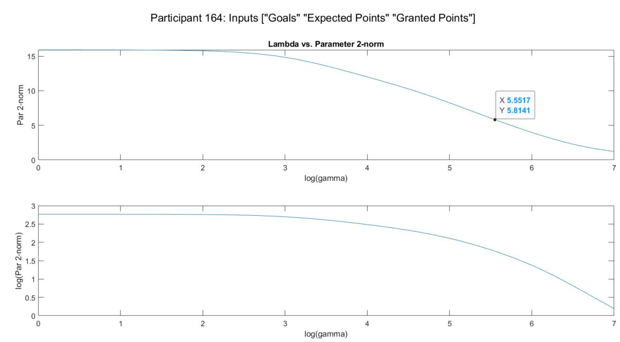
Figure 3.2: Plots for determining regularization constant γ for participant 180



(a) Plot of MSE and Loss Function versus $\log \gamma$

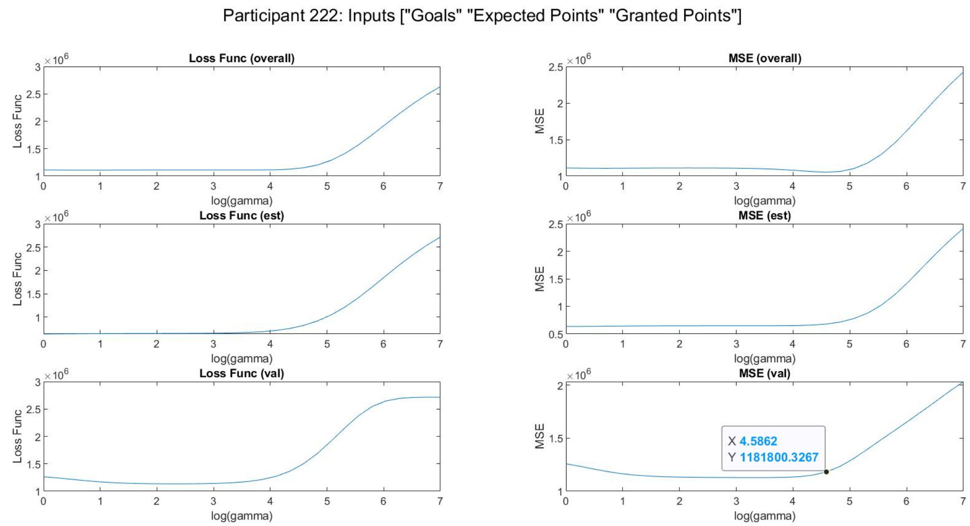


(b) Plot of % Fit versus $\log \gamma$

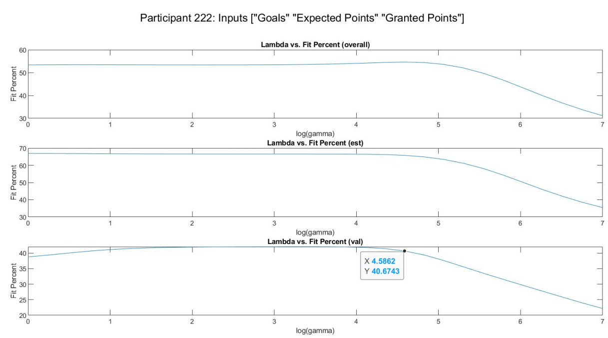


(c) Plot of parameter 2-norm and its log value versus $\log \gamma$

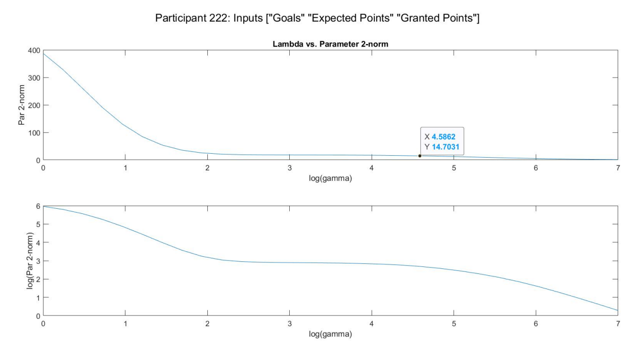
Figure 3.3: Plots for determining regularization constant γ for participant 164



(a) Plot of MSE and Loss Function versus $\log \gamma$

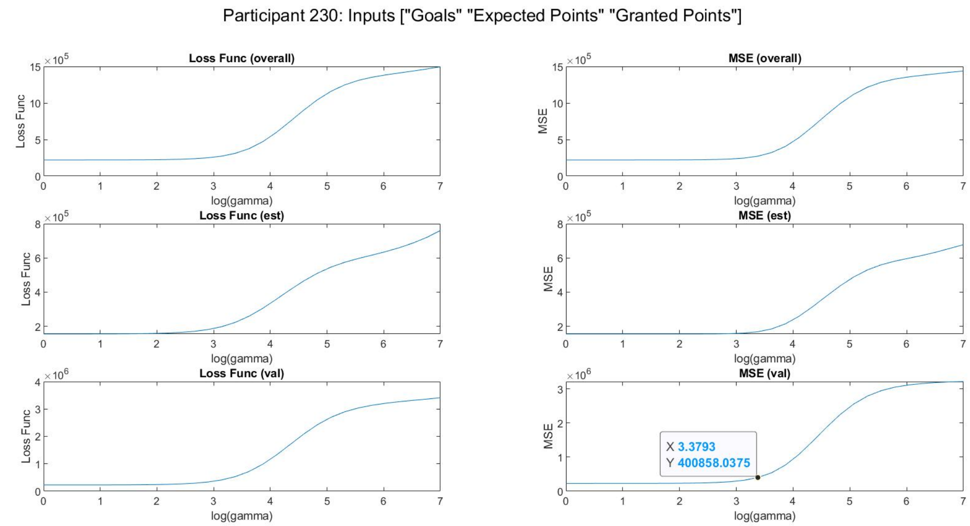


(b) Plot of % Fit versus $\log \gamma$

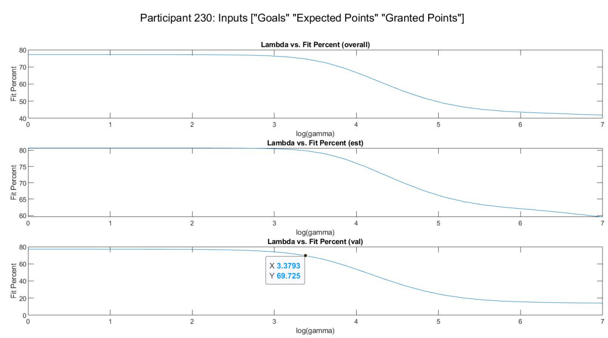


(c) Plot of parameter 2-norm and its log value versus $\log \gamma$

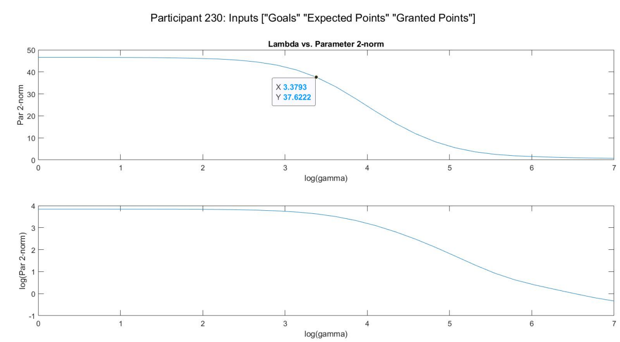
Figure 3.4: Plots for determining regularization constant γ for participant 222



(a) Plot of MSE and Loss Function versus $\log \gamma$



(b) Plot of % Fit versus $\log \gamma$



(c) Plot of parameter 2-norm and its log value versus $\log \gamma$

Figure 3.5: Plots for determining regularization constant γ for participant 230

For this particular case, as seen from the plots, the value for $\log_{10}\gamma = 5.069$. Hence the regularization constant $\gamma = 1.1722 \times 10^5$. For participant 180, it is seen that when Temperature is an input, the optimum value for the regularization constant γ changes to a different value 67235.7. A similar approach approach has been used for determining the regularization constants for all the four participants that have been studied, for all input combinations. This has been shown in Table 3.3.

Table 3.3: Values for regularization constant (γ) for all participants

Participant	**log(Lambda)**	**Lambda**
164	5.5517	3.56205×10^5
180	5.069	1.1722×10^5
180 (when T is added input)	4.8276	67235.7
222	4.5862	38565.6
230	3.3793	2394.97

3.2.2.2 Kernel-based Methods

For the kernel-based methods, the MATLAB function "arxRegul" was used, which automatically determined the regularization constants for the model estimation. The regularization options were configured to specify a regularization kernel. The values obtained for the regularization constant (that determines the bias-variance trade-off), and the weighting matrix (which is a positive definite matrix), are then used for estimating the regularized ARX models. This process is repeated using the kernels 'TC', 'SE', 'SS', 'HF', 'DI' and 'DC', all of which have been described in Chapter 2. The models were estimated and they were analyzed in detail to obtain results and conclusions described in the following section 3.3.

3.3 Results and Conclusions

For all four participants, the following models were estimated, for different input combinations:

1. Black-box ARX model (using order selection)

2. Regularized FIR models (using ridge regression and kernel-based methods)

Initially, a basic 3-input (inputs: Goals, Expected Points, and Granted Points) model, was estimated using the above mentioned methods. To get an insight into how each estimated basic model performs, the step responses are observed, and a *residual analysis* was run on each model. Residuals are differences between the one-step-predicted output from the model and the measured output from the validation data set. Hence, residuals represent the portion of the validation data that is not explained by the model. Residual analysis consists of two tests [7]:

- *Whiteness test*: This checks if the model has the residual autocorrelation function inside the confidence interval of the corresponding estimates, which indicates that the residuals are uncorrelated, and helps in assessing if the estimated model is good.

- *Independence test*: This is to see if the model residuals are uncorrelated with past inputs. If correlation exists, it indicates that the model does not describe how that part of the output relates to the corresponding input. For example, a peak outside the confidence interval for lag k means that the output $y(t)$ that originates from the input $u(t-k)$ is not properly described by the model.

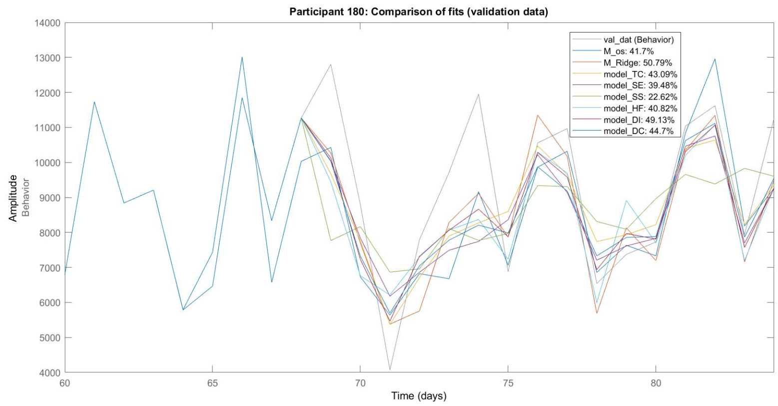
After analysing the basic 3-input model, additional inputs are added to get a set of all possible combinations of 4 and 5 inputs, which include the basic 3 inputs as well.

Model estimation is performed for these, in a similar way to the 3-input case. The models with the best results, in terms of the NRMSE fits are then shortlisted, and from these results the model estimation methods which best work for that particular participant is decided upon. High order FIR ($n_b = 10$) models are estimated for the 3-input case for each participant, and compared with the regularized models to analyze if regularization gives better results. After examining all four participants, a conclusion can be obtained based on which model estimation method might work best for all cases.

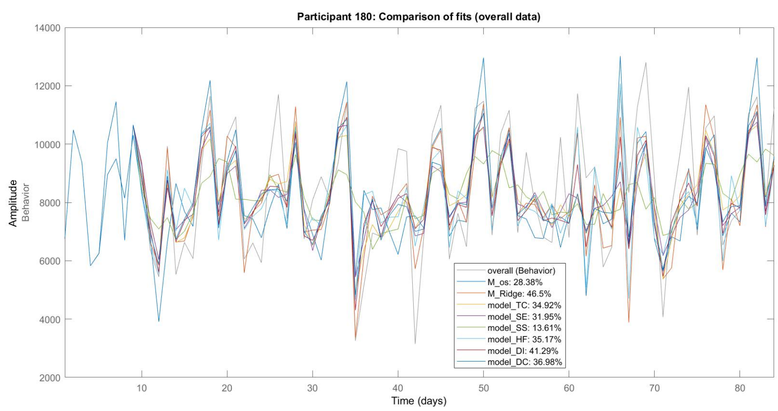
3.3.1 Participant 180

The models were estimated for Participant 180, and the % Fit with respect to the validation, estimation and overall data were computed. These have been tabulated along with the MSE, and parameter 2-norm in Table 3.4a. The blue, orange, and green columns contain the value for the % Fit, when the output of the estimated model is compared with the estimation, validation, and the overall data respectively. The boxes with the best % Fit values for each category have been highlighted using a darker color, from which it can be seen that the regularized ARX models using ridge regression, and kernel-based methods - DI and DC, give a pretty good estimate for the 3-input behavioral model for participant 180.

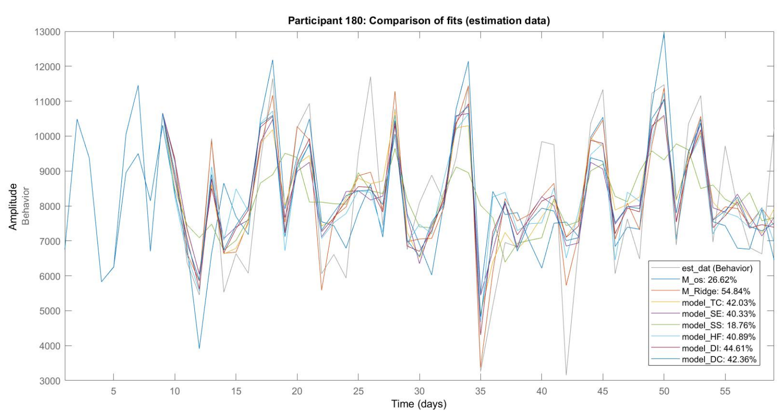
The % fit comparisons can be seen from the plots in Fig 3.6. From Table 3.4a, and the plots, it can be observed that some of the regularized models perform better than the ARX model estimated using order selection. It is also seen that for the high order FIR model estimated, the parameter 2-norm is very high, which can be an indication of a higher variance, the distance between the absolute values of the parameter coefficients being higher. The regularized methods, using the penalty term, minimize the parameter coefficient estimates, and minimize the 2-norm value. The worst of all



(a) Comparison of % fit with respect to validation data



(b) Comparison of % fit with respect to overall data



(c) Comparison of % fit with respect to estimation data

Figure 3.6: Compare fit percentage of estimated models, participant 180

the models is the one estimated using the SS Kernel, which a comparatively lower fit when compared to the remaining estimated models. To get a better idea of the estimated models for this participant, the step responses are plotted, as shown in Fig 3.7. From the step responses of participant 180, the following can be inferred:

- For the unit step goal, looking at the responses it is seen that the step response for order selection is abrupt, whereas the regularized methods give a more gradual response.

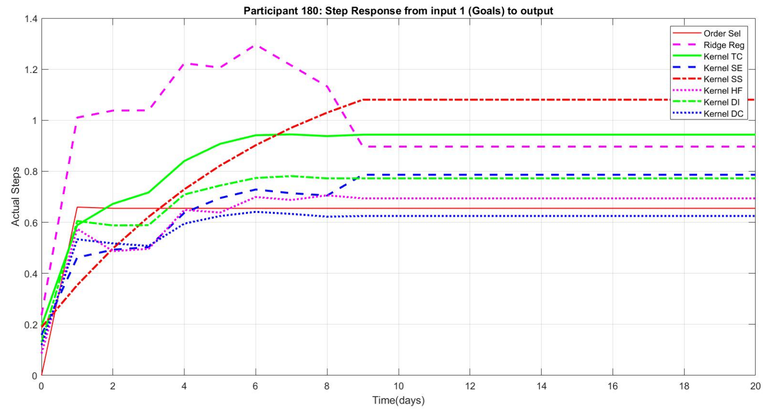
- When the step input is a unit positive change in expected points, order selection shows no change to the input, whereas the responses for all regularized models show a change in the negative direction.

- For the third step response, where the step input is a unit change in granted points, most of the regularized approaches give a much smoother step response, whereas, for order selection spikes can be seen.

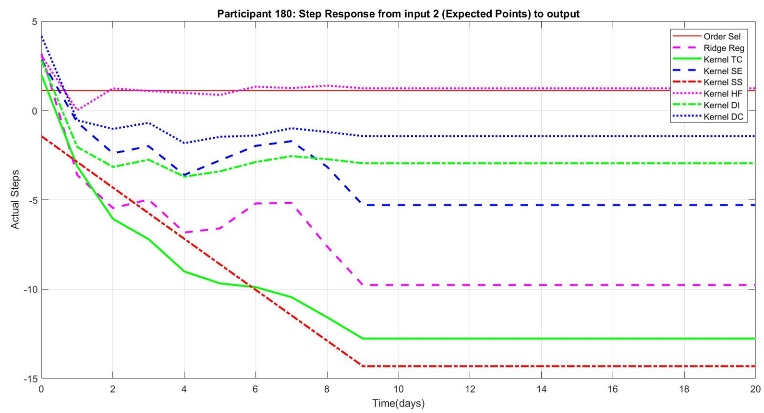
- The above point is true for all 3 responses, for some of the regularization approaches.

A residual analysis was run for all the estimated 3-input models, for participant 180. The autocorrelation plot of the residuals, as well as the cross correlation of the inputs with the residuals was within the 99% confidence bound, which proved that the all the estimated models were good. The residual analysis results have been shown in Fig. 3.8.

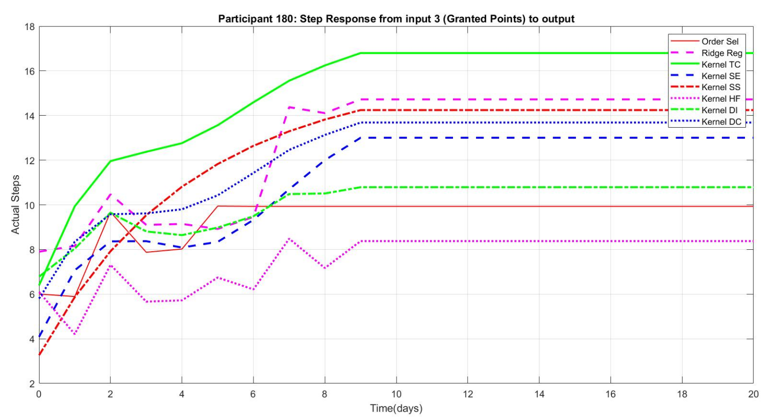
Next, as described previously, the % fit results are obtained for models estimated for all possible 4 and 5 input combinations. The results have been tabulated in Table 3.4b and Table 3.4c. The models having a good fit for a particular input combination have been highlighted in the table. Looking at Tables 3.4a, 3.4b and



(a) Goals to Behavior



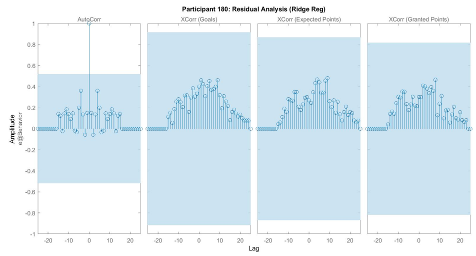
(b) Expected Points to Behavior



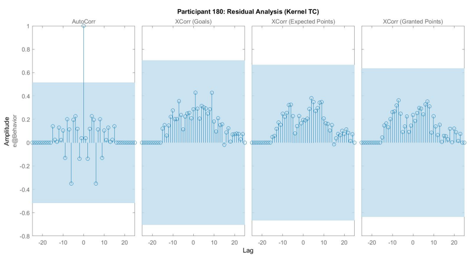
(c) Granted Points to Behavior

Figure 3.7: Step response for participant 180 for all 3-input models

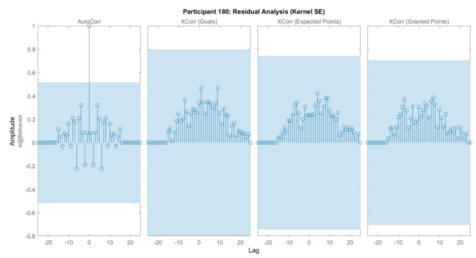
(a) Order Selection



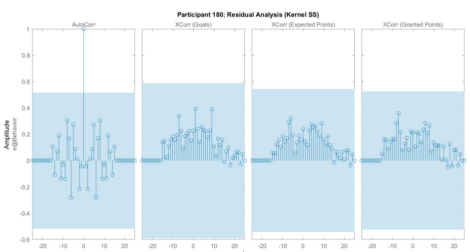
(b) ridge regression



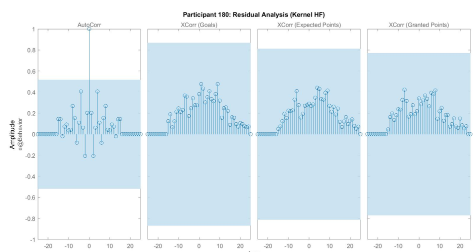
(c) Kernel TC



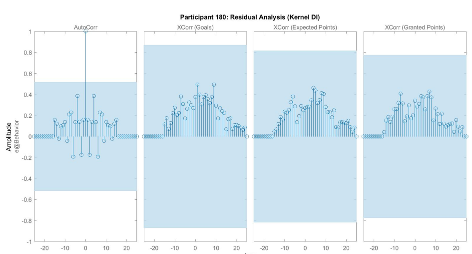
(d) Kernel SE



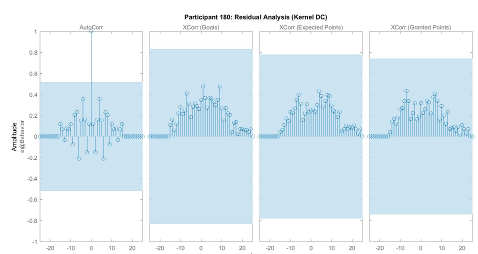
(e) Kernel SS



(f) Kernel HF



(g) Kernel DI



(h) Kernel DC

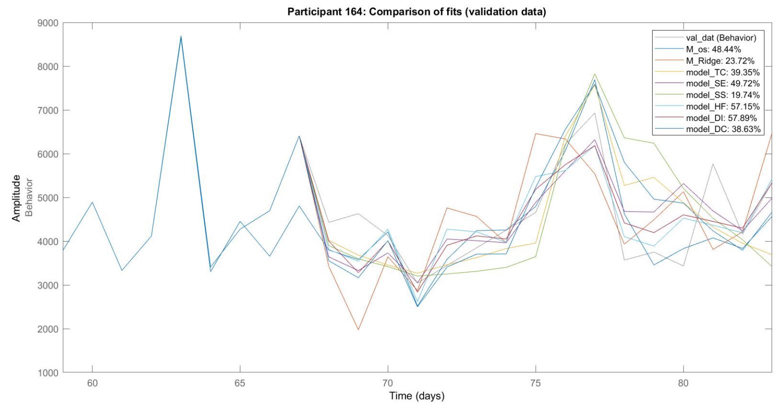
Figure 3.8: Residual analysis for estimated models for participant 180

3.4c, it can be seen that regularized models estimated using “ridge regression” and “kernel DI” always performs better than the other estimation methods that have been implemented for participant 180. Even though a slightly higher bias and variance is observed for ridge regression in general from the step responses, it can be considered as a good method for model estimation, due to its simplicity, and considerably high fit to the measured data. The model estimated using kernel DI performs better in terms of step responses, as can be seen from the plots in Fig 3.7. This can be primarily because the penalty term is made more effective in kernel-based methods, by using a positive definite matrix, which allows for the weighting and/or rotation of the parameter vector [8]. A similar approach has been used for the analysis of the remaining four participants, 164, 222 and 230, the results for which have been explained in the following sections.

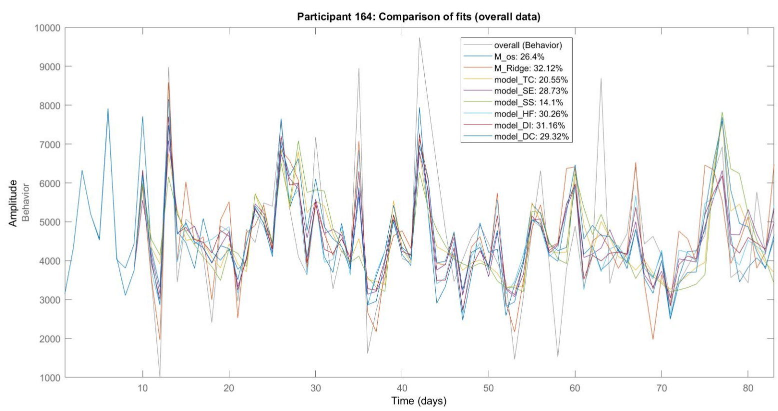
3.3.2 Participants 164, 222, 230

The results indicating the quality of the estimated models (in terms of fit percentages) for participants 164, 222 and 230, for all input combinations have been documented in Tables 3.5, 3.6 and 3.7 respectively. The blocks highlighted in dark colors help in locating the model estimation method, and the corresponding input combinations used, that give a reasonably good fit with respect to the estimation, validation, and the overall data. This selection was basically done by first shortlisting the best validation fits. From the previous subset, the ones with reasonable overall fits were then selected.

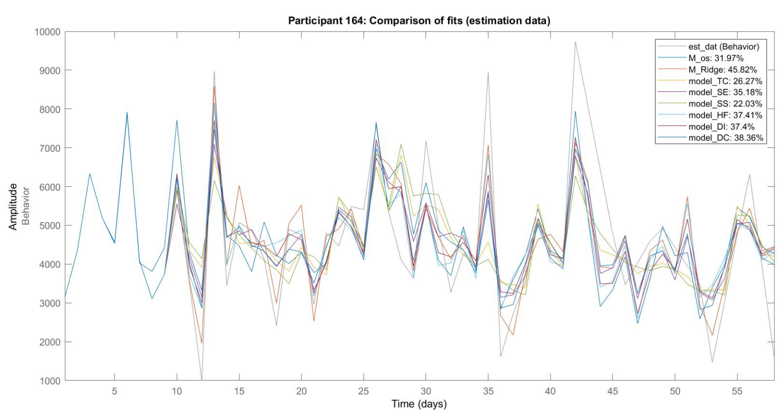
For participant 164, from the Table 3.5 it can be seen that for most input combinations, the order selection and kernel-based regularization methods HF and DI perform well. In most cases the regularization methods perform better than the order selection. The high order FIR model fits very well to the estimation data, but



(a) Comparison of % fit with respect to validation data



(b) Comparison of % fit with respect to overall data



(c) Comparison of % fit with respect to estimation data

Figure 3.9: Compare fit percentage of estimated models, participant 164

looking at the poor fit to the validation data, it can be seen that this model does not generalize well to new data. Also, the parameter 2-norm is very high. Both these factors indicates the presence of a high variance, which is taken care of by the regularized models.

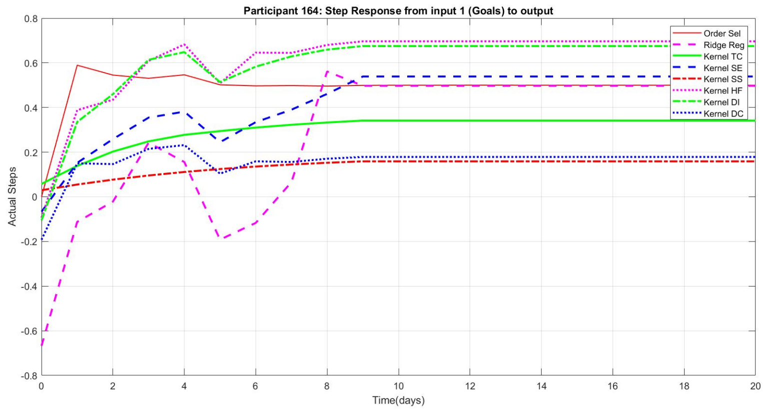
From the step responses for participant 164, the following observations are made:

- For a unit step change in step goals, for order selection, the step response rises sharply, and then gradually decreases to settle at a value of about 0.5. Regularized models estimated using kernels DI and HF perform well, with a gradual increase in step response, which settles at a value of about 0.7, that is closer to unity.

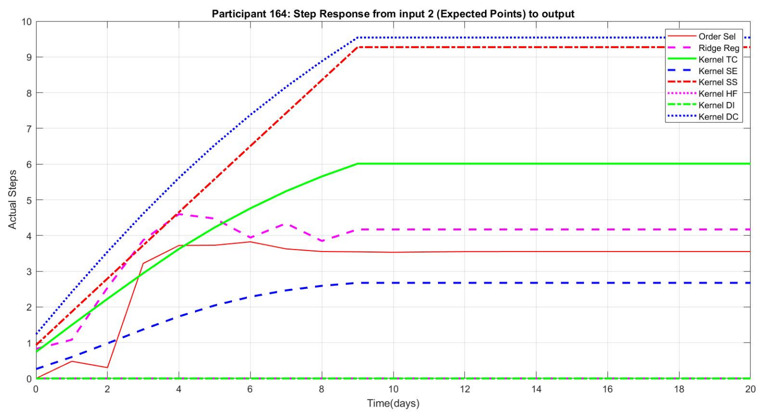
- For a unit change in expected points, this participant seems to behave differently. Increasing reward targets encourages the participant to engage in PA. For kernels DI and HF, the response of the models show that changing reward targets has little, or no effect on this participant. Whereas, the order selection, as well as the other regularization methods, indicate a significant positive change in behavior.

- For a unit step granted point, the regularized models using kernels DI, DC, SE, TC and SS give a gradually increasing smooth step response, that shows a slight overshoot, after which it settles. For the order selection a sharp peak is seen, followed by a slight undershoot, before settling down.

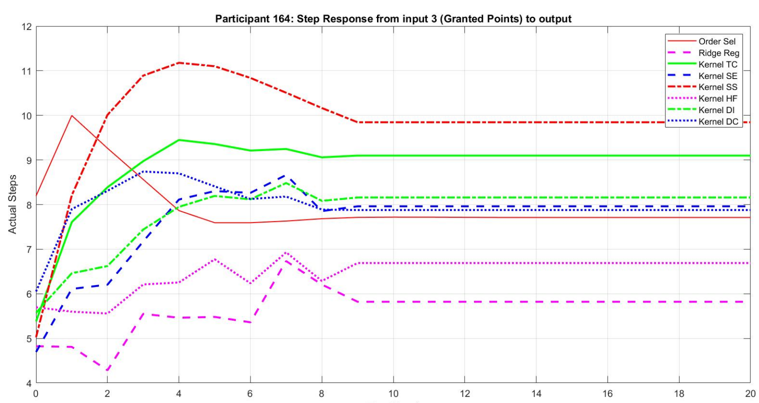
In a similar way, on examining the results for participant 222 in Table 3.6, it can be seen that for almost all input combinations, regularized models estimated using kernels HF and DI give very good fit percentages. Kernel TC gives good results in many of the cases as well. For this participant, regularized models perform better than the ones estimated using order selection in general. For the high order FIR



(a) Goals to Behavior

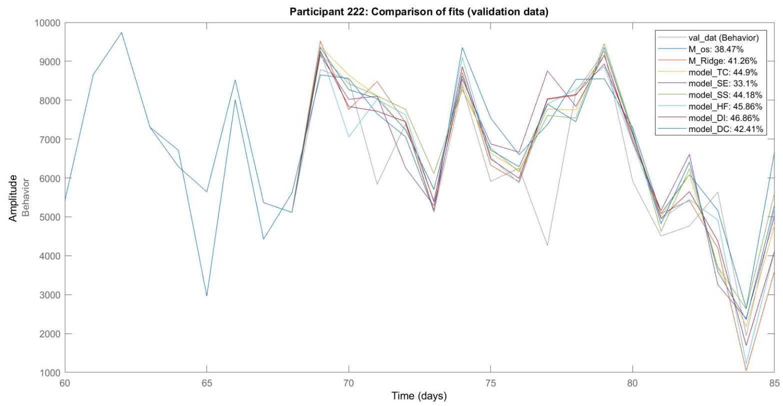


(b) Expected Points to Behavior

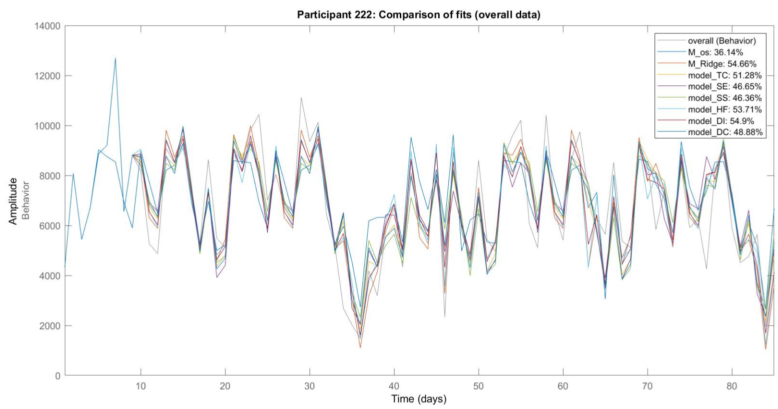


(c) Granted Points to Behavior

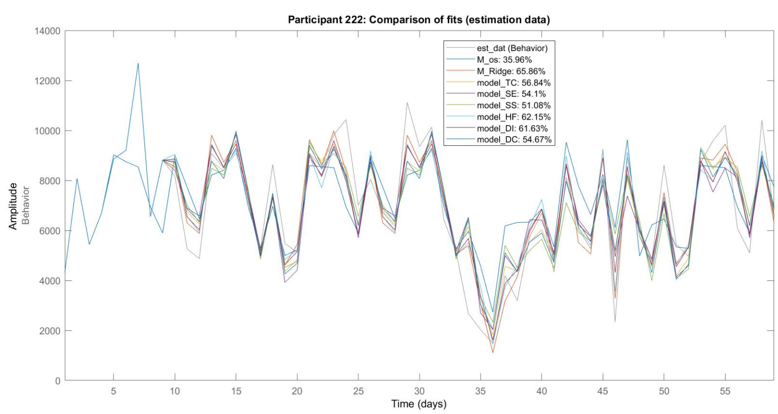
Figure 3.10: Step response for participant 164 for all 3-input models



(a) Comparison of % fit with respect to validation data



(b) Comparison of % fit with respect to overall data



(c) Comparison of % fit with respect to estimation data

Figure 3.11: Compare fit percentage of estimated models, participant 222

model, even though the fits are quite good, the parameter 2-norm value is extremely high. This is significantly reduced using the regularization approaches.

For participant 222, on observing the step responses, the following conclusions can be made:

- For a unit positive step goal, for order selection, the response increases, and then stabilizes at a value of 0.8. For kernels DI, DC and TC there is a slight overshoot, followed by an undershoot, after which it stabilizes.

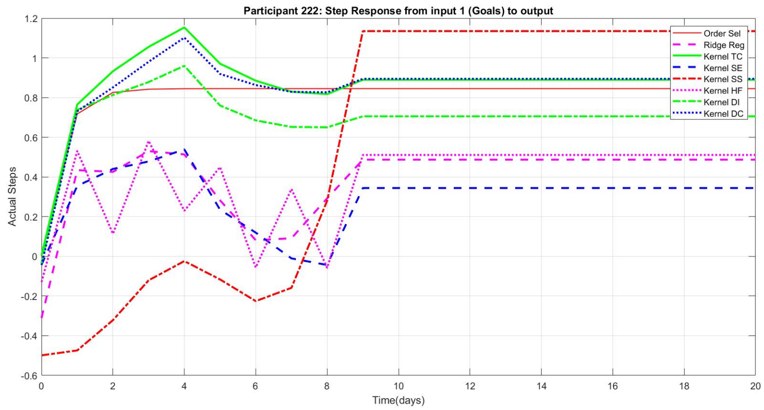
- For a unit positive step increase in expected points, order selection and kernel SE give constant response, which indicated that this input has almost zero influence on participant. The other regularized models show a decrease in response.

- For unit step granted points, the regularized models give much smoother response with higher gains. Whereas the model obtained using order selection gives a slightly spiked response, which settles at a lower value.

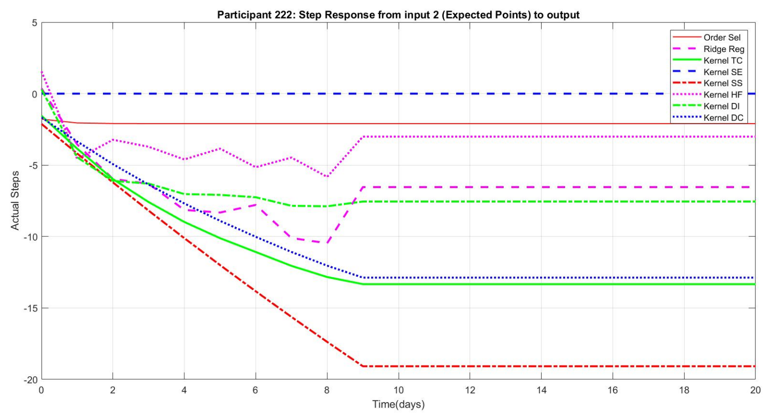
For participant 230, referring to the % fit results in Table 3.7, it is evident that models estimated using order selection and ridge regression show the best performance, and for most of the cases, except for one 5-input case where the additional inputs are predicted stress and work, the kernel DI based regularized models performs well too. For this participant, the high order FIR does well in terms of fit percentages. The parameter 2-norm value is high, close to the one for order selection. The regularized models bring it down from the range of 50 to 30.

Examining the step responses, the following observations are made:

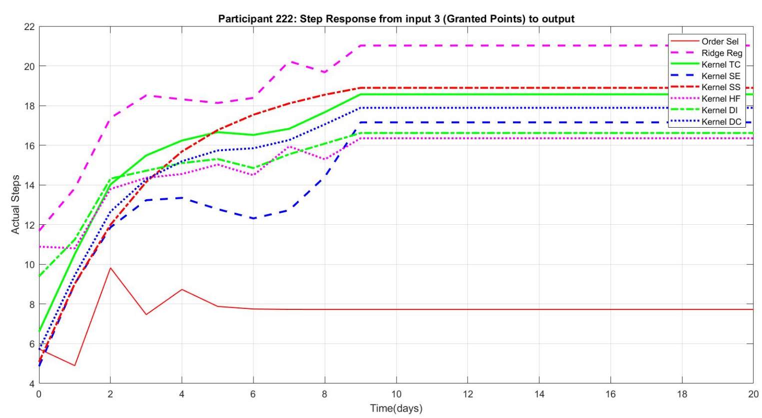
- For a unit step increase in goals, for order selection, and regularized methods using kernels TC, DI and DC show a positive change, and the responses are very similar.



(a) Goals to Behavior

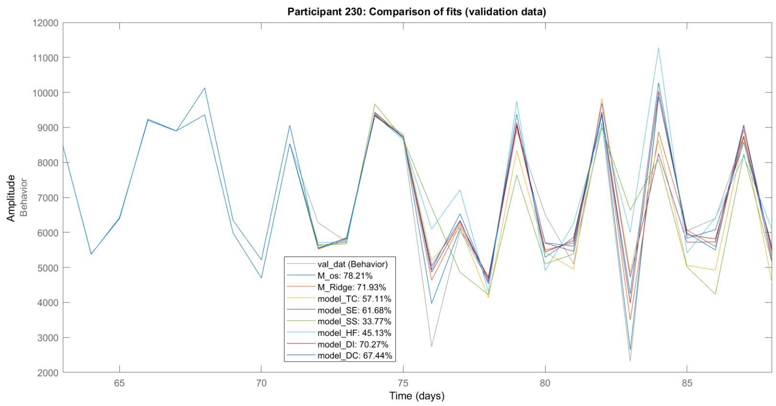


(b) Expected Points to Behavior

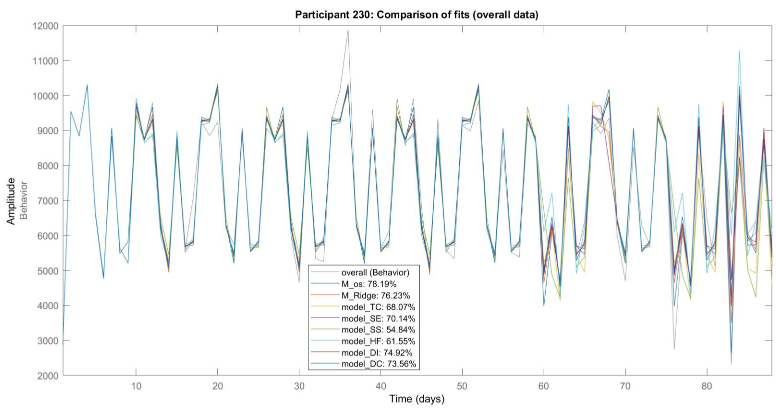


(c) Granted Points to Behavior

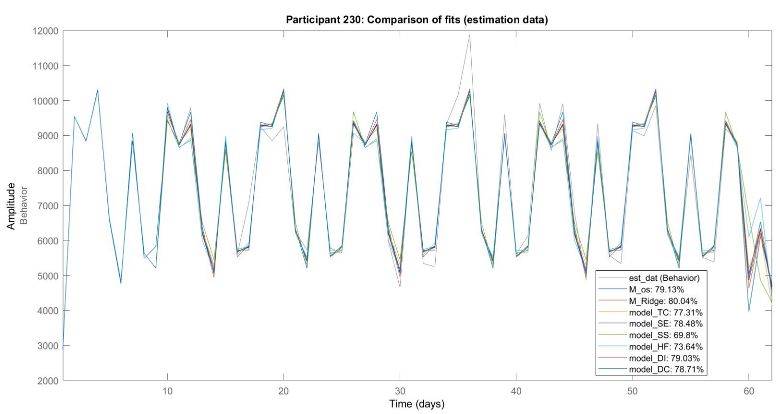
Figure 3.12: Step response for participant 222 for all 3-input models



(a) Comparison of % fit with respect to validation data

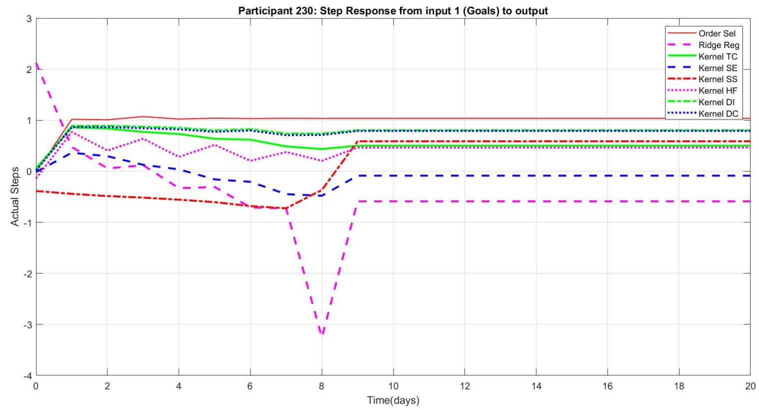


(b) Comparison of % fit with respect to overall data

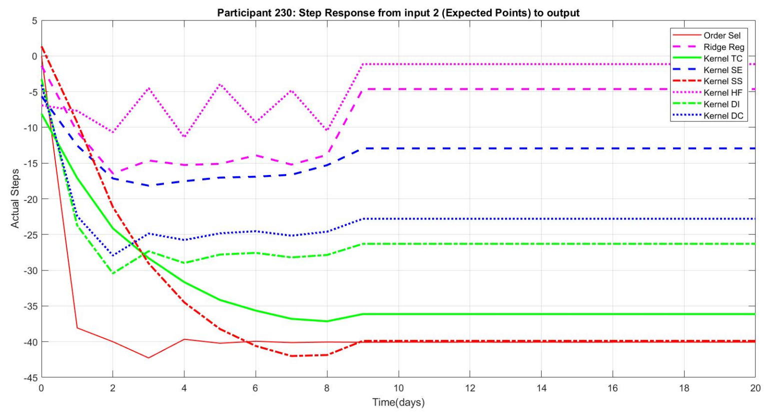


(c) Comparison of % fit with respect to estimation data

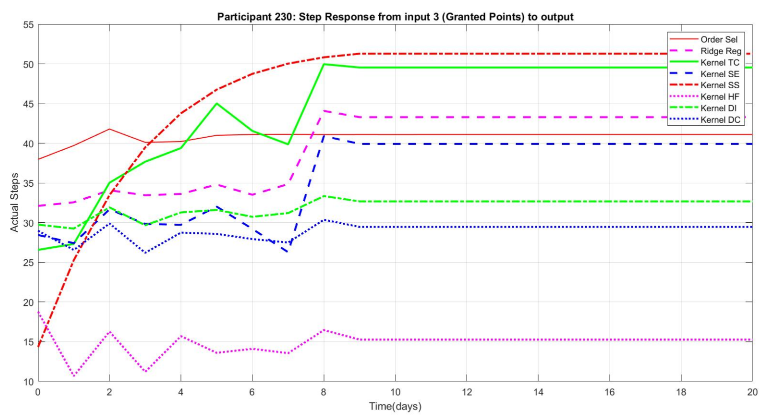
Figure 3.13: Compare fit percentage of estimated models, participant 230



(a) Goals to Behavior



(b) Expected Points to Behavior



(c) Granted Points to Behavior

Figure 3.14: Step response for participant 230 for all 3-input models

- For a unit positive change in reward targets, all models give negative step response, with order selection and kernel SS having a maximum negative settling value of -40.

- For the step response from granted points to behavior for this participant, models estimated using order selection, kernels DI and DC, and ridge regression show a near constant response, differing in the final settling value. The other regularized models show a positive response.

3.3.3 Conclusion

After extensively analyzing the results for all four participants, it can be concluded that models estimated using Kernel DI perform consistently well throughout. Hence, this could be considered as a reliable method for estimating regularized models for other participants who were a part of this study. The one that did not perform well in for all four participants was the SS kernel. These have not been tried out yet on the other 18 participants that were a part of the study in order to confirm it.

Another interesting observation is that the results vary significantly in many cases, depending on the combination of inputs used for model estimation. This gives an insight into how addition of certain input/inputs to the basic 3-input model improves the % fit of the estimated model significantly. Examining results from the perspective of % fit change when certain inputs are added can help in determining what factors play an important role in deciding the PA of a person. This in turn can contribute significantly towards designing personalized interventions. Chapter 4 elaborates more on this aspect.

Table 3.4: Analysis results and % fit to estimation (E), validation (V) and overall (O) data for estimated models for participant 180

(a) 3-input (Analysis Results)

Participant 180: 3-input (G,EP,GP)					
Model	Fit_est	Fit_val	Fit_ovr	MSE	par_2norm
Order Sel	26.62	41.70	28.38	2,580,499.21	18.39
High Order	59.44	45.92	45.89	786,784.70	22.43
Ridge Reg	54.84	50.79	46.50	975,648.47	13.10
Kernel TC	42.03	43.09	34.92	1,607,708.69	10.40
Kernel SE	40.33	39.48	31.95	1,703,489.47	8.19
Kernel SS	18.76	22.62	13.61	3,157,825.13	7.01
Kernel HF	40.89	40.82	35.17	1,671,407.92	9.14
Kernel DI	44.61	49.13	41.29	1,467,610.43	9.31
Kernel DC	42.36	44.70	36.98	1,589,479.20	9.31

(b) 4-input (%fit)

Model	G-EP-GP-PB			G-EP-GP-PS			G-EP-GP-PT			G-EP-GP-W			G-EP-GP-T		
	E	V	O	E	V	O	E	V	O	E	V	O	E	V	O
Order Sel	48.89	56.67	43.84	39.96	54.01	40.32	45.81	63.56	44.29	44.14	62.43	47.34	45.53	72.63	48.57
Ridge Reg	54.54	50.72	46.36	54.54	50.72	46.36	54.54	50.72	46.36	54.54	50.72	46.36	57.37	52.52	48.16
Kernel TC	45.86	46.60	38.70	34.39	32.00	25.85	50.10	48.97	41.98	26.35	36.41	24.83	43.83	47.58	38.24
Kernel SE	49.81	38.56	33.14	30.92	34.39	24.68	61.94	36.51	37.23	30.95	38.72	28.65	40.27	39.45	31.91
Kernel SS	20.25	21.31	13.73	15.06	20.83	11.31	34.71	26.89	22.74	19.31	30.50	18.09	18.66	22.63	13.56
Kernel HF	43.76	41.82	37.27	37.14	41.39	33.56	53.04	48.36	44.55	35.77	42.25	33.18	40.86	40.80	35.14
Kernel DI	49.02	51.63	44.01	39.71	43.71	35.99	58.87	54.55	49.27	39.70	45.24	37.43	44.61	49.13	41.29
Kernel DC	48.40	48.68	41.06	35.77	37.57	29.72	59.01	49.43	45.28	38.90	47.46	39.12	42.32	44.65	36.94

(c) 5-input (%fit)

Model	G-EP-GP-PB-PS			G-EP-GP-PB-PT			G-EP-GP-PB-W			G-EP-GP-PB-T			G-EP-GP-PS-PT		
	E	V	O	E	V	O	E	V	O	E	V	O	E	V	O
Order Sel	48.62	61.48	46.54	54.76	63.44	50.94	51.14	71.18	47.69	49.77	79.35	51.54	46.43	71.47	47.08
Ridge Reg	50.58	54.79	45.22	50.85	54.90	45.34	50.85	54.90	45.34	53.16	56.96	46.93	50.85	54.90	45.34
Kernel TC	23.28	30.51	20.92	43.14	47.51	38.72	27.81	38.18	26.78	60.10	44.22	50.92	46.03	40.42	35.23
Kernel SE	34.29	33.64	28.02	48.14	41.95	38.19	36.56	40.07	32.61	76.76	13.26	32.26	48.46	34.02	37.32
Kernel SS	14.17	18.07	10.57	17.02	20.14	12.89	19.18	28.40	17.80	52.22	44.49	46.59	15.67	19.27	11.81
Kernel HF	36.48	41.81	34.01	41.09	45.94	37.60	39.11	43.54	36.74	62.97	49.75	36.61	39.18	44.01	35.92
Kernel DI	39.34	42.29	36.09	49.93	52.31	45.52	40.34	44.72	38.08	71.64	40.16	48.00	49.34	40.26	41.07
Kernel DC	36.17	38.39	30.72	49.09	45.61	41.55	41.32	48.16	40.50	73.20	31.61	44.31	47.06	37.12	36.60

Model	G-EP-GP-PS-W			G-EP-GP-PS-T			G-EP-GP-PT-W			G-EP-GP-PT-T			G-EP-GP-W-T		
	E	V	O	E	V	O	E	V	O	E	V	O	E	V	O
Order Sel	43.42	60.71	46.65	46.23	73.23	48.84	48.65	62.13	47.15	47.81	75.29	49.68	49.58	77.65	53.74
Ridge Reg	50.85	54.90	45.34	53.16	56.96	46.93	50.85	54.90	45.34	53.16	56.96	46.93	53.16	56.96	46.93
Kernel TC	35.29	34.96	30.82	34.89	39.70	30.86	42.93	35.19	34.53	32.51	34.73	27.20	24.30	33.81	23.01
Kernel SE	32.79	39.45	30.43	34.29	34.09	28.72	40.70	41.36	36.50	32.47	23.43	23.78	27.70	34.41	25.39
Kernel SS	18.88	28.12	17.49	14.47	18.60	10.93	18.96	28.21	17.54	14.01	17.93	10.42	18.46	27.41	16.94
Kernel HF	38.74	41.70	36.06	50.57	48.75	43.03	40.33	44.16	37.60	36.50	43.12	34.17	34.26	41.36	32.70
Kernel DI	38.67	44.05	36.76	41.54	43.80	38.08	41.55	45.33	39.15	36.12	42.09	34.06	34.57	41.88	33.19
Kernel DC	38.98	47.91	39.10	38.98	42.21	35.70	43.42	42.84	38.31	37.16	35.41	30.26	30.76	37.75	27.67

Table 3.5: Analysis results and % fit to estimation (E), validation (V) and overall (O) data for estimated models for participant 164

(a) 3-input (Analysis Results)

Participant 164: 3-input (G,EP,GP)					
Model	**Fit_est**	**Fit_val**	**Fit_ovr**	**MSE**	**par_2norm**
Order Sel	31.97	48.44	26.40	1,438,036.33	7.21
High Order	51.17	-5.24	28.36	777,711.54	16.06
Ridge Reg	45.82	23.72	32.12	957,651.20	5.89
Kernel TC	26.27	39.35	20.55	1,773,284.57	6.24
Kernel SE	35.18	49.72	28.73	1,370,393.84	5.26
Kernel SS	22.03	19.74	14.10	1,982,749.94	6.96
Kernel HF	37.41	57.15	30.26	1,277,938.24	5.90
Kernel DI	37.40	57.89	31.16	1,278,354.67	5.78
Kernel DC	38.36	38.63	29.32	1,239,281.34	7.09

(b) 4-input (%fit)

Model	**G-EP-GP-PB**			**G-EP-GP-PS**			**G-EP-GP-PT**			**G-EP-GP-W**			**G-EP-GP-T**		
	E	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**
Order Sel	37.11	61.71	31.18	30.82	39.67	23.46	33.29	37.81	23.90	30.86	43.18	25.01	35.67	64.45	32.14
Ridge Reg	45.82	23.72	32.12	45.82	23.72	32.12	45.82	23.72	32.12	45.82	23.72	32.12	46.01	23.87	32.32
Kernel TC	26.01	39.17	20.33	24.57	36.53	18.81	25.78	27.92	18.40	21.18	41.10	16.92	50.93	-9.65	28.91
Kernel SE	38.80	45.68	30.23	27.61	30.28	20.24	33.76	49.39	27.71	17.18	36.86	13.30	50.55	21.75	33.78
Kernel SS	21.93	19.74	14.02	20.48	25.81	14.01	21.42	20.43	13.92	15.94	35.36	11.92	25.05	15.96	14.98
Kernel HF	23.13	43.52	17.79	36.86	53.23	29.67	37.89	51.43	30.22	31.73	56.59	26.40	50.48	31.22	35.78
Kernel DI	39.38	56.37	32.45	35.71	59.43	30.29	36.70	58.65	30.82	31.74	60.10	27.64	51.48	25.57	35.56
Kernel DC	39.61	39.01	30.17	36.11	38.68	27.78	37.36	38.67	28.64	29.59	38.64	23.17	53.90	27.13	38.69

(c) 5-input (%fit)

Model	**G-EP-GP-PB-PS**			**G-EP-GP-PB-PT**			**G-EP-GP-PB-W**			**G-EP-GP-PB-T**			**G-EP-GP-PS-PT**		
	E	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**
Order Sel	35.78	46.68	26.19	38.63	59.16	30.14	36.41	52.62	30.13	31.14	13.96	27.39	34.17	47.61	25.59
Ridge Reg	45.80	25.76	32.27	45.80	25.76	32.27	45.80	25.76	32.27	45.99	25.86	32.46	45.80	25.76	32.27
Kernel TC	26.00	16.19	17.84	24.90	15.65	16.98	25.04	19.05	17.83	61.91	-29.08	29.49	27.60	16.42	18.99
Kernel SE	32.29	21.93	22.74	26.97	20.52	19.09	28.06	20.51	19.81	71.79	-72.98	15.57	20.65	22.09	15.17
Kernel SS	23.61	-6.89	10.51	22.52	4.60	12.70	23.09	-5.95	10.44	53.96	-69.01	10.47	21.15	11.16	13.33
Kernel HF	36.22	52.89	29.12	34.40	53.27	28.04	35.45	52.27	29.27	65.42	5.04	37.08	32.59	52.74	26.73
Kernel DI	37.03	52.55	31.06	34.59	52.66	29.50	34.96	52.80	29.78	65.37	-13.57	33.51	32.33	53.01	27.90
Kernel DC	38.02	28.37	27.63	34.54	27.32	25.29	35.01	27.68	25.65	68.09	-26.49	31.23	31.62	23.05	22.49

Model	**G-EP-GP-PS-W**			**G-EP-GP-PS-T**			**G-EP-GP-PT-W**			**G-EP-GP-PT-T**			**G-EP-GP-W-T**		
	E	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**
Order Sel	33.51	42.50	24.94	31.95	8.31	26.75	32.74	52.71	25.44	29.65	15.23	26.67	29.73	13.50	26.05
Ridge Reg	45.81	25.75	32.28	45.99	25.86	32.46	45.80	25.76	32.27	45.99	25.86	32.46	45.99	25.86	32.46
Kernel TC	23.94	21.66	17.64	49.59	17.58	34.10	24.81	13.35	16.38	48.49	17.89	33.38	41.49	24.28	29.93
Kernel SE	21.22	21.31	15.47	54.91	19.78	37.70	26.72	18.11	18.48	54.39	16.29	36.58	45.09	31.17	33.98
Kernel SS	21.91	5.97	12.72	38.33	-13.36	19.43	22.92	-1.68	11.41	36.25	12.50	23.85	24.54	-4.47	11.77
Kernel HF	33.42	52.05	27.74	38.28	31.15	25.03	34.78	52.46	28.93	35.41	23.50	20.42	40.20	48.07	32.55
Kernel DI	32.59	53.08	28.08	55.22	21.02	37.63	33.91	53.31	29.01	51.85	23.30	35.59	47.00	34.13	35.12
Kernel DC	32.62	25.84	23.87	54.76	21.44	37.16	33.76	25.51	24.31	56.20	20.66	37.98	48.53	34.19	36.66

Table 3.6: Analysis results and % fit to estimation (E), validation (V) and overall (O) data for estimated models for participant 222

(a) 3-input (Analysis Results)

Participant 222: 3-input (G,EP,GP)					
Model	**Fit_est**	**Fit_val**	**Fit_ovr**	**MSE**	**par_2norm**
Order Sel	35.96	38.47	36.14	2,265,855.28	8.86
High Order	66.92	37.5	53.23	634,770.87	670.75
Ridge Reg	65.86	41.26	54.66	676,423.71	14.41
Kernel TC	56.84	44.90	51.28	1,080,900.88	9.86
Kernel SE	54.10	33.10	46.65	1,222,511.97	7.90
Kernel SS	51.08	44.18	46.36	1,388,286.70	9.87
Kernel HF	62.15	45.86	53.71	831,385.38	13.77
Kernel DI	61.63	46.86	54.90	854,131.49	11.40
Kernel DC	54.67	42.41	48.88	1,192,419.98	8.95

(b) 4-input (%fit)

Model	**G-EP-GP-PB**			**G-EP-GP-PS**			**G-EP-GP-PT**			**G-EP-GP-W**			**G-EP-GP-T**		
	E	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**
Order Sel	32.60	43.74	35.09	63.70	53.83	53.88	33.96	46.64	35.34	36.17	41.58	36.82	31.86	45.68	34.82
Ridge Reg	65.86	41.27	54.67	65.87	41.26	54.67	65.87	41.26	54.67	65.86	41.26	54.67	67.20	41.40	55.20
Kernel TC	52.82	42.98	48.70	62.78	44.38	54.95	55.10	44.58	49.84	52.72	44.72	48.17	58.36	12.05	42.64
Kernel SE	55.84	33.77	48.44	69.49	41.48	58.46	56.58	18.36	44.04	52.94	33.98	46.14	59.39	17.82	45.22
Kernel SS	50.63	44.92	47.04	55.17	41.87	49.14	55.30	23.30	44.18	49.76	45.07	45.72	57.06	19.81	44.01
Kernel HF	52.79	42.81	47.87	64.84	44.60	54.49	59.66	45.92	52.13	52.49	42.79	47.63	60.46	47.65	53.02
Kernel DI	57.58	46.39	52.45	70.98	45.47	59.61	61.64	42.53	55.08	57.08	46.27	52.10	61.04	40.36	53.52
Kernel DC	53.12	40.77	48.06	70.43	42.58	59.48	54.89	17.97	42.46	51.83	41.62	46.75	57.56	17.30	43.40

(c) 5-input (%fit)

Model	**G-EP-GP-PB-PS**			**G-EP-GP-PB-PT**			**G-EP-GP-PB-W**			**G-EP-GP-PB-T**			**G-EP-GP-PS-PT**		
	E	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**
Order Sel	64.69	55.27	55.35	36.77	42.59	37.45	40.65	39.32	39.88	33.60	45.85	35.93	35.43	48.57	36.70
Ridge Reg	62.90	41.57	52.82	62.91	41.56	52.82	62.90	41.57	52.81	64.25	41.82	53.20	62.91	41.56	52.82
Kernel TC	54.99	42.80	49.85	59.02	40.49	52.57	52.16	43.40	48.09	57.52	18.28	44.05	55.51	43.41	49.50
Kernel SE	56.60	24.79	45.35	56.30	13.34	40.96	51.84	28.81	43.63	57.66	10.56	41.53	60.86	18.09	45.46
Kernel SS	53.01	41.72	47.88	42.17	30.41	37.17	38.85	39.20	35.96	55.75	18.79	42.88	56.21	30.00	47.57
Kernel HF	56.42	44.16	49.23	52.04	47.05	46.13	51.52	46.97	46.49	53.84	46.65	47.33	55.05	44.26	46.90
Kernel DI	62.67	48.20	56.29	53.53	38.92	48.38	51.93	44.86	48.27	59.26	39.51	50.99	66.80	43.27	58.05
Kernel DC	56.67	37.97	49.71	55.19	34.28	48.31	52.59	41.13	47.47	58.25	20.21	44.88	57.99	29.62	48.81

Model	**G-EP-GP-PS-W**			**G-EP-GP-PS-T**			**G-EP-GP-PT-W**			**G-EP-GP-PT-T**			**G-EP-GP-W-T**		
	E	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**	**E**	**V**	**O**
Order Sel	38.70	40.61	38.97	32.99	45.85	35.75	37.27	44.29	37.58	34.69	49.12	36.62	38.96	49.62	40.49
Ridge Reg	62.90	41.56	52.81	64.26	41.81	53.20	62.91	41.55	52.82	64.26	41.79	53.21	64.25	41.81	53.20
Kernel TC	51.68	43.78	47.07	59.40	38.18	50.76	53.14	40.67	48.22	54.97	41.95	49.10	67.09	39.34	56.83
Kernel SE	45.75	34.46	40.63	60.46	31.29	48.68	53.00	13.47	39.49	55.59	31.07	46.69	63.65	35.07	52.72
Kernel SS	40.08	39.84	37.10	56.53	35.16	47.69	51.96	34.49	45.98	43.31	41.30	39.84	57.98	32.95	48.23
Kernel HF	52.09	47.01	46.84	61.52	46.03	51.60	53.17	46.93	47.44	51.29	45.92	45.88	66.78	40.94	51.85
Kernel DI	54.46	47.04	50.40	66.67	47.53	56.83	57.47	42.37	52.24	54.68	37.84	48.44	68.94	44.04	57.00
Kernel DC	52.21	41.03	46.73	57.83	30.83	47.59	55.33	32.11	47.85	56.30	30.12	46.62	60.99	32.11	50.05

Table 3.7: Analysis results and % fit to estimation (E), validation (V) and overall (O) data for estimated models for participant 230

(a) 3-input (Analysis Results)

Participant 230: 3-input (G,EP,GP)					
Model	Fit_est	Fit_val	Fit_ovr	MSE	par_2norm
Order Sel	79.13	78.21	78.19	181,721.21	53.88
High Order	80.63	79.45	78.71	155,277.81	56.48
Ridge Reg	80.04	71.93	76.23	164,857.10	36.89
Kernel TC	77.31	57.11	68.07	212,983.43	34.07
Kernel SE	78.48	61.68	70.14	191,620.32	34.33
Kernel SS	69.80	33.77	54.84	377,487.02	28.93
Kernel HF	73.64	45.13	61.55	287,615.30	29.65
Kernel DI	79.03	70.27	74.92	181,927.80	37.40
Kernel DC	78.71	67.44	73.56	187,598.51	35.88

(b) 4-input (%fit)

Model	G-EP-GP-PB			G-EP-GP-PS			G-EP-GP-PT			G-EP-GP-W			G-EP-GP-T		
	E	V	O	E	V	O	E	V	O	E	V	O	E	V	O
Order Sel	80.75	79.08	79.26	79.47	79.86	78.20	79.88	78.85	79.00	79.66	80.04	79.01	80.56	78.06	79.10
Ridge Reg	80.06	71.94	76.25	80.05	71.93	76.24	80.06	71.91	76.24	80.05	71.95	76.25	82.17	66.64	75.45
Kernel TC	75.12	46.68	62.66	71.19	36.24	56.70	75.78	47.77	63.57	60.60	12.83	41.63	76.94	32.22	56.33
Kernel SE	77.15	57.59	68.17	76.63	55.39	67.17	77.17	54.38	67.04	73.54	47.74	62.92	78.34	57.69	69.06
Kernel SS	61.83	15.66	43.53	62.03	16.42	43.95	64.63	18.18	45.60	60.38	12.98	41.57	62.64	14.48	43.22
Kernel HF	71.52	36.29	56.47	71.27	35.51	56.00	71.45	36.07	56.34	70.72	33.94	55.04	72.25	38.86	58.16
Kernel DI	78.16	66.21	72.95	77.72	64.30	71.99	78.36	65.60	72.76	76.05	57.60	68.46	79.38	64.04	73.17
Kernel DC	67.24	25.15	50.24	66.19	20.61	47.52	68.16	18.45	47.05	64.82	19.81	46.61	67.84	25.18	50.61

(c) 5-input (%fit)

Model	G-EP-GP-PB-PS			G-EP-GP-PB-PT			G-EP-GP-PB-W			G-EP-GP-PB-T			G-EP-GP-PS-PT		
	E	V	O	E	V	O	E	V	O	E	V	O	E	V	O
Order Sel	80.96	80.86	80.19	81.42	80.77	80.15	82.46	80.21	79.62	80.81	78.55	79.65	80.54	80.50	79.30
Ridge Reg	79.60	73.03	75.93	79.62	73.00	75.93	79.61	73.05	75.94	81.71	67.87	75.49	76.65	59.35	69.35
Kernel TC	76.93	49.07	64.31	74.04	38.51	58.35	63.07	16.71	44.56	74.71	25.90	51.72	73.34	36.72	57.31
Kernel SE	79.28	72.51	75.70	76.56	53.76	66.76	63.33	17.03	44.80	78.78	62.93	72.05	65.38	8.39	41.44
Kernel SS	77.77	62.67	71.58	61.37	13.59	42.29	60.80	14.15	42.37	67.73	19.50	47.20	61.33	11.13	40.91
Kernel HF	72.30	41.97	59.72	71.90	19.68	49.31	70.47	33.38	55.39	69.64	31.01	53.88	65.20	16.39	45.32
Kernel DI	79.26	71.78	75.69	78.26	64.07	72.22	76.91	61.17	70.43	79.48	63.96	73.21	77.87	62.82	71.53
Kernel DC	79.30	70.54	75.13	68.77	18.37	47.67	65.61	21.38	47.90	71.23	30.30	54.09	68.81	17.96	47.36

Model	G-EP-GP-PS-W			G-EP-GP-PS-T			G-EP-GP-PT-W			G-EP-GP-PT-T			G-EP-GP-W-T		
	E	V	O	E	V	O	E	V	O	E	V	O	E	V	O
Order Sel	80.58	80.31	78.87	80.23	80.00	78.93	80.62	80.88	79.57	81.36	80.31	80.27	82.72	78.27	79.20
Ridge Reg	79.60	73.04	75.93	81.70	67.86	75.48	79.62	73.01	75.93	81.71	67.86	75.49	81.70	67.88	75.49
Kernel TC	60.22	12.42	41.30	75.40	37.08	58.02	61.85	11.03	41.19	75.65	38.81	59.03	63.47	16.31	44.50
Kernel SE	62.01	13.12	42.36	78.07	60.48	70.71	64.33	8.32	41.01	78.88	63.66	72.43	63.58	17.25	45.02
Kernel SS	59.86	11.09	40.50	61.80	13.58	42.51	60.86	12.04	41.21	69.35	26.65	51.22	61.73	13.67	42.51
Kernel HF	66.44	22.26	48.57	70.60	32.43	54.69	72.19	25.27	51.93	69.72	32.73	54.49	69.36	33.22	54.63
Kernel DI	62.50	14.69	43.40	79.09	62.50	72.42	75.17	51.96	65.71	79.57	64.22	73.36	77.55	57.57	69.59
Kernel DC	64.08	18.13	45.68	70.69	29.17	53.48	66.87	16.57	45.96	72.72	25.67	52.61	66.02	22.26	48.45

Chapter 4

ESTIMATION OF PERSONALIZED BEHAVIORAL MODELS

4.1 Overview

In the previous chapters, the main focus was on estimation of dynamical system models using system ID techniques from the available input-output data. The aim was to estimate models using various regularization approaches, and from the results obtained, try to determine which is the can be a good technique that can be used for model estimation, based on the step responses, as well as how well the model fits to the data.

In this chapter, the focus changes from getting a good model estimate for the participant to using regularization techniques to determine which inputs affect the behavior of the participant the most. The aim is to explore the relationships between the 'inputs', which includes the manipulated variables or intervention components (e.g., daily step goal and points) and time-varying disturbance variables (e.g., perceived stress, weather), and the predicted 'output' or outcome of interest (e.g., actual steps taken per day) for a given individual over time. Having this kind of knowledge using the idiographic approach can help in tailoring person-specific variables in order to personalize decision rules and intervention dosages for individuals [28].

4.2 Using Model Fits to Assess Participant Behavior

Initially a basic 3-input model (Goals, Expected Points and Granted Points) was estimated using all identification methods. Then additional inputs were added to this set of basic inputs to observe how each input affected the model estimate, in a positive or negative way. This was measured in terms of change in % Fit, for each additional input that was added. The additional inputs were: Predicted Busyness, Predicted Stress, Predicted Typical, Weekday/Weekend and Temperature. The results were plotted for each participant in order to visualize what predictors impact most the % Fit, which can be useful for developing personalized interventions. Also, this gives an idea regarding what model estimation methods are best suited for estimating behavioral models for a particular participant. This has been implemented and studied for all four participants, and is discussed in the following subsections.

Table 4.1: % Improvement in fit with specific additional inputs for each model estimation method

(a) Participant 164

Methods	Validation		Overall		Estimation	
Order Sel.	T	16%	T	6%	PB-T	6%
Ridge Regr.	PB-T	2%	PB-T	0.30%	T	0.20%
Kernel TC	W	1%	PS-T	14%	PB-T	34%
Kernel SE	-	-	PS-T	8%	PB-T	37%
Kernel SS	W	20%	PT-T	10%	PB-T	31%
Kernel HF	-	-	PB-T	7%	PB-T	28%
Kernel DI	W	1%	PS-T	6%	PB-T	29%
Kernel DC	-	-	T	9%	PB-T	29%

(b) Participant 180

Methods	Validation		Overall		Estimation	
Order Sel.	PB-T	22%	W-T	10%	PB-PT	12%
Ridge Regr.	PB-T	6%	T	1.80%	T	1.80%
Kernel TC	PT	6%	PB-T	18%	PB-T	16%
Kernel SE	PB-T	2%	PB-T	5%	PB-T	35%
Kernel SS	PB-T	22%	PB-T	32%	PB-T	32%
Kernel HF	PB-T	9%	PT	9%	PB-T	22%
Kernel DI	PT	5%	PT	8%	PB-T	25%
Kernel DC	PT	4%	PT	8%	PB-T	31%

(c) Participant 222

Methods	Validation		Overall		Estimation	
Order Sel.	PB-PS	17%	PB-PS	19%	PB-PS	30%
Ridge Regr.	PB-T	1%	T	0.50%	T	1.20%
Kernel TC	-	-	W-T	5%	W-T	10%
Kernel SE	PS	8%	PS	12%	PS	16%
Kernel SS	W	1%	PS	3%	W-T	7%
Kernel HF	T	2%	PS	0%	W-T	4%
Kernel DI	PB-PS	2%	PS	5%	PS	9%
Kernel DC	-	-	PS	11%	PS	17%

(d) Participant 230

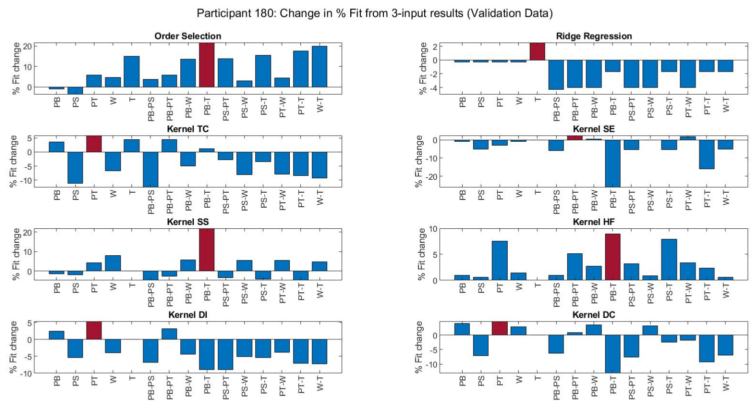
Methods	Validation		Overall		Estimation	
Order Sel.	PT-W	3%	PT-T	2%	PT-W	4%
Ridge Regr.	-	-	-	-	-	-
Kernel TC	-	-	-	-	-	-
Kernel SE	PB-PS	10%	PB-PS	5%	PB-PS	1%
Kernel SS	PB-PS	28%	PB-PS	16%	PB-PS	8%
Kernel HF	-	-	-	-	-	-
Kernel DI	PB-PS	2%	PB-PS	1%	PB-PS	1%
Kernel DC	PB-PS	4%	PB-PS	2%	PB-PS	1%

4.2.1 Participant 180

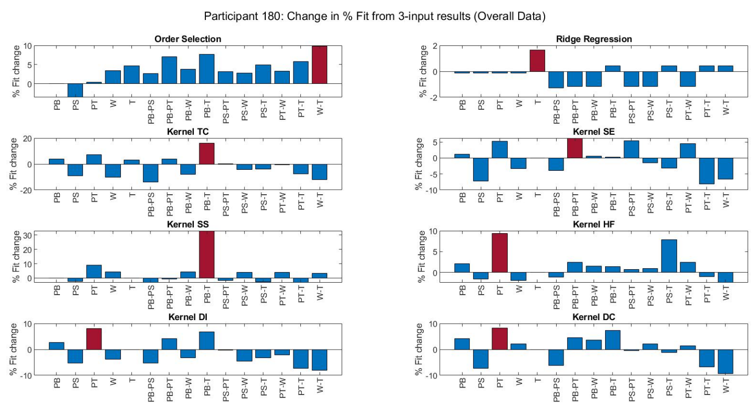
The "% fit change" results for participant 180 in Fig 4.1 can be analyzed in a manner, where the maximum % change in fit (for a particular additional input combination) is noted down for each model estimation method. This can be seen in Table 4.1b. The fits highlighted in light brown show that there is a significant positive change in the % Fit when the corresponding inputs are used in addition with the basic 3-input combination, for model estimation. The best results for Participant 180 is displayed by the Regularized 5-input ARX model which is estimated using the SS kernel, with PB (Predicted Busyness), and T (Temperature) as additional inputs. The results for this are highlighted in "green" in the table. From the highlighted blocks, as well as many of the other results in the Table 4.1b, it is seen that in most cases, for this participant, Predicted Busyness and Temperature contribute towards an increased fit %, which indicates that addition of these inputs can have a significant impact on the walking behavior of the participant. This result helps in providing an important insight into personalization of behavioral models for the individual.

The data in Table 4.1 is basically a numerical summary obtained from the plots which depict the % change in fit, which respect to the 3-input models, when a certain combination of additional inputs are added. For example, for participant 180, from the plots in 4.1, the maroon bar indicates the highest positive change in fit when the corresponding combination of inputs are added to the basic 3-input model, for that particular estimation method. The entries in Table 4.1b, are basically the magnitude of those maroon bars. If for any participant, and any estimation method, and input combination, the maximum percent change is the fit is either negative, zero, or negligible, then the corresponding table entry in Table 4.1 is left blank.

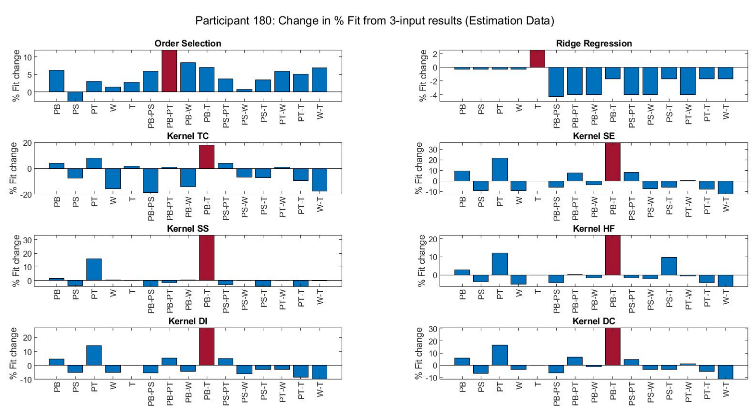
The result obtained related to personalization, obtained from the % change in



(a) % Change in fit (Validation Data)



(b) % Change in fit (Overall Data)



(c) % Change in fit (Estimation Data)

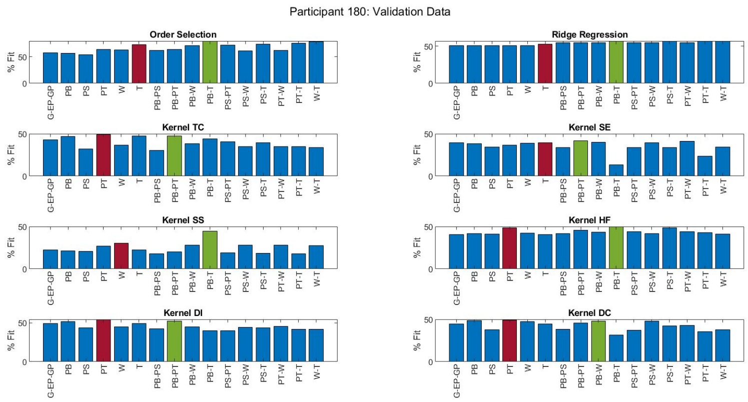
Figure 4.1: Change in % fit with respect to 3-input models for participant 180

fit plots are then verified with the plots in Fig. 4.2. These plots basically helps us visualize addition of which inputs leads to estimation of models with significantly high fit values. The input combination contributing to the highest % fit value among the 4-input models, for each of the estimated models, has been highlighted in maroon, and the best among the 5-input cases has been highlighted in green.The plots in Fig. 4.2, further reinforce that Predicted Busyness and Temperature play an important role in deciding whether the participant will walk or not.

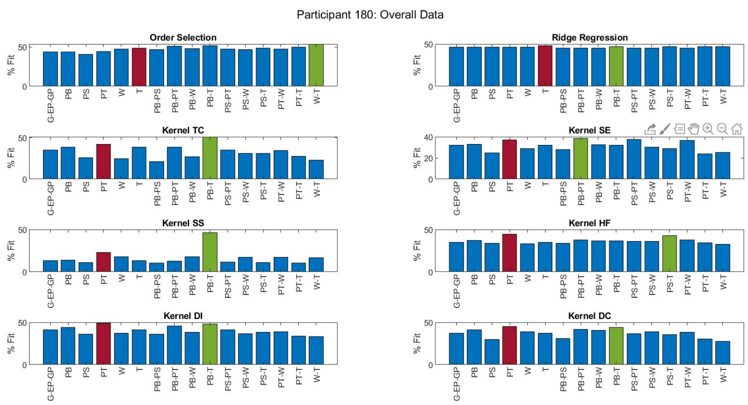
Now for the regularized model estimated using the SS kernel, a residual analysis is done, and the step response is plotted to visualize the model response. This is to ensure that it is in agreement with the intuition, as well as to see if it has the right gain and time constant. The step responses (Fig. 4.3) for the selected model look good. From the plot of the autocorrelation of residuals in Fig. 4.4, very slight correlation is observed. but since it is only very slightly outside the 99% confidence bound, this can be ignored. There is no correlation between the inputs and the residuals, which lends credibility to the model.

4.2.2 Participants 164, 222, 230

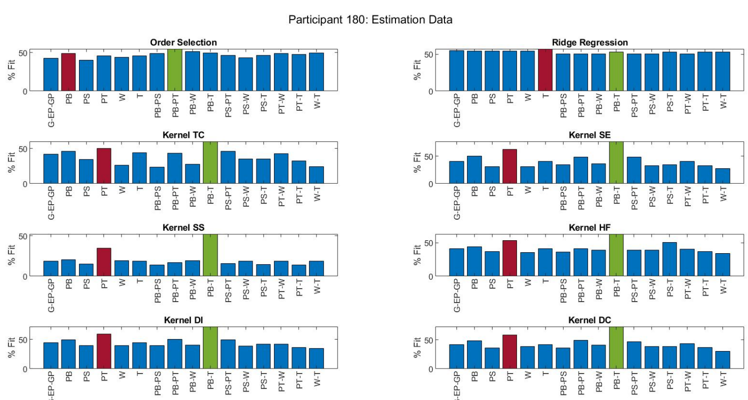
For Participant 164, referring to Table 4.1a, it can be seen from the results of % fit wit overall data that additional inputs PS-T and PT-T result in improved fits when compared to the basic 3-input case. There is no uniformity in the results, and hence it is difficult to predict if any of the inputs apart from G, EP and GP affect participant behavior. However, the % fit results with the estimation data show some uniformity. For almost all cases (except for a couple of them), the input combination PB-T when added to the basic 3-inputs while model estimation, show significant improvements in the fit percentages. From this observation, a conclusion can be made that predicted busyness and temperature affect the PA levels for this



(a) Validation Data)

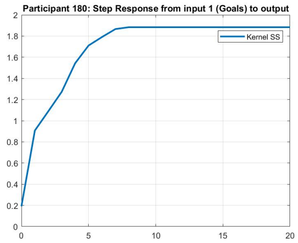


(b) Overall Data

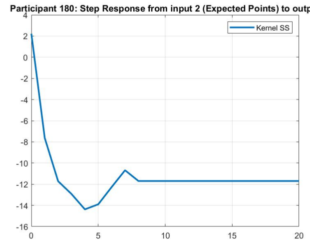


(c) Estimation Data

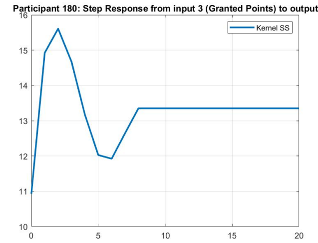
Figure 4.2: % Fit for each estimated model with respect to measured data for participant 180



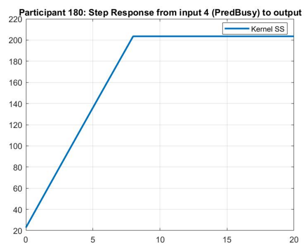
(a) Goals



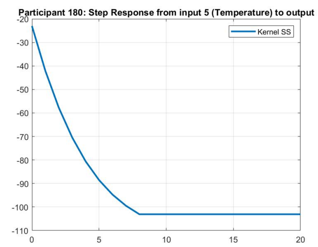
(b) Expected Points



(c) Granted Points



(d) Predicted Busy



(e) Temperature

Figure 4.3: % Step response for model estimated using SS kernel for participant 180, additional inputs: PB, T

participant. This can be further verified from the plots in Fig. 4.6, where it can be seen that for many of the model estimation methods, the combination of added inputs PB-T show higher fit percentages (ones highlighted in green). For getting this result, kernel based regularization methods TC, SE SS, HF, DI and DC can be used. Order selection and ridge regression do not work well in this case.

Next participant 222 is analyzed, where it can be seen from Table 4.1c that adding input PS (when order selection is used) or PB-PS (for regularized FIR using kernel SE) result in improvement in the fit percentages. These % fits are with respect to estimation, validation and overall data. From this data it is difficult to come to a conclusion with regards to which input combination influence participant behavior the most. Hence, it is necessary to examine the plots in Fig. 4.8, where it can be seen that in general, predicted stress as an input does influence participant behavior more

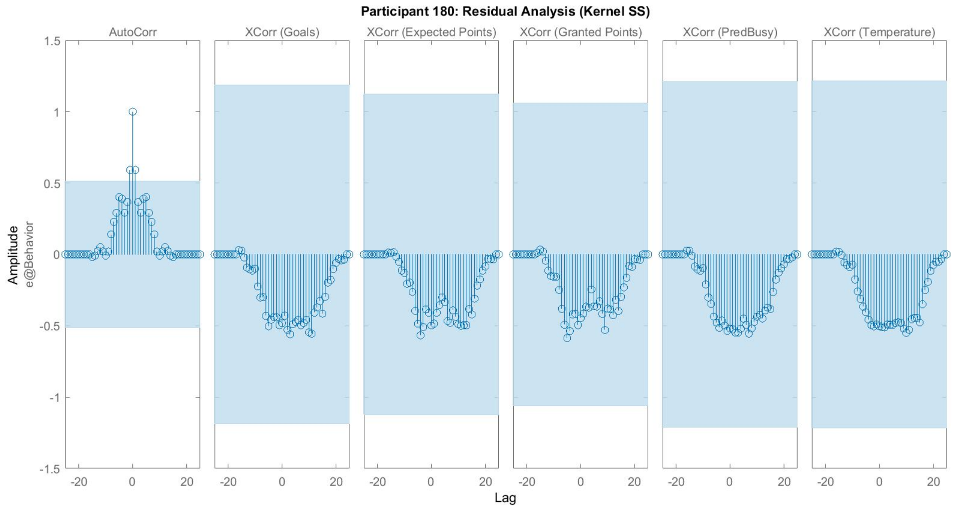
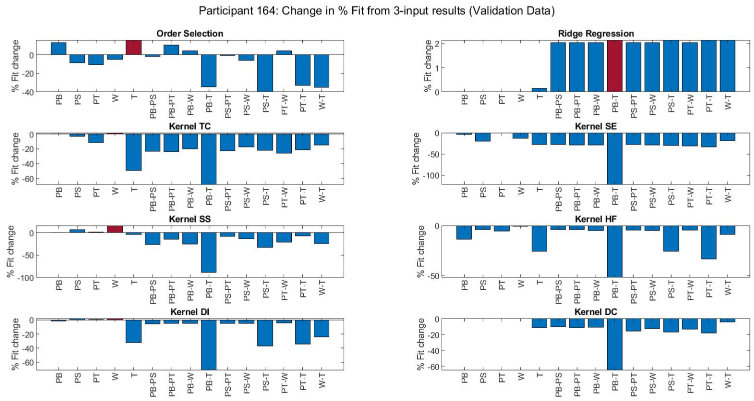


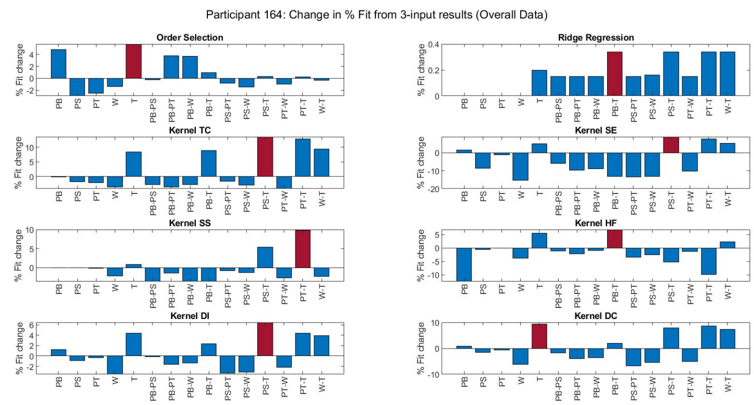
Figure 4.4: Residual analysis for the model estimated using SS kernel for participant 180, additional inputs: PB, T

that the PB-PS combination. But the question is whether the change in significant enough to be taken into consideration. For this the results in Table 4.1c needs to be re-examined. It is seen that for regularized models estimated using kernels SE, DI and DC, do show improved fits in the range of 9-16% when the output is compared to the estimation data, and an improvement in the range of 5-12% can be seen when compared to the overall data. Even though these numbers may not be very high, it can still be said that predicted stress does influence participant behavior to some extent.

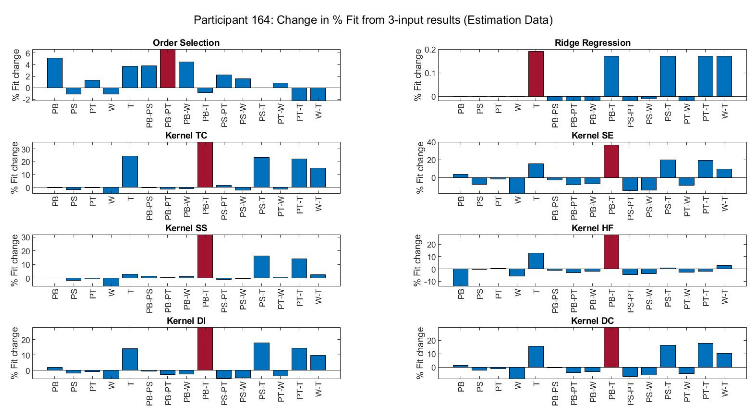
Similarly, for participant 230, from Table 4.1d, it is clearly seen that the addition of input combination PB-PS has a positive effect on the % fit change for while using kernel-based regularization methods SE, SS, DI and DC. The best result, consistent with the estimation, validation and overall data is observed for the regularized model estimated using kernel SS. Kernel SE gives satisfactory results as well. Now, taking look at Fig. 4.10, it is evident that when predicted busy and predicted stress are added, the results do show a higher fit percent than the rest of the cases. The extent



(a) % Change in fit (Validation Data)

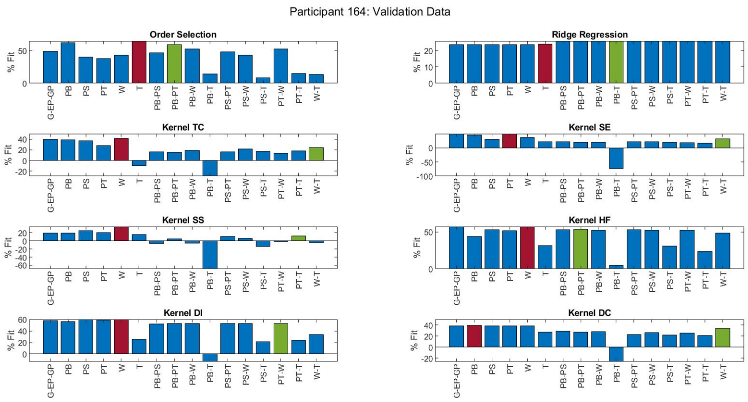


(b) % Change in fit (Overall Data)

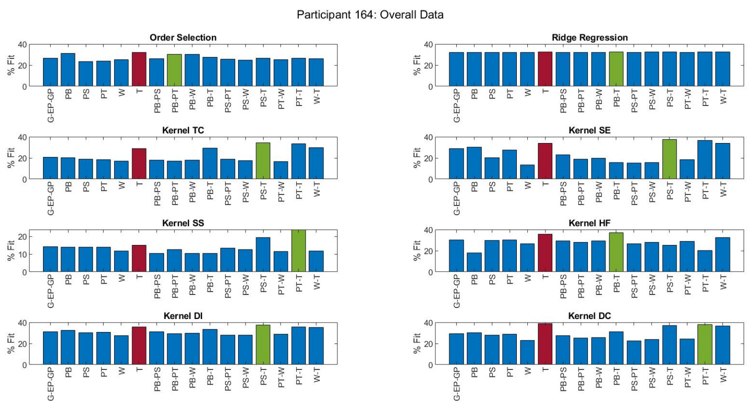


(c) % Change in fit (Estimation Data)

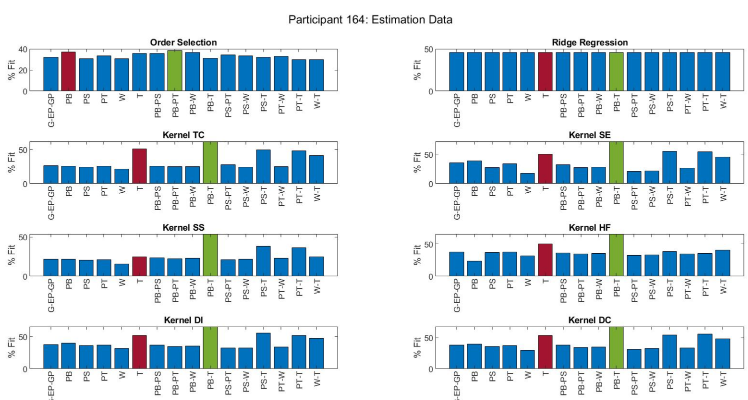
Figure 4.5: Change in % fit with respect to 3-input models for participant 164



(a) Validation Data)

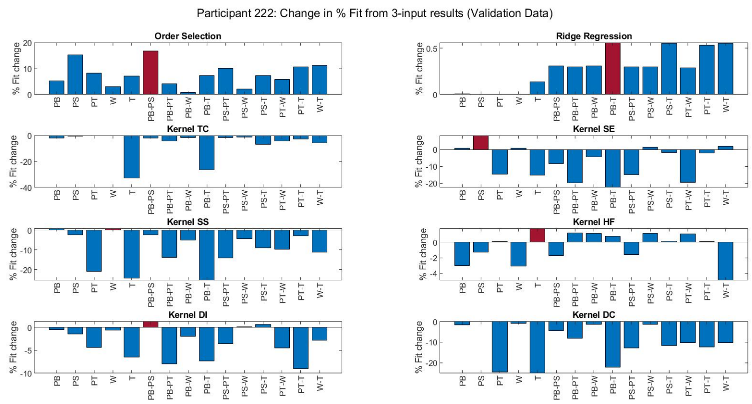


(b) Overall Data

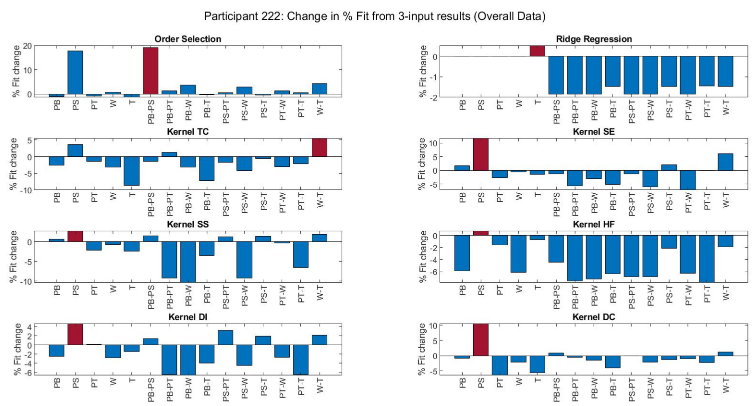


(c) Estimation Data

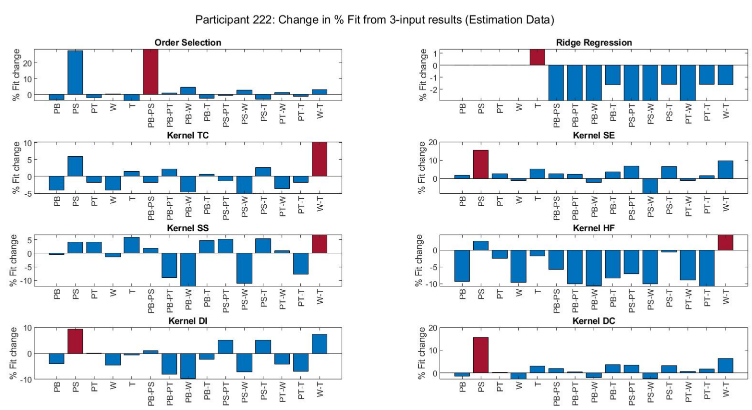
Figure 4.6: % Fit for each estimated model with respect to measured data for participant 164



(a) % Change in fit (Validation Data)

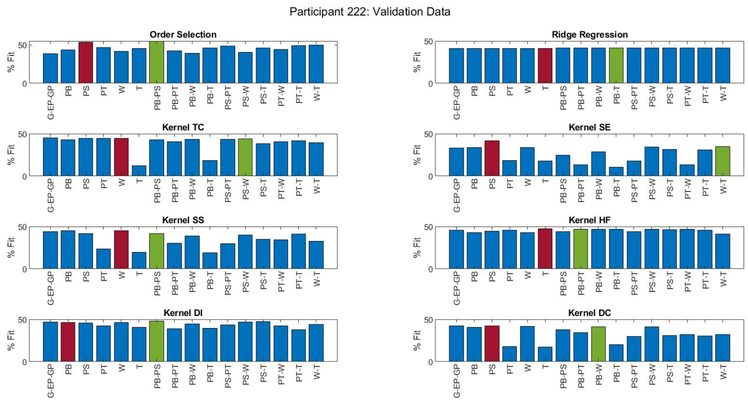


(b) % Change in fit (Overall Data)



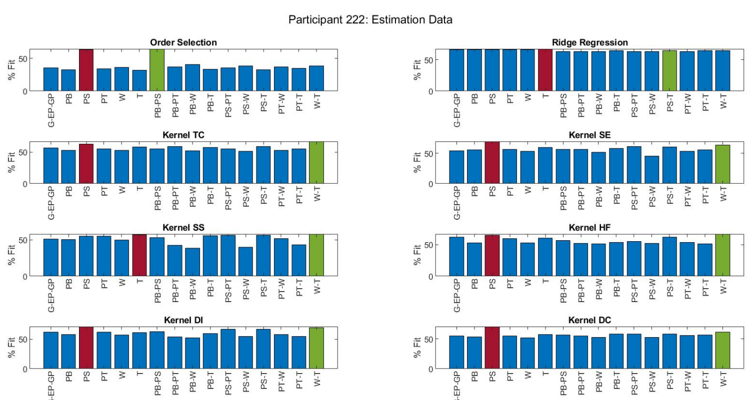
(c) % Change in fit (Estimation Data)

Figure 4.7: Change in % fit with respect to 3-input models for participant 222



(a) Validation Data)

(b) Overall Data



(c) Estimation Data

Figure 4.8: % Fit for each estimated model with respect to measured data for participant 222

to which these inputs impact a participant can be inferred from the magnitude of improvement in model quality (% fit), when compared to the basic model with 3 inputs.

4.3 Conclusions

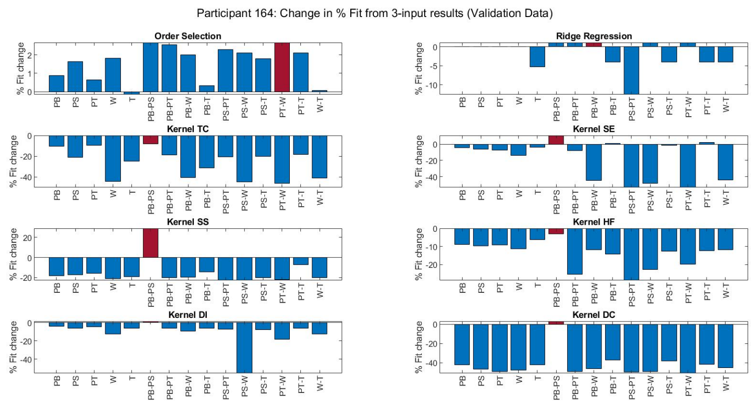
In general, the results of this chapter can be summarized as follows:

- From the four participants that were studied, it can be seen that the factors affecting the behavior of the individual, that is the measured PA depends mainly on predicted busyness, predicted stress and temperature. These factors, either individually, or when combined can have an impact on the level of activity of the person.

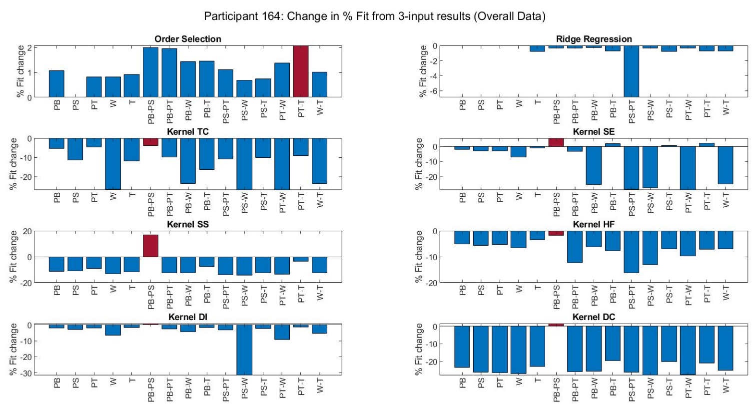
- The graphical results help in predicting which inputs determine the individual's behaviour, as well as the magnitude of the impact that they have on the person. This can be seen from the data which shows the change in % fit from the basic 3-input model.

- Addition of certain input combinations, and using some model estimation methods can worsen the model quality, when compared to the 3-input model. This can mean that either the inputs under consideration are not useful in determining participant behavior, or else the model estimation selected is not the right one.

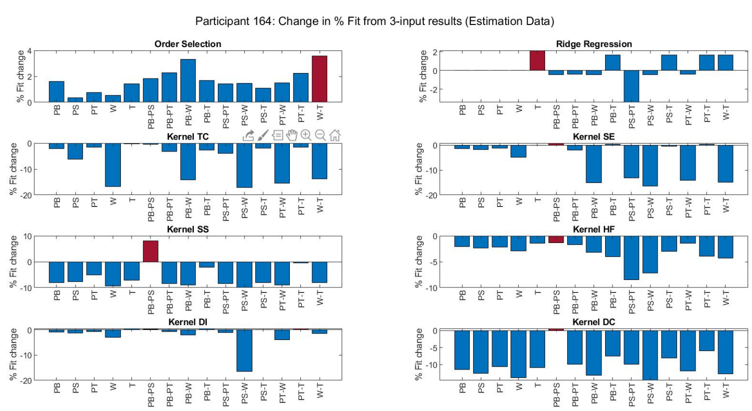
- In general, it is seen that regularized models estimated using the SS kernel can be a good method for the purpose of personalization of the interventions. Even though the model quality might not be very good in terms of the fit percentages, when compared to the other estimation methods, it is good enough for selecting inputs which matter the most.



(a) % Change in fit (Validation Data)

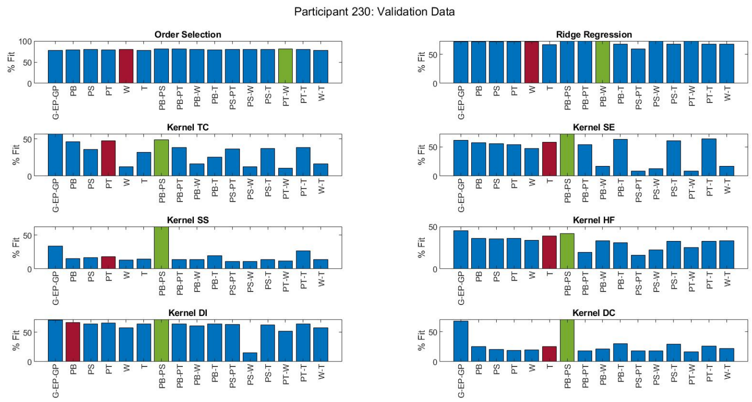


(b) % Change in fit (Overall Data)

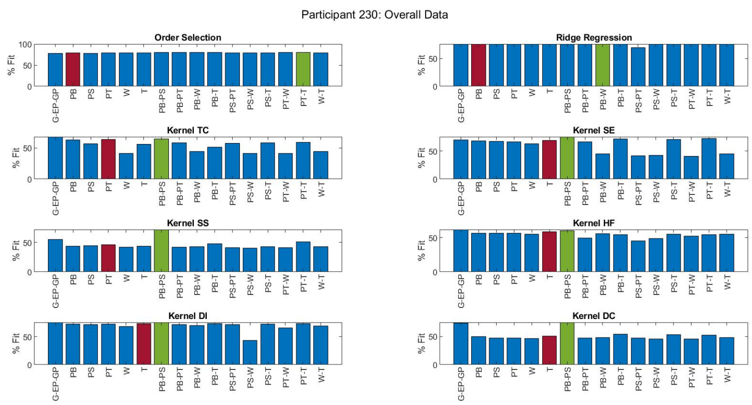


(c) % Change in fit (Estimation Data)

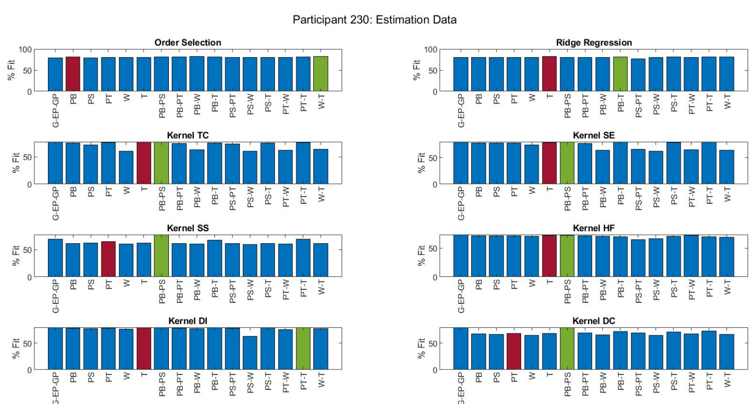
Figure 4.9: Change in % fit with respect to 3-input models for participant 230



(a) Validation Data)



(b) Overall Data



(c) Estimation Data

Figure 4.10: % Fit for each estimated model with respect to measured data for participant 230

- For most cases, the regularized models perform better than the one estimated using order selection, from the perspective of personalization.

Chapter 5

CONCLUSION AND FUTURE WORK

5.1 Conclusions

Control system methods, in particular system identification, play an important role in building dynamical mathematical models for behavioral interventions. Estimating behavioral models is essential for designing personalized interventions for individual participants. These behavioral models can be helpful in providing an insight into what factors can impact an individual's behavior, in terms of physical activity. Since every individual is different, it is important that this analysis is done on an idiographic level, and not a generalized nomothetic one.

The *Just Walk* study originally used traditional ARX black box system identification methods for model estimation. In classical system identification, one of the most important factors on which the quality of the estimated model depends, is the selection of the model structure. If that is not done properly it can lead to two types of errors. A bias, or systematic error is seen if the order is too low. This happens because the model predicted is not able to capture all features of the system. In case of a very high order model variance errors are caused due to overfitting. In this case, the model may fit very well to the data used for estimation. However, it may not generalize well to new data. What is required for a good model is a balance between the bias and variance, so that the main features of the system are captured by the model, without random errors.

The above can be achieved using an order selection, where models are estimated over a range of orders (low to high). The best model structure is then evaluated using model validation techniques, i.e. the one with the least MSE is selected (since both bias and variance contribute to the MSE). This can be a good method when the number of inputs and outputs are less. However in MIMO systems, as the number of inputs or outputs increases, the computational complexity of this method increases as well. In this case, there is a need to explore a method which will be robust, and computationally less expensive. In this case, there is a need to explore alternative methods from the machine learning domain. This is where regularization comes into picture. Regularization helps in taking care of the bias variance trade-off by adding a penalty terms to high order models. This method may increase the bias by a slight amount, but the variance errors are significantly reduced.

The goal of this thesis was to examine if regularization would be benefiial in the identification of behavioral models for a physical activity intervention, like the *Just Walk* study. Also, another factor that was examined was if the regularized models provide better results for determining the factors that affect individual behavior, which in turn would be helpful for designing personalized interventions. The results using order selection was compared with the results obtained using regularized models. These results have been described in detail in Chapter 3. The study was carried out on four participants (164, 180, 222, 230). The performance of the methods implemented for these four participants, in terms of personalization has been described in Chapter 4.

From this study it was seen that the performance in terms of the fit percentages varied for different estimation methods, from person to person. Methods that work well for one participant, may not work well for others. However, it was seen that the regularized model estimated using the DI kernel gave consistently good results for all four participants. The next part of the study dealt with determining the inputs, which when added to the standard 3-input model, help in describing the individual better by improving significantly the model quality, which is indicated by a large increase in the % fit results. For this portion of the study, it is not important to have really good fits; the important task is to figure out the combination of inputs which improve model quality. After analyzing the results for all four participants, it can be seen that regularized models using the kernel SS method work well for all participant.

From this study it can be concluded that even if in some cases, regularized models may not give the best results when compared to order selection, and some regularized models might not perform well at all, the ones that do well are reasonably good. Hence regularization can be considered as a good option, keeping in mind the advantage of reduced computational complexity.

5.2 Future Work

Identification of good models depends on the following factors:

- Having a model structure with optimal order (not too low, or not too high)
- Deciding on the predictors that need to be considered for model estimation
- Designing the input in such a ways that it can capture the important dynamics of the system under consideration

While using traditional system identification methods, if the above points are not taken care of well, then the identification process may end up in estimating a model of poor quality, which might have bias or variance issues or low fit percentages. This model may not be a very accurate representation of the true system in this case.

The ideas used in this thesis, to overcome the above mentioned issues in a behavioral study, *Just Walk*, could be possibly used in future for other studies, and application, in the field of behavioral medicine, or any other application of system identification. As a part of this thesis, what has been implemented on the four participant in the *Just Walk* Study, could be further extended to the remaining participants involved in this study in order to get a much larger view of any new approach that could be taken. This could involve, designing inputs in a way which can contribute to better results, or trying to determine whether scaling them to be in a particular range would have a positive impact on the quality of model estimated. In order to get further insights, these approaches could be extended so that they can be evaluated over multiple cycles that will involve estimation and validation at different times (as was done in *Just Walk* with standard ARX). Also, getting a bigger picture can help in developing a tool, which will automatically select the model estimation methods, based on various factors, internal and external to the participant.

The use of machine learning ideas, like regularization, along with system identification principles could be something that can be implemented in the future, for identification for complex systems, where in many cases traditional system identification methods alone, would not give satisfactory results. Also, controls criteria could be used as a means to include closed-loop considerations in the regularization problem, which will lead to different and possibly improved results. One possible approach is to apply regularization to prefiltered data, using control-relevant prefilters as described in Rivera *et al.* [37]. These are some ideas that could be explored further in extensions of this research.

REFERENCES

[1] *System Identification Overview*, https://www.mathworks.com/help/ident/gs/about-system-identification.html (accessed June 04, 2020).

[2] *Regularized Identification of Dynamic Systems*, https://www.mathworks.com/help/ident/examples/regularized-identification-of-dynamic-systems.html (accessed June 05, 2020).

[3] *Model Structure Selection: Determining Model Order and Input Delay*, https://www.mathworks.com/help/ident/ug/model-structure-selection-determining-model-order-and-input-delay.html (accessed June 11, 2020).

[4] *Select model order for single-output ARX models - MATLAB selstruc*, https://www.mathworks.com/help/ident/ref/selstruc.html (accessed June 11, 2020).

[5] *Compare identified model output and measured output - MATLAB*, https://www.mathworks.com/help/ident/ref/compare.html (accessed June 17, 2020).

[6] *Loss Function and Model Quality Metrics - MATLAB & Simulink*, https://www.mathworks.com/help/ident/ug/model-quality-metrics.html (accessed June 17, 2020).

[7] *What Is Residual Analysis? - MATLAB & Simulink*, https://www.mathworks.com/help/ident/ug/what-is-residual-analysis.html (accessed June 18, 2020).

[8] *Regularized Estimates of Model Parameters*, https://www.mathworks.com/help/ident/ug/regularized-estimates-of-model-parameters.html (accessed May 28, 2020).

[9] Bandura, A., *Social foundations of thought and action: a social cognitive theory* (Prenctice Hall, Englewood Cliffs, NJ, 1995).

[10] Carlin, B. P. and T. A. Louis, *Bayes and empirical Bayes methods for data analysis* (Chapman & Hall, London, 1996).

[11] Cesar, A. M. M., *A system identification and control engineering approach for optimizing mHealth behavioral interventions based on social cognitive theory*, Ph.D. thesis, Arizona State University (2016).

[12] Chen, T., "On kernel design for regularized lti system identification", Automatica **90**, 109–122 (2018).

[13] Chen, T., M. S. Andersen, L. Ljung, A. Chiuso and G. Pillonetto, "System identification via sparse multiple kernel-based regularization using sequential convex optimization techniques", IEEE Transactions on Automatic Control **59**, 11, 2933–2945 (2014).

[14] Chen, T. and L. Ljung, "On kernel structures for regularized system identification (i): a machine learning perspective", IFAC-PapersOnLine **48**, 28, 1035–1040 (2015).

[15] Chen, T., H. Ohlsson and L. Ljung, "On the estimation of transfer functions, regularizations and gaussian processes—revisited", Automatica **48**, 8, 1525–1535 (2012).

[16] Freigoun, M. T., C. A. Martin, A. B. Magann, D. E. Rivera, S. S. Phatak, E. V. Korinek and E. B. Hekler, "System identification of just walk: A behavioral mhealth intervention for promoting physical activity", 2017 American Control Conference (ACC) (2017).

[17] Hekler, E. B., D. E. Rivera, C. A. Martin, S. S. Phatak, M. T. Freigoun, E. Korinek, P. Klasnja, M. A. Adams and M. P. Buman, "Tutorial for using control systems engineering to optimize adaptive mobile health interventions", Journal of Medical Internet Research **20**, 6 (2018).

[18] Korinek, E. V., S. S. Phatak, C. A. Martin, M. T. Freigoun, D. E. Rivera, M. A. Adams, P. Klasnja, M. P. Buman and E. B. Hekler, "Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention", Journal of Behavioral Medicine **41**, 1, 74–86 (2017).

[19] Ljung, L., *System identification: Theory for the User* (Prentice Hall, New Jersey, 1987).

[20] Ljung, L. and T. Chen, "Convexity issues in system identification", 2013 10th IEEE International Conference on Control and Automation (ICCA) (2013).

[21] Ljung, L. and T. Chen, "What can regularization offer for estimation of dynamical systems?", IFAC Proceedings Volumes **46**, 11, 1–8 (2013).

[22] Ljung, L., T. Chen and B. Mu, "A shift in paradigm for system identification", International Journal of Control **93**, 2, 173–180, URL `https://doi.org/10.1080/00207179.2019.1578407` (2020).

[23] Ljung, L., R. Singh and T. Chen, "Regularization features in the system identification toolbox", IFAC-PapersOnLine **48**, 28, 745–750 (2015).

[24] Ljung, L. and B. Wahlberg, "Asymptotic properties of the least-squares method for estimating transfer functions and disturbance spectra", Advances in Applied Probability **24**, 2, 412–440 (1992).

[25] Martin, C. A., D. E. Rivera, W. T. Riley, E. B. Hekler, M. P. Buman, M. A. Adams and A. C. King, "A dynamical systems model of social cognitive theory", 2014 American Control Conference (2014).

[26] Martín, C. A., D. E. Rivera and E. B. Hekler, "Design of informative identification experiments for behavioral interventions", IFAC-PapersOnLine **48**, 28, 1325–1330 (2015).

[27] Martín, C. A., D. E. Rivera, E. B. Hekler, W. T. Riley, M. P. Buman, M. A. Adams and A. B. Magann, “Development of a control-oriented model of social cognitive theory for optimized mhealth behavioral interventions”, IEEE Transactions on Control Systems Technology **28**, 2, 331–346 (2020).

[28] Phatak, S. S., M. T. Freigoun, C. A. Martín, D. E. Rivera, E. V. Korinek, M. A. Adams, M. P. Buman, P. Klasnja and E. B. Hekler, “Modeling individual differences: A case study of the application of system identification for personalizing a physical activity intervention”, Journal of Biomedical Informatics **79**, 82–97 (2018).

[29] Pillonetto, G., *Regularized Kernel-based Approaches to System Identification*, 18^{th} IFAC Symposium on System Identification (SYSID 2018), Stockholm, plenary presentation retrieved from: `https://www.kth.se/social/files/5b716a5b56be5bd565e07e4f/gianluigi_pillonetto_sysid2018.pdf` (2018).

[30] Pillonetto, G., F. Dinuzzo, T. Chen, G. D. Nicolao and L. Ljung, “Kernel methods in system identification, machine learning and function estimation: A survey”, Automatica **50**, 3, 657–682 (2014).

[31] Pillonetto, G. and G. D. Nicolao, “A new kernel-based approach for linear system identification”, Automatica **46**, 1, 81–93 (2010).

[32] Rallo, G., S. Formentin, A. Chiuso and S. M. Savaresi, “Virtual reference feedback tuning with bayesian regularization”, 2016 European Control Conference (ECC) (2016).

[33] Rallo, G., S. Formentin, A. Chiuso and S. M. Savaresi, *Covariance functions*, `http://evelinag.com/Ariadne/covarianceFunctions.html` (accessed June 26, 2020).

[34] Rasmussen, C. E. and C. K. I. Williams, *Covariance Functions* (MIT Press, Cambridge, MA, 2008).

[35] Rivera, D. E., E. B. Hekler, J. S. Savage and D. S. Downs, “Intensively adaptive interventions using control systems engineering: Two illustrative examples”, Optimization of Behavioral, Biobehavioral, and Biomedical Interventions Statistics for Social and Behavioral Sciences p. 121–173 (2018).

[36] Rivera, D. E., H. Lee, H. D. Mittelmann and M. W. Braun, “Constrained multisine input signals for plant-friendly identification of chemical process systems”, Journal of Process Control **19**, 4, 623–635 (2009).

[37] Rivera, D. E., J. F. Pollard and C. E. Garcia, “Control-relevant prefiltering: a systematic design approach and case study”, IEEE Transactions on Automatic Control **37**, 7, 964–974 (1992).

[38] Sjöberg, J., T. Mckelvey and L. Ljung, “On the use of regularization in system identification”, IFAC Proceedings Volumes **26**, 2, 75–80 (1993).