Dynamic Modeling, System Identification, and Control Engineering Approaches for Designing Optimized and Perpetually Adaptive Behavioral Health Interventions

by

Mohammad T. Freigoun

A Dissertation Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

Approved July 2021 by the Graduate Supervisory Committee:

Gregory B. Raupp, Co-Chair
Konstantinos S. Tsakalis, Co-Chair
Andreas S. Spanias
Erica S. Forzani
Christopher L. Muhich

ARIZONA STATE UNIVERSITY
August 2021
ABSTRACT

Behavior-driven obesity has become one of the most challenging global epidemics since the 1990s, and is presently associated with the leading causes of death in the U.S. and worldwide, including diabetes, cardiovascular disease, strokes, and some forms of cancer. The use of system identification and control engineering principles in the design of novel and perpetually adaptive behavioral health interventions for promoting physical activity and healthy eating has been the central theme in many recent contributions. However, the absence of experimental studies specifically designed with the purpose of developing control-oriented behavioral models has restricted prior efforts in this domain to the use of hypothetical simulations to demonstrate the potential viability of these interventions. In this dissertation, the use of first-of-a-kind, real-life experimental results to develop dynamic, participant-validated behavioral models essential for the design and evaluation of optimized and adaptive behavioral interventions is examined.

Following an intergenerational approach, the first part of this work aims to develop a dynamical systems model of intrauterine fetal growth with the prime goal of predicting infant birth weight, which has been associated with subsequent childhood and adult-onset obesity. The use of longitudinal input-output data from the “Healthy Mom Zone” intervention study has enabled the estimation and validation of this fetoplacental model. The second part establishes a set of data-driven behavioral models founded on Social Cognitive Theory (SCT). The “Just Walk” intervention experiment, developed at Arizona State University using system identification principles, has lent a unique opportunity to estimate and validate both black-box and semiphysical SCT models for predicting physical activity behavior. Further, this dissertation addresses some of the model estimation challenges arising from the limitations of “Just Walk”, including the need for developing nontraditional modeling
approaches for short datasets, as well as delivers a new theoretical and algorithmic framework for structured state-space model estimation that can be used in a broader set of application domains. Finally, adaptive closed-loop intervention simulations of participant-validated SCT models from “Just Walk” are presented using a Hybrid Model Predictive Control (HMPC) control law. A simple HMPC controller reconfiguration strategy for designing both single- and multi-phase intervention designs is proposed.
To the spirit of my mother, Samia.

To my father, Tarig.

To my siblings, Hala, Sara, and Omer.

To Rasitti.

To my beloved fiancé, Douana.
ACKNOWLEDGEMENTS

First and foremost, I would like to express my utmost gratitude to my advisors Professor Gregory B. Raupp and Professor Konstantinos S. Tsakalis for all their time and support, and ultimately for being instrumental in bringing the work of this dissertation to light. I would like to specifically thank them for entrusting me to continue with my research with reasonable autonomy, and provide nothing but constructive feedback, always in the most positive spirit. My special thanks goes to Dr. Paulo Lopes dos Santos whose collaboration and revelations were invaluable for my professional growth in general and for the completion of this work in particular.

Further, I would like to share that it is my greatest honor and privilege to have been taught by Dr. Andreas Spanias, Dr. Erica Forzani and Dr. Christopher Muhich, who also served as members in my dissertation committee, and have their blessings to join the community of scholars. To them I say thank you for your support, valuable insights, and mostly for always putting the interests of your students first. My appreciation to Dr. Bin Mu for serving in my committee before leaving ASU. I am also grateful to the Director of the School for Engineering of Matter, Transport & Energy (SEMTE), Dr. Lenore Dai, the graduate program chair for chemical engineering, Dr. Marylaura Lind Thomas, and the graduate advising manager within SEMTE, Ms. Tiffany Wingerson, for their support throughout my journey in the program. A personal and warm thank you to Ms. Christine Quintero for all her trust and help in navigating through the program, particularly in the most uncertain of times. She will always inspire me to treat all the vulnerable with respect and kindness, and fulfil my job responsibilities with endless dedication and professionalism.

My appreciation goes beyond and includes all co-authors of my publications, with a particular acknowledgment of the guidance from Dr. Daniel Rivera, Dr. Eric Hekler, Dr. Jennifer Savage, and Dr. Danielle Downs during my work in the *Just Walk* and
Healthy Mom Zone projects. A special appreciation to Professor Hans D. Mittelmann who delivered my first and most favorite ASU course of all: APM 523 Optimization (Applied Mathematics).

I also wish to express all the best to my former colleagues, Dr. César Martín and Dr. Penghong Guo, for all the great times and fun conversations we had; thank you for enriching my graduate school experience on many levels. To my dearest friend from ASU, Alicia Magann, who remained in touch even after her graduation: your friendship has always been a source of joy and pride; I sincerely wish you all the best in your professional and personal life.

Last but certainly not least, I would like to all my lifelong friends in Arizona and abroad who continue to inspire me every day. To Mr. John Wayne Mercure, thank you for your generosity. To the dear Nedal Abdeen, thank you for all your support, inspiration, and continued friendship.

Research work presented in this proposal was, in part, funded by the NIH grants R56-HL126799 & R01-HL119245, NSF grant IIS-1449751, and Arizona State University. Opinions expressed in this dissertation are the author’s own and do not necessarily reflect the views of any mentioned individual or entity.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
</tbody>
</table>

## CHAPTER

### 1 INTRODUCTION .................................................. 1

1.1 Motivation ...................................................... 1

1.2 Modeling Behavioral Interventions ...................... 6

1.3 Research Goals ................................................ 11

1.3.1 Developing and Validating a Fetal Model Using Healthy Mom Zone ........................................... 12

1.3.2 Estimating and Validating Social Cognitive Theory Models Using Just Walk .................................. 12

1.3.3 Closed-loop Simulations of Participant-Validated Models Using Hybrid Model Predictive Control .......... 14

1.4 Contributions of the Dissertation ....................... 14

1.5 Dissertation Outline ........................................... 17

1.6 Publications .................................................... 18

1.6.1 Conference Proceedings .................................... 18

1.6.2 Journal Articles ............................................. 19

### 2 A DYNAMICAL SYSTEMS MODEL OF INTRAUTERINE FETAL GROWTH ....................................................... 21

2.1 Background ....................................................... 21

2.2 Fetal Energy Balance Model .................................... 22

2.2.1 Initial Assumptions ........................................ 23

2.2.2 Energy Balance Equation ................................. 23
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.3</td>
<td>Efficiency of Energy Conversion &amp; Energy Balance Reformulation</td>
<td>26</td>
</tr>
<tr>
<td>2.2.4</td>
<td>Rate of Fetal Fat Mass Deposition</td>
<td>32</td>
</tr>
<tr>
<td>2.2.5</td>
<td>Placental Volume Growth Model</td>
<td>33</td>
</tr>
<tr>
<td>2.3</td>
<td>Parameter Estimation and Model Validation</td>
<td>35</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Estimated Fetal Weight</td>
<td>35</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Estimated Placental Volume</td>
<td>36</td>
</tr>
<tr>
<td>2.3.3</td>
<td>Fetal Body Composition</td>
<td>37</td>
</tr>
<tr>
<td>2.3.4</td>
<td>Glycemic Index</td>
<td>37</td>
</tr>
<tr>
<td>2.3.5</td>
<td>Maternal Physical Activity</td>
<td>38</td>
</tr>
<tr>
<td>2.3.6</td>
<td>Maternal Energy Intake</td>
<td>39</td>
</tr>
<tr>
<td>2.3.7</td>
<td>Model Estimation Problem Formulation</td>
<td>39</td>
</tr>
<tr>
<td>2.3.8</td>
<td>Relative Weights &amp; Initialization</td>
<td>41</td>
</tr>
<tr>
<td>2.3.9</td>
<td>Estimation Results</td>
<td>42</td>
</tr>
<tr>
<td>2.3.10</td>
<td>Model Validation</td>
<td>49</td>
</tr>
<tr>
<td>2.4</td>
<td>Chapter Summary</td>
<td>51</td>
</tr>
<tr>
<td>3</td>
<td>BLACK-BOX IDENTIFICATION OF JUST WALK: TOWARD ESTIMATING DYNAMICAL SYSTEMS MODELS OF SOCIAL COGNITIVE THEORY</td>
<td>53</td>
</tr>
<tr>
<td>3.1</td>
<td>Background</td>
<td>53</td>
</tr>
<tr>
<td>3.2</td>
<td>Description of the Just Walk Intervention</td>
<td>54</td>
</tr>
<tr>
<td>3.3</td>
<td>Input Signal Design of Just Walk</td>
<td>55</td>
</tr>
<tr>
<td>3.4</td>
<td>ARX Model Estimation &amp; Validation</td>
<td>58</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Data Pre-Processing and Model Structure Considerations</td>
<td>59</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>3.4.2</td>
<td>Model Parameter Estimation and Validation .......... 61</td>
<td></td>
</tr>
<tr>
<td>3.4.3</td>
<td>Overall Fit Analysis and Assessment of Individual Participant Characteristics ..................... 63</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>Chapter Summary ............................................. 65</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>A SPECTRAL DECOMPOSITION IDENTIFICATION FORMULATION FOR STRUCTURED STATE-SPACE MODELS: ESTIMATING SEMIPHYSICAL MODELS OF SOCIAL COGNITIVE THEORY .......... 66</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>Background .................................................. 66</td>
<td></td>
</tr>
<tr>
<td>4.2</td>
<td>Spectral Decomposition Formulation ................. 68</td>
<td></td>
</tr>
<tr>
<td>4.2.1</td>
<td>Main Formulation ............................................ 69</td>
<td></td>
</tr>
<tr>
<td>4.2.2</td>
<td>Existence, Uniqueness, and Identifiability ............ 71</td>
<td></td>
</tr>
<tr>
<td>4.2.3</td>
<td>Model Realization: An LMI Approach .................. 74</td>
<td></td>
</tr>
<tr>
<td>4.2.4</td>
<td>Standard SD Loss Function ............................... 75</td>
<td></td>
</tr>
<tr>
<td>4.3</td>
<td>Numerical Examples ........................................ 77</td>
<td></td>
</tr>
<tr>
<td>4.3.1</td>
<td>DC Servo Motor ( (p = n,</td>
<td>C</td>
</tr>
<tr>
<td>4.3.2</td>
<td>OCSE Fluid Analogy Model of Social Cognitive Theory ... 78</td>
<td></td>
</tr>
<tr>
<td>4.4</td>
<td>Just Walk: OCSE-SCT Semiphysical Identification Results .......... 81</td>
<td></td>
</tr>
<tr>
<td>4.4.1</td>
<td>Model Validation ............................................ 82</td>
<td></td>
</tr>
<tr>
<td>4.4.2</td>
<td>OCSE Structure: Identifiability Analysis ............... 88</td>
<td></td>
</tr>
<tr>
<td>4.5</td>
<td>Chapter Summary ............................................. 90</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>HYBRID MODEL PREDICTIVE CONTROL FOR CLOSED-LOOP INTERVENTION SIMULATIONS OF PARTICIPANT-VALIDATED SOCIAL COGNITIVE THEORY MODELS ..................... 92</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>Background .................................................. 92</td>
<td></td>
</tr>
<tr>
<td>Chapter</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>5.2</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>5.2.1</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>5.2.2</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>5.2.3</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>5.3</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>5.3.1</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>5.3.2</td>
<td>103</td>
<td></td>
</tr>
<tr>
<td>5.4</td>
<td>105</td>
<td></td>
</tr>
<tr>
<td>5.4.1</td>
<td>105</td>
<td></td>
</tr>
<tr>
<td>5.4.2</td>
<td>107</td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td>107</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>6.1</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>6.1.1</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td>6.1.2</td>
<td>112</td>
<td></td>
</tr>
<tr>
<td>6.1.3</td>
<td>114</td>
<td></td>
</tr>
<tr>
<td>6.2</td>
<td>114</td>
<td></td>
</tr>
<tr>
<td>6.2.1</td>
<td>117</td>
<td></td>
</tr>
<tr>
<td>6.3</td>
<td>119</td>
<td></td>
</tr>
<tr>
<td>6.3.1</td>
<td>119</td>
<td></td>
</tr>
<tr>
<td>6.3.2</td>
<td>120</td>
<td></td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Glossary of the Fetal Model Constants, Parameters, and Variables</td>
</tr>
<tr>
<td>2.2</td>
<td>Mean and Standard Deviation Values of Daily Glycemic Index Estimations for Four Representative HMZ Participants</td>
</tr>
<tr>
<td>2.3</td>
<td>Initialization Points for Four Representative HMZ Participants</td>
</tr>
<tr>
<td>2.4</td>
<td>Estimated Model Parameter Values for Four Representative Hmz Participants with Mean and Standard Deviation (SD)</td>
</tr>
<tr>
<td>2.5</td>
<td>Performance Summary of the Model Developed in Thomas et al., 2008</td>
</tr>
<tr>
<td>2.6</td>
<td>Summary of the Goodness-of-fit from Various Metrics for Four Representative HMZ Participants</td>
</tr>
<tr>
<td>3.1</td>
<td>Intermediate Results for a 4-input ARX Model of a Selected Participant from <em>Just Walk</em></td>
</tr>
<tr>
<td>A.1</td>
<td>Key for Scaling to Reported Ticks from retrieved Graphical Sets 1-15 Used for Chapter 4 (Participants A &amp; B) and Chapter 5 (Participant B)</td>
</tr>
<tr>
<td>A.2</td>
<td>Final Processed Set (1/3) of Published <em>Just Walk</em> Plot from Mercere, 2017 (Participant A) Used in Chapter 4: Intervention Days 0-40</td>
</tr>
<tr>
<td>A.3</td>
<td>Final Processed Set (2/3) of Published <em>Just Walk</em> Plot from Mercere, 2017 (Participant A) Used in Chapter 4: Intervention Days 41-80</td>
</tr>
<tr>
<td>A.4</td>
<td>Final Processed Set (3/3) of Published <em>Just Walk</em> Plot from Mercere, 2017 (Participant A) Used in Chapter 4: Intervention Days 81-87</td>
</tr>
<tr>
<td>A.5</td>
<td>Final Processed Set (1/2) of Published <em>Just Walk</em> Plot from Freigoun et al., 2017 (Participant B) Used in Chapters 4 and 5: Intervention Days 0-50</td>
</tr>
</tbody>
</table>
Table 1

A.6 Final Processed Set (2/2) of Published Just Walk Plot from Freigoun et al., 2017 (Participant B) Used in Chapters 4 and 5: Intervention Days 51-87 ......................................................... 149

A.7 Presentation of Round Off Accuracy and Confidence in the Recovery of Published Graphical Data in 1-15 from Cited Sources ($N_A = N_B = 88$ measurements). .............................................................. 150
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 A Simplified Dynamical Systems Model of Social Cognitive Theory (Freigoun et al., 2017; Hekler et al., 2018).</td>
<td>5</td>
</tr>
<tr>
<td>1.2 A Conceptual Block Diagram for an Integrated, Open-loop Behavioral Intervention Model for Managing GWG (and Fetal Growth) And/Or Obesity.</td>
<td>6</td>
</tr>
<tr>
<td>1.3 A Depiction of an Interdisciplinary Approach for Using Behavioral Theories to Estimate Dynamical Models Using Path Diagrams and Fluid Analogy.</td>
<td>7</td>
</tr>
<tr>
<td>1.4 A Time-domain Visualization of the <em>Goals</em> and <em>Expected Points</em> Input Signals (See Figure 1.1) Using a “Zippered” Spectra Multisine Design for a Representative <em>Just Walk</em> Participant. <em>Goals</em> Are Measured in Steps/Day; <em>Expected Points</em> Are Measured over an Arbitrarily Constructed Point Scale (with 2,500 Points Redeeming a $5 Gift Card).</td>
<td>8</td>
</tr>
<tr>
<td>1.5 Scope (Shaded in Gray) of Modeling Work in This Dissertation with Respect to Selected Relevant Literature.</td>
<td>9</td>
</tr>
<tr>
<td>2.1 Representative Placental Volume Growth Profile.</td>
<td>34</td>
</tr>
<tr>
<td>2.2 Example of an Ultrasound Report for Establishing Estimated Measurements of Estimated Fetal Weight (EFW), Estimated Placental Volume (EPV), and Fetal Body Composition of a Representative HMZ Participant.</td>
<td>36</td>
</tr>
<tr>
<td>2.3 Time-domain Response (Fetal Weight, Placental Volume, and Fetal % Body Fat) with Energy Intake and Physical Activity for a Representative HMZ Intervention Participant (Participant A). Simulation Starts at the Day of First Ultrasound Measurement and Ends at Birth.</td>
<td>44</td>
</tr>
</tbody>
</table>
2.4 Time-domain Response (Fetal Weight, Placental Volume, and Fetal % Body Fat) with Energy Intake and Physical Activity for a Representative HMZ Control Participant (Participant B). Simulation Starts at the Day of First Ultrasound Measurement and Ends at Birth. ........... 44

2.5 Time-domain Response (Fetal Weight, Placental Volume, and Fetal % Body Fat) with Energy Intake and Physical Activity for a Representative HMZ Intervention Participant (Participant C). Simulation Starts at the Day of First Ultrasound Measurement and Ends at Birth. ........ 45

2.6 Time-domain Response (Fetal Weight, Placental Volume, and Fetal % Body Fat) with Energy Intake and Physical Activity for a Representative HMZ Control Participant (Participant D). Simulation Starts at the Day of First Ultrasound Measurement and Ends at Birth. ........ 45

2.7 Fetal Fat Mass and Fat-free Mass Growth Profiles over Time for Representative HMZ Participants (Participants A and B). ......................... 46

2.8 Fetal Fat Mass and Fat-free Mass Growth Profiles over Time for Representative HMZ Participants (Participants C and D). ......................... 46

2.9 Time-varying $\tau_f$ and $e_f$ for Representative HMZ Participants (Participants A and B). ................................................................. 47

2.10 Time-varying $\tau_f$ and $e_f$ for Representative HMZ Participants (Participants C and D). ................................................................. 47

2.11 Time-varying Gain and the Establishment of Positive Fetal Energy Balance for Representative HMZ Participants (See equation (2.23)). ... 48

2.12 Predicted Time-domain Profile of Fetal Energy Intake $I_f(t)$ for Representative Hmz Participants (See equation (2.15)). ......................... 48
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Screenshots of the <em>Just Walk</em> Mobile Application (Phatak et al., 2018).</td>
<td>54</td>
</tr>
<tr>
<td>3.2</td>
<td>Conceptual Representation of a “Zippered” Spectra Design for $n_u = 2$ Design Inputs, and $n_s = 6$ Harmonic Frequencies (Freigoun et al., 2017).</td>
<td>56</td>
</tr>
<tr>
<td>3.3</td>
<td>Time Series Plot Showing Seven Selected Input Sequences (Manipulated Inputs &amp; Measured Disturbances), Predicted Behavior (from an ARX Black-box Model), Actual Behavior, Model Overall Fit, and Estimation &amp; Validation Cycles (1st, 2nd, and 5th for Estimation; 3rd and 4th for Validation) for a Selected <em>Just Walk</em> Participant.</td>
<td>60</td>
</tr>
<tr>
<td>3.4</td>
<td>Average Validation % Fits of Individualized ARX Models from Black-box System Identification for Three Individuals: Goals-Expected Points-Granted Points Model (Base Inputs); B: Predicted Busyness; S: Predicted Stress; T: Predicted Typical; W: Weekday-Weekend.</td>
<td>64</td>
</tr>
<tr>
<td>3.5</td>
<td>Step Responses of a 4-input Model (Including Predicted Busyness) for a Selected Participant.</td>
<td>65</td>
</tr>
<tr>
<td>4.1</td>
<td>Simulation of the SD-identified DC Motor Model (Continuous-Time) From (4.15) Contrasted Against the Deterministic N4SID Model.</td>
<td>77</td>
</tr>
<tr>
<td>4.2</td>
<td>Further Simplified Fluid Analogy of Social Cognitive Theory (OCSE Model).</td>
<td>79</td>
</tr>
<tr>
<td>4.3</td>
<td>Visualization of the Partial Identifiability in $B(\theta_p)$ of the OCSE Model in (4.25) under Three Different Signal-to-noise Ratio (SNR) Scenarios Per Corollary 4.2.2.1. Solver Settings and Random Elements ($T_f$ and output noise) Retained in the Estimated Hypothetical Models for Verified Reproducibility.</td>
<td>81</td>
</tr>
</tbody>
</table>
Figure 4.4 Simulations of the optimal ARX, Standard N4SID, Eigenvalue-constrained N4SID (Realizable) From (4.17) and (4.26), and SD-initialized OCSE-SCT Models for Two Representative Just Walk Participants In Freigoun et al., 2017; Mercere, 2017 (See Appendix A). The Simulations Feature the Predicted Output $\eta_4(t)$ (Behavior) Against Input-Output Data in the Estimation, Validation, and Overall Data. Improving from Freigoun et al., 2021b, Optimal ARX Models were Added and Other Models for Participant A were Re-estimated Following improved settings. The Reader May Use the “zoom” Functionality to Enlarge This Page and View Values in the Electronic Version of This Dissertation.

Figure 4.5 Bland-Altman Agreement Plots for the Four Estimated Behavioral Models of Participant B from Just Walk (Figure 4.4): Optimal ARX, Unconstrained N4SID, Eigenvalue-constrained N4SID, and the Semiphysical OCSE Model of Social Cognitive Theory. Software: Klein, 2014.

Figure 4.6 Residual Analysis Plots for Model Validation. Auto-correlation of Residuals and Cross-Correlations with Input Signals are Plotted Against 99% Confidence Intervals for the Four Estimated Behavioral Models of Participant B from Just Walk (Figures 4.4 and 4.5): Optimal ARX, Unconstrained N4SID, Eigenvalue-constrained N4SID, and the Semiphysical OCSE Model of Social Cognitive Theory. Frames in Red (Color) Emphasize Model ‘Falsification’ by Data.

Figure 5.1 Conceptual Application of the Receding Horizon Control Strategy to the Physical Activity Behavioral Problem (Hekler et al., 2018).
5.2 Time Series Plot Illustrating Input Signals, Measured Changes in Behavior (Output) [Steps/Day], and the Predicted Behavior from the OCSE Model Using Retrieved Input-Output Participant Data in Freigoun et al., 2017 (see Appendix A) with \( n_d = 4 \). NRMSE Fits: Estimation (34.2%, Green Region); Validation (73.65%, Cyan Region); Overall Data (41.5%). .............................................. 98

5.3 Step Responses of the Semiphysical OCSE Model for a Just Walk Participant over a 7-day (Week) Prediction Horizon. ...................... 99

5.4 HMPC Closed-loop Performance Simulations (Single-phase Intervention) of a Participant-validated OCSE-SCT Model. HMPC Settings: Sampling Time \( T_s = 1 \) day, \( N = 3 \), \( u_{\min} = \begin{bmatrix} 4000 & 0 & 0 \end{bmatrix}^T \), \( u_{\max} = \begin{bmatrix} 10000 & 500 & 500 \end{bmatrix}^T \), \( y_{\min} = 0 \), \( Q_y = 1 \), \( Q_u = \text{diag}\{0, [0.005, 0.25, 2], 0\} \), \( c_8 = 4000 \), \( \nu_j|_{U_8} = \{1, 1^1/8, 1^3/16, 1^1/4, 1^3/8, 1^1/2\} \), \( c_9 = 100 \), \( \nu_j|_{U_9} = \{0, 1, 2, 3, 4, 5\} \).................................................. 106

5.5 HMPC Closed-loop Performance Simulations (Multi-phase Intervention) of a Participant-validated OCSE Model. HMPC Settings Indicated in the Caption of Figure 5.4 Apply with the Following Exceptions: Only the Response Of \( Q_u = \text{diag}\{0, 0.1, 0\} \) is simulated; \( u_8^{\max} = 6000 \forall t \in [0, 80] \); \( u_8^{\max} = 8000 \forall t \in [81, 160] \); \( u_8^{\max} = 10000 \forall t \in [161, 240] \). .............................................................. 109
Chapter 1

INTRODUCTION

1.1 Motivation

System identification and control are among the most mature fields in engineering that offer rigorous mathematical framework for the characterization of complex dynamical behaviors of causal systems, and ultimately utilize predictive models to establish optimized real-time solutions in a wide array of applications. Beyond process design and other traditional industrial settings, system identification and control research have been expanding and now include novel applications in new areas such as behavioral health and medicine. Despite the ever-growing body of scientific literature and subsequent efforts devoted to public awareness, some of the most pressing challenges in behavioral health remain unresolved. In the United States, according to current reports from Centers for Disease Control and Prevention (CDC), among the top ten public health problems are obesity, HIV, drug abuse, and tobacco use, with obesity strongly tied to other problems in this category such as heart disease and stroke. Over the recent decades, following smoking, the obesity pandemic was the second leading cause of preventable deaths estimated at 300,000 mortalities per year in the United States alone (Flegal et al., 2004). Moreover, being associated with type 2 diabetes, hypertension, heart disease, asthma, depression, and even cancer among other conditions, the United States national direct costs of obesity were estimated at $147 billion in 2008 (Finkelstein et al., 2009; Cawley and Meyerhoefer, 2012). Indirect costs including obesity-related absenteeism have ranged between $3.38 and $6.38 billion (Trogdon et al., 2008).
While it is more difficult to prevent some of the risk factors such as age, family history and genetic factors, regulation of one’s habits and behavior concerning physical activity, diet, and tobacco use is indeed more approachable, and can remarkably increase life expectancy. For example, substantial evidence shows that physical activity reduces chronic disease risk (Bauer et al., 2014; Owen et al., 2008; Haskell et al., 2009). With national guidelines suggesting 30-60 minutes of moderate physical activity per day, often in the form of walking (U.S. Department of Health and Human Services, 2008); research furthermore demonstrates a 20-30% reduced risk of breast cancer when guidelines are met (Thune et al., 1997). Despite the well-known benefits of physical activity, a large segment of the U.S. population does not meet these guidelines (Troiano et al., 2008).

In addition to sedentary behavior, Gestational Weight Gain (GWG) and high birth weight have also been independently associated with subsequent childhood and adult-onset obesity, cardiovascular disease, and some forms of cancer (Rich-Edwards et al., 1997; Hillier et al., 2007; Spracklen et al., 2014; Savage et al., 2014; Qiao et al., 2015; O’Neill et al., 2015). While it is known that these associations can impact both women and their offspring, adherence of pregnant women in the United States to guidelines from the Institute of Medicine for appropriate GWG is only at 30%, with over 50% exceeding guidelines toward overweight and morbid obesity (Dudenhausen et al., 2015). These low levels of adherence to recommended guidelines indeed call for alternative interdisciplinary strategies that can invite a broader set of tools and research methods.

Only over the past decade, following the successful use of fluid analogy descriptions in production-inventory systems (Schwartz et al., 2006), the Rivera et al., 2007 article introduced the system identification and control engineering framework as a viable and promising approach for the design of novel adaptive interventions in the
context of behavioral health. In the article, it was pointed out that tools acquired from dynamical systems modeling and control have “the potential to significantly inform the analysis, design, and implementation of adaptive interventions, leading to improved adherence, better management of limited resources, a reduction of negative effects, and overall more effective interventions.” Admittedly, over the last decade, this proposition has gained a sizable attraction from the behavioral science and medicine society, and has led to the development of new dynamical models of behavior change, which in turn inspired a family of adaptive intervention designs using control systems engineering ideas, each targeting a specific behavioral health challenge.

Toward addressing obesity and meeting a balanced dietary and physical activity lifestyle, similar to Schwartz et al., Navarro-Barrientos et al., 2011 proposed fluid analogy as a suitable tool for mechanistic modeling of complex dynamical systems, leading to the proposal of a dynamic behavior change model based on the Theory of Planned Behavior (TPB) (Ajzen, 1991). This TPB model was then integrated with a three-compartment, energy balance model developed in Hall and Jordan, 2008; Chow and Hall, 2008; Hall, 2009, 2010 and validated by the Minnesota Semi-Starvation Experiment (Keys et al., 1950) to establish predictions of human weight gain/loss that are simultaneously driven by human psychology and physiology. Using the integrated behavioral-energy balance model, a classical feedback control system design was used to produce simulations illustrating optimized assignments of various intervention dosages over time. Moreover, the work of Dong et al., 2012, 2013, 2014; Savage et al., 2014 presented a receding horizon control strategy and a more elaborate intervention modeling that includes self-regulation and intervention delivery dynamics for designing improved GWG interventions and regulating infant birth weight; the potential in the proposed intervention strategies were illustrated by evaluating sim-
ulations of hypothetical human subjects. More recently, these efforts were followed by Guo et al., 2016, 2017, 2018; Freigoun et al., 2018; Pauley et al., 2018; Symons Downs et al., 2018 in which models of real human participants are estimated and validated from experimental data; these works have contributed in both modeling and improved intervention design for managing GWG and regulating infant birth weight.

Furthermore, alternative to traditional static (i.e., steady-state) modeling methods and “one-size-fits-all” strategies for smoking cessation, contributions from Timms et al., 2013, 2014d,a,c,b presented an alternative dynamic modeling and control engineering-based intervention design using a mediational understanding of smoking cessation dynamics. An additional novel application of system identification and control was in the domain of fibromyalgia pain management (Deshpande et al., 2011, 2012, 2014b,a). Indeed, the use of behavioral (or integrated behavioral-energy balance) models in cited literature has delivered an entirely novel conceptual paradigm for the design of personalized and perpetually adaptive behavioral health interventions. Without a doubt, this was realized by inviting system identification and control engineering principles to the design of behavioral health interventions that seek to improve and sustain adherence to recommended dietary and physical activity guidelines. This appeal to such a design is further magnified by the explosion of advanced mobile health (mHealth) sensor technologies (Nilsen et al., 2017) that not only offer a cost-effective deployment of adaptive interventions at the largest of scales, but also facilitate the evaluation and improvement of existing behavioral theory models by leveraging intensive longitudinal data (Riley et al., 2011).

In light of the discussed challenges, this dissertation is primarily motivated by the need to use real-life experimental studies to advance the system identification, control, and behavioral science research, including the development of new practical approaches and algorithms in these domains. To achieve this, two recent novel
Figure 1.1: A Simplified Dynamical Systems Model of Social Cognitive Theory (Freigoun et al., 2017; Hekler et al., 2018).

open-loop intervention studies, Healthy Mom Zone (Symons Downs et al., 2018), and Just Walk (Hekler et al., 2018), are utilized in developing, estimating, and validating behavioral and energy balance models; an established participant-validated model will serve as basis for closed-loop intervention design. More specifically, this dissertation aims to answer calls from two 2014 papers: Savage et al. and Martín et al. In the former, authors called for the need to utilize experimental data from a real-life intervention (such as Healthy Mom Zone) to improve, estimate, and validate their proposed maternal-fetal intervention model. As a result, chapter 2 is entirely devoted for the development of a dynamical systems model of intrauterine fetal growth, validated by data from Healthy Mom Zone.

The rest of this dissertation efforts are motivated by calls from Martín et al., 2014, 2016a for estimating and validating the semiphysical model of Social Cognitive Theory (Bandura, 1986) in Figure 1.1 using data from real human subjects, and ultimately
utilize these models to further support their hypothetical simulations of closed-loop physical activity interventions. In this work, these objectives are approached in the context of *Just Walk*, a more detailed description of which is provided in Chapter 3.

1.2 Modeling Behavioral Interventions

In the design of adaptive behavioral health interventions, depending on the target outcome (e.g., adherence to a prescribed level of daily physical activity, healthy eating, weight gain/loss, healthy infant birth weight), a model-based approach must utilize *a priori* knowledge of the underlying psychological and/or physiological mechanisms driving the particular outcomes of interest. For example, an intervention that aims to use a human psychology approach in order to achieve and maintain a prescribed level of physical activity will ultimately need to rely on a particular cognitive basis for predicting individual responses to various external cues and other contextual factors that influence behavior. A behavioral theory such as Social Cognitive Theory (SCT) can indeed lend a strong theoretical basis for establishing dynamical systems models that suit this purpose. Furthermore, if the intervention simultaneously targets weight gain/loss and promoting consistent physical activity, then it is natural to include an understanding for metabolic and other physiological mechanisms alongside the behavioral model on some level; this is also feasible by using an energy balance ap-
Typically, energy balance models can be developed by applying the laws of thermodynamics, as well as the incorporation of empirical functions for some of the established variables. This will be visited with more depth in the context of developing a dynamical systems model of intrauterine fetal growth in Chapter 2. On the other hand, the modeling of behavioral theories such as SCT (Martín, 2016) or TPB (Guo, 2018) has proven to be a more challenging involved task in many respects. Figure 1.3 features a pathway from using behavioral theories to obtain control-oriented, dynamical models useful for adaptive intervention design. First, a behavioral theory in its abstract form is adopted, followed by harnessing all constructs (e.g., self-efficacy, subjective norm) originating from that theory. In the case of modeling and TPB, recent efforts (Navarro-Barrientos et al., 2011; Dong, 2014; Martín, 2016; Guo, 2018) have collaborated in multidisciplinary research teams that included experts from the fields of psychology and behavioral medicine. Second, the considered behavioral theory is used to construct a path diagram characterizing the causal relations among the

Figure 1.3: A Depiction of an Interdisciplinary Approach for Using Behavioral Theories to Estimate Dynamical Models Using Path Diagrams and Fluid Analogy.
Figure 1.4: A Time-domain Visualization of the Goals and Expected Points Input Signals (See Figure 1.1) Using a “Zippered” Spectra Multisine Design for a Representative Just Walk Participant. Goals Are Measured in Steps/Day; Expected Points Are Measured over an Arbitrarily Constructed Point Scale (with 2,500 Points Redeeming a $5 Gift Card).

listed constructs (e.g., mediation, moderation, feedback) with the ultimate goal of describing a pathway from each measurable predictor (system input) to a measurable outcome (system output).

Next, the established path diagram can be used to construct a fluid analogy model (similar to Figure 1.1), which is most valuable for introducing the dynamical nature of all causalities framed by the obtained path diagram. For example, Figure 1.1 presents a fifth-order SCT model comprised of five interacting subsystems, each characterizing first-order dynamics. Using this fluid analogy model, a system of ordinary linear differential equations (ODEs) can be derived using the conservation of mass principle (i.e., the First Law of Thermodynamics) and, subsequently, a state-space model. Note that, conceptually, the assumed first law of thermodynamics need not be observed following the determination of a behavioral state-space representation.

Finally, with a state-space representation at hand, a “patient-friendly” experiment
design based on system identification principles may be in order. In Just Walk, as will be further described in Chapter 3, the input signal design used sinusoidal excitations for the two input-output channels (goals and expected points) of the SCT model in Figure 1.1. The designed multisine perturbations feature a “zippered” spectra design for the purposes of generating time-domain signals that are orthogonal in frequency (Freigoun et al., 2017). A time-domain visualization of the goals and expected points input signals using a “zippered” spectra multisine design for a representative Just Walk participant is shown in Figure 1.4. Finally, with a produced set of measured input-output data obtained from the identification experiment, parameter estimation and model validation follow; this is the necessary, final step that will be explored in depth for the SCT model in Figure 1.1 in the context of Just Walk. Some of the key contributions of this work is concerned with SCT model estimation and validation.

To further clarify the scope of this work, Figure 1.5 highlights areas of specific contributions in this application domain with respect to selected recently-produced body of work. The modeling of intrauterine fetal growth presented in Chapter 2 follows a white-box identification strategy, while the identification of participant-validated
SCT models using Just Walk relies on both, black- and grey-box estimation techniques. Since this terminology will be intensively used in the description of this work, the following introduction for the various types of dynamical systems models is borrowed from mainstream system identification (Ljung, 1994; Lindskog and Ljung, 1995; Lindskog, 1996; Ljung, 1999):

- **White-box (physically parameterized) models.** These are typically models of physical or biological systems that are derived from first-principles modeling (i.e., using laws of thermodynamics, motion, etc.), and incorporate all insights and knowns about the system behavior, usually in state-space form. They often contain known and unknown parameters; parameters have a physical significance of their own (i.e., meaning, units, etc.), such as unknown physical constants that are estimated from experimental data. While it is certainly possible to utilize these models in control system design (in fact are preferred if available), these are typically general-purpose, theoretical formulations that result from laborious modeling aimed at developing a better understanding and intuition for certain “truths” about the system in question.

- **Black-box models.** This type of models are data-driven, usually in the form of linear differential or difference equations. Neither the structure nor the parameters of a black-box model necessarily have physical significance. With the objective of fitting observed data to reproduce the past behavior of the system as accurate as possible, black-box models are typically easy to produce and used for prediction/simulation and/or control purposes. These types of models may also be estimated and studied in a preliminary step preceding structured modeling (Ljung, 1994).

- **Gray-box (semiphysical) models.** As the name implies, these models fall between
white- and black-box models, where prior knowledge about the concerned system is used in building a parsimonious model structure that is (i) intuitively descriptive of the system’s components, and (ii) predictive of the system’s dynamical behavior. However, building this structure is not carried out to the extent that a formal physically parametrized model is constructed since parameters may not have a direct physical interpretation. Gray-box system identification is particularly attractive in modeling and understanding causality in nonphysical systems and between abstract constructs. Indeed, in lieu of white-box models, gray-box models are also often employed in practice as “pragmatic” models (Ljung, 1999) when it is necessary or more practical to simplify the modeling of complex physical systems. Semiphysical models typically have the potential for serving all modeling purposes: theoretical, control, and simulation/prediction.

Upon the conclusion of all identification efforts, the established dynamical models can be used for control system design. Ultimately, a model-based intervention design that uses control systems engineering offer the ability to optimize intervention dosages according to individual characteristics (e.g., goal-driven, weekend warrior) and the ever-changing contextual factors (e.g., busyness, weather), while simultaneously respect economical and a host of other constraints.

1.3 Research Goals

At the highest level, the goal of this dissertation is to utilize experimental data drawn from recent novel behavioral interventions to provide a data-driven basis for supporting a number of recent works in this application domain. Almost at the same level, a main objective is to develop practical methods and glean insights that can prove useful for the future researcher or user interested in translating behavioral theories into dynamical models amenable to system identification and control system
design. More specifically, the goals of this dissertation include developing and validating a fetal model using Healthy Mom Zone, estimating and validating semiphysical Social Cognitive Theory models using Just Walk, and the evaluation of closed-loop simulations of participant-validated models using Hybrid Model Predictive Control.

1.3.1 Developing and Validating a Fetal Model Using Healthy Mom Zone

The underlying mechanisms for how maternal perinatal obesity and intrauterine environment influence fetal development are not well understood and thus require further understanding. In this dissertation, the goal is to use first principles modeling (i.e., energy balance and entropy concepts) for developing a comprehensive dynamical systems model for fetal growth that illustrates how maternal factors (energy intake and physical activity) influence fetal weight and related components (fat mass, fat-free mass, and placental volume) over time. Using Healthy Mom Zone (HMZ), a novel intervention for managing gestational weight gain in obese/overweight women, a more specific goal is to estimate and validate the developed fetoplacental model given intensive measurements of fetal weight and placental volume obtained from sonographic imaging technology. Ultimately, we aim to deliver a parsimonious system of equations that can reliably predict fetal weight gain and birth weight based on a sensible number of assessments, with a proven ability to inform clinical care recommendations and show how adaptive, HMZ-like interventions can influence fetal growth and birth outcomes.

1.3.2 Estimating and Validating Social Cognitive Theory Models Using Just Walk

Following the delivery of a more detailed description of Just Walk, an intensively adaptive physical activity intervention that has been designed on the basis of system identification and control engineering principles, a second major goal of this disserta-
tion is to further establish the viability of dynamical systems framework in capturing the key factors that influence and predict behavior change over time. To accomplish this, we aim to develop an unconventional ARX estimation-validation procedure that better suits *Just Walk* and seeks to balance predictive ability over validation data segments with overall goodness of fit. Black-box models (i.e., ARX models) come with the value of providing important clues to individual participant characteristics that influence physical activity; these insights will prove to be critical in building semiphysical models lending their theoretical basis from Social Cognitive Theory.

Further, while the semiphysical SCT model in Figure 1.1 can rise with favorable statistical properties compared to its fully-parametrized black-box counterpart, solver initialization of classical methods and structural identifiability often pose a challenge to the user seeking satisfactory results. Thus, a crucial step in this work is to apply a judicious model reduction thinking to propose a lower-complexity version of the SCT model in Figure 1.1; one goal is to show that this could be approached from both a conceptual and data-driven perspectives. Moreover, By assuming distinct poles and zero-order hold intersample behavior of the underlying system (similar to *Just Walk*), an important technical goal of this dissertation is show that the typical grey-box constrained optimization problem can be formulated into an easier one by solving linearly-constrained eigenvalue problems. Here, the proposed strategy is to follow the trend of existing literature and develop a formulation that relies on a consistent discrete-time black-box model (e.g., N4SID) to solve for a structured, continuous-time one in the absence of prior knowledge. Following the development of this grey-box identification algorithm, the goal is to establish semiphysical, participant-validated models of Social Cognitive Theory.
1.3.3 Closed-loop Simulations of Participant-Validated Models Using Hybrid Model Predictive Control

A final goal of this work is to produce and evaluate closed-loop intervention simulations of a participant-validated semiphysical SCT model from the previous section using published data from *Just Walk*. Consistent with real-world requirements, including the need for hybrid decision rules policies that factor in logical, physical, and financial constraints, it is only natural that considered closed-loop intervention designs in this dissertation would follow a Hybrid Model Predictive Control formulation. Since it is conceivable that clinical or psychological considerations may relax or require the gradual change of behavior, it is desired to examine simulations that feature single-phase as well as multi-phase interventions.

1.4 Contributions of the Dissertation

As highlighted earlier, the contributions from this work can be viewed under two major themes: First, harnessing experimental data to deliver an intrauterine fetal growth model relying on first principles. Second, the proven advancements in the domain of designing adaptive physical activity behavioral health interventions using Social Cognitive Theory models. In terms of the former part, the contributions of this dissertation are summarized as follows:

1. *A single-output energy balance model*. Building from the model in (Thomas *et al.*, 2008), the proposed energy balance model in this work features a single, easy-to-measure output (total fetal weight). In addition to grounding a better theoretical understanding of external factors and pre-existing conditions directly influence intrauterine fetal weight growth, this reformulation highlights less expensive and invasive requirements for estimating individualized model
parameters; measurements of total fetal weight can be far more reliable than measurements of body composition (more so in the first trimester; Bernstein and Catalano, 1991).

2. **Application of the Second Law of Thermodynamics.** The fetal energy balance model in this work provides a succinct, well-established accounting for the impact of entropy on fetal growth. Despite that the idea of entropy of new tissue formation has originated in Christiansen et al., 2005, the work in Christiansen et al., 2005 features an obesity model and the formulation cannot be directly re-purposed for quantifying fetal growth dynamics. Part of the contribution in this work is to merge efforts from Christiansen *et al.* and Thomas *et al.* to produce a more rigorous and complete reformulation of fetal growth.

3. **Use of HMZ study data.** Utilizing data from the HMZ study (Symons Downs *et al.*, 2018), the developed fetal model presents a method for quantifying the impact of daily changes of physical activity on fetal growth. Moreover, using intensive, longitudinal participant data from HMZ, it is possible to estimate and validate the general first-principles fetal model structure developed in this work, as well as estimate a logistic profile of fetal fat mass accretion whose structure is supported by the literature.

4. **An improved placental volume model formulation.** As is discussed in Section 2.2.5, in this work, the curvature of the proposed placental volume model is more independently parameterized, which gives a more intuitive and easier model to estimate. This model also implicitly enforces the initial condition at conception; hence, for model estimation and simulation, the proposed model does not require a placental volume measurement for establishing an initial condition.
In terms of advancing intervention design using Social Cognitive Theory, the contributions of this work are:

1. **Black-box modeling of Just Walk.** Using input-output participant data, a library of ARX models were estimated and validated following a nontraditional, exhaustive modeling strategy. This contributions has proven useful as the same conceptual approach was applied in estimating and validating a host of MoliZoft models in dos Santos *et al.*, 2017, 2018, as well as in harnessing models that help improve a cocoa biofertilization process in Gallino *et al.*, 2018. Results and insights from this work has been cited in a relevant application setting that explore “modeling human-in-the-loop behavior and interactions with HVAC systems” (Kane, 2018).

2. **Semiphysical modeling of Just Walk.** In a necessary step mandated by the limitations of Just Walk, this work has introduced a further simplified, third-order semiphysical model of Social Cognitive Theory that is supported by experimental data. This further simplified model structure aims to capture behavior-change dynamics resulting from operant conditioning and self-efficacy. Above all, this dissertation claims to deliver the first estimated semiphysical models of Social Cognitive Theory using real input-output data drawn from an single-subject experiment that was designed based on system identification principles. The estimated models were validated using classical cross-validation tools from system identification and other disciplines.

3. **Identification framework for structured (grey-box) state-space models.** Inspired by challenges arising from Just Walk, a spectral decomposition identification algorithm was developed, along with its theoretical framework, and was intended to relieve the user from the burden of judicious solver initialization in
the absence of sufficient prior knowledge. Conditions relevant to the existence, uniqueness, and identifiability of linear grey-box structures under the proposed framework were provided. A numerical example illustrating the effectiveness of this formulation is provided. An extension to the main formulation that aims to accommodate quadratic structures was presented.

4. Closed-loop simulations of participant-validated models. Using HMPC control law together with a participant-validated SCT model from published Just Walk data, it is also claimed that the work of this dissertation is first to deliver closed-loop simulations of a data-driven participant model. By applying a simple controller reconfiguration, evaluated simulations in this work feature both single- and multi-phase intervention designs to provide more flexibility to the user.

1.5 Dissertation Outline

Following this introduction, the dissertation continues with Chapter 2 which introduces the Healthy Mom Zone study and utilizes it in the process of estimating and validating a quasi-linear parameter-varying fetoplacental model. The First and Second Laws of Thermodynamics are used to develop a dynamical systems, energy balance model of intrauterine fetal growth. Details of the estimation and validation steps are discussed, with simulations of multiple Healthy Mom Zone participants over the second and third trimesters presented.

Next, this dissertation moves to introduce the modeling and identification efforts using Just Walk in Chapter 3. Further details about the Just Walk intervention is provided, including the experimental design methodology. Further, results obtained from black-box modeling and estimation are presented, underscoring the amenability of the dynamical systems framework to capturing and predicting human behavior
change over time. Results from human subject models highlighting the idiosyncratic
tonature of human behavior are outlined.

In Chapter 4, a spectral decomposition identification formulation for structured
state-space models is delivered along with its theoretical foundation. This general-
purpose formulation is intended to address one of the long-existing grey-box identifi-
cation challenges that presented a barrier to achieving some of the goals of this work.
The developed algorithm is utilized in the process of estimating data-driven semi-
physical models of Social Cognitive Theory. Results establishing a basis for model
validation are also included.

Using a semiphysical model of Social Cognitive Theory developed in Chapter 4,
Chapter 5 features an evaluation for the time-domain responses of a Just Walk par-
ticipant. A brief overview on the general and application-specific HMPC formulations
is presented. Further, the participant-validated SCT model is used to produce and
evaluate HMPC-governed closed-loop simulations of a physical activity intervention.
Both single- and multi-phase intervention designs are proposed and simulated.

Finally, the dissertation concludes with Chapter 6, providing a summary and
some of the key conclusions stemming from this work. The chapter includes a brief
discussion of some of the potential extensions to this work.

1.6 Publications

The following are lists of conference proceedings and journal papers highlighting
published contributions over the course of this research.

1.6.1 Conference Proceedings

mHealth Intervention for Promoting Physical Activity”, in “American Control Conference (ACC), 2017”, pp. 116-121 (2017). Chapter 3 is adapted from this publication.


1.6.2 Journal Articles


A DYNAMICAL SYSTEMS MODEL OF INTRAUTERINE FETAL GROWTH

2.1 Background

Prior work has described a conceptual framework for managing GWG among overweight/obese women (Symons Downs et al., 2018) and for regulating infant birth weight (Savage et al., 2014); this framework relies on methods from control systems engineering to develop decision policies that optimize the adaptation for participant response. The implementation of such a framework calls for developing advanced control systems which rely on dynamical models that are able to predict individualized responses to different intervention components and subsequently predict GWG, the intrauterine growth profile, and infant birth weight (Savage et al., 2014; Thomas et al., 2012; Guo et al., 2016). In particular, one important end use of a dynamical systems model of intrauterine fetal growth is as the internal model in a model-based controller that accomplishes an optimized, adaptive intervention (Rivera et al., 2007, 2017, 2018).

Energy balance for modeling weight and body composition change has been examined extensively, including among pregnant women (Hall, 2014; Thomas et al., 2012). Modeling intrauterine growth has received some prior examination Thomas et al., 2008; however, further modeling efforts are needed to better understand how prenatal status ‘programs’ fetal growth (Chandler-Laney and Bush, 2011; Catalano and Ehrenberg, 2006). To address this gap, we use intensive longitudinal data from Healthy Mom Zone (HMZ) (Symons Downs et al., 2018), an ongoing trial, which is an individually-tailored, adaptive intervention to manage weight gain in overweight and obese preg-
nant women. While the model developed in this work extends from prior work (Thomas et al., 2008; Christiansen et al., 2005), it grounds a more complete theoretical understanding for how external maternal factors (e.g., daily energy intake and physical activity) influence fetal growth profiles.

In this chapter, we present parameter estimation and model validation results drawn from four representative HMZ participants. The final fetal energy balance model parameters are estimated by solving a nonlinear least squares optimization problem; the set of estimated model parameters is then used to generate simulations for model validation.

This chapter is organized as follows: Section 2.2 presents the underlying modeling assumptions and describes the derivation of the proposed fetal energy balance model. Section 2.3 features the optimization problem that accomplishes model parameter estimation from ultrasound measurements, followed by a presentation of the metrics and criteria used for model validation. Section 2.4 summarizes conclusions and future work.

2.2 Fetal Energy Balance Model

We begin by outlining important assumptions and simplifications leading to the final proposed fetal energy balance model. Next, building on insights from prior researchers (Christiansen et al., 2005; Thomas et al., 2008), we establish a first-principles energy balance model of fetal growth. Following the first law of thermodynamics, this fetal energy balance model applies the conservation of energy principle. Further, the presented derivation explicitly accounts for the energy loss due to new fetal tissue formation, as dictated by the second law of thermodynamics from which it follows that the conversion of energy requires energy. Finally, explicitly defined logistic growth functions are established to estimate the rate of fetal fat mass deposition.
and the placental volume.

2.2.1 Initial Assumptions

The following initial assumptions and simplifications lead to the proposed fetal model (equation (2.21)):

1. Fetal body mass is divided into two main components: fat and fat-free tissues.

2. Fetal energy expenditure due to diet-induced thermogenesis is negligible.

3. Fetal physical activity in the womb is negligible.

4. The rate of fetal fat mass deposition is only regulated by the total fetal body mass (Christiansen et al., 2005).

5. The contribution of daily maternal diet to fetal nutrition substantially exceeds additional nutrient supply originating from maternal body components.

6. Fetal energy imbalances are always positive and follow from the diet of a healthy, well-nourished mother.

7. The proportion of fetal body fat that contributes to expenditure is equal to that of fetal fat-free tissues.

Following the derivation of the fetal model in Sections 2.2.2 and 2.2.3, a discussion of the rationale for assumptions 6. and 7. is provided.

2.2.2 Energy Balance Equation

The basis for determining fetal growth is a daily energy balance based on the First Law of Thermodynamics that takes into account the metabolizable energy intake $I_f$ (provided by the mother) and fetal energy expenditure $E_{ef}(t)$ to define a rate of
Table 2.1: Glossary of the Fetal Model Constants, Parameters, and Variables

<table>
<thead>
<tr>
<th>Constants</th>
<th>Definition</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_{FM_f}$</td>
<td>Energy stored per unit fetal fat mass</td>
<td>kcal/kg</td>
</tr>
<tr>
<td>$\lambda_{FFM_f}$</td>
<td>Energy stored per unit fetal fat-free mass</td>
<td>kcal/kg</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Daily energy expenditure per unit fetal body mass</td>
<td>kcal/kg/d</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Definition</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_{FM_f},e_{FFM_f}$</td>
<td>Efficiencies of conversion of excess energy to new fat and fat-free tissues</td>
<td>[1]</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Proportionality constant</td>
<td>d/kcal/ml</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Conversion coefficient</td>
<td>ml$^{-1}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variables</th>
<th>Definition</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>Gestational age</td>
<td>days</td>
</tr>
<tr>
<td>$C_f(t)$</td>
<td>Daily energy accumulation in the fetus</td>
<td>kcal</td>
</tr>
<tr>
<td>$I_f(t)$</td>
<td>Daily fetal energy intake resulting from maternal energy intake</td>
<td>kcal/d</td>
</tr>
<tr>
<td>$E_{e_f}(t)$</td>
<td>Total fetal energy expenditure</td>
<td>kcal/d</td>
</tr>
<tr>
<td>$E_{M_f}(t)$</td>
<td>Energy required to maintain the fetus life</td>
<td>kcal/d</td>
</tr>
<tr>
<td>$E_{e_f}(t)$</td>
<td>Energy required for the conversion of excess energy into new fetal tissue</td>
<td>kcal/d</td>
</tr>
<tr>
<td>$FM_f(t)$</td>
<td>Fetal fat mass</td>
<td>kg</td>
</tr>
<tr>
<td>$FFM_f(t)$</td>
<td>Fetal fat-free mass</td>
<td>kg</td>
</tr>
<tr>
<td>$W_f(t)$</td>
<td>Total fetal weight</td>
<td>kg</td>
</tr>
<tr>
<td>$f_r(W_f)$</td>
<td>Rate of fetal fat mass deposition</td>
<td>$1=kg_{FM_f}/kg_{W_f}$</td>
</tr>
<tr>
<td>$W_m(t)$</td>
<td>Total maternal weight</td>
<td>kg</td>
</tr>
<tr>
<td>$E_f(t)$</td>
<td>Total energy to build the fetal tissue up to day $t$</td>
<td>kcal</td>
</tr>
<tr>
<td>$E_{FM_f}(t)$</td>
<td>Total energy to build the fetal fat tissue up to day $t$</td>
<td>kcal</td>
</tr>
<tr>
<td>$E_{FFM_f}(t)$</td>
<td>Total energy to build the fetal fat-free tissue up to day $t$</td>
<td>kcal</td>
</tr>
<tr>
<td>$m(t)$</td>
<td>Maternal energy intake</td>
<td>kcal/d</td>
</tr>
<tr>
<td>$PA(t)$</td>
<td>Maternal physical activity</td>
<td>kcal/d</td>
</tr>
<tr>
<td>$P(t)$</td>
<td>Placental volume</td>
<td>ml</td>
</tr>
<tr>
<td>$g(t)$</td>
<td>Glycemic impact of intake</td>
<td>[1]</td>
</tr>
<tr>
<td>$K_f(t)$</td>
<td>Fetal gain coefficient from intake</td>
<td>kg·d/kcal/ml</td>
</tr>
<tr>
<td>$\tau_f(W_f)$</td>
<td>Time constant of fetal weight growth</td>
<td>d</td>
</tr>
<tr>
<td>$e_f(W_f)$</td>
<td>Overall efficiency of energy conversion</td>
<td>[1]</td>
</tr>
</tbody>
</table>
accumulation of the total fetal energy $C_f(t)$. Considering the fetus as the system of
interest, we have

$$\frac{dC_f(t)}{dt} = I_f(t) - E_{e_f}(t) \quad (2.1)$$

with

$$E_{e_f}(t) = E_{e_f} = E_{M_f}(t) + E_{c_f}(t) \quad (2.2)$$

accounting for fetal energy expenditure towards maintaining and sustaining life ($E_{M_f}$)
and the energy required for the conversion of excess energy into new tissue ($E_{c_f}$). A
full glossary of model components is presented in Table 2.1.

Considering a two-compartment energy balance model (i.e., total body mass di-
vided into fat and fat-free mass components), the positive rate of change of the total
combustible fetal energy content, $dC_f/dt$, can also be calculated by accounting for
changes of fetal body components Thomas et al., 2008, giving

$$\frac{dC_f(t)}{dt} = \lambda_{FM_f} \frac{dM_f(t)}{dt} + \lambda_{FFM_f} \frac{dFFM_f(t)}{dt} \quad (2.3)$$

which, in turn, when combined with (2.1), yields

$$\lambda_{FM_f} \frac{dM_f(t)}{dt} + \lambda_{FFM_f} \frac{dFFM_f(t)}{dt} = I_f(t) - E_{e_f}(t) \quad (2.4)$$

with $FM_f(t)$ and $FFM_f(t)$ denoting the total fetal fat and fat-free masses, re-
respectively; $\lambda_{FM_f}$ and $\lambda_{FFM_f}$ are the energy densities of the fetal fat and fat-free com-
ponents, respectively (i.e., energy content per unit fat/fat-free mass). As depicted
in equation (2.4), both $\lambda_{FM_f}$ and $\lambda_{FFM_f}$ are assumed time-invariant.

Equation (2.4) presents the basic fetal energy balance result following directly from
the first law of thermodynamics, as similarly highlighted in the Thomas et al., 2008
model. However, equation (2.4) involves terms that need to be further defined, are difficult to measure experimentally, or expensive to track in an intervention setting. More specifically, in this work, profiles describing the evolution of the fetal body composition (FM and FFM), portion of maternal energy intake contributing to fetal nutrition, and the influence of maternal physical activity are all terms that are explored and expanded further from (2.4). Moreover, the expenditure term, $E_{ef}(t)$, requires estimates for the efficiency of energy conversion into new fetal fat and fat-free tissues (energy loss due to entropy); these efficiencies are difficult and expensive to measure experimentally. Furthermore, given current imaging technologies that build from well-studied sonographic methods to estimate total fetal weight, it is advantageous to reformulate the basic fetal energy balance equation shown as (2.4) in terms of the total fetal body mass (also referred to as total fetal weight, $W_f(t)$). In the following section, the primary aim is to establish a parsimonious fetal energy balance model that proves to overcome these challenges.

2.2.3 Efficiency of Energy Conversion & Energy Balance Reformulation

The goal of this section is to formulate equation (2.4) in terms of total fetal weight. To achieve this outcome, we built from concepts used to develop human obesity models by Christiansen et al., 2005. The time-varying rate of fetal fat mass deposition (with respect to total fetal weight) is defined as follows:

$$f_r(W_f) \overset{\text{def}}{=} \lim_{\Delta W_f \to 0} \frac{\Delta FM_f}{\Delta W_f} = \frac{dFM_f}{dW_f} \quad (2.5)$$
which leads to the following expressions for the rate of change of \(FM_f\) and \(FFM_f\) in terms of total fetal weight \(W_f\),

\[
\frac{dFM_f}{dt} = \frac{dFM_f}{dW_f} \frac{dW_f}{dt} \triangleq f_r(W_f) \frac{dW_f}{dt} \quad (2.6a)
\]

\[
\frac{dFFM_f}{dt} \triangleq \frac{d}{dt} (W_f - FM_f) = \left[1 - f_r(W_f)\right] \frac{dW_f}{dt} \quad (2.6b)
\]

With an explicitly defined \(f_r(W_f)\), the components \(FM_f(t)\) and \(FFM_f(t)\) become explicit functions of the total fetal weight, \(W_f(t)\). Using equations (2.3) and (2.6), we now have

\[
\frac{dC_f}{dt} = \left(\lambda_{FM_f} f_r(W_f) + \lambda_{FFM_f} [1 - f_r(W_f)]\right) \frac{dW_f}{dt} \quad (2.7)
\]

Second, we also define the efficiencies of new fetal tissue formation arising from the second law of thermodynamics as follows (Çengel and Boles, 2005):

\[
\text{efficiency of fat mass deposition} = e_{FM_f} \overset{\text{def}}{=} \lambda_{FM_f} \frac{dFM_f}{dE_{FM_f}} \quad (2.8a)
\]

\[
\text{efficiency of fat-free mass deposition} = e_{FFM_f} \overset{\text{def}}{=} \lambda_{FFM_f} \frac{dFFM_f}{dE_{FFM_f}} \quad (2.8b)
\]

where \(dE_f \overset{\Delta}{=} dE_{FM_f} + dE_{FFM_f}\) captures the total energy required to increase the total fetal body energy content by \(dC_f\). The efficiencies in (2.8) provide a useful parametric representation for the energy loss due to new fetal fat and fat-free tissue formation, respectively. Thus, from equations (2.6) and (2.8) we have

\[
\frac{dE_f}{dt} \overset{\Delta}{=} \frac{dE_{FM_f}}{dt} + \frac{dE_{FFM_f}}{dt} = \frac{\lambda_{FM_f}}{e_{FM_f}} \frac{dFM_f}{dt} + \frac{\lambda_{FFM_f}}{e_{FFM_f}} \frac{dFFM_f}{dt}
\]

\[
= \left(\frac{\lambda_{FM_f} f_r(W_f)}{e_{FM_f}} + \frac{\lambda_{FFM_f} [1 - f_r(W_f)]}{e_{FFM_f}}\right) \frac{dW_f}{dt} \quad (2.9)
\]

As first realized by Christiansen et al., 2005, the dynamic rate of change of \(E_f\) can be calculated by establishing the available energy for new fetal tissue deposition; i.e., the difference between the fetal energy intake and the energy expenditure required
for sustaining and maintaining life of existing fetal tissues, thus

\[
\frac{dE_f}{dt} \triangleq I_f - E_{M_f} \tag{2.10}
\]

Combining (2.9) and (2.10) gives

\[
\frac{dW_f}{dt} = \frac{dW_f}{dE_f} \frac{dE_f}{dt} = \frac{I_f - E_{M_f}}{\lambda_{FM_f} f_r(W_f)/e_{FM_f} + \lambda_{FFM_f} [1 - f_r(W_f)]/e_{FFM_f}} \tag{2.11}
\]

which, when substituted into (2.7), gives

\[
\frac{dC_f}{dt} = \frac{\lambda_{FM_f} f_r(W_f) + \lambda_{FFM_f} [1 - f_r(W_f)]}{\lambda_{FM_f} f_r(W_f)/e_{FM_f} + \lambda_{FFM_f} [1 - f_r(W_f)]/e_{FFM_f}} \left( I_f - E_{M_f} \right) \tag{2.12}
\]

where now the ratio between \( dC_f/dt \) and \( (I_f - E_{M_f}) \) represents the overall time-varying thermodynamic efficiency of energy conversion into new fetal tissue, \( 0 \leq e_f(W_f) \leq 1 \); this was first similarly established by Christiansen \textit{et al.}, 2005, however, with a constant \( f_r \) assumed. Inserting equation (2.2) into equation (2.1) and contrasting with (2.12) provides an accounting method for the energy loss due to new tissue formation

\[
E_{cj}(t) = (1 - e_f(W_f)) (I_f(t) - E_{M_f}(t)) \tag{2.13}
\]

with \( e_f(W_f) \) per equation (2.12).

To establish the \( I_f(t) \) term in (2.12), it is known that the fetal energy intake through the placenta (whose volume is denoted by \( P(t) \)) originates mainly from two nutritional sources: maternal diet, \( m(t) \), and maternal body components (e.g., muscles, fats, bones, etc.) Barker, 2008, hence giving

\[
\hat{I}_f(t) = \gamma(t) [g(t)m(t) + \alpha_W(t)W_m(t)] P(t) \tag{2.14}
\]

where \( W_m(t) \) is daily total maternal weight; \( \alpha_W(t) \) is a function that captures the daily fraction of maternal body mass directly contributing to fetal nutrition (\( \alpha \))
kcal/kg/d). \( g(t) \) denotes the daily glycemic impact of intake (ranges from 0 to 1); \( \gamma(t) \) is a conversion coefficient that is associated with maternal physical activity, as postulated in equation (2.22) (and discussed later in the chapter). However, for the case of a healthy, non-fasting and well-nourished mother, it may be accepted to assume that the basic nutritional needs for fetal growth can be met by daily maternal diet alone (Langhoff-Roos et al., 1987; Cetin et al., 2009). Hence, it is assumed that \( g(t)m(t) \gg \alpha_{W}(t)W_{m}(t) \quad \forall t \), giving (identical to Thomas et al., 2008)

\[
I_{f}(t) = \gamma(t)m(t)g(t)P(t)
\]  

(2.15)

The fetal energy expenditure term in (2.12) \( (E_{Mf}(t)) \) can be considered, for simplicity, as a direct proportion of total fetal body mass Thomas et al., 2008:

\[
E_{Mf}(t) = \mu [FM_{f}(t) + FFM_{f}(t)] \approx \mu W_{f}(t)
\]  

(2.16)

where \( \mu \) is the daily energy expenditure per unit fetal body mass. Hence, from (2.7) and (2.12) we have

\[
\left( \lambda_{FM_{f}}f_{r}(W_{f})/e_{FM_{f}} + \lambda_{FFM_{f}} [1 - f_{r}(W_{f})]/e_{FFM_{f}} \right) \frac{dW_{f}(t)}{dt} = I_{f}(t) - E_{Mf}(t)
\]  

(2.17)

Applying equations (2.15) and (2.16) gives

\[
\left( \lambda_{FM_{f}}f_{r}(W_{f})/e_{FM_{f}} + \lambda_{FFM_{f}} [1 - f_{r}(W_{f})]/e_{FFM_{f}} \right) \frac{dW_{f}(t)}{dt} = \gamma(t)m(t)g(t)P(t) - \mu W_{f}(t)
\]  

(2.18)

Furthermore, dividing equation (2.18) by \( \mu \) and defining

\[
K_{f}(t) = \frac{\gamma(t)g(t)}{\mu}
\]  

(2.19)

\[
\tau_{f}(W_{f}) = \frac{\lambda_{FM_{f}}f_{r}(W_{f})/e_{FM_{f}} + \lambda_{FFM_{f}} [1 - f_{r}(W_{f})]/e_{FFM_{f}}}{\mu}
\]  

(2.20)

yields a final fetal energy balance equation in terms of the total fetal weight:

\[
\tau_{f}(W_{f}) \frac{dW_{f}(t)}{dt} + W_{f}(t) = K_{f}(t)m(t)P(t)
\]  

(2.21)
Equation (2.21) features an intrauterine fetal weight growth model that conforms with the description of a first-order quasi Linear Parameter-Varying (quasi-LPV) system whose scheduling variable is the output, i.e., the total fetal weight, $W_f(t)$. In equation (2.21), the growth parameter $\tau_f$ is the time constant (Ogunnaike and Ray, 1994) which characterizes the speed of response. The parameters $\gamma(t)$ and $g(t)$ appearing in equation (2.19) are discussed in the explanation of equations (2.22) and (2.30), respectively.

In addition to achieving the goal of reformulating equation (2.4) in terms of a single, measurable output variable (i.e., the total fetal weight), equation (2.21) features an intuitive, well-understood first-order dynamical systems model structure that is more amenable to system identification and control. A further outcome following from the development of (2.21) is that estimates for $e_{FM_f}$ and $e_{FFM_f}$ can be determined directly from ultrasound measurements.

Following the development of the model in equation (2.21) we make the following remarks:

- Given that the exact mechanism governing the influence of maternal physical activity on fetal weight is yet to become sufficiently understood, we follow Thomas et al., 2008 in assuming that maternal physical activity influences the placenta function, and thereby influences the fetus’ nutrition. This is further established in Clapp, 2003 from which it is known that the effect of maternal exercise on fetal growth depends on numerous factors such as type, frequency, intensity, and the time point in pregnancy when the exercise is performed. Hence, for simplicity, we assume that, over a baseline, maternal physical activity is proportional to placental function, which is captured via the $\gamma(t)$ parameter in (2.21); i.e.,

$$\gamma(t) = \alpha PA(t) + \beta$$

(2.22)
where \( PA(t) \) denotes the daily maternal physical activity, \( \alpha \) is the proportionality constant, and \( \beta \) is the established baseline. Following the literature review presented by Thomas et al., 2008, we further assume that \( \alpha \leq 0 \) and \( \beta \geq \gamma > 0 \) \( \forall t \geq 0 \).

- It follows from assumption 6. that \( dW_f/dt \geq 0 \) during gestation; thus, from equation (2.21) we have

\[
K_f(t) = \frac{\gamma(t)g(t)}{\mu} \geq \frac{W_f(t)}{m(t)P(t)} \quad \forall t \text{ during gestation (2.23)}
\]

providing one important criterion for model validation. Additionally, the inequality in (2.23) can serve as an approximate (yet useful) diagnostic tool indicating rate of fetal growth (as will be discussed later in Figure 2.11).

- Kennaugh and Hay Jr, 1987 reported estimates where \( \mu_{FM_f} \) and \( \mu_{FFM_f} \) need not be averaged into a single proportion of energy expenditure (energy requirement); in which case, contrary to assumption 7., if \( \mu_{FM_f} \neq \mu_{FFM_f} \), it can be shown that equation (2.21) becomes

\[
\tau_f(W_f) \frac{dW_f(t)}{dt} + W_f(t) = K'_f(t)m(t)P(t) + \left( \frac{\mu_{FFM_f} - \mu_{FM_f}}{\mu_{FFM_f}} \right) \int_0^{W_f(t)} f_r(W_f) dW_f
\]

(2.24)

with

\[
K'_f(t) = \frac{\gamma(t)g(t)}{\mu_{FFM_f}}, \quad \tau'_f(W_f) = \frac{\lambda_{FM_f} f_r(W_f)/e_{FM_f} + \lambda_{FFM_f} [1 - f_r(W_f)]/e_{FFM_f}}{\mu_{FFM_f}}
\]

where \( \mu_{FM_f} \) and \( \mu_{FFM_f} \) are the proportions of energy expenditure corresponding to maintaining and sustaining life of fetal fat and fat-free tissues, respectively. Nonetheless, given the desire for a parsimonious model, we continue to assume that \( \mu_{FM_f} = \mu_{FFM_f} = \mu \), where equation (2.21) applies.
• The model parameters $\gamma(t)$, $e_{FM_f}$, and $e_{FFM_f}$ are assumed to vary on an individual level. According to the formulation of (2.21), these parameters may capture between- and/or within-group differences (e.g., genetic variations (Noblet et al., 1999), exercising vs. non-exercising).

2.2.4 Rate of Fetal Fat Mass Deposition

From the previous discussion, the importance of understanding the rate of fetal fat deposition, $f_r(W_f)$, as a key variable to attaining a predictive fetal energy balance model is now clear. From data presented and analyzed in a fetal body composition study by Demerath et al., 2016, good a priori knowledge is now available to establish the dependence of $f_r$ on $W_f(t)$. Literature also strongly suggests that the accretion of fetal fat starts accumulating after 26-30 weeks gestation (Schwartz and Galan, 2003; Thomas et al., 2008); to this effect, the following piecewise modified logistic equation can be considered

$$
FM_f(t) = \begin{cases}
\frac{c_{fr}}{1 + e^{-a_{fr}[W_f(t)-b_{fr}]}} + C & t \geq t_0 \\
0 & t < t_0
\end{cases}
$$

(2.25)

with identifiable parameters $a_{fr}$, $b_{fr}$, $c_{fr}$, and initial time, $t_0$, estimated as described in Section 2.3; $C$ is a constant. When $FM_f(W_{f0}) = 0$ at $t_0$, we get

$$
C = -\frac{c_{fr}}{1 + e^{-a_{fr}(W_{f0}-b_{fr})}}
$$

where $W_{f0}$ is the initial weight at the initial time $t_0$. From equations (2.5) and (2.25), $f_r$ is now an explicitly defined function; namely,

$$
f_r(W_f) = \begin{cases}
\frac{a_{fr}c_{fr}e^{-a_{fr}(W_f(t)-b_{fr})}}{1 + e^{-a_{fr}(W_f(t)-b_{fr})}}^2 & t \geq t_0 \\
0 & t < t_0
\end{cases}
$$

(2.26)
For simplicity, in this work we will assume that \( t_0 = 0 \) (or \( Wf_0 = 0 \)). Finally, given a well-defined \( f_r \) (equal to (2.26) or otherwise), estimations of the fetal body components readily follow from equation (2.6); namely,

\[
FM_f(t) = \int_0^{W_f(t)} f_r(W_f)dW_f
\]
\[
FFM_f(t) = W_f(t) - \int_0^{W_f(t)} f_r(W_f)dW_f
\]

### 2.2.5 Placental Volume Growth Model

Following Thomas et al. (2008) Thomas et al., 2008; Azpurua et al., 2010, we consider the placental volume \( P(t) \) as the most suitable variable to characterize placental growth. There is a substantial literature where placental development and growth profiles are presented and characterized throughout gestation for humans (Pitkin, 1976; Thompson et al., 2007; Wallace et al., 2013) and animals (Mu et al., 2008). The placenta grows in three phases: first, a ‘lag’ phase in which cells begin to form; second, an exponential growth phase where cells continue to form and rapidly divide; and finally, due to space restrictions, a deceleration in the growth rate is expected in the final weeks towards birth. These three growth phases are adequately captured with a logistic growth function. Figure 2.1 features a standard logistic growth profile where these three phases are depicted.

Similar to equation (2.25), the ‘modified’ logistic function is considered

\[
P(t) = c_P \left[ \frac{1}{1 + e^{-a_P(t-b_P)}} - \frac{1}{1 + e^{a_Pb_P}} \right] \quad \forall \ t \geq 0
\]

with \( P(0) = 0 \) and \( \lim_{a_Pb_P \to \infty} \lim_{t \to \infty} P(t) = c_P \lim_{a_Pb_P \to \infty} \left( \frac{e^{a_Pb_P}}{1 + e^{a_Pb_P}} \right) = c_P \), where \( a_P, b_P \), and \( c_P \) (\( c_P \) is the ‘ultimate’ carrying capacity) are identifiable model parameters from the estimation procedure described in Section 2.3.7; it follows from Figure 2.1 that
Figure 2.1: Representative Placental Volume Growth Profile.

\( a_P, b_P, c_P \geq 0 \). The algebraic model in equation (2.29) differs from the placental volume in Thomas et al., 2008 in that its curvature features are more independently parameterized: the parameter \( a_P \) assigns the rate of growth, \( b_P \) assigns the inflection point, and \( c_P \) assigns the ultimate carrying capacity or the scale of the profile (note \( c_P = 1 \) in Figure 2.1 for illustration). In Thomas et al., 2008, the proposed model does not apply when the initial condition is \( P(0) = 0 \), and hence, requires an additional estimated ultrasound measurement of EPV. Moreover, the parameters (including the initial condition) of the model in Thomas et al., 2008 play simultaneous role in determining its final curvature features, which makes it less intuitive.

It has been reported that the size and growth rates of the placenta are associated with physical activity (Thomas et al., 2008; Clapp, 2003) and additional genetic factors (Regnault et al., 2001). In the presence of more intensive ultrasound measurements, the carrying capacity parameter \( c_P \) can be further investigated such that moderations of placenta size over time by physical activity or genetic differences are more understood; this also applies to the growth rate \( a_P \) and mid-point \( b_P \) parameters. In our parameter estimation, we assume constant parameters \( a_P, b_P, \) and \( c_P \)
such that averaged, fixed-effects are captured.

2.3 Parameter Estimation and Model Validation

The Healthy Mom Zone (HMZ) study (Symons Downs et al., 2018) is an individually-tailored, adaptive behavioral intervention for managing weight in pregnant women with overweight and obesity. The target sample is 30 pregnant women who are randomized to either the intervention or control group from approximately 8 to 36 weeks gestation. Study measures including weight, physical activity, and energy intake are obtained at baseline, throughout the course of the intervention (e.g., daily, weekly, or monthly), and at follow-up. The detailed intervention protocols that includes eligibility, recruitment, intervention description, dosages, and measurement schedule have been published elsewhere (Symons Downs et al., 2018). In addition, an ancillary project provides six ultrasound measures used to estimate fetal weight, placental volume, and fetal body composition. In this section, further discussion of each estimated measurement is presented; four representative completed participants (n = 4; three overweight, one obese; mean age=30.3 years, two intervention, two control) are considered.

2.3.1 Estimated Fetal Weight

An estimated fetal weight (EFW) can be drawn from ultrasounds when specific biomarkers are measured as displayed in the example Figure 2.2. Using these biomarkers, one of the best known and well-established correlations that can be applied is the Hadlock estimation. For our model estimation, we use a set of six EFW measurements in addition to birth weight, as is described in more detail in Section 2.3.7. The first EFW is used to establish the upper and lower bounds for the initial condition ($\tilde{t}_0, \tilde{W}_{f0}$) used for solving equation (2.21). In this study, on average, the first
ultrasound measurement was taken at 14 weeks gestation, followed by five additional measurements each every four weeks through 34 weeks gestation. Infant’s birth weight was measured immediately after delivery.

2.3.2 Estimated Placental Volume

As the case with EFW, up to six ultrasound measurements are used to obtain the estimated placental volume (EPV) measurements using the Azpurua et al., 2010 approximation method, for which a number of simplifying assumptions have been made. As detailed in Section 2.3.8, EPV measurements are incorporated in the estimation cost function with lower emphasis than EFW measurements. This is justified given the following:

1) Absence of EPV measurements at or near birth. The bias that can result from equally emphasizing EPV measurements with EFW in (2.32) given the missing value at birth is crucial since fitting to earlier measurements only will tend to produce an exponential growth profile that can be, misleadingly, well-captured with the modified logistic equation in (2.29).

2) The Azpurua et al., 2010 EPV estimation method using 2D ultrasound mea-
measurements (similar to Figure 2.2) provides a rather simplified approximation (assumes spherical topology) that is mainly targeted for establishing EPV in the first and second trimesters of pregnancy for patients with normal BMI index (in this study, participants are either overweight or obese); this approximation can become exceedingly inaccurate at advanced gestational ages due to remaining technical difficulties associated with existing ultrasound technology.

3) The EPV approximation method in Azpurua et al. does not estimate standard errors; only 10th, 50th, and 90th percentile trajectories are given.

2.3.3 Fetal Body Composition

Studying ultrasound reports similar to Figure 2.2 (namely, anterior abdominal wall thickness and abdominal circumference) also produced at least two acceptable estimates per participant for the fetal % body fat using the correlation presented by Bernstein and Catalano, 1991. While the number of estimates can be as many as available ultrasounds, it is known that this estimate becomes more reliable at advanced gestational ages, and therefore we only consider a subset of the ultrasound measurements for the estimation of fetal % body fat.

2.3.4 Glycemic Index

Glycemic index (GI) was estimated using food items and portion sizes reported in a smart phone application. For each food, carbohydrate content (g) of the reported portion size was determined using the USDA Food and Nutrient Database for Dietary Studies (FNDDS) 2013-2014 data set. Next, a GI value for each food was determined by matching foods to the database generated by Flood et al., 2006. For foods without an exact match, the GI value of the closest matching item was used. Estimated GI
of each day was then calculated as the average GI of all foods consumed in a day, weighted by their contribution to total carbohydrate intake for that day, i.e.

\[ g(t) = \frac{\sum_{i=1}^{n} \text{food } i \text{ GI } \times \text{ food } i \text{ carbohydrate [g]}}{\sum_{i=1}^{n} \text{ food } i \text{ carbohydrate [g]}} \tag{2.30} \]

In equation (2.21), the glycemic index, \( g(t) \), is understood as a key variable for estimating fetal energy intake. One can generally presume a time-varying profile of \( g(t) \) on a daily scale; however, given that the collected \( g(t) \) time series shows to be stationary with a low variance, one can assume a constant \( \bar{g} \) value drawn from the average of all available average daily estimates, \( \bar{g} \). Table 2.2 lists the fractional average and standard deviation of estimated daily glycemic index values for the four participants presented in this study.

### Table 2.2: Mean and Standard Deviation Values of Daily Glycemic Index Estimations for Four Representative HMZ Participants

<table>
<thead>
<tr>
<th>Participant</th>
<th>( \bar{g} )</th>
<th>( \sigma_g )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.5478</td>
<td>0.0461</td>
</tr>
<tr>
<td>B</td>
<td>0.5441</td>
<td>0.0653</td>
</tr>
<tr>
<td>C</td>
<td>0.5806</td>
<td>0.0644</td>
</tr>
<tr>
<td>D</td>
<td>0.5763</td>
<td>0.0835</td>
</tr>
</tbody>
</table>

#### 2.3.5 Maternal Physical Activity

As noted in section 2.2.3, equation (2.22) characterizes the assumed simple, linear dependence of the fetal energy balance model in (2.21) on maternal physical activity. It is assumed that physical activity moderates the energy intake to the fetus by regulating the placental function (e.g., through blood flow; Ferraro et al., 2012). In the HMZ study, intensive objective assessment of physical activity is carried out using wrist-worn activity tracker. Missing and implausible physical activity measurements are imputed with mean replacement. These data are also used to establish the estimated daily maternal energy intake in equation (2.31).
2.3.6 Maternal Energy Intake

The daily maternal energy intake variable, \( m(t) \), can be reliably estimated with the availability of daily maternal weights and estimated energy expenditure data; the latter are estimated by correlating with daily physical activity and estimated/measured resting metabolic rates. As presented in Guo et al., 2016, back-calculated maternal energy intake from measured daily maternal weights and physical activity measures is considered:

\[
m(t) = -\frac{W_m(t + 2) + 8W_m(t + 1) - 8W_m(t - 1) + W_m(t - 2)}{12TK_1} - \frac{K_2}{K_1} [PA(t) + RMR(t)]
\] (2.31)

where \( K_1 \) and \( K_2 \) are gains (coefficients) that map changes of daily energy intake and physical activity, respectively, into maternal weight gain/loss; \( T \) is the sampling time; \( PA(t) \) and \( RMR(t) \) are the maternal daily physical activity and resting metabolic rate, respectively. To reduce the significant variability in equation (2.31), it is necessary to smooth the weight measurement \( W_m(t) \). A 9-day moving average filter is considered for all participants, except for participant D where a 13-day moving average filter is considered.

2.3.7 Model Estimation Problem Formulation

In this section, we establish a problem formulation for the least squares objective from which, with the presence of sufficient estimation and validation data, model parameters can be estimated and validated using nonlinear regression. Next, we describe in more detail how emphasis is split between different measured variables, and how the nonlinear optimization solver is initialized. Finally, in the results section, we present simulations of the estimated individual models and list the mean value and standard deviation associated with all model parameters.
The parameter estimation problem statement is formulated as a constrained optimization problem. The prediction error is minimized over estimation data using a non-linear least squares objective. For model estimation, using a total of $N$ EFW measurements inferred from ultrasound reports (similar to the example sonographic images shown in Figure 2.2, including birth weight), $M$ EPV measurements, and $L$ estimated body composition data points, the approach considered is to solve

$$
\min_\theta \quad \varepsilon^T Q \varepsilon \\
\text{subject to} \quad \theta_{lb} < \theta < \theta_{ub} \quad (2.32)
$$

where

$$
\varepsilon = [\Delta W_f(1) \cdots \Delta W_f(N) \quad \Delta P(1) \cdots \Delta P(M) \quad \Delta F M_f(1) \cdots \Delta F M_f(L)]^T
$$

$$
Q = \begin{bmatrix}
\lambda_1 \\
\vdots \\
\lambda_{N+M+L}
\end{bmatrix}
$$

$$
\theta = [\alpha \quad e_{FM_f} \quad e_{FFM_f} \quad a_p \quad b_p \quad c_p \quad a_{fr} \quad b_{fr} \quad c_{fr} \quad \tilde{W}_{fo}]^T
$$

$$
\Delta W_f(t) = EFW(t) - W_f(t), \quad \Delta P(t) = EPV(t) - P(t), \quad \text{and} \quad \Delta F M_f(t) = \hat{F} M_f(t) - F M_f(t)
$$

with $EFW(t)$, $EPV(t)$, and $\hat{F} M_f(t)$ denoting the estimated ultrasound measurements of the fetal weight, placental volume, and fetal fat mass at day $t$, respectively; $Q$ is a positive semi-definite weighting matrix used to establish the desired emphasis for model estimation. $W_f(t)$ is obtained from the numerical solution of the following fetal model

$$
\tau_f(W_f) \frac{dW_f(t)}{dt} + W_f(t) = K_f(t) m(t) P(t), \quad W_f(\tilde{t}_0) = \tilde{W}_{fo} \quad (2.33a)
$$

$$
\tau_f(W_f) = \frac{\lambda_{FM_f} f_r(W_f) / e_{FM_f} + \lambda_{FFM_f} [1 - f_r(W_f)] / e_{FFM_f}}{\mu} \quad (2.33b)
$$

$$
K_f(t) = \frac{\gamma(t) \tilde{g}}{\mu}, \quad \gamma(t) = \alpha PA(t) + \beta, \quad \alpha \leq 0, \quad \beta \geq \gamma > 0 \quad \forall \ t \geq 0 \quad (2.33c)
$$

$$
P(t) = c_p \left[ \frac{1}{1 + e^{-a_p(t-b_p)}} - \frac{1}{1 + e^{a_{pb}b_p}} \right] \quad (2.33d)
$$

$$
f_r(W_f) = a_{fr} c_{fr} \frac{e^{-a_{fr}(W_f(t)-b_{fr})}}{\left[1 + e^{-a_{fr}(W_f(t)-b_{fr})}\right]^2} \quad \forall \ t \geq 0 \quad (2.33e)
$$
with $\tilde{t}_0 > 0$ (initial time of simulation) and $m(t)$ per equation (2.31). Lower and upper parameter bounds, $\theta_{lb}$ and $\theta_{ub}$, are known a priori. The physical activity parameter $\alpha$ is constrained as shown in equation (2.33c). In this work, for purposes of simplicity, the value of the $\beta$ parameter is fixed at 0.000234 ml$^{-1}$, which is equal to the estimated nominal value of $\gamma$ in Thomas et al., 2008. Thermodynamic efficiencies $e_{FM_f}$ and $e_{FFM_f}$, by definition, range from 0 to 1. Also, given the strict growth of both $P(t)$ and $FM_f(t)$ profiles, the parameters $a_P, b_P, c_P, a_f, b_f,$ and $c_f$ are bounded below at 0, and are unbounded above.

The optimization is initialized using nominal parameter values/ranges drawn from literature. For example, Christiansen et al. Christiansen et al., 2005; Noblet et al., 1999 reports values for thermodynamic efficiencies drawn from animal studies; Thomas et al., 2008 gives an estimate for the conversion parameter, $\gamma(t)$; Demerath et al., 2016 provides fat and fat-free mass profiles from preterm infants that are used for initializing $a_f, b_f,$ and $c_f$ using standard regression; finally, also by similar means, EPV measurements calculated from our ultrasound data are used for initializing $a_P, b_P,$ and $c_P$. In the following section, we report in additional detail on the final set of parameter values used for solver initialization.

2.3.8 Relative Weights & Initialization

In this section, the specific relative weights ($\lambda_i$ in the diagonal $Q$ matrix in equation (2.32)) are presented for each participant. In addition, the specific initialization points (initial guesses) are also established in this section. It must be noted that given the limited amount of estimation data and the non-convexity of the optimization problem, the non-linear least squares solver becomes increasingly sensitive to relative weights and proper initialization as multiple local minima are expected. To avoid undesired solutions, solver features such as multistart can be used (Ugray
Judicious selection of $\lambda_i$ values is important for establishing an effective estimation cost function for each of the HMZ participants evaluated with this method. In the selection of $\lambda_i$ values, output emphasis, scaling, number of measurements, and measurement standard errors are all taken into consideration. While each data point can have its specific assigned $\lambda_i$ weight, we group measurements per model state (i.e., EFW, EPV, fetal FM) with one relative weight as $\lambda_{EFW} : \lambda_{EPV} : \lambda_{FMf}$. For participant A, the established ratios are $1 : 0.5 : 1$, whereas for participants B, C, and D the ratios are $1 : 0.3 : 1$.

Table 2.3 lists established initialization points for the studied HMZ participants. In the selection of these initializations, approximations from the literature, actual measurement values, and multiple iterations are all influencing factors. More specifically, initial guesses for $e_{FMf}$ and $e_{FFMf}$ were drawn from Christiansen et al., 2005 followed by multiple iterations (multiple solutions); $a_P$ was drawn from Thomas et al., 2008; $b_P$, $c_P$, $a_{fr}$, $b_{fr}$, and $c_{fr}$ were initialized from examining the actual measurements followed by multiple iterations; finally, the initialization of $\alpha$ was established after multiple iterations.

<table>
<thead>
<tr>
<th></th>
<th>$\alpha \times 10^8$</th>
<th>$e_{FMf}$</th>
<th>$e_{FFMf}$</th>
<th>$a_P$</th>
<th>$b_P$</th>
<th>$c_P$</th>
<th>$a_{fr}$</th>
<th>$b_{fr}$</th>
<th>$c_{fr}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$-0.5$</td>
<td>0.77</td>
<td>0.11</td>
<td>0.03</td>
<td>175</td>
<td>1281</td>
<td>0.47</td>
<td>9.56</td>
<td>9.47</td>
</tr>
<tr>
<td>B</td>
<td>$-0.5$</td>
<td>0.77</td>
<td>0.11</td>
<td>0.03</td>
<td>175</td>
<td>864</td>
<td>0.47</td>
<td>9.56</td>
<td>12.30</td>
</tr>
<tr>
<td>C</td>
<td>$-0.5$</td>
<td>0.44</td>
<td>0.15</td>
<td>0.03</td>
<td>175</td>
<td>774</td>
<td>0.47</td>
<td>9.56</td>
<td>8.99</td>
</tr>
<tr>
<td>D</td>
<td>$-0.5$</td>
<td>0.77</td>
<td>0.11</td>
<td>0.03</td>
<td>175</td>
<td>760</td>
<td>0.47</td>
<td>9.56</td>
<td>10.98</td>
</tr>
</tbody>
</table>

2.3.9 Estimation Results

In this section, for each of the examined HMZ participants, qualitative and quantitative model fit to data are presented from simulations when actual measured inputs
are applied to the model. In addition to the intrauterine fetal weight (primary model state), other model states (i.e., placental volume, body composition) and the evolution of intermediate constructs over time (e.g., $e_f(t)$, $\tau_f(t)$, and $K_f(t)$) are also shown. Finally, estimated model parameters tabulated in Table 2.4 are discussed.

Figures 2.3-2.7 feature simulations of the estimated models for one intervention participant (participant A) and one control participant (participant B). Overall, the goodness of fit does not appear to differ across intervention and control participants. In Figures 2.3-2.7, the simulation start time is selected to match the day of the first ultrasound measurement; the simulation is carried out through the reported actual day of birth. In these simulations, measurements of the two model inputs, i.e., maternal energy intake (back-calculated EI) and maternal PA (direct measurements), are displayed. In addition, the model states, i.e., fetal weight, placental volume, and body composition, are plotted and contrasted against estimated ultrasound measurements to qualitatively demonstrate the goodness of fit. Moreover, in Figure 2.9 and Figure 2.11 featuring the time-varying profiles of $\tau_f$, $e_f$, and $K_f$; it can be seen that, across all individuals, both $\tau_f$ and $e_f$ appear to exponentially increase over time as the fetus continues to grow. It is noted that Thomas et al., 2008 provides a significantly higher estimate for the overall efficiency ($e_f = 0.799$) than the estimated ranges from our data (approximately, in the 0.1-0.4 range). Finally, Table 2.4 summarizes the estimated model parameters with mean and standard deviation values for the examined participants.
Figure 2.3: Time-domain Response (Fetal Weight, Placental Volume, and Fetal % Body Fat) with Energy Intake and Physical Activity for a Representative HMZ Intervention Participant (Participant A). Simulation Starts at the Day of First Ultrasound Measurement and Ends at Birth.

Figure 2.4: Time-domain Response (Fetal Weight, Placental Volume, and Fetal % Body Fat) with Energy Intake and Physical Activity for a Representative HMZ Control Participant (Participant B). Simulation Starts at the Day of First Ultrasound Measurement and Ends at Birth.
Figure 2.5: Time-domain Response (Fetal Weight, Placental Volume, and Fetal % Body Fat) with Energy Intake and Physical Activity for a Representative HMZ Intervention Participant (Participant C). Simulation Starts at the Day of First Ultrasound Measurement and Ends at Birth.

Figure 2.6: Time-domain Response (Fetal Weight, Placental Volume, and Fetal % Body Fat) with Energy Intake and Physical Activity for a Representative HMZ Control Participant (Participant D). Simulation Starts at the Day of First Ultrasound Measurement and Ends at Birth.
Figure 2.7: Fetal Fat Mass and Fat-free Mass Growth Profiles over Time for Representative HMZ Participants (Participants A and B).

Figure 2.8: Fetal Fat Mass and Fat-free Mass Growth Profiles over Time for Representative HMZ Participants (Participants C and D).
Figure 2.9: Time-varying $\tau_f$ and $e_f$ for Representative HMZ Participants (Participants A and B).

Figure 2.10: Time-varying $\tau_f$ and $e_f$ for Representative HMZ Participants (Participants C and D).
Figure 2.11: Time-varying Gain and the Establishment of Positive Fetal Energy Balance for Representative HMZ Participants (See equation (2.23)).

Figure 2.12: Predicted Time-domain Profile of Fetal Energy Intake $I_f(t)$ for Representative Hmz Participants (See equation (2.15)).
Table 2.4: Estimated Model Parameter Values for Four Representative Hmz Participants with Mean and Standard Deviation (SD)

<table>
<thead>
<tr>
<th></th>
<th>$\alpha \times 10^8$</th>
<th>$e_{FM_f}$</th>
<th>$e_{FFM_f}$</th>
<th>$a_P$</th>
<th>$b_P$</th>
<th>$c_P$</th>
<th>$a_{f_P}$</th>
<th>$b_{f_P}$</th>
<th>$c_{f_P}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-13.70</td>
<td>0.44</td>
<td>0.24</td>
<td>0.027</td>
<td>206.2</td>
<td>1519.1</td>
<td>0.655</td>
<td>7.86</td>
<td>11.33</td>
</tr>
<tr>
<td>B</td>
<td>-2.82</td>
<td>0.60</td>
<td>0.07</td>
<td>0.039</td>
<td>154.3</td>
<td>1026.8</td>
<td>0.441</td>
<td>9.71</td>
<td>12.30</td>
</tr>
<tr>
<td>C</td>
<td>-0.31</td>
<td>0.82</td>
<td>0.12</td>
<td>0.028</td>
<td>182.3</td>
<td>1031.0</td>
<td>0.267</td>
<td>12.10</td>
<td>6.84</td>
</tr>
<tr>
<td>D</td>
<td>-0.06</td>
<td>0.42</td>
<td>0.06</td>
<td>0.024</td>
<td>174.8</td>
<td>1101.7</td>
<td>0.577</td>
<td>8.22</td>
<td>11.27</td>
</tr>
<tr>
<td>Mean</td>
<td>-4.22</td>
<td>0.57</td>
<td>0.12</td>
<td>0.030</td>
<td>179.4</td>
<td>1169.7</td>
<td>0.485</td>
<td>9.47</td>
<td>10.44</td>
</tr>
<tr>
<td>SD</td>
<td>6.44</td>
<td>0.18</td>
<td>0.08</td>
<td>0.007</td>
<td>21.4</td>
<td>235.5</td>
<td>0.170</td>
<td>1.93</td>
<td>2.44</td>
</tr>
</tbody>
</table>

2.3.10 Model Validation

The work of Thomas et al., 2008 established their model performance using compiled data from various sources in the literature as summarized in Table 2.5; plots evaluating their placental volume model performance against three data points for both exercising and non-exercising groups were also provided in the paper. However, with longitudinal data from Healthy Mom Zone, the performance and validation of the improved model in this work are determined by goodness-of-fit metrics as well as contrasting diverse estimated model features such as structure, parameter ranges, and output profiles against prior knowledge from existing literature.

Table 2.5: Performance Summary of the Model Developed in Thomas et al., 2008

<table>
<thead>
<tr>
<th>Prediction (Birth Weight)</th>
<th>Prediction (Fat Mass)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Glycemic Diet</td>
<td>96.40%</td>
</tr>
<tr>
<td>Runner Group</td>
<td>98.55%</td>
</tr>
<tr>
<td>Non-exercising Group</td>
<td>100%</td>
</tr>
</tbody>
</table>

First, all simulated fetal weight and placental volume growth profiles in Figures 2.3, 2.4, 2.5, and 2.6 are plausible and consistent with expected growth profiles from literature (Hadlock et al., 1984, 1985, 1991; Arleo et al., 2014). Comparing model predictions against the experimentally observed data (ultrasounds), a summary of individualized model outputs fit against available data is presented in Table 2.6. The Normalized Root-Mean-Square Error (NRMSE) is defined as
Table 2.6: Summary of the Goodness-of-fit from Various Metrics for Four Representative HMZ Participants

<table>
<thead>
<tr>
<th>Participant</th>
<th>NRMSE$_{EFW}$</th>
<th>$R^2_{EFW}$</th>
<th>$R^2_{EPV}$</th>
<th>$R^2_{FMf}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.9619</td>
<td>0.9986</td>
<td>0.8953</td>
<td>1.0000</td>
</tr>
<tr>
<td>B</td>
<td>0.9763</td>
<td>0.9994</td>
<td>0.9008</td>
<td>0.9925</td>
</tr>
<tr>
<td>C</td>
<td>0.9608</td>
<td>0.9985</td>
<td>0.8969</td>
<td>0.9997</td>
</tr>
<tr>
<td>D</td>
<td>0.9763</td>
<td>0.9994</td>
<td>0.9174</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

\[
\text{NRMSE}_{EFW} = 1 - \frac{\|EFW(t) - W_f(t)\|_2}{\|EFW(t) - \bar{EFW}\|_2} \tag{2.34}
\]

is considered as the primary metric for establishing the model goodness of fit against the HMZ data. $W_f(t)$ is the simulated output, $EFW(t)$ is the measured output, $\bar{EFW}$ is the mean of all measured $EFW(t)$ values, and $\|\cdot\|_2$ denotes the $l_2$-norm. $R^2_{EFW}$, $R^2_{EPV}$, and $R^2_{FMf}$ denote the coefficients of determination for fitting to estimated fetal weights, placental volumes, and fetal fat mass in utero, respectively. For a qualitative evaluation of model fit, the reader may refer to Figures 2.3-2.7.

Further giving validity to our model is that the estimated $e_{FMf}$ values are consistently larger than $e_{FFMf}$, in agreement with reported patterns only available from animal studies (Christiansen et al., 2005). Moreover, the mean estimated value of the placental volume growth rate parameter $a_P$ (with a narrow standard deviation of 0.003) matches the reported and validated value in Thomas et al., 2008; Orzechowski et al., 2014: $r = 0.03$. Furthermore, from Figures 2.3, 2.4, 2.5, and 2.6, predicted % body fat at birth approximately ranges from 10 to 18%, which fall into the typical ranges reported in literature (Widdowson and Spray, 1951; Schwartz and Galan, 2003; Demerath and Fields, 2014; Bernstein and Catalano, 1991). In agreement with Demerath et al. (2016) Demerath et al., 2016, Figures 2.7 and 2.8 show that predicted $FFM_f(t)$ profiles can be described as linear, while the $FM_f(t)$ are curvilinear (linear-exponential).

Finally, Figure 2.11 confirms that, except for only two brief instances in Participant A’s simulations, all estimated models satisfy the constraint in equation (2.23)
and hence validates the positive energy balance assumption throughout gestation. From Figure 2.12, one can observe the estimated rate of fetal energy intake $I_f(t)$ (note the negative $I_f(t)$ values in the two instances where Participant A’s positivity constraint is violated); comparing this to the maternal energy intake $m(t)$ (‘Estimated EI’ in Figures 2.3, 2.4, 2.5, and 2.6) provides support for the assumption of a well-nourished mother.

2.4 Chapter Summary

In conclusion, a dynamical systems model of intrauterine growth has been developed from first-principles, relying on the first and second laws of thermodynamics. This proposed model provides a rigorous yet more simple formulation than the fetal energy balance model of current literature. In the parameter estimation of this model, a non-linear least squares, constrained multi-objective optimization problem was formulated and guided by \textit{a priori} knowledge of ranges of model parameters. For the first time (to the authors’ knowledge), estimates (and an estimation method) for the thermodynamic efficiencies governing the formation of new tissues of human fetuses are established. This developed model has been estimated and validated against ultrasound measurements provided from the \textit{Healthy Mom Zone} study; despite the explained challenges with the estimation measurements, predictions follow from this model show good agreement with the data.

The availability of more intensive estimation and validation datasets (i.e., datasets with more frequent measurements) in a future study should create opportunities for parameter refinement and increased model understanding. More intensive measurements will allow for further investigation of the contribution of maternal body components in fetal nutrition (see equation (2.14)). Moreover, additional ultrasound measurements (particularly closer to delivery) may allow estimation of a less biased fetal
model described in equation (2.24). A better theoretical understanding of mechanisms behind the evolution of placental volume and the rate of fetal fat mass deposition is needed. This work considered a linear dependence of placental function on maternal physical activity; in future work, a more developed characterization of the influence of maternal physical activity may generate more resilient models: models with good predictions when input levels are far from those used in model estimation. Furthermore, a broader future goal is to use a combination of more experimental data and increased physiological understanding to reduce the modeling assumptions (outlined in Section 2.2) as much as possible.

Finally, the aim of this work (achieved using a limited number of HMZ intervention and control participants) was to develop a more comprehensive energy balance model for fetal weight gain derived from first-principles modeling that can be validated through data. These aims were facilitated by the availability of intensive, longitudinal participant data from the HMZ intervention. Model estimation and validation efforts for the additional HMZ participants ($N = 32$) could enable making conclusions regarding participant differences and intervention versus control outcomes, which was not the scope of this work. However, studying group differences (intervention vs. control) is a subject of current and future research.
Chapter 3

BLACK-BOX IDENTIFICATION OF *JUST WALK*: TOWARD ESTIMATING DYNAMICAL SYSTEMS MODELS OF SOCIAL COGNITIVE THEORY

3.1 Background

One recent emerging application of system identification and control theory is the design of optimized interventions in health behavior. In designing adaptive interventions (Rivera et al., 2007), a key consideration is the ability to estimate *personalized* behavior models that identify both individual-invariant and individual-variant dynamics. Parsimonious modeling, guided by *a priori* knowledge, is therefore crucial to accomplishing this. Using control systems engineering principles, behavior change theory can be utilized to develop models and decision frameworks for interventions that promote physical activity (PA) among sedentary individuals. One example is Social Cognitive Theory (SCT), proposed by Bandura (Bandura, 1986), which is among the leading theories of behavior change. The work of Martín et al. (Martín et al., 2014) established a dynamical systems fluid analogy model that captures key SCT concepts. Figure 1.1 represents a simplified fluid analogy dynamical system model of Social Cognitive Theory. SCT includes an extended list of potential constructs associated with predicting complex behavior dynamics, and whose effect needs to be accounted for during modeling. Consider, for example, the *Environmental Context* construct in Figure 1.1; this can include weather, busyness, stress, weekday, mood, and several other known and unknown variables.

Following an experimental design methodology based on system identification principles (Martín et al., 2015), a unique single-subject intervention study, *Just Walk*,
Section 3.2 introduces the *Just Walk* intervention and experiment execution, followed by a brief description on the input signal design approach using orthogonal multisine excitations in Section 3.3.

Section 3.4 discusses the use of an unconventional black-box approach that provides key insights into the dynamics of the intervention participants, and will ultimately be instrumental in accomplishing semi-physical identification of SCT models (Figure 1.1). Section 3.5 outlines a number of important conclusions and future directions on input signal design, modeling, identification, and intervention design.

### 3.2 Description of the *Just Walk* Intervention

*Just Walk* was developed as an adaptive walking intervention app for sedentary, overweight adults. It was designed primarily as a tool to generate individualized computational models for understanding PA behavior via system identification. The intervention system included a front-end Android app, *Just Walk* (Figure 3.1), a backend server, and an activity tracker (Fitbit Zip) to objectively measure PA and
automatically sync with the smartphone application. Participants were recruited nationally to partake in a walking intervention and receive daily step goals via the Just Walk app, and daily announced points were granted if the goals were achieved that day; granted points were converted into Amazon gift cards after a certain threshold was reached. Participants were also required to complete a series of daily morning and evening ecological momentary assessment (EMA; Shiffman et al., 2008) measures (e.g. confidence in achieving goal, predicted business for that day, previous night’s sleep quality, etc.) for the entire duration of the study.

The study duration was 14 weeks, including an initial two-week baseline period in which no step goals were delivered. Each participant’s step goals were then based on their median daily step value as calculated from the 14-day baseline period. The step goals were designed to establish a mechanism for individualizing the definition of an “ambitious, but doable” step range. All PA data were collected from the Fitbit Zip (provided to participants as a part of the study) and stored both locally and in Fitabase (Small Steps Labs, San Diego, CA, USA). Participants were generally healthy, inactive, 40-65 years old, with a body mass index (BMI) of 25-45 kg/m², who currently owned an Android phone capable of connecting to a Fitbit Zip via Bluetooth 4.0, and were willing to engage with the mHealth intervention for 14 weeks.

3.3 Input Signal Design of Just Walk

The input signal design procedure utilized in the Just Walk study was designed using deterministic yet “pseudo-random” signals that are orthogonal in the frequency domain. The procedure is described in detail in Martín et al., 2015. In Just Walk, Goals establish the desired behavior in a quantitative form, while Expected Points are the daily available points announced each morning that are granted upon goal achievement. Goals and Expected Points are two manipulated input signals $u_n$ gen-
Figure 3.2: Conceptual Representation of a “Zippered” Spectra Design for $n_u = 2$ Design Inputs, and $n_s = 6$ Harmonic Frequencies (Freigoun et al., 2017).

\[
\begin{align*}
    u_n(k) &= \lambda_n \sum_{j=1}^{N_s/2} \sqrt{2\alpha_{[n,j]}} \cos(\omega_j k T_s + \phi_{[n,j]}) \\
    \omega_j &= \frac{2\pi j}{N_s T_s}, \quad k = 1, \ldots, N_s
\end{align*}
\]  

(3.1)

where $\lambda_n$ 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 each variable has the following meaning: $\alpha_{[n,j]}$ is a factor used to specify the relative power of the harmonic, $\omega_j$ 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; this is called a “zippered” spectra design, an idea introduced in Rivera et al., 2009. A conceptual representation of the “zippered” design is presented in Figure 3.2. For $n_u$ design inputs and $n_s$ independently excited sinusoids the Fourier
coefficients are specified as

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

(3.2)

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

\[ N_s \geq 2n_s \]  

(3.3)

If \( n_s = 6 \) excited sinusoids are selected for the \( n_u = 2 \) design inputs, then from applying (3.3), \( N_s = 16 \) days (selected) is a feasible option. Phases \( \phi_{[n,j]} \) are selected to minimize the crest factor of the signal using the approach proposed by Guillaume et al., 1991.

In applying this design methodology for Just Walk, amplitudes for input signals (\( u_8 \) and \( u_9 \) in Figure 3.3) were chosen relying on experiences from previous studies (King et al., 2013; Adams et al., 2013) designed to obtain an expected profile of PA. The maximum number of step goals was selected as a factor of the initial baseline level of PA. For most cases in this experimental design, this factor was equal to 2; however, it was varied if the actual baseline step level of individuals was too high or low. Specifically, if participant’s baseline median steps were below 3,000, then the range for the goals was between 1 and 2.5 of their baseline median steps, to increase the likelihood of “ambitious” goals. If baseline median steps were greater than 7,500 steps, then the range was set between 1 and 1.75 (to reduce the likelihood of overly ambitious goals, such as 15,000 steps in one day). In addition to the two manipulated input channels, a large set of disturbances were also measured using mHealth technologies.
Overall experimental duration beyond the baseline varied between five to six cycles for each participant. A time series plot for a representative participant that depicts the behavior and seven inputs is shown in Figure 3.3.

3.4 ARX Model Estimation & Validation

In this section, black-box modeling strategies used for *Just Walk* are outlined, and results from fitting Auto Regressive with eXogenous input (ARX) parametric models (Ljung, 1999) are presented. As noted, identifying optimal ARX models (decisions on model inputs and model order) will play a pivotal role in ultimately identifying personalized semi-physical (grey-box) models that are informed by well-established behavior theories (Martín et al., 2014). Prior to ARX estimation, standard nonparametric modeling tasks such as correlation analysis have been informative. Because the *Just Walk* study included a wide array of input/output measurements, results from input-output and input-input correlation analyses have been useful (Phatak et al., 2016; Hekler, 2015; Korinek et al., 2018); for brevity, these are not included in this work. Incorporating all measured disturbances for estimating an SCT behavior model (particularly the Environmental Context construct in Figure 1.1) can be computationally demanding, may pose identifiability challenges, and will require large informative datasets that are typically difficult to gather from a practical standpoint in research involving human subjects.

Preprocessed data are fitted to an ARX model structure ARX-\([n_a, n_{k_1}, \ldots, n_{k_n}, n_{b_1}, \ldots, n_{b_u}]\), which can be expressed in the following concise form:

\[
y(k) + \sum_{l=1}^{n_a} a_l y(k - l) = \sum_{j=1}^{n_u} \sum_{i=1}^{n_{b_j}} b_{ij} u_j(k - n_{k_j} - i) + e(k) \tag{3.4}
\]

where \(y(k)\) is the measured output (e.g., steps/day), \(u_j(k)\) is the measured input \(j\), \(e(k)\) is the prediction error, all measured/estimated at day \(k\). The ARX model
in (3.4) is estimated by using regression. ARX parameter estimation constitutes a
linear least-squares regression problem (Ljung, 1999) and has attractive statistical
properties such as consistency. Figure 3.3 illustrates an example contrasting the
difference between actual output measurements and the prediction from a 7-input
ARX model with the structure in (3.4); a detailed discussion of black-box modeling
strategies used in this work follows. To quantify model fits, the normalized root mean
square error (NRMSE) fit index is used

\[
\text{model fit (\%)} = 100 \times \left(1 - \frac{\|y(k) - \hat{y}(k)\|_2}{\|y(k) - \bar{y}\|_2}\right) \quad (3.5)
\]

\(y(k)\) 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.

3.4.1 Data Pre-Processing and Model Structure Considerations

Data pre-processing tasks include interpolation (to account for missing data),
mean subtraction, and shifting Actual Steps and Granted Points by one sample to re-
fect temporal precedence. Model structure selection decisions consist of determining,
for each participant, the input signals to be included, and corresponding ARX model
orders for the output and each input, in accordance with (3.4). Taking advantage of
the computational simplicity associated with ARX modeling, the approach taken here
is to exhaustively examine a range of model orders, and use model validation proce-
dures to determine the most suitable structure. For this case study, ARX model order
ranges for \(n_a\) and \(n_b\) from 1 to 3 (i.e., \(\max(n_a) = 3\), and \(\max(n_b) = 3 \quad \forall j = 1, \ldots, n_u\))
seemed reasonable. A priori knowledge of the SCT fluid analogy model developed in
Martín et al., 2014 implies that very high order models should not be necessary to
characterize these behavior-change dynamics. From inspecting the intervention data,
it was reasonable to assume a basic unit input delay (i.e., \(n_{k_j} = 1 \quad \forall j\)).
Figure 3.3: Time Series Plot Showing Seven Selected Input Sequences (Manipulated Inputs & Measured Disturbances), Predicted Behavior (from an ARX Black-box Model), Actual Behavior, Model Overall Fit, and Estimation & Validation Cycles (1st, 2nd, and 5th for Estimation; 3rd and 4th for Validation) for a Selected Just Walk Participant.
The absence of drifts and trends in the data leads to assume stationary (though potentially time-varying) noise characteristics over the course of the intervention period. In determining the inputs to be considered, the approach is 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 4 additional measured inputs (Predicted Busyness, Predicted Stress, Predicted Typical, and Weekday-Weekend) to this basic model. All possible combinations of these inputs are estimated. Model validation following estimation ultimately determines which of these inputs are most important in describing individual behavior. Nonetheless, in the pre-processing stage, correlation analysis can be used to determine inputs that may be significantly crosscorrelated with each other or to identify inputs that appear to have no significant effect on the output. In both scenarios, the number of inputs that needs to be considered in the parameter estimation procedure can be reduced, ultimately leading to parsimonious models that can be generated with less effort.

3.4.2 Model Parameter Estimation and Validation

Model estimation and concomitant validation with the Just Walk intervention data is now considered. As mentioned earlier, first, a basic 3-input model was estimated and evaluated, followed by the addition and combination of 4 more inputs, leading to estimation of all possible combinations of these additional inputs. At an individual level, the full dataset was segmented into informative 16-day cycles for model estimation/validation. The cycle length was defined by the multisine input signal described in Section 3.3.

Cross-validation (the process of evaluating model fit over data not used for estimation) represents one of the most valuable aspects of system identification Ljung, 1994. The conventional approach in system identification is to assign a certain percentage
of data for estimation, followed by validation (e.g., 50% estimation, 50% validation). Such an approach assumes that the noise characteristics of the problem remain unchanged during the course of the intervention. However, it is reasonable to expect that noise and disturbance characteristics will vary over long-duration interventions such as *Just Walk*. In the analysis, each data cycle was assigned to either estimation or validation; all combinations of data cycles involving at least two cycles for validation were generated and evaluated.

Table 3.1 summarizes results of this procedure for a 4-input model (*Goals, Expected Points, Granted Points* and *Predicted Busyness*) of a selected participant. The fit index from Equation (3.5) was calculated for each cycle and averaged for estimation and validation data, respectively. All data cycle combinations that feature at least two cycles for validation or estimation (twenty candidate ARX models) were evaluated. For each of these combinations of estimation and validation cycles (corresponding to a specific row in Table 3.1), ARX orders were determined from an exhaustive search routine that selects a stable ARX model with highest predictive ability (based on the maximum average validation fit). This step provides a safeguard against overparametrization. The final chosen model should reflect, in addition to a good fit to validation data, a good fit for the entire data set (consisting of both estimation and validation cycles). This suggests that the final model choice should correspond to the model that yields highest overall fit (the “Overall NRMSE Fit” column in Table 3.1). Incorporating the overall fit criterion with the fit to cross-validation data balances good prediction with model accuracy over the entire data set. Note that using this analysis, the best results for the specific participant occur in the model resulting from row 18 (cycles 1, 2, and 5 for estimation; 3 and 4 for validation) with an overall NRMSE index at 46.03% for a model with structure $n_a = 2$, $n_{b_1} = 3$, $n_{b_2} = 1$, $n_{b_3} = 2$, and $n_{b_4} = 3$. This model performs close to the model with best fit.
Table 3.1: Intermediate Results for a 4-input ARX Model of a Selected Participant from *Just Walk*

<table>
<thead>
<tr>
<th>$E^*$</th>
<th>$V^*$</th>
<th>NRMSE Fit (%)</th>
<th>Average Estimation NRMSE Fit (%)</th>
<th>Average NRMSE Validation Fit (%)</th>
<th>Overall NRMSE Fit (%)</th>
<th>ARX Order (4-input)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycle 1</td>
<td>Cycle 2</td>
<td>Cycle 3</td>
<td>Cycle 4</td>
<td>Cycle 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1,2]</td>
<td>[3,4,5]</td>
<td>77.40%</td>
<td>85.44%</td>
<td>79.27%</td>
<td>27.68%</td>
<td>13.70%</td>
</tr>
<tr>
<td>[1,3]</td>
<td>[2,4,5]</td>
<td>77.39%</td>
<td>82.25%</td>
<td>81.30%</td>
<td>26.88%</td>
<td>15.36%</td>
</tr>
<tr>
<td>[1,4]</td>
<td>[2,3,5]</td>
<td>64.82%</td>
<td>71.25%</td>
<td>67.27%</td>
<td>45.89%</td>
<td>21.04%</td>
</tr>
<tr>
<td>[1,5]</td>
<td>[2,3,4]</td>
<td>61.36%</td>
<td>59.51%</td>
<td>60.96%</td>
<td>40.14%</td>
<td>24.47%</td>
</tr>
<tr>
<td>[2,3]</td>
<td>[1,4,5]</td>
<td>70.46%</td>
<td>90.25%</td>
<td>84.15%</td>
<td>25.00%</td>
<td>11.19%</td>
</tr>
<tr>
<td>[2,4]</td>
<td>[1,3,5]</td>
<td>49.06%</td>
<td>71.94%</td>
<td>67.25%</td>
<td>52.39%</td>
<td>22.98%</td>
</tr>
<tr>
<td>[2,5]</td>
<td>[1,3,4]</td>
<td>54.89%</td>
<td>61.75%</td>
<td>60.36%</td>
<td>47.35%</td>
<td>24.47%</td>
</tr>
<tr>
<td>[3,4]</td>
<td>[1,2,5]</td>
<td>45.97%</td>
<td>61.27%</td>
<td>69.24%</td>
<td>51.46%</td>
<td>24.02%</td>
</tr>
<tr>
<td>[3,5]</td>
<td>[1,2,4]</td>
<td>63.11%</td>
<td>66.96%</td>
<td>52.29%</td>
<td>41.52%</td>
<td>19.47%</td>
</tr>
<tr>
<td>[4,5]</td>
<td>[1,2,3]</td>
<td>36.37%</td>
<td>52.47%</td>
<td>50.06%</td>
<td>49.24%</td>
<td>25.88%</td>
</tr>
<tr>
<td>[3,4,5]</td>
<td>[1,2]</td>
<td>53.63%</td>
<td>64.61%</td>
<td>49.26%</td>
<td>46.59%</td>
<td>19.93%</td>
</tr>
<tr>
<td>[2,4,5]</td>
<td>[1,3]</td>
<td>50.12%</td>
<td>59.76%</td>
<td>59.36%</td>
<td>49.92%</td>
<td>23.64%</td>
</tr>
<tr>
<td>[2,3,5]</td>
<td>[1,4]</td>
<td>58.63%</td>
<td>66.76%</td>
<td>64.91%</td>
<td>49.62%</td>
<td>27.28%</td>
</tr>
<tr>
<td>[2,3,4]</td>
<td>[1,5]</td>
<td>59.43%</td>
<td>76.99%</td>
<td>70.11%</td>
<td>41.51%</td>
<td>22.32%</td>
</tr>
<tr>
<td>[1,4,5]</td>
<td>[2,3]</td>
<td>57.91%</td>
<td>61.11%</td>
<td>60.18%</td>
<td>45.69%</td>
<td>24.92%</td>
</tr>
<tr>
<td>[1,3,5]</td>
<td>[2,4]</td>
<td>66.34%</td>
<td>66.02%</td>
<td>67.24%</td>
<td>42.13%</td>
<td>22.57%</td>
</tr>
<tr>
<td>[1,3,4]</td>
<td>[2,5]</td>
<td>68.39%</td>
<td>77.75%</td>
<td>73.46%</td>
<td>41.86%</td>
<td>18.78%</td>
</tr>
<tr>
<td>[1,2,5]</td>
<td>[3,4]</td>
<td>61.85%</td>
<td>56.05%</td>
<td>68.43%</td>
<td>44.82%</td>
<td>35.02%</td>
</tr>
<tr>
<td>[1,2,4]</td>
<td>[3,5]</td>
<td>71.99%</td>
<td>73.18%</td>
<td>72.36%</td>
<td>43.28%</td>
<td>20.40%</td>
</tr>
<tr>
<td>[1,2,3]</td>
<td>[4,5]</td>
<td>75.95%</td>
<td>87.02%</td>
<td>80.67%</td>
<td>26.39%</td>
<td>13.36%</td>
</tr>
</tbody>
</table>

* $E^*$ = Estimation Cycles (magenta), $V^*$ = Validation Cycles (cyan)

over the validation data (average validation fit of 56.63% for row 18 vs 60.65% in row 15); however, the model with the best fit to validation data does not exhibit the best fit to data overall (38.81% in lieu of 46.03%).

3.4.3 Overall Fit Analysis and Assessment of Individual Participant Characteristics

Similar analyses to those presented in Table 3.1 can be performed with additional inputs, for all possible combinations. For example, for a total of 7 inputs, 16 different input models can be generated for each participant (since Goals ($u_8$), Expected Points ($u_9$), and Granted Points ($u_{10}$) are always grouped). Evaluating these 16 input combinations allows us to draw conclusions on participant characteristics that resulted from the intervention.

Figure 3.4 depicts model validation % fit results from three different participants from *Just Walk*. The Y-axis indicates the % fit of the 3, 4, 5, 6, and 7-input models,
and the X-axis corresponds to the psychosocial measures (busyness, stress, weekday, typical) measured daily. Here, it is seen that Participant A’s walking behavior is largely driven by stress (highest % fit seen for the stress bar in the 4-input model), Participant B’s behavior is driven by whether it is a weekday or weekend, while Participant C has the highest % fit for the 3-input model, indicating that the daily step goal had the greatest impact on walking behavior. Step responses from the individual ARX models can be used to reveal more precise directionality and magnitude information; for example, from Figure 3.5, one can predict that the selected participant will typically reach 80% of the desired daily step goals within the first day of goal announcement. Responsiveness to other inputs and disturbances can be determined similarly. This strategy has significant implications for personalized and adaptive behavior change interventions; if one can determine the inputs that are most meaningful for a given individual in a given context, it is possible then to optimize the target behavior over a specified time (hours, days, weeks, months).
3.5 Chapter Summary

This chapter presented a system identification modeling strategy for a physical activity (e.g., walking) intervention delivered via a smartphone application. The results from this dataset represent an important accomplishment in understanding behavior-change from a data-driven perspective. Predictive and consistent black-box models are crucial for validating behavioral theory (such as the SCT model). It is shown that segmenting and evaluating the data at a per-cycle level gives the most valid results to date. These types of models are necessary to effectively model behavior, which is highly complex, idiosyncratic, and dynamic in nature. In addition, drawn from the analysis of the estimated models, it is important in experimental design to consider capturing more low-frequency dynamics (e.g., using pseudo-random binary sequences), to draw more decisive conclusions on participant long-term (steady-state) responses. Finally, the enhanced identification testing monitoring procedure in Martín et al., 2016b can be considered in future experiment design.
4.1 Background

The grey-box identification problem, i.e., the estimation of physical/semiphysical state-space models with an imposed structure, is a topic of increasing interest and a growing number of applications. For physical, semiphysical, and compartmental/network models, the resulting state-space realizations often emerge in continuous-time with some sort of a structure. Furthermore, in the vast majority of structured state-space models, an imposed or emerged model structure is often linear in the sense that some entries of the system matrices are determined \( \text{a priori} \), and can be expressed as a set of linear equations (e.g., \( a_{ij} = 0, a_{ij} + c_1 b_{11} = c_2, b_{ij} = c_3 \) for some indices \( i \) and \( j \); \( c_{1,2,3} \in \mathbb{R} \) are some constants). We refer to such models as linearly-structured. A subset of this class of structures has been presented as affinely parametrized structures (Yu et al., 2019). As expressed in Ljung, 2019, considering linear structures is not very restrictive; most of structured models belong either readily or after a reformulation. Motivated by the real-life behavioral experiment Just Walk, and the need for data-driven semiphysical models of Social Cognitive Theory (SCT), we present a spectral decomposition (SD) identification algorithm for estimating this specific class of models, i.e., linearly-structured models.

While it is recognized that the general prediction-error method (PEM), which estimates the constrained model directly from Input/Output (I/O) data, has the best
possible asymptotic properties Yu et al., 2018, it is also known that due to the non-convexity of the problem, “the domain of attraction of the global minimum is not very forgiving” (Ljung, 2019), and the method is thus crucially dependent on good initial estimates for its practical success in problems of realistic sizes. This particular difficulty has been illustrated by the numerical experiment presented in Ljung, 2019 from Parrilo and Ljung, 2003 and will be further illustrated with an additional numerical example in this chapter (see Section 4.3.2). Indeed, this solver initialization challenge is amplified when the structure in question is inherently unidentifiable, leading to the fact that even if a global minimum is found, it need not be unique. In fact, in this case, an infinity of global minima solutions may be on offer, and hence the physical significance of model parameters need not be preserved. Note that, here, model identifiability is defined per, Ljung, 1999, Definition 4.6. Nevertheless, in this chapter, we present a formulation that aims at eliminating or at least significantly reducing the burden on the user with the challenge of judicious solver initialization, particularly in the absence of sufficient prior knowledge. This may be possible at the expense of increased computational load while solving multiple random initializations/restarts that may be needed in harder problems.

The notion of initializing PEM for grey-box model estimation from black-box models that are fully-parameterized (i.e., “total models” in Ljung, 2019) has been suggested and standardized in numerous settings before; the reader is referred to Yu et al., 2019; Yu et al., 2018; Parrilo and Ljung, 2003; Mercère et al., 2014 and references therein for elaboration. In particular, subspace identification methods (e.g., N4SID, MOESP) seemed favorable given their asymptotic consistency (Yu et al., 2018; Parrilo and Ljung, 2003). Similarly, building from an identified black-box model, the main premise of the proposed SD formulation is to exploit the known linear structure and formulate an easier similarity transformations search, restricting the search space to
only of that which is fundamentally unknowable \textit{a priori}.

The rest of this chapter is organized as follows: Section 4.2 presents the SD formulation for estimating linearly-structured, continuous-time state-space models. Sufficient conditions for the existence and uniqueness of SD models is presented with additional conclusions on identifiability. Next, Section 4.3 provides two numerical examples and introduces a further simplified SCT model, followed by a brief report on results from \textit{Just Walk} in Section 4.4. Finally, conclusions and potential future directions are summarized in Section 4.5.

4.2 Spectral Decomposition Formulation

Using measured input-output sequences \{U_k, Y_k\}_{k=1}^N, consider a minimal, fully-parametrized discrete-time model \(\hat{S}_d(\hat{A}_d, \hat{B}_d, \hat{C}_d, \hat{D}_d, \hat{K}_d)\) estimated using a classical subspace method (i.e., N4SID, Van Overschee and De Moor, 2012), satisfying the following set of linear equations

\[
\begin{bmatrix}
    \dot{X}_{k+1} \\
    Y_k
\end{bmatrix} =
\begin{bmatrix}
    \hat{A}_d & \hat{B}_d \\
    \hat{C}_d & \hat{D}_d
\end{bmatrix}
\begin{bmatrix}
    \dot{X}_k \\
    U_k
\end{bmatrix} +
\begin{bmatrix}
    \hat{K}_d \\
    I
\end{bmatrix}
\begin{bmatrix}
    \varepsilon_k
\end{bmatrix}
\tag{4.1}
\]

where \(\dot{X}_k \in \mathbb{R}^{n \times N}, Y_k \in \mathbb{R}^{p \times N},\) and \(\varepsilon_k \in \mathbb{R}^{p \times N}\) is a column-wise sequence of noise in innovations form (i.e., \(\varepsilon \sim \mathcal{N}(0, \sigma_i) \forall i \in \{1, \ldots, p\}\)). With respect to I/O data \{U_k, Y_k\}, the identified states from the subspace method are modulo a similarity transformation \(T_f\) (i.e., \(\hat{X}_k = T_f X_k\)) with \(\hat{S}_d(T_f A_d T_f^{-1}, T_f B_d, C_d T_f^{-1}, D_d, T_f K_d)\). For the identified system \(\hat{S}_d\) of McMillan degree \(n\), \(m\) inputs, and \(p\) measured outputs, the minimal realization system matrices are \(A_d \in \mathbb{R}^{n \times n}, B_d \in \mathbb{R}^{n \times m}, C_d \in \mathbb{R}^{p \times n}, D_d \in \mathbb{R}^{p \times m}, K_d \in \mathbb{R}^{n \times p},\) and \(T_f \in \mathbb{R}^{n \times n}\). For simplicity, and without loss of generality, we proceed in the various parts of this chapter with a strictly proper, disturbance-free model of \(\hat{S}_d(T_f A_d T_f^{-1}, T_f B_d, C_d T_f^{-1})\).
4.2.1 Main Formulation

Through the remainder of this chapter, two main assumptions follow: First, we assume that $\hat{A}_d$ is of $n$ distinct eigenvalues. Second, we assume Zero-Order Hold (ZOH) intersample behavior. The ZOH assumption becomes more practical in experiments with smaller sampling times (i.e., $T_s \to 0$). Benefiting from assuming ZOH, we recall that the continuous-time image of $\hat{S}_d$, $\hat{S}(A, B, C)$, can be obtained from the well-known exact discretization equations from linear systems theory

$$A_d = e^{AiTs}$$  \hfill (4.2a)

$$B_d = \int_0^{T_s} e^{Ai\tau} Bu(\tau) d\tau \overset{ZOH}{=} A^{-1} (A_d - I_n) B$$  \hfill (4.2b)

$$C_d = C$$  \hfill (4.2c)

where $T_s$ is the sampling time. By assuming $n$ distinct poles of the identified $\hat{S}_d$, we allow for the following spectral decomposition

$$A_d = T\Lambda_d T^{-1} \Rightarrow A = T\Lambda T^{-1}, \quad \Lambda = \log(\Lambda_d)/T_s$$  \hfill (4.3)

Thus, an estimate for the state-transition matrix subject to a linear structure imposed by the pair $(P_A, d_A) := \{(P_A, d_A) : P_A\theta_A = d_A, \ \theta_A = \text{vec}\{A\}\}$ is obtained by solving the linearly-constrained eigenvalue problem

$$T^{-1}AT = \Lambda$$  \hfill (4.4)

subject to the structure pair $(P_A, d_A)$, from which the following quadratic program (QP) follows

$$\min_{\theta_A} \frac{1}{2} \theta_A^T H_A \theta_A + f_A^T \theta_A$$

s.t. \quad $P_A\theta_A = d_A$  \hfill (4.5)
where

\[ H_A = 2(T^T \otimes T^{-1})^T (T^T \otimes T^{-1}) \]
\[ f_A = [-2\varphi_A^T(T^T \otimes T^{-1})]^T \]
\[ \varphi_A = \text{vec}\{\Lambda\} \]

and \( \otimes \) denoting the Kronecker product. Similarly, by assuming ZOH, one can use (4.2b) and (4.3) to establish

\[ T^{-1}B = (\Lambda_d - I_n)^{-1} \Lambda T^{-1}B_d \quad (4.7) \]

for estimating the input gain matrix \( B \) subject to the structure \((P_B, d_B)\), from which it also follows

\[ \min_{\theta_B} \frac{1}{2} \theta_B^T H_B \theta_B + f_B^T \theta_B \]
\[ \text{s.t.} \quad P_B \theta_B = d_B \quad (4.8) \]

with

\[ H_B = 2 \left( I_m \otimes T^{-1} \right)^T \left( I_m \otimes T^{-1} \right) \]
\[ f_B = -2 \left[ \varphi_B^T \left( I_m \otimes T^{-1} \right) \right]^T \]
\[ \varphi_B = \text{vec}\{\Gamma^{-1}T_0^{-1}B_0\} \]
\[ \Gamma = \Lambda^{-1} (\Lambda_d - I_n) \]

Note that the characteristic matrix \( \Gamma \) is system-invariant (i.e., \( \Gamma \) is unique and does not depend on the applied similarity transformation); this extends the invariance to \( \varphi_B \) (similar to \( \varphi_A \)) once \( S_0^0 \) of \( A_0^0 \) (thereby \( T_0 \)) and \( B_0^0 \) is initially established. By defining the Kronecker eigenvector matrix

\[ T = \begin{bmatrix} T \otimes T^{-T} & 0_{n^2 \times nm} \\ 0_{nm \times n^2} & I_m \otimes T^{-T} \end{bmatrix} \quad (4.10) \]

70
and using the Kronecker product properties, it can be shown that in a more compact form, (4.5) and (4.8) can be solved simultaneously by the following augmentation

$$\begin{align*}
\min_{\theta} & \quad \frac{1}{2} \theta^T H \theta + f^T \theta \\
\text{s.t.} & \quad P \theta = d
\end{align*}$$

(4.11)

with

$$H = 2 \begin{bmatrix} T T^T \otimes (T T^T)^{-1} & 0_{nm \times n^2} \\ 0_{nm \times n^2} & I_m \otimes (T T^T)^{-1} \end{bmatrix} = 2 T T^T$$

$$f = -2 \begin{bmatrix} T \otimes T^{-T} & 0_{n^2 \times nm} \\ 0_{n^2 \times nm} & I_m \otimes T^{-T} \end{bmatrix} \varphi = -2 T \varphi$$

$$\varphi = \begin{bmatrix} \varphi_A^T \\ \varphi_B^T \end{bmatrix}$$

$$\theta = \begin{bmatrix} \theta_A^T \\ \theta_B^T \end{bmatrix}$$

$$P = \begin{bmatrix} P_A & 0_{p_a \times nm} \\ 0_{p_a \times n^2} & P_B \end{bmatrix}$$

$$d = \begin{bmatrix} d_A^T \\ d_B^T \end{bmatrix}$$

where $\circ^{-T} = (\circ^T)^{-1} = (\circ^{-1})^T$. Note that in cases where some linear relations exist between parameters of the structured $A$ and $B$, the augmented QP in (4.11) is used. However, the pairs $(P_A, d_A)$ and $(P_B, d_B)$ would no longer separable, but constructed as a single $(P, d)$ instead; $P$ will no longer be block-diagonal.

4.2.2 Existence, Uniqueness, and Identifiability

With the main formulation delivered in (4.11), we now proceed to establish the existence and uniqueness of an SD-identified model, as well as present a corollary on identifiability.

**Theorem 4.2.1** (Existence). For the minimal discrete-time model $S_d$ in (4.1) with $n$ distinct eigenvalues of $A_d$ and ZOH intersample behavior on all inputs, a sufficient condition for the existence of $S_{[P, d]}$ with a linearly-structured $\theta_{SD}^*$ subject to $(P, d)$ in (4.11) is $|P H^{-1} P^T| \neq 0$. 

71
Proof. The explicit solution of (4.11) is known (Boyd and Vandenberghe, 2004)

\[
\begin{bmatrix}
\theta_{SD}^* \\
\lambda^*
\end{bmatrix} =
\begin{bmatrix}
H & P^T \\
P & 0
\end{bmatrix}^{-1}
\begin{bmatrix}
-f \\
d
\end{bmatrix} 
\]  
(4.13)

and exists for some Lagrange multipliers \(\lambda^*\) if the Karush-Kuhn-Tucker (KKT) matrix is nonsingular. Using block matrix inversion, i.e.,

\[
\begin{bmatrix}
H & P^T \\
P & 0
\end{bmatrix}^{-1} =
\begin{bmatrix}
H^{-1} - H^{-1}P^T(PH^{-1}P^T)^{-1}PH^{-1} & H^{-1}P^T(PH^{-1}P^T)^{-1} \\
(PH^{-1}P^T)^{-1}PH^{-1} & -(PH^{-1}P^T)^{-1}
\end{bmatrix} 
\]  
(4.14)

yields

\[
\theta_{SD}^* = H^{-1}P^T \Psi^{-1} d + (H^{-1}P^T \Psi^{-1} P - I) H^{-1} f 
\]  
(4.15)

where \(\Psi = PH^{-1}P^T\). \(A_d\) of \(n\) distinct eigenvalues implies \(T\) is nonsingular, which establishes the existence of \(H^{-1}\). Hence, a sufficient condition for the existence of \(\theta_{SD}^*\) (thereby \(\Psi^{-1}\)) is \(|PH^{-1}P^T| \neq 0\), i.e., the imposed structure \(P\) is feasible. \(\Box\)

Note that \(H = 2TT^T\) in (4.11) implies that this QP is convex. Further, by extension, if the quadratic structural constraints in (6.1) are convex, the resulting QCQP remains convex.

**Theorem 4.2.2** (Uniqueness). For the minimal discrete-time model \(S_d\) in (4.1) of order \(n\) and \(p\) outputs, the linearly-structured, SD-identified continuous-time model \(S_{[P,d]}\) subject to \((P,d)\) in (4.11) such that \(|PH^{-1}P^T| \neq 0\) is unique when \(p = n\), \(|C| \neq 0\).

*Proof.* Given \(|PH^{-1}P^T| \neq 0\) establishes the nonsingularity of the KKT matrix in (4.14) and implies that \([HT \ P^T]^T\) has linearly independent columns. It follows that \(\theta_{SD}^*\) of (4.11) is unique; full proof in Boyd and Vandenberghe, 2004. Hence, a sufficient condition for the uniqueness of the SD model follows from the uniqueness of the range space of \(T\) in \(\mathbb{R}^p\), \(\mathcal{R}(T)\), which is achieved when \(p = n\), \(|C| \neq 0\); see, Bellman and Åström, 1970, Sec. 6 and, Delforge, 1982, Sec. 3. \(\Box\)
Corollary 4.2.2.1 (Identifiability). For a minimal discrete-time model $\mathcal{S}_d$ of order $n$ and $p$ outputs, assume ZOH intersampling behavior on all inputs, and define the set of similarity transformations $\mathcal{T}_f$

$$\mathcal{T}_f := \left\{ T_f : T_f = \begin{bmatrix} I_p & 0_{n-p} \\ -H_f & - \end{bmatrix}, H_f \in \mathbb{R}^{(n-p)\times n}, |T_f| \neq 0 \right\}$$

The identifiability of the continuous-time $\mathcal{S}$ is as follows:

- $p = n$, $|C| \neq 0$, and $|PH^{-1}P^T| \neq 0$: the continuous-time $\mathcal{S}_{[P,d]}$ subject to $(P,d)$ is globally identifiable; the estimator (4.15) yields a unique estimate for all system matrices.

- $p < n$, $C = (I_p \ 0_{n-p})$: at least the first $p$ rows of $B$ and $K$ are uniquely identifiable, modulo $\mathcal{E}_k$, in the fully-parametrized $\mathcal{S}$ (i.e., $(P,d) = \{\emptyset\}) \forall T_f \in \mathcal{T}_f$.

Proof. When $p = n$, the SD-identified model is unique, and the proof follows immediately from Theorem 4.2.2. When $p < n$, recall that from (4.1) we have

$$\dot{X}_{k+1} = T_f A_d T_f^{-1} \dot{X}_k + T_f B_d U_k + T_f K_d \mathcal{E}_k$$

$$Y_k = C T_f^{-1} \dot{X}_k + D_d U_k + \mathcal{E}_k$$

which by substitution gives

$$Y_k = C A_d T_f^{-1} \dot{X}_{k-1} + C B_d U_{k-1} + C K_d \mathcal{E}_{k-1} + D_d U_k + \mathcal{E}_k$$

(4.16)

With $C = (I_p \ 0_{n-p})$, (4.16) shows that entries of the first $p$ rows of $B_d$ and $K_d$ must be unique $\forall T_f \in \mathcal{T}_f$; noting that $B = T_f T \Gamma^{-1} T^{-1} B_d$, $T_f \in \mathcal{T}_f \implies T_f^{-1} \in \mathcal{T}_f$ by block inversion, and the augmentation $B'_d = [B_d \ K_d]$ and $U'_k = [U_k^T \ \mathcal{E}_k^T]^T$ extends the result to $B$ and $K$. \qed
4.2.3 Model Realization: An LMI Approach

In the identification of structured state-space models of physical and other types of systems, the model order is often determined \textit{a priori} and the model must be physically realizable. Hence, when opting in for SD-identified models, we know from (4.3) that $\Lambda_d$ must at least be positive semidefinite. However, in many cases when estimating discrete-time models, the presence of process/measurement noise or unmodeled dynamics may suffice to explain the emergence of negative eigenvalues, i.e., physically unrealizable models of ZOH inversion. To circumvent this potential obstacle for the SD user, we include a brief review on the work of Miller \textit{et al.}, 2012; Miller and De Callafon, 2013 that ensures $\Lambda_d \succeq 0$ by applying Linear Matrix Inequalities (LMI) constraints in the earlier step of estimating fully-parametrized subspace models in (4.1).

Per Miller and De Callafon, 2013, the main theorem in Chilali and Gahinet, 1996 establishes that the eigenvalues of a matrix $A_d$ lie within an LMI region defined by the characteristic function

$$f_D(z) = \alpha + \beta z + \beta^T \bar{z}$$

with

$$D = \{z \in \mathbb{C} : f_D(z) \geq 0\}$$

if and only if $\exists P_d \in \mathbb{R}^{n \times n}$ such that

$$P_d = P_d^T > 0, \quad \mathcal{M}_D(Q_d, P_d) \succeq 0$$

where

$$\mathcal{M}_D(Q_d, P_d) = \alpha \otimes P_d + \beta \otimes Q_d + \beta^T \otimes Q_d^T$$

with $\alpha$ (symmetric) and $\beta$ (square) are generally not unique matrices establishing the feasible LMI region $D$ in the $z$-plane. Given matrices $R_1, R_2,$ and $W$, depending on the underlying subspace method, the new eigenvalue-constrained $A^*_d$ can then be
found by solving the following convex, semidefinite program

\[
\min_{Q_d, P_d} \| R_1 Q_d - W R_2^\dagger P_d \|_F \\
\text{s.t.} \quad \mathcal{M}_D(Q_d, P_d) \succeq 0 \\
P_d = P_d^T > 0 \\
\text{tr}(P_d) = l
\]  

(4.17)

where \( Q_d \overset{\Delta}{=} A_d P_d, l > 0 \) is a positive scalar \((l = 1 \text{ in Miller et al., 2012})\) that ensures \( P_d \) does not become arbitrarily small during the minimization procedure; \( \dagger \) and \( \| \cdot \|_F \) denoting the Moore-Penrose pseudoinverse and the Frobenius norm, respectively. The eigenvalue-constrained \( A_d^* \) is then found as \( A_d^* = Q_d^* P_d^{*-1} \). When the used subspace method is the deterministic N4SID, \( R_1, R_2, \) and \( W \) are defined below:

\[
R_1 = I_n, \quad R_2 = \hat{X}_k, \quad W = \hat{X}_{k+1} - B_d U_k
\]

(4.18)

For additional subspace methods, including the stochastic N4SID, corresponding values for \( R_1, R_2, \) and \( W \) are tabulated in Miller et al., 2012.

4.2.4 Standard SD Loss Function

As noted in Section 4.2, if \( \hat{C}_d \) in (4.1) is full-rank, the SD model is unique and obtained analytically by evaluating (4.15). Otherwise, a search for a suitable similarity transformation may be in order. Consider

\[
J_0 = (Y - \Phi_0 \theta_d^0)^T (Y - \Phi_0 \theta_d^0)
\]

(4.19)

where \( \Phi_0 = [X^T_{0:k-1} \ U^T_{k-1}] \otimes C_0 \); \( Y \) and \( \theta_d^0 \) denote the vectorized measured outputs and the initial discrete-time parameter vector, respectively (note the typological error in Freigoun et al., 2021b). \( C_0 \) is structurally feasible and is estimated in discrete-time (e.g., using PEM) if unknown \textit{a priori}. When applying a similarity transform \( T_f \),
mapping from $S_d^0$ to $S_d^1$ (both belonging to the set of structurally feasible models), it can be shown that

$$\Phi_1 = \Phi_0 T_f^T, \quad \theta_d^1 = T_f^{-T} \theta_d^0 = T_f \Gamma T_f^{-1} \theta_d^1$$

where

$$T_f = \begin{bmatrix} (T_f T_0)^{-T} \otimes I_n & 0_{n^2 \times nm} \\ 0_{nm \times n^2} & I_m \otimes T_f T_0 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} \Lambda d \Lambda^{-1} \otimes I_n & 0_{n^2 \times nm} \\ 0_{nm \times n^2} & I_m \otimes \Gamma \end{bmatrix} \quad (4.20)$$

and $T_0$ is a well-conditioned initial eigenvector matrix in (4.3). By using (4.15), $\theta_{SD}$ is obtained in terms of $T_f$ as follows

$$\theta_{SD} = T_f^{-T} H_0^{-1} T_f^{-1} P^T \left( P T_f^{-T} H_0^{-1} T_f^{-1} P^T \right)^{-1} d$$

$$- T_f^{-T} H_0^{-1} T_f^{-1} P^T \left( P T_f^{-T} H_0^{-1} T_f^{-1} P^T \right)^{-1} P T_f^{-T} T_0^{-T} \varphi$$

$$+ T_f^{-T} T_0^{-T} \varphi \quad (4.21)$$

where $H_0^{-1} = 0.5 T_0^{-T} T_0^{-1}$. Finally, with $\hat{\Phi}(T_f) \overset{\text{def}}{=} \Phi_0 T_f^T T_f \Gamma T_f^{-1}$, the standard SD loss function $J_{SD}$ is minimized with respect to $T_f$, i.e.,

$$\min_{T_f} \left( Y - \hat{\Phi} \theta_{SD} \right)^T \left( Y - \hat{\Phi} \theta_{SD} \right) \quad \text{s.t.} \quad C_0 T_f = C_0$$

(4.22)

When $Y$ is replaced with $Y_{\text{sim}} = \Phi_0 \theta_d^0$, minimizing (4.22) translates into a more direct model-matching problem. It is recognized that the objective function in (4.22) describes an approximate estimate of the actual simulation error; the appeal is in its quasi-linear form (hence referred to as “standard”). The SD formulation can, however, be applied to the standard PEM formulation for an exact description of the simulation error during the minimization procedure. Furthermore, while also nonlinear, in more practical settings it may be interesting to also consider applying a direct NRMSE objective in 3.5 jointly with (4.21) in lieu of (4.22) and classical PEM to arrive at an SD model for PEM initialization.
Figure 4.1: Simulation of the SD-identified DC Motor Model (Continuous-Time) From (4.15) Contrasted Against the Deterministic N4SID Model.

4.3 Numerical Examples

4.3.1 DC Servo Motor ($p = n, |C| \neq 0$)

In this illustration, we borrow the classical DC servo motor example presented in Ljung, 1999 by loading the demo `iddemo7` of the `ident` toolbox in MATLAB® Online R2021a. While this is a very basic example with near ideal conditions, it serves as a suitable illustration for using (4.15) since, here, a full-rank $C$ is given and both assumptions of Theorem 4.2.1 are satisfied. The simulation of the deterministic N4SID model (with $A_d \succeq 0$) is shown in Figure 4.1 (blue: 31.79% & 83.81% NRMSE fits). Given the physical model
\[ \dot{x}(t) = \begin{bmatrix} 0 & 1 \\ 0 & \theta_1 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ \theta_2 \end{bmatrix} u(t) \] (4.23a)

\[ y(t) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} x(t) + e(t) \] (4.23b)

one can observe a slight dislocation of one eigenvalue from the origin to \(-0.1578\). To calculate the structured SD model, the pair \((P,d)\) is constructed from (4.23a)

\[ P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad d = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \] (4.24)

and the solution is obtained analytically from (4.15) in a single step; simulations are in Figure 4.1 (red: 98.3\% & 84.47\%). This improvement can only be explained by the use of the structural information known \textit{a priori} to correct for fitting to noise. It is also noted that the estimated SD model from the stochastic N4SID (using the \texttt{n4sid} command in MATLAB®) gives slightly improved results (98.32\% & 84.47\%) that level with the reported PEM solution (98.35\% & 84.43\%).

4.3.2 OCSE Fluid Analogy Model of Social Cognitive Theory

The purpose of this example is to present a numerical experiment on a newly introduced, further simplified fluid analogy model of Social Cognitive Theory Martín \textit{et al.}, 2014; Freigoun \textit{et al.}, 2017; the Operant Conditioning-Self-Efficacy (OCSE) loops model in Figure 4.2. A visualization of the results is provided to illustrate the partial identifiability case of Corollary 4.2.2.1. In Section 4.4, we use the OCSE structure in this example to estimate individual semiphysical SCT models using input-output participant data retrieved from the \textit{Just Walk} study (see Appendix A). Using the conservation of mass principle, the OCSE model in Figure 4.2 can be represented by the following state-space model
Figure 4.2: Further Simplified Fluid Analogy of Social Cognitive Theory (OCSE Model).

\[
\begin{align*}
\dot{x}(t) &= A(\theta_p)x(t) + B(\theta_p)u(t) \\
y(t) &= Cx(t) + e(t)
\end{align*}
\]

where \(x(t) = [\eta_4(t) \quad \eta_3(t) \quad \eta_5(t)]^T \in \mathbb{R}^3\), \\
u(t) = [\xi_8(t) \quad \xi_9(t) \quad \xi_{10}(t) \quad \xi_{7_1}(t) \quad \ldots \quad \xi_{7_{n_d}}(t)]^T \in \mathbb{R}^{3+n_d}, y(t) = \eta_4(t) \in \mathbb{R}. \text{ With } \eta_{11}(t) = \eta_4(t) - \xi_8(t), \text{ we have }
\]

\[
A(\theta_p) = \begin{bmatrix}
-1/\tau_4 & \beta_{43}/\tau_4 & \beta_{45}/\tau_4 \\
(\gamma_{311} + \beta_{34})/\tau_3 & -1/\tau_3 & 0 \\
\beta_{54}/\tau_5 & 0 & -1/\tau_5
\end{bmatrix}
\]

\[
B(\theta_p) = \begin{bmatrix}
\gamma_{48}/\tau_4 & \gamma_{49}/\tau_4 & 0 & 0 & \ldots & 0 \\
-\gamma_{311}/\tau_3 & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & \gamma_{510}/\tau_5 & \gamma_{\tau_1}/\tau_5 & \ldots & \gamma_{\tau_5}/\tau_5
\end{bmatrix}
\]

\[
C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}
\]

with \(n_d = 5\), the parameter vector \(\theta_p \in \mathbb{R}^{16}\) is defined below

\[
\theta_p = [\tau_4, \tau_3, \tau_5, \beta_{43}, \beta_{45}, \beta_{34}, \beta_{54}, \gamma_{48}, \gamma_{49}, \gamma_{311}, \gamma_{510}, \gamma_{\tau_1}, \gamma_{\tau_2}, \gamma_{\tau_3}, \gamma_{\tau_4}, \gamma_{\tau_5}]^T
\]

In this example, 83 output sequences from the model in (4.25) with the hypothetical parameters

79
were generated by applying a subset of input-output data from a Just Walk Participant (published data were digitized according to the procedure detailed in Appendix A); see Freigoun et al., 2017; Mercere, 2017; Phatak et al., 2018. Next, similar to the actual implementation of Just Walk study, where only \( \eta_4(t) \) is measured, the hypothetically generated outputs sequences \( \eta_3(t) \) and \( \eta_5(t) \) were discarded. An additional input was constructed (similar to Weekday/Weekend) for the purposes of this example. Following Corollary 4.2.2.1 and assuming no model mismatch (here, we know this is the case), one expects to recover the true first row of \( B(\theta_p) \), including \( \gamma_{48}/\tau_4 = 3.75 \) and \( \gamma_{49}/\tau_4 = 2.5 \), as \( e(t) \to 0 \). This notion of partial identifiability is best visualized in Figure 4.3, where the poor initialization (intentional)

\[
A(\theta_{p0}) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad B(\theta_{p0}) = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}
\]

was used to obtain the classical PEM models (\texttt{ssest} in MATLAB) in the first column. For each SNR scenario, 1000 different zero mean Gaussian noise realizations were made to corrupt the ‘true’ output; a random \( T_f \in \mathcal{T}_f \) is applied to the full N4SID model with every noise realization. The first column shows results from PEM when poorly initialized; the second column shows estimates from the N4SID method; the third column shows PEM estimates when initialized by the N4SID model in the second column. The first two rows feature two identifiable parameters in \( B(\theta_p) \) per Corollary 4.2.2.1; the third row features an unidentifiable parameter in \( B(\theta_p) \). Results of this experiment and Corollary 4.2.2.1 are in agreement with conclusions drawn from the classic identifiability analysis technique of matching the transfer function of the model with an estimated one (see Section 4.4.2).
Figure 4.3: Visualization of the Partial Identifiability in $B(\theta_p)$ of the OCSE Model in (4.25) under Three Different Signal-to-noise Ratio (SNR) Scenarios Per Corollary 4.2.2.1. Solver Settings and Random Elements ($T_f$ and output noise) Retained in the Estimated Hypothetical Models for Verified Reproducibility.

4.4 Just Walk: OCSE-SCT Semiphysical Identification Results

This section reports new results extending from the prior work in Freigoun et al. and dos Santos et al.. Input-output data are used to estimate individual OCSE-SCT models. Simulation response comparisons against used participants data (data retrieved from Freigoun et al., 2017; Mercere, 2017 with high accuracy; see Appendix A) are presented in Figure 4.4. For each participant, three models are estimated to generate and compare predictions of individual behaviors [steps/day] across dif-
ferent datasets: First, the standard N4SID model is estimated. Second, using the YALMIP toolbox J. Löfberg, 2004, this first model is applied to (4.17) for estimating an eigenvalue-constrained model over the three common LMI regions in Miller and De Callafon, 2013 with $\alpha$ and $\beta$ ensuring stability, damped response, and physical realization, i.e.,

$$
\alpha = \begin{bmatrix}
(1 - \delta_s)I_2 & 0 & 0 \\
0 & 2\delta_r I_2 & 0 \\
0 & 0 & 2\delta_p I_2
\end{bmatrix}, \quad \beta = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

Finally, with the absence of prior knowledge on parameter ranges, the constrained model and the featured SD formulation in Section 4.2 are used to initialize the standard PEM estimator to enforce the OCSE structure in Example 4.3.2. The NRMSE standard defined in (3.5) is used to quantify the goodness of fit.

4.4.1 Model Validation

Once a certain model is developed and estimated (particularly black- and grey-box models), a central step in system identification is model validation. Ideally, the experimental data are split into two segments (or concatenated into two larger segments) for model estimation and validation.

Typically, the ultimate deliverable in model validation is to show that, at some level, the model in question has not been falsified by experimental data (Ljung, 1999). When comparing models and their predictions against measured input-output data, from an agreement point of view, one may be curious to run a simple Bland-Altman (BA) analysis. In areas such as analytical chemistry, the BA plot is often used to examine the agreement between measures of the same construct using two distinct methods/devices (note the use of the term agreement and not correlation; see Jinyuan
et al., 2016). As a word of caution, we cite O’Connor et al., 2011 in that “the Bland-Altman method should not be used in regression cross-validation studies.” Nonetheless, since the BA plot is ultimately a difference plot (i.e., a scatter plot of model error in this context), the BA plot can give a crude assessment for the bias and variance across different models as they are compared against measured data. Figure 4.5 presents the BA plots for each of the established models in Figure 4.4 (participant B). It is noted that the main ensuing observations from BA agreement plots are reproducible from participant A.

In Figure 4.5, the four models of Participant B are considered (four columns): the Optimal ARX model, initial N4SID subspace model, the eigenvalue-constrained derivative of the first N4SID subspace model, and structured SCT-OCSE model. In the two rows, the corresponding difference and agreement BA plots are provided with a summary of numerical values. It is clearly shown that the initial subspace model is drastically improved following model reduction and the incorporation of structural modeling insights imported from Social Cognitive Theory using the SD-PEM algorithm. It is also evident the structured OCSE delivers the best bias-variance trade-off, and the highest agreement with measured data (i.e., steps/day).

In terms of using statistical correlation in control-oriented model validation, the conventional approach in system identification includes the use the cross-validation dataset for residual analysis. With an established degree of confidence, it can be determined if the concerned model is falsified by the data via examining auto-correlation of the residuals and their cross-correlation with measured input signals. In Figure 4.6, we rely on 99% confidence intervals (i.e., 2.576σ) for a residual analysis on the validation data to conclude that the identified OCSE models are not falsified by the data (as opposed to ARX and unconstrained N4SID models for participant B). In other words, we cannot say that the OCSE models “have not picked up all dynamics” from
$u(t)$ on $\eta(t)$ (note the double negation as pointed in Ljung, 1999).

Overall, using simulation plots, BA plots, and residual analysis in Figures 4.4, 4.5, and 4.6 (respectively) we note the sensible improvements when prior knowledge is incorporated into the model (e.g., when LMI constraints in (4.17) are applied to the N4SID subspace model) and, ultimately, best improvements when models are reduced to conform with the OCSE structure in Figure 4.2, estimated via SD-PEM.
Figure 4.4: Simulations of the optimal ARX, Standard N4SID, Eigenvalue-constrained N4SID (Realizable) From (4.17) and (4.26), and SD-initialized OCSE-SCT Models for Two Representative Just Walk Participants In Freigoun et al., 2017; Mercere, 2017 (See Appendix A). The Simulations Feature the Predicted Output \( \eta_4(t) \) (Behavior) Against Input-Output Data in the Estimation, Validation, and Overall Data. Improving from Freigoun et al., 2021b, Optimal ARX Models were Added and Other Models for Participant A were Re-estimated Following improved settings. The Reader May Use the “zoom” Functionality to Enlarge This Page and View Values in the Electronic Version of This Dissertation.
Figure 4.5: Bland-Altman Agreement Plots for the Four Estimated Behavioral Models of Participant B from *Just Walk* (Figure 4.4): Optimal ARX, Unconstrained N4SID, Eigenvalue-constrained N4SID, and the Semiphysical OCSE Model of Social Cognitive Theory. *Software*: Klein, 2014.
Figure 4.6: Residual Analysis Plots for Model Validation. Auto-correlation of Residuals and Cross-Correlations with Input Signals are Plotted Against 99% Confidence Intervals for the Four Estimated Behavioral Models of Participant B from Just Walk (Figures 4.4 and 4.5): Optimal ARX, Unconstrained N4SID, Eigenvalue-constrained N4SID, and the Semiphysical OCSE Model of Social Cognitive Theory. Frames in Red (Color) Emphasize Model ‘Falsification’ by Data.
4.4.2 OCSE Structure: Identifiability Analysis

In this section, an identifiability analysis on the OCSE model in Figure 4.2 is performed. First, we reduce the unknown state-space variables in (5.2b) as follows:

\[
A = \begin{bmatrix}
a_1 & a_2 & a_3 \\
a_4 & a_5 & 0 \\
a_6 & 0 & a_7 \\
\end{bmatrix}, \quad B = \begin{bmatrix}
b_1 & b_2 & 0 & 0 & 0 & 0 & 0 \\
b_3 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & b_4 & b_5 & b_6 & b_7 & b_8 \\
\end{bmatrix}, \quad C = [1 \ 0 \ 0] \quad (4.27)
\]

with the aim of establishing the degree(s) of freedom in the model. A classical technique is to obtain the transfer function matrix (input-output unique) in terms of the variables \(\{a_i\}_{i=1:7} \& \{b_j\}_{j=1:9}\), and then verify if these can be retrieved uniquely from input-output data. Using

\[
H(s) = C(sI - A)^{-1}B + D
\]

determines the following transfer function matrix \(H(s)\):

\[
H(s) = \begin{bmatrix}
-a_1s^2 + (a_3b_1 + a_5b_1 - a_2b_3)s + (a_2a_7b_3 - a_5a_7b_1) \\
-(a_1 + a_5 + a_7)s^3 + (a_1a_5 + a_4a_7 + a_5a_7 - a_2a_4 - a_3a_6)s + (a_2a_4a_7 + a_3a_5a_6 - a_1a_5a_7) \\
-s^3 - (a_1 + a_5 + a_7)s^2 + (a_1a_5 + a_4a_7 + a_5a_7 - a_2a_4 - a_3a_6)s + (a_2a_4a_7 + a_3a_5a_6 - a_1a_5a_7) \\
-s^3 - (a_1 + a_5 + a_7)s^2 + (a_1a_5 + a_4a_7 + a_5a_7 - a_2a_4 - a_3a_6)s + (a_2a_4a_7 + a_3a_5a_6 - a_1a_5a_7) \\
-s^3 - (a_1 + a_5 + a_7)s^2 + (a_1a_5 + a_4a_7 + a_5a_7 - a_2a_4 - a_3a_6)s + (a_2a_4a_7 + a_3a_5a_6 - a_1a_5a_7) \\
-s^3 - (a_1 + a_5 + a_7)s^2 + (a_1a_5 + a_4a_7 + a_5a_7 - a_2a_4 - a_3a_6)s + (a_2a_4a_7 + a_3a_5a_6 - a_1a_5a_7) \\
-s^3 - (a_1 + a_5 + a_7)s^2 + (a_1a_5 + a_4a_7 + a_5a_7 - a_2a_4 - a_3a_6)s + (a_2a_4a_7 + a_3a_5a_6 - a_1a_5a_7) \\
-s^3 - (a_1 + a_5 + a_7)s^2 + (a_1a_5 + a_4a_7 + a_5a_7 - a_2a_4 - a_3a_6)s + (a_2a_4a_7 + a_3a_5a_6 - a_1a_5a_7) \\
\end{bmatrix}^T
\]

and, consequently, the following set of equations follow:

\[
b_1 = -\alpha_1 \quad (4.28)
\]

\[
(a_5 + a_7)b_1 - a_2b_3 = \alpha_2 \quad (4.29)
\]

\[
a_2a_7b_3 - a_5a_7b_1 = \alpha_3 \quad (4.30)
\]

\[
b_2 = -\alpha_4 \quad (4.31)
\]

\[
(a_5 + a_7)b_2 = \alpha_5 \quad (4.32)
\]

\[
a_5a_7b_2 = -\alpha_6 \quad (4.33)
\]
\begin{align*}
a_3b_4 &= -\alpha_7 \quad (4.34) \\
a_3a_5b_4 &= \alpha_8 \quad (4.35) \\
a_3b_5 &= -\alpha_9 \quad (4.36) \\
a_3a_5b_5 &= \alpha_{10} \quad (4.37) \\
a_3b_6 &= -\alpha_{11} \quad (4.38) \\
a_3a_5b_6 &= \alpha_{12} \quad (4.39) \\
a_3b_7 &= -\alpha_{13} \quad (4.40) \\
a_3a_5b_7 &= \alpha_{14} \quad (4.41) \\
a_3b_8 &= -\alpha_{15} \quad (4.42) \\
a_3a_5b_8 &= \alpha_{16} \quad (4.43) \\
a_3b_9 &= -\alpha_{17} \quad (4.44) \\
a_3a_5b_9 &= \alpha_{18} \quad (4.45) \\
a_1 + a_5 + a_7 &= -\alpha_{19} \quad (4.46) \\
a_1a_5 + a_1a_7 + a_5a_7 - a_2a_4 - a_3a_6 &= \alpha_{20} \quad (4.47) \\
a_2a_4a_7 + a_3a_5a_6 - a_1a_5a_7 &= \alpha_{21} \quad (4.48)
\end{align*}

where \(\alpha_i\) are the identified numerical coefficients of the estimated \(H(s)\). It is highlighted that, in alignment with Corollary 4.2.2.1, \(b_1 = B_{11} = \frac{\gamma_{48}}{r_4}\) and \(b_2 = B_{12} = \frac{\gamma_{49}}{r_4}\) are indeed uniquely identifiable from input-output data per (4.28) and (4.31) (also see Figure 4.3). In addition, a moment’s glance at equations (4.34)-(4.45) renders \(a_5\) known, which when plugged into (4.32) given (4.31) also determines \(a_7\). Further, with known \(a_5\) and \(a_7\), \(a_1\) is immediately obtained from (4.46).

Given known \(a_{(1,5,7)}\) and \(b_{(1,2)}\), and unknowns \(a_{(2,3,4,6)}\) and \(b_{(3-9)}\), it can be noted that the pairs \(\{a_2, b_3\}, \{a_3, b_{(4-9)}\}, \{a_2, a_4\}, \text{ and } \{a_3, a_6\}\) are coupled in equations (4.29), (4.30), (4.34)-(4.45), (4.47), and (4.48). Hence, for a unique determination of the structured model which yields global identifiability, at least two model parameters must be known \textit{a priori}: at least one parameter must belong in the set \(\{a_2, a_4, b_3\}\), and at least one another must belong in the set \(\{a_3, a_6, b_{(4-9)}\}\). The following are a few examples:

1. Given \(a_2\) and \(a_3\): given \(a_2, b_3\) is obtained from (4.29). Also, given \(a_3, b_{(4-9)}\) are obtained from (4.34)-(4.45). Finally, with \(a_2\) and \(a_3\), equations (4.47) and (4.48)
yield $a_4$ and $a_6$. Hence, the system becomes identifiable given $(a_2, a_3)$.

ii. **Given $a_4$ and $a_6$:** Plugging known $a_4$ and $a_6$ into equations (4.47) and (4.48) determines $a_2$ and $a_3$; the rest follows similar to (i.).

iii. **Given $b_3$ and any of $b_{(4-9)}$:** Equation (4.29) is used to determine $a_2$, and any of (4.34)-(4.45) can be used to determine $a_3$; the rest follows similarly from (i.).

This analysis makes quite clear that there are two degrees of freedom in the OCSE model, hence the model structure is unidentifiable, which entails an infinite number of models with identical and different structures that may be observationally equivalent. It is also clear that (iii.) corroborates Corollary 4.2.2.1 in that when all model states are directly observed (measured), the semiphysical OCSE model structure is indeed globally identifiable.

4.5 Chapter Summary

In this chapter, it was shown how a spectral decomposition of identified black-box models enables the formulation of constrained eigenvalue problems for estimating grey-box identification problems of a linear structure. The partial identifiability of a specific class of structures was discussed and a visualization from a numerical experiment was provided. Moreover, a further simplified OCSE model of Social Cognitive Theory was introduced as an extension from prior work in Chapter 3. This OCSE model was estimated and validated using input-output data. As is the case with existing literature, it may follow good practice if an explicit constraint is considered in (4.22) for managing the condition number of $T_f$ during the search. While the proposed $J_{SD}$ is still non-convex, local solutions with well-conditioned $T_f$ matrices still
provide sub-optimal matches for the underlying system characteristics and thus can produce good candidate models for initializing PEM. With randomly-initialized $J_{SD}$, it is crucial to evaluate results from multiple runs, or use the multistart functionality of existing global optimization software. Finally, it is noted that the presented method can be extended to include and handle quadratic structures; the reader is referred to Section 6.3.4 for more insights.
5.1 Background

There is strong evidence highlighting the association between sedentary lifestyle and the increased risk for type 2 diabetes, cardiovascular disease, and cardiovascular and all-cause mortality (Wilmot et al., 2012). Furthermore, McTiernan, 2008 not only attributed 25% of cancer cases worldwide to obesity and sedentarism, but also pointed out findings in several types of cancer indicating improved prognoses among individuals diagnosed with cancer and remained physically active. Motivated by the increasing access to affordable high-precision pedometers (i.e., motion sensors), the goal of maintaining 10,000 steps/day has gained more popularity and was recommended for “apparently” healthy adults (Tudor-Locke and Bassett, 2004). A sedentary lifestyle index is suggested as < 5,000 steps/day, whereas activity in the range of 5,000-7,499 steps/day is considered low-active, 7,500-9,999 is classified as somewhat active, and 10,000-12,499 steps/day is active. The aforementioned serious health ramifications of the sedentary way of life, together with the abundance of advanced mHealth technologies, have created an impetus to the development of improved interventions that are individually-tailored, adaptive, scalable, and cost-effective.

Just Walk, an innovative single-subject intervention experiment (open-loop) designed based on system identification principles, provided a unique opportunity for the
estimation and validation of individual dynamical systems (grey-box) models of Social Cognitive Theory (SCT) using input-output participant data (Hekler et al., 2018; Freigoun et al., 2017); see Chapters 3 and 4 for more elaboration on Just Walk.

Further, featured closed-loop physical activity intervention simulations in this work rely on a Hybrid Model Predictive Control (HMPC) controller design strategy. The general HMPC formulation in Bemporad and Morari, 1999; Bemporad, 2004 and hybrid decision rules in Martín et al., 2016a are considered in the scope of this work. Similar to behavior “initiation” and “maintenance” phases in Martín et al., 2016a, a controller reconfiguration strategy is proposed for the design of multi-phase interventions that optimally drive positive behavior change gradually over time. Generated closed-loop performance simulations in this work provide a proof of concept for the amenability of established methods in system identification and control to behavior-change problems.

The rest of this chapter is organized as follows: Section 5.2 features a reduced dynamical systems model of Social Cognitive Theory and reviews relevant prior system identification work. Section 5.3 reviews an existing HMPC formulation used for closed-loop intervention design. Finally, closed-loop performance simulations of a participant-validated model under different scenarios and controller configurations are provided in Section 5.4.

5.2 Grey-box Identification of Just Walk

In this section, a brief background of prior control-relevant identification work leading to participant-validated behavioral models is presented. With the development of a dynamic process model being an inherent requirement for MPC design, a reduced behavioral model lending its theoretical basis from Social Cognitive Theory is considered. The estimation and validation of this model from retrieved previously
published *Just Walk* input-output data is featured (see Appendix A).

### 5.2.1 Behavioral Process Model

In an attempt to translate a narrated presentation of SCT into a simple state-space realization amenable to theory-testing and capable of predicting behavior to a practically useful extent, Martín *et al.*, 2014 developed a sixth-order grey-box (i.e., semiphysical) SCT model using fluid analogy to capture reciprocal determinism and other key elements of the theory. A simplified version of this model (fourth-order) was published in Martín *et al.*, 2015; another fifth-order variant was featured in Freigoun *et al.*, 2017 and dos Santos *et al.*, 2018. Later, in Freigoun *et al.*, 2021b, a number of practical challenges with data and other factors imposed by the design and subsequent analyses of the pilot *Just Walk* study propelled the need for a more ‘pragmatic’ and parsimonious model by applying further simplifications leading to the ‘core,’ Operant Conditioning–Self-efficacy (OCSE) model of Social Cognitive Theory illustrated in Fig. 4.2. To name a few, these challenges and considerations stem from measurement design, data length, frequency content of relevant signals, estimated model order by inspection of singular values of a projection matrix computed intermediately in subspace analysis (suggesting third-order dynamics), optimal black-box model structures (also favoring third-order models; see Freigoun *et al.*, 2017 & dos Santos *et al.*, 2018), and the lack of sufficient prior knowledge with respect to parameter values in the original model. While most of these challenges can be alleviated through improved future *Just Walk*-like experiments, we meanwhile consider the third-order OCSE model in the scope of this chapter for the purpose of HMPC-based closed-loop intervention design. Model order determination using subspace analysis (e.g., N4SID) can be found in Theorems 2 and 8 of Van Overschee and De Moor, 2012. Further, featured closed-loop physical activity intervention simulations in this work rely on a
Figure 5.1: Conceptual Application of the Receding Horizon Control Strategy to the Physical Activity Behavioral Problem (Hekler et al., 2018).

Hybrid Model Predictive Control (HMPC) controller design. A conceptualization of the HMPC, receding horizon strategy applied to physical activity interventions is illustrated in Figure 5.1. Similar to behavior “initiation” and “maintenance” phases in Martín et al., 2016a, a controller reconfiguration strategy is proposed for the design of multi-phase interventions that optimally drive positive behavior change gradually over time. Generated closed-loop performance simulations in this work provide a proof of concept for the amenability of behavior-change problems to established methods in system identification and control.

OCSE Model

The semiphysical OCSE model in Figure 4.2 features a two-loop model consisting of operant conditioning (i.e., positive reinforcement/“negative punishment” learning) and self-efficacy loops. The reader is referred to Martín et al., 2014 and Riley
et al., 2015 for more elaborate definitions of SCT constructs and signals relevant to this model. Applying the conservation of mass principle to the fluid flow model in Figure 4.2 yields the following set of first-order differential equations:

\[
\begin{align*}
\tau_4 \frac{d\eta_4(t)}{dt} &= \gamma_{48} \xi_8(t) + \gamma_{49} \xi_9(t) + \beta_{43} \eta_3(t) + \beta_{45} \eta_5(t) - \eta_4(t) \\
\tau_3 \frac{d\eta_3(t)}{dt} &= \gamma_{311} \xi_{11}(t) + \beta_{34} \eta_4(t) - \eta_3(t) \\
\tau_5 \frac{d\eta_5(t)}{dt} &= \gamma_{510} \xi_{10}(t) + \sum_{j=1}^{n_d} \gamma_{7j} \xi_{7j}(t) + \beta_{54} \eta_4(t) - \eta_5(t)
\end{align*}
\]

which can be represented by the following state-space equations

\[
\begin{align*}
\dot{x}(t) &= Ax(t) + Bu(t) \\
y(t) &=Cx(t)
\end{align*}
\]

with system matrices defined as in (4.25).

5.2.2 Just Walk: Input Signal Design

As a part of the Just Walk open-loop intervention design, orthogonal-in-frequency multisine excitations of independent intervention dosages (i.e., \(u_n(t) = \xi_n(t); n = \{8, 9\}\)) were generated for each individual subject based on their baseline physical activity. For more details on this “zippered” spectra design, the reader may review Section 3.3 for further elaboration, as well as Martín et al., 2015 in which an input signal design procedure motivated by “patient-friendly” constraints is provided. The input signal granted points, \(u_{10}(t) = \xi_{10}(t)\), is constructed by using an “If-Then” rule defined by the function

\[
u_{10}(k) \triangleq \begin{cases} u_9(k - 1) & y(k) \geq u_8(k - 1) \\ 0 & y(k) < u_8(k - 1) \end{cases}
\]

where \(k \in \mathbb{N}\) is the discrete-time sampling instance.
5.2.3 Model Estimation and Validation

**Just Walk Identifiability Limitations**

In experimental settings where the structure of the model itself is of a particular significance, or when unknown model parameters carry some physical significance or certain implications about the underlying phenomenon, one potential requirement is the acquisition of informative measurements of all states of the minimal model representation. This requirement enables the estimation of a globally-optimal, unique model as a result of the guaranteed global identifiability. Freigoun *et al.*, 2021b presented a subspace-based spectral decomposition algorithm for estimating structured continuous-time state-space models conforming with (5.2). Maintaining two common practical assumptions, the proposed method can be of a particular value for data-driven theory testing experiments since the convexity of the estimation problem is extended from obtaining a fully-parameterized subspace model to a structured continuous-time one such as the OCSE model in Figure 4.2. Thus, in the absence of prior knowledge concerning relevant parameter values, an important experiment design requirement in future *Just Walk*-like experiments is to reliably measure responses of all SCT states in the considered SCT model. Hence, for a more complete validation of the OCSE model structure itself, measurements capturing *self-efficacy* $\eta_3(t)$ and *behavioral outcomes* $\eta_5(t)$ are as essential as the measurable *behavior* $\eta_4(t)$, i.e., $C = I$.

**Selected Participant Results**

Despite being unidentifiable in the context of *Just Walk*, individual OCSE models are estimated and validated using an SD-PEM method in Freigoun *et al.*, 2021b and digitized published *Just Walk* participant data in Freigoun *et al.*, 2017 (see Appendix
Figure 5.2: Time Series Plot Illustrating Input Signals, Measured Changes in Behavior (Output) [Steps/Day], and the Predicted Behavior from the OCSE Model Using Retrieved Input-Output Participant Data in Freigoun et al., 2017 (see Appendix A) with \( n_d = 4 \). NRMSE Fits: Estimation (34.2%, Green Region); Validation (73.65%, Cyan Region); Overall Data (41.5%).
A simulation of a selected *Just Walk* participant model (OCSE structure) is depicted in Figure 5.2. As indicated in Figure 5.2, this participant’s behavior is best predicted with the following disturbance signals: *predicted busyness*, *predicted stress*, *predicted typical*, and *weekday-weekend*. The first 27 input-output samples were reserved for model validation, and the following 61 samples were used for model estimation. An overall model NRMSE fit of 41.5% is reported in Freigoun *et al.*, 2021b, which highlights that the identified OCSE models using SD-PEM outperformed classical subspace models (i.e., standard and eigenvalue-constrained N4SID models).

**Step Responses**

While goodness of fit over validation data and residual analysis provide important criteria for model validation, step responses that are in keeping with common intuition are desirable. For the identified, participant-validated OCSE model featured
in Figures 4.2 and 5.2, step responses for manipulated inputs are shown in the top row of Figure 5.3, followed by responses to unit step changes in disturbance signals in the bottom row. Understandably, the *Just Walk* experiment length and input signals power and resolution, constrained by “patient-friendly” design considerations, can negatively impact confidence intervals. As a result, Section 5.4 is restricted to considering lower prediction and move horizons in the HMPC loop design; this is consistent with the short horizon controller tuning in Martín *et al.*, 2016a. Furthermore, as an additional value drawn from inspecting step responses, the likely directions and initial amplitudes depicted in Fig. 5.3 can guide the design of disturbance signals for the purposes of closed-loop performance evaluation (not in this chapter).

### 5.3 HMPC Framework

In this section, the main HMPC controller formulation adopted in this work is reviewed and relevance to the particular physical activity application setting is emphasized. With an HMPC design strategy with an embedded OCSE behavioral model, *setpoint tracking* (i.e., reaching and maintaining desired physical activity levels) is achieved by directly manipulating issued goals $u_8(t)$, expected points $u_9(t)$ for positive reinforcement, and indirectly via granted points per (5.3). *Measured disturbance rejection* is similarly designed for mitigating the effects from measured environmental context signals such as measured disturbances $\xi_{\tau_1}(t)$ featured in Figs. 5.2 and 5.3; correction for unmeasured disturbances and/or unmodeled dynamics is accomplished through state feedback. For the latter, Martín *et al.*, 2016a gives the example of “sickness of a family member” for a potentially immeasurable isolated (i.e., discrete) event. An additional real-life example from the *Just Walk* experiment is that one participant had reported migraine, accounting for one of the two sharpest drops in steps towards the end of the open-loop intervention as depicted in Fig. 5.2.
5.3.1 Main HMPC formulation.

In this section, we present a brief recapitulation of the optimal control problem formulation for hybrid systems in Bemporad and Morari, 1999 that describes the mechanism for the used hybcon controller object of the Hybrid Toolbox (Bemporad, 2004) in MATLAB® is presented. Moreover, operational constraints developed in Martín et al., 2016a are also incorporated in the implementation of this particular application setting.

A hybrid linear system with \( n_u \) inputs (discrete and continuous), real and integer states, and \( n_y \) outputs subject to logical/discrete decisions can be described with the following Mixed Logical and Dynamical (MLD) representation (Bemporad and Morari, 1999):

\[
\begin{align*}
    x_{k+1} &= Ax_k + B_1u_k + B_2\delta_k + B_3z_k + B_dd_k \\
    y_k &= Cx_k + D_1u_k + D_2\delta_k + D_3z_k + d'_k \\
    E_2\delta_k + E_3z_k &\leq E_1u_k + E_4x_k + E_5
\end{align*}
\]

where \( x = [x_c^T x_\ell^T]^T \) is the hybrid system state vector, \( x_c \in \mathbb{R}^{n_c} \) and \( x_\ell \in \{0,1\}^{n_\ell} \); \( u = [u_c^T u_\ell^T]^T \) is the system input with continuous and discrete/logical elements \( u_c \in \mathbb{R}^{n_c} \) and \( u_\ell \in \{0,1\}^{n_\ell} \); \( y \in \mathbb{R}^{n_y} \) is the system output; \( d \) and \( d' \) are measured and unmeasured disturbances, respectively, in \( \mathbb{R}^* \). \( \delta \in \{0,1\}^{n_\delta} \) and \( z \in \mathbb{R}^{n_z} \) are discrete/logical and continuous auxiliary variables that establishes the character of the concerned hybrid system.

With a hybrid system expressed in MLD structure, similar to classical MPC, the used hybrid optimal control problem in Bemporad, 2004 is formulated to optimize the sequence of control actions \( \{u,\delta,z\}_0^N \) over an interval \( N \). The given control law
minimizes the $p$-norm cost function $J$, i.e.,

$$\min_{\{u, \delta, z\}_0^N} J \left( \{u, \delta, z\}_0^N, x(t) \right)$$

(5.5)

where

$$J \triangleq \|Q_{xN} (x(N|t) - x_r)\|_p + \sum_{k=1}^{N} \|Q_x (x(k) - x_r)\|_p + \sum_{k=1}^{N} \|Q_u (u(k) - u_r)\|_p$$

$$+ \|Q_z (z(k|t) - z_r)\|_p + \|Q_y (y(k|t) - y_r)\|_p$$

subject to the mixed integer constraints established in (5.4c) and the following hard bounds:

$$x_{\min} \leq x_{k+i|k} \leq x_{\max}, \quad i \in [1, N] \subset \mathbb{N}$$

(5.6a)

$$u_{\min} \leq u_{k+i} \leq u_{\max}, \quad i \in [0, N-1] \subset \mathbb{I}$$

(5.6b)

$$y_{\min} \leq y_{k+i} \leq y_{\max}, \quad i \in [0, N-1] \subset \mathbb{I}$$

(5.6c)

In (5.5), $*_{r}$ denote the reference signals/levels for states, inputs, continuous auxiliary variables, and outputs, respectively; $Q_* \triangleq Q_*^T \succeq 0$ are penalty weight matrices on the feedback error, energy/control signal, auxiliary continuous variables, and outputs, respectively. The reader is referred to Bemporad and Morari, 1999; Bemporad, 2004 for more elaboration and detailed notation of the HMPC problem in (5.5).

Similar to Nandola and Rivera, 2013, for setpoint tracking and measured disturbance rejection, the implemented formulation utilizes an adjustable Type-I discrete-time filter structure from Morari and Zafiriou, 1989, i.e.,

$$f(q, \alpha_{[r,d]}) = \frac{(1 - \alpha_{[r,d]})q}{q - \alpha_{[r,d]}}$$

(5.7)

where $q$ is the forward-shift operator and $\alpha_{[r,d]} \in [0, 1)$ are adjustable tuning parameters governing the speed of response; a lower $\alpha_{[r,d]}$ drives a faster response and vice versa. While some measured disturbances of the OCSE model may be forecasted
(e.g., *weekday-weekend* signal \( \xi_7(t) \)), we assume for simplicity that all \( d_k \) signals are not forecasted in featured closed-loop performance simulations of Section 5.4.

Finally, it is noted that in the absence of a move suppression term in (5.5) that penalizes move sizes (useful in avoiding rapid change of intervention dosages), it is possible to mimic that effect in the Hybrid Toolbox via online programmatic controller reconfiguration at each time step. This may be achieved with an adjustment of hard constraints in (5.6b) or by considering soft constraints (i.e., \( \rho \) parameter of the Hybrid Toolbox, \( Q.rho \)).

### 5.3.2 HMPC-OCSE design considerations.

Particular to the design of the hybrid closed-loop system at hand, i.e., assignment of ‘optimal’ dosages of daily step goals and reward points, Martín *et al.*, 2016a established a set of operational constraints for predefined finite sets that include available intervention dosages; constraints for the realization of granted points per (5.3) were also given. To review, for the featured performance simulations in Section 5.4, the following sets of possible/feasible intervention dosages are defined

\[
\begin{align*}
\bar{u}_8(k) &\in U_8 \overset{\text{def}}{=} \left\{ c_8 \nu_1, \ldots, c_8 \nu_{n_{u_8}} \right\}, \\
\bar{u}_9(k) &\in U_9 \overset{\text{def}}{=} \left\{ c_9 \nu_{n_{u_8}+1}, \ldots, c_9 \nu_{n_{u_8}+n_{u_9}} \right\}.
\end{align*}
\]

where \( c_8 \) and \( c_9 \) are some established constants (e.g., \( c_8 \) is baseline behavior in (5.8a), \( c_9 \) is 100 points in (5.8b)) that are scaled up (or down) by value options \( \nu_i \). These constraints can be incorporated into the HMPC formulation in (5.4c) by introducing the logical and continuous auxiliary variables \( \delta_j(k) \) and \( z_j(k) \), respectively, such that

\[
z_j(k) = c_{[8,9]} \nu_j \delta_j(k)
\]
and
\[ \sum_{j=1}^{n_{u_8}} \delta_j(k) = 1, \quad u_8(k) = \sum_{j=1}^{n_{u_8}} z_j(k) \quad (5.10) \]
\[ \sum_{j=n_{u_8}+1}^{n_{u_8}+n_{u_9}} \delta_j(k) = 1, \quad u_9(k) = \sum_{j=n_{u_8}+1}^{n_{u_8}+n_{u_9}} z_j(k) \quad (5.11) \]

Furthermore, in Martín et al., 2016a, additional logical and continuous auxiliary variables \( \delta_{10}(k) \) and \( z_{10}(k) = u_{10}(k) \) are introduced to enforce the "If-Then" definition of the \textit{granted points} input \( u_{10}(k) \) per (5.3). 'Big-M' formulations in Bemporad and Morari, 1999 are used, i.e., the high-level descriptions

\[
[f_1(x) \overset{\text{def}}{=} y(k) - u_8(k-1) \leq 0] \rightarrow [\hat{\delta}_{10} = 1], \quad \hat{\delta}_{10}(k) \overset{\text{def}}{=} 1 - \delta_{10}(k),
\]

\[
\{[\delta_{10}(k) = 1] \rightarrow [z_{10}(k) = u_9(k-1)]\} \land \{[\delta_{10}(k) = 0] \rightarrow [z_{10}(k) = 0]\}
\]

establish the following set of constraints:

\[
y(k) - u_8(k-1) \leq \delta_{10}(k)[y_{\text{max}} - u_{8\text{min}}] \quad (5.12)
\]
\[
y(k) - u_8(k-1) \geq [1 - \delta_{10}(k)][y_{\text{min}} - u_{8\text{max}}] \quad (5.13)
\]
\[
u_9(k-1) - z_{10}(k) \leq [1 - \delta_{10}(k)][u_{9\text{max}} - u_{10\text{min}}] \quad (5.14)
\]
\[
u_9(k-1) - z_{10}(k) \geq [1 - \delta_{10}(k)][u_{9\text{min}} - u_{10\text{max}}] \quad (5.15)
\]

and

\[
z_{10}(k) \leq \delta_{10}(k)u_{10\text{max}}, \quad z_{10}(k) \geq \delta_{10}(k)u_{10\text{min}} \quad (5.16)
\]

Consistent with the notation in Bemporad and Morari, 1999, constraints (5.14) and (5.15) are produced with \( f_2(x) \overset{\text{def}}{=} u_9(k-1) - u_{10}(k) \), \( M_2 = u_{9\text{max}} - u_{10\text{min}} \), and \( m_2 = u_{9\text{max}} - u_{10\text{max}} \), i.e., \([\delta_{10}(k) = 1] \rightarrow [y_2 \overset{\text{def}}{=} \delta_{10}(k)f_2(x) \overset{\Delta}{=} 0]\). Constraints (5.16) are directly constructed to reflect the logic in \([\delta_{10}(k) = 0] \rightarrow [z_{10}(k) = 0]\). For simulations presented in Section 5.4, the HYSDEL software tool (Torrisi and Bemporad, 2004) was used to conveniently build (5.4c) via the construction of (5.9)-(5.16) from the established high level descriptions of the hybrid OCSE closed-loop system.
5.4 Closed-loop HMPC Simulations

In this section we present HMPC closed-loop performance simulations for single- and multi-phase intervention strategies using the participant-validated OCSE model from Just Walk (featured in Fig. 5.2). Results following different HMPC configurations and tuning are illustrated and discussed. For simulations performed in this section, the response of the ‘true’ plant (human subject) is set to match that of the model’s noise-corrupted output. In both cases (i.e., single- and multi-phase), unmeasured disturbances are Gaussian processes with \( d'_k \sim \mathcal{N}(0, 400) \). Tuning parameters in setpoint and disturbance rejection (measured and unmeasured) are \( \alpha_r = 0.95 \), \( \alpha_d = [0.1]_{nd} \). In all presented simulations, measured disturbance signals (i.e., environmental context signals in Fig. 4.2) are set to equal the values from the identified Just Walk participant featured in Fig. 5.2 in both intervention scenarios. These signals are extended (with mirroring) to proceed beyond the Just Walk experiment duration with as much resemblance to reality as possible.

5.4.1 Single-phase intervention.

Fig. 5.4 features a single-phase, 90-day simulation in which the intervention path is direct from sedentary to active with no intermediate transitions. Responses of three \( w_{u9} \) tunings of \( Q_u \) are shown; the higher \( w_{u9} \), the more suppression on issued expected points (and, consequently, total granted points) is applied throughout the intervention. When the system of points-to-money conversion in Just Walk is used (i.e., 500 points \( \equiv \$1 \); Hekler et al., 2018), Fig. 5.4 shows estimates for the granted/expected dollar amounts corresponding to each suppression setting of \( w_{u9} \). For example, when \( w_{u9} = 0.25 \), the potential expected reward of the corresponding intervention is \$18, of which the individual is actually granted \$2.2 by ‘cashing’ the monetary equivalent

105
Figure 5.4: HMPC Closed-loop Performance Simulations (Single-phase Intervention) of a Participant-validated OCSE-SCT Model. HMPC Settings: Sampling Time $T_s = 1$ day, $N = 3$, $u_{\min} = [4000 \ 0 \ 0]^T$, $u_{\max} = [10000 \ 500 \ 500]^T$, $y_{\min} = 0$, $Q_y = 1$, $Q_u = \text{diag}\{0, [0.005, 0.25, 2], 0\}$, $c_8 = 4000$, $\nu_j|\mathcal{U}_8 = \{1, 1^1/8, 1^3/16, 1^1/4, 1^3/8, 1^1/2\}$, $c_9 = 100$, $\nu_j|\mathcal{U}_9 = \{0, 1, 2, 3, 4, 5\}$.
of accumulated granted points. With a fixed value of $u^\text{max}_8$, it is intuitive to conclude from Fig. 5.4 that when incentive is reduced (via the suppression of expected points) the intervention’s outcome quality can be impacted.

5.4.2 Multi-phase intervention.

Contrary to a single-phase intervention design, a gradual increase of intervention intensity/dosages over time may be desired in many application settings. In the physical activity behavioral problem, Fig. 5.5 illustrates a multi-step intervention simulation that gradually drives an individual from a sedentary baseline state to an active state. In this case, the HMPC controller is reconfigured through an online adjustment of $u^\text{min}_8$ and $u^\text{max}_8$ at predefined stages or instances so that all phase requirements are met.

5.5 Chapter Summary

In this chapter, a more elaborate discussion on the semiphysical identification of a lower-complexity SCT model (OCSE model) was provided, followed by an outline summarizing some of the challenges and limitations arising in the context of Just Walk. Next, a participant-validated, semiphysical SCT model identified using techniques from the previous chapters was featured. More specifically, a time series plot and step responses of the OCSE model, highlighting some of the most important individual (i.e., participant) behavioral characteristics, were presented.

Further, the receding horizon strategy used in the the design of closed-loop intervention simulations was introduced; an outline for the main HMPC formulation followed. Application-specific design considerations relevant to the physical activity behavioral problem were provided and analytically expressed using the MLD framework. HMPC-governed closed-loop simulations of the identified SCT-OCSE model
were evaluated with simultaneous setpoint tracking and disturbance rejection; a simple online controller reconfiguration approach was proposed to allow for both single- and multi-phase intervention designs.
Figure 5.5: HMPC Closed-loop Performance Simulations (Multi-phase Intervention) of a Participant-validated OCSE Model. HMPC Settings Indicated in the Caption of Figure 5.4 Apply with the Following Exceptions: Only the Response Of $Q_u = \text{diag}\{0, 0.1, 0\}$ is simulated; $u_{8}^{\text{max}} = 6000 \text{ } \forall t \in [0, 80]$; $u_{8}^{\text{max}} = 8000 \text{ } \forall t \in [81, 160]$; $u_{8}^{\text{max}} = 10000 \text{ } \forall t \in [161, 240]$. 
6.1 Dissertation Summary

This dissertation has continued to demonstrate the viability of using system identification and control systems engineering frameworks in the design of optimized, perpetually adaptive behavioral health interventions. In particular, the work of this dissertation featured the use of real-life, single-subject experimental data in the estimation and validation of both behavioral and energy balance models. While applicable in other domains, the scope of presented work has remained in the domain of designing behavioral health interventions that help prevent or treat behavior-driven and “intergenerational” obesity, conceptually tracing back to intrauterine growth. More specifically, this work was thematically split into two main parts: First, an intergenerational approach included the utilization of the Healthy Mom Zone (HMZ) study (providing longitudinal experimental data) in the development, estimation, and validation of a dynamical systems model for regulating infant birth weight was presented in Chapter 2. The second part (Chapters 3-5) followed from calls in recent literature for the estimation and validation of dynamic, control-oriented Social Cognitive Theory (SCT) models using longitudinal data from experiments such as Just Walk to facilitate the promotion of physical activity among sedentary populations. Of course, the journey of fulfilling these goals has illuminated a number of technical challenges that highlighted and approached in multiple parts of this contribution.

The rest of this chapter provides an executive summary of this dissertation, as well as some conclusions and potential future directions in this research domain.
6.1.1  Intrauterine Fetal Growth Model

Using first principles modeling, particularly the laws of thermodynamics, a first-order, parameter-varying differential equation was developed and estimated in Chapter 2. Following assumptions outlined in Section 2.2.1 and fundamental concepts such as conservation of energy and entropy, this quasi-LPV model delivers an energy balance weight growth profile as a dynamic function of daily energy intake and expenditure in the forms of maternal dietary intake and physical activity, respectively. A positivity constraint was presented to establish the continuous fetal growth in utero, providing not only a model validation criterion, but also a potential diagnostic simulation tool for the early detection and prevention of small-for-gestational Age (SGA) and large-for-gestational age (LGA) growth rates. Drawing from existing literature, the presented model includes a modified, intuitive algebraic placental volume equation that can easily be estimated from experimental data guided by prior knowledge.

As opposed to classical cross-sectional studies, the featured HMZ study in this work (Chapter 2) provided a unique opportunity for estimating individual fetoplacental models using longitudinal, single-subject experimental data. Measures and estimates during the second and third trimesters (and birth weight) included daily maternal dietary intake and physical activity, estimated fetal weight and placental volume (from ultrasound measures), fetal body composition (i.e., % body fat), daily glycemic impact of food, and other measures outlined in Symons Downs et al., 2018. The estimated models were validated using a number of validation arguments, including the goodness of fit, contrasting against knowledge from existing literature, and the developed positivity constraint validation criterion. The developed model has been published and cited in recent works (Baller et al., 2019).
6.1.2 System Identification of Just Walk: Social Cognitive Theory Models

As outlined in the introductory chapter, the design of optimized behavioral health interventions using the control system engineering framework typically requires the development of a plant (‘human’) model; see Figure 1.2. Intuitively, it is clear that a more complete human model must incorporate an energy balance component that account for physiological outcomes (e.g., weight gain/loss), as well as a behavioral component that enable the prediction (and correction) of the ever-changing psychological states (e.g., self-efficacy). Similar to HMZ, the Just Walk pilot study also presented a first-of-a-kind experimental design methodology that was founded on a strong theoretical basis (i.e., Social Cognitive Theory) while simultaneously guided by system identification principles. The Just Walk study utilized in this work has also generated longitudinal individual datasets that include input-output measures of behavior and environmental context signals over the course of approximately 14 weeks, posing a unique opportunity for this contribution to estimate and validate pragmatic, control-oriented SCT models. Further, some of the challenges emerging from this pilot Just Walk experiment has inspired the development of a new system identification framework for initializing state-of-the-art grey-box solvers in difficult problems.

Using previously published Just Walk input-output data, the modeling effort was first segmented into two main consecutive parts: black-box model identification, and grey-box model estimation. With an ultimate goal of estimating and validating control-oriented SCT models, a natural first step is to identify good black-box models using participant data. Thus, individual participant datasets were utilized in establishing input-output causality, as well as in detecting potential input-input correlations (or co-linearity). With an individually identified (i.e., personalized) input-
output dataset for each participant, Auto-Regressive with eXogenous inputs (ARX) structure was considered. The choice of this structure originates from its attractive theoretical properties as further explained in Chapter 3. Next, an exhaustive search approach for the optimal ARX model was carried out across all possible configurations that include different model orders, input combinations, and estimation/validation data segmentation. A simple penalty weight approach was proposed to underscore models with most favorable statistical properties.

Further, gleaned insights from black-box identification efforts were used to pave the way to semiphysical identification of Just Walk. First, a pragmatic, further simplified SCT structure stemming from the semiphysical model in Martín et al., 2016a was introduced and featured as an Operant Conditioning–Self-Efficacy (OCSE) model in Chapter 4. With the absence of prior knowledge of parameter values, combined with the unidentifiable OCSE structure in the context of Just Walk, a new method for judicious grey-box solver initialization was proposed as part of this contribution. This Spectral Decomposition (SD) formulation relies on identifying fully-parametrized, physically realizable subspace (black-box) models. Theoretical results include the proposal of a sufficient condition for the existence of a given structure under the proposed formulation; a discussion highlighting sufficient conditions for uniqueness and identifiability followed.

Eigenvalue-constrained subspace models were established and used for the estimation of semiphysical (grey-box) OCSE-SCT models. The estimated OCSE models were validated using the NRMSE goodness of fit criterion over cross-validation data segments, in addition to observations drawn from the classic Bland-Altman (agreement) and residual analysis (correlation) plots. Results from model validation clearly marked the statistical improvements (i.e., bias and variance) as structural, SCT-driven insights are incorporated into the initial crude black-box model.
6.1.3 HMPC Loop Evaluation Using Participant-Validated Models

Following the delivery of participant-validated SCT models, it was possible to accomplish the central goal of this dissertation, which is to demonstrate (using data-driven SCT models) the viability of the dynamical systems and control approach for designing optimized and perpetually adaptive behavioral health interventions. In keeping with the trend of recent works in this domain, Chapter 5 adopted the Hybrid Model Predictive Control (HMPC) framework for adaptive intervention design. A brief overview of the Bemporad HMPC formulation followed by a review of application-specific controller configuration requirements were presented. Finally, a real-life participant-validated model was used to produce and evaluate HMPC-governed closed-loop simulations of single-phase and multi-phase intervention designs. Closed-loop performances, including simultaneous setpoint tracking and (un)measured disturbance rejection, were evaluated under different HMPC tuning settings.

6.2 Conclusions

A number of main conclusions, including conceptual, experimental, and modeling conclusions are outlined below for the interested future researcher. Furthermore, in a following subsection, a brief address to the concerned behavioral science and medicine societies is provided.

- In lieu of using hypothetical models and simulations, the work of this dissertation relied on experimental data drawn from real-life human participants and well-established system identification approaches to demonstrate the efficacy and amenability of the dynamical systems framework in importing and capturing some of the key concepts of the most popular behavioral theories in Psychology.
• Even at the very basic level of correlation analysis and black-box modeling, results from the *Just Walk* did in fact underscore the *idiosyncratic* nature of human behavior, and that a conceivably effective behavioral health intervention must consider the individual environmental context as well as *adapt* to the changing human needs. For example, using the ARX estimator, it was demonstrated that some individuals may be goal-driven, while a “busy” day may predict higher or lower levels of physical activity.

• In a collaboration with the authors of dos Santos *et al.*, 2018, a clear limitation from the pilot *Just Walk* experiment was shown to be that low-frequency characteristics (i.e., steady-state information) were not effectively captured due to insufficient low-frequency excitation, which in turn would require longer and thus more expensive studies of the same sample size. Step and frequency responses of multiple identified MoliZoft models in dos Santos *et al.*, 2018 were plotted to demonstrate this fact; agreement was only found in high frequency (transient responses), with little to no agreement in low frequency (steady-state) responses. While this may be acceptable (to an extent) for a closed-loop intervention design operating on a daily scale, it may be unacceptable for drawing conclusions or issuing predictions at steady-state, which has a value of its own. In retrospect, the designer of a *Just Walk*-like experiment may consider the incorporation of Pseudo Random Binary Sequence (PRBS) excitations (with longer pulses) in combination with the used multisine inputs used in Equation 3.1, resulting in a longer experiment at the expense of a lower sample size, or even resorting to a close-loop identification design that can typically lend more tolerance for longer studies. Moreover, longer experiments may be needed
to incorporate more variability in environmental context signals, some of which only produce sufficient variability over longer periods of time. The designer may also want to consider the feasibility of adapting a higher resolution (i.e., a finer time scale) for measuring inputs (intervention dosages), disturbances (environmental context), and output (behavior); it has been analytically shown in Chapter 4 that lower sampling time will naturally result into more accurate estimates of SCT models.

- The process of estimating and validating semiphysical models of Social Cognitive Theory in the context of Just Walk has revealed the importance of measuring all possible states, whether directly or via a strong proxy. Unfortunately, in the case of validating the OCSE model of Social Cognitive Theory in the context of Just Walk, daily measurements of self-efficacy and behavioral outcomes were absent. This has resulted in the structure being unidentifiable, which invited a host of limitations both in model estimation and validation as outlined in Chapters 4 and 5. The interested future researcher may want to consider as many model states as possible to reduce the dimensionality of the estimation problem, or be able to glean more a priori knowledge about pivotal model parameter values (as illustrated in Section 4.4.2).

- The HMPC framework remains to provide the most suitable strategy for closed-loop design of optimized, adaptive behavioral health interventions for promoting healthy levels of physical activity. Equipped with the ability to issue optimal hybrid decisions, HMPC control can be configured to incorporate logical, physical, environmental, and financial constraints that define the character of the needed intervention.
6.2.1 Remarks for the Behavioral Science and Medicine Communities

As mentioned in previous chapters, this dissertation work has been motivated by prior efforts in this application domain. The full-blown SCT model was proposed in Martín et al., 2014 noting that “testing these models with actual data is critical” and “[t]he model structure must be more deeply validated, via data that [come] from experiments”. In their final remarks, Riley et al., 2015 noted that “[o]nly by testing various components of the model with actual data will we be able to determine if this complexity is necessary or if the model can be further simplified and streamlined”, concluding that rigorous computational approaches are “needed for health behavior theory testing and intervention development.”

The work of Chapters 3 and 4 attempted to explore the quoted remarks in the context of the pilot Just Walk study that was designed using a number of principles from system identification. In Chapter 3, some of the most important conclusions drawn from exploring simple, black-box ARX models on “actual data” (i.e., Just Walk) pointed that the SCT model can potentially be reduced and further simplified. For example, in addition to reduction by established idiosyncrasy (i.e., personalization of inputs/predictors; Figure 3.4), Table 3.1 and similar efforts with the MoliZoft modeling methods in dos Santos et al., 2018 all point out to the potential viability of a reduction to a third-order SCT model. Based on these insights, the OCSE model “structure” in Figure 4.2 was proposed and studied in Chapters 4 and 5.

Now that a further simplified, third-order OCSE model of Social Cognitive Theory is established, it is only natural to estimate and validate such a model in the context of Just Walk experimental data. However, there are two types of challenges associated with that: First, as discussed in Chapter 4, the prediction-error method (PEM) typically used for estimating grey-box models such as OCSE from input-output data
(used in Martín et al., 2014) is nonconvex and thus principally depends on good initial guesses for parameter values for its success in reaching a global optimum solution or at least a “good enough” local one (see first column “PEM (Poor Initialization)” in Figure 4.3). Second, it was shown in Chapter 4 that in the absence of reliable and informative measurements for both self-efficacy and behavioral outcomes and prior knowledge concerning model parameter values (at least the ‘pivotal’ ones discussed in Section 4.4.2), the OCSE model becomes fundamentally unidentifiable and cannot be formally validated (or invalidated) using input and behavior data. This was also the case with the proposed SCT model in Martín et al., 2014 in the context of the used MILES data. Nonetheless, the effort in Chapter 4 was concerned with proposing this reduced OCSE structure with preliminary model validation using data in Appendix A.

In future Just Walk-like experiments, once actual informative and reliable data are available for all model states such as behavior, self-efficacy, and behavioral outcomes for the OCSE model in Figure 4.2 (also the Cue to Action and Outcome Expectancy states for the model in Figure 1.1), the SCT model in question becomes identifiable from data. The contribution in Chapter 4 distilled in Equation (4.15) delivered a convexification of the grey-box estimation problem (see Example 4.3.1), providing a valuable formal model validation (or invalidation) and “theory testing” tool that is ‘immune’ to potential “local minima” challenges provided that both assumptions in Theorem 4.2.1 are held. When and if a formal OCSE model validation is established in future Just Walk-like studies (with all model states reliably measured), strategies informed by the various constructs (i.e., self-efficacy, etc.) can be developed and utilized in an HMPC-based closed-loop intervention design scheme.
6.3 Potential Future Directions

As the scope of work in this dissertation explored a library of dynamical systems models useful for designing optimized and adaptive behavioral health interventions, an initial effort constrained by experimental limitations; the following are a few interesting directions of potential future work relevant to this research.

6.3.1 Further developments in the fetal growth model

In the work of Chapter 2, we postulated an empirical, ‘modified’ logistic function characterizing the phases of placental volume growth from conception through birth. Unfortunately, this function (Equation (2.29)) does not incorporate nor produce any further physiological insights other than the general growth profile shaped by the known cell multiplication and spacial constraints. In future developments, it may prove to be useful having a model that characterizes all significant influences on placental growth, other than just time. This model can stem either from first principles modeling or by coupling known physiological factors with experimental results. Further, following observations from Thomas et al., 2008, the impact of maternal physical activity on fetal growth was assumed to be mediated by the placental volume, i.e.,

\[ \gamma(t) = \alpha PA(t) + \beta \]  \hspace{1cm} (2.22)

This needs to be further established using future studies and/or new analyses of potentially existing data; see Baller et al., 2019 for a more elaborate reference of existing studies.

Finally, while it might have been clear to justify the assumed positivity (or non-negativity) of fetal growth rate throughout gestation, it is however, not clear in a mathematical sense how to capture events such as starvation or malnutrition. The potential for developing a more complete model in that respect may lie in revealing
the functional form of the $\alpha_W(t)$ in

$$\dot{I}_f(t) = \gamma(t) [g(t)m(t) + \alpha_W(t)W_m(t)]P(t) \quad (2.14)$$

6.3.2 LPV Modeling of Social Cognitive Theory

While identification results from earlier chapters show real promise for the potency of Linear Time-Invariant (LTI) dynamical systems in predicting and explaining human behavior, it is not only conceivable but in fact known that underlying dynamics driving behavior change can indeed be time-varying and nonlinear (Hayes et al., 2007; Korinek et al., 2018). Given the well-understood theory and properties of linear systems in both identification and control, one may be interested in exploring nonlinearities and time-varying characteristics driving human behavior-change using the LTI framework, which is on offer by LPV identification. Similar to the provided reasoning that resulted in choice of ARX modeling in Chapter 3, and following the general advice from Ljung, 1999 on “try simple things first!”, an LPV-ARX model structure is thus the recommended choice of initial structure for identification, followed by introducing and exploring practical LPV extensions to the semiphysical OCSE-SCT model. The attractive properties of the LPV-ARX model structure such as consistency (Cox, 2018, Theorem 6.1) and convexity (existence of a unique analytical solution) continue to extend from LTI-ARX.

6.3.3 Application of HMPC Design in ‘Real-life’ Intervention Settings

Despite that all models produced in this contribution represent real-life participants from Just Walk, one is indeed welcome to question the efficacy of the estimated behavioral models or proposed intervention designs over an extended periods of time not only in the scale of months, but several years. Guo has written a similar section in a recent dissertation relevant to this application domain, stating that “it would
be useful to examine how models could be obtained for the intervention in practice,” and that, similar to the case with this dissertation, “HMPC-based control was not performed online.” As such, in agreement with Guo, 2018, an adaptive, closed-loop identification strategy is proposed to approach the ‘longevity’ argument. One may consider such a design in future Just Walk-like experiments by starting with an initial model estimated from a short baseline, open-loop dataset or from an “averaged” participant response data from prior analyses as in Chapters 3 and 4.

6.3.4 Toward the ‘Convexification’ of Semiphysical Identification & Extensions to the SD Method

The proposed Spectral Decomposition (SD) identification algorithm in this work benefited two main assumptions outlined in Section 4.2. Namely, it was assumed that the initially obtained black-box model possesses a number of distinct eigenvalues (i.e., poles) that is equal to the order of the minimal state-space representation of that system. The second assumption is that input excitations in the data-generating experiment follow a zero-order hold intersampling behavior. A powerful extension may consider to deliver further derivations and/or explicit conditions that enable the elimination of one (or both) of these assumptions.

Additionally, the SD algorithm introduced a loss function that is nonlinear and is thus nonconvex with respect to the similarity transformation $T_f$. In order to become amenable to the larger family of optimization methods and solvers, one may be interested in deriving exact or at least good gradient and Hessian approximations for solving (4.22).

Further, on a theoretical level, the delivered SD formulation has only considered linear grey-box structures. It is noted that one may also consider the extension of delivered theoretical results in Chapter 4 for quadratic structures which may be
needed for further model specification. A more general quadratic structure quintuplet \((P, d, Q, q_i, r_i)\) can result into the following quadratically constrained quadratic program (QCQP)

\[
\begin{align*}
\min_{\theta} & \quad \frac{1}{2} \theta^T H \theta + f^T \theta \\
\text{s.t.} & \quad \frac{1}{2} \theta^T Q_i \theta + q_i^T \theta + r_i \leq 0 \quad \forall i = 1, \ldots, l \\
& \quad P \theta = d
\end{align*}
\]

(6.1)

with variables defined in (4.11) and \(Q, q_i, r_i\) enforcing the quadratic structural constraints; \(l\) is the number of established quintuplets. Note the amenability of this formulation to the incorporation of integer and/or binary constraints. Finally, it is known from Chapter 4 that for an estimation problem with \(H\) and \(Q\) being positive semidefinite (i.e., \(H \succeq 0\) and \(Q_i \succeq 0\) \(\forall i = 1, \ldots, l\)), the resulting QCQP in (6.1) remains convex.


APPENDIX A

PUBLISHED JUST WALK DATA DIGITIZATION
A.1 Overview

In Chapters 2 and 3, this dissertation had access to the original datasets from the *Healthy Mom Zone* and *Just Walk* studies, respectively. However, for unrestricted publication purposes, the rest of this dissertation strictly resorted to recovering the *Just Walk* data used in the completion of Chapters 4 and 5 from published graphical sources. The purpose of this appendix is to describe the data digitization process in support of results in Chapters 4 and 5 and provide the necessary information and arguments that underline the estimated accuracy and confidence in the obtained final sets.

The primary concern in Chapter 4 was to deliver an identification formulation for estimating structured state-space (grey-box) models such as the Operant Conditioning–Self-Efficacy (OCSE) model introduced in Figure 4.2. In order to preliminarily validate the OCSE model using experimental data of real human participants as well as test the efficacy of the developed spectral decomposition formulation relative to “off-the-shelf” software and methods, input-output data for Participants A and B published in Mercere, 2017 and Freigoun et al., 2017 (respectively) were digitized and recovered with high ‘accuracy’ (and full confidence in 13 out of 14 total signals; see Table A.7). To promote further inquiry in this area, the final recovered sets from the cited sources are given in Tables A.2-A.6 for the interested researcher.
A.2 Known Facts & Graphical Sets

To support statements of confidence given in Table A.7, the following Just Walk information must be highlighted:

1. Recovery of data points from the published electronic versions of Mercere, 2017; Freigoun et al., 2017 was carried out using the Adobe Illustrator vector graphics editor, producing descaled data with a precision of 7 decimal points (see Graphical Sets 1-15). Fortunately, both cited sources provided the concerned plots with high accuracy information in the form of line, polyline, and polygon graphical objects that maintain the relative distance between all data points established by the published plot.

Both line and polyline graphical objects contain ‘unique’ descaled data points with precision of 7 decimal places. The one and only case containing a polygon object was treated by ‘cutting’ the polygon to create two polyline sets (see Graphical Sets 6 and 7). It is known from these objects that the ‘true’ value must lie within the maximum possible gap between the two polyline sets originating from the original polygon object.

2. In both signals, Goals and Expected Points, it is known from Freigoun et al., 2017; Phatak et al., 2018; Korinek et al., 2018 that orthogonal-in-frequency, 16-day repeating cycles were used in the design of these signals (see Table 1 in Korinek et al., 2018 for a real representative sample). Using the known “IF-THEN” rule, the Granted Points signal value at day $k$ must be either zero or identical to the Expected Points value at day $k - 1$ (see Equation 5.3).

3. It is also known from Korinek et al., 2018; Phatak et al., 2018 (which was also visually and numerically verified) that Expected Points are issued on a 100-500
point scale, restricted to multiples of 25 only (see Table 1 in Korinek et al., 2018 for a real representative sample).

4. By definition, all other signal values are strictly whole numbers (i.e., 0 or positive integers) on predefined scales in Freigoun et al., 2017; Korinek et al., 2018; Phatak et al., 2018 as follows. **Behavior**: whole number; **Goals**: whole number; **Predicted Busyness**: 1-4; **Predicted Stress**: 1-5; **Predicted Typical**: 1-4; **Weekday - Weekend**: 0 for weekday, 1 for weekend.

5. The total number of available data points in both sources are 88 (for 88 days), starting from a time index of 0. This is verified visually as well as can be inferred from the retrieved Graphical Sets 1-15.

The following Graphical Sets 1-15 include retrieved data points to arbitrary scales from Mercere, 2017; Freigoun et al., 2017 using a high-precision software.

---

**Graphical Set 1: Participant A (Goals)**

```xml
<line id="XMLID_7_" class="stb" x="1.3.716952" y="18.2731934" y2="8" y2="18.2731934"/>
```

---

137
Graphical Set 5: Participant A (Predicted Stress)

Graphical Set 6: Participant A (Actual Steps: Top Curve)

Graphical Set 7: Participant A (Actual Steps: Bottom Curve)
Graphical Set 12: Participant B (Predicted Stress)

Graphical Set 13: Participant B (Predicted Typical)

Graphical Set 14: Participant B (Weekday - Weekend)
Table A.1: Key for Scaling to Reported Ticks from retrieved Graphical Sets 1-15
Used for Chapter 4 (Participants A & B) and Chapter 5 (Participant B)

<table>
<thead>
<tr>
<th>Just Walk</th>
<th>Graphical Set#</th>
<th>Tick ID</th>
<th>Reported Real Value</th>
<th>Tick ID</th>
<th>Reported Real Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>XMLID_7_</td>
<td>6000</td>
<td>XMLID_5_</td>
<td>8000</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>XMLID_3_</td>
<td>6000</td>
<td>XMLID_2_</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>XMLID_6_</td>
<td>200</td>
<td>XMLID_5_</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>XMLID_2_</td>
<td>0</td>
<td>XMLID_5_</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>XMLID_2_</td>
<td>0</td>
<td>XMLID_5_</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>XMLID_6_</td>
<td>6000</td>
<td>XMLID_5_</td>
<td>8000</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>XMLID_6_</td>
<td>6000</td>
<td>XMLID_5_</td>
<td>8000</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>XMLID_6_</td>
<td>6000</td>
<td>XMLID_5_</td>
<td>8000</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>XMLID_5_</td>
<td>0</td>
<td>XMLID_3_</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>XMLID_4_</td>
<td>0</td>
<td>XMLID_3_</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>XMLID_5_</td>
<td>0</td>
<td>XMLID_4_</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>XMLID_5_</td>
<td>0</td>
<td>XMLID_4_</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>XMLID_4_</td>
<td>0</td>
<td>XMLID_3_</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>XMLID_4_</td>
<td>0</td>
<td>XMLID_3_</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>XMLID_3_</td>
<td>6000</td>
<td>XMLID_2_</td>
<td>8000</td>
</tr>
</tbody>
</table>

A.3 Final Processed Sets, ‘Accuracy’ & Confidence

Since cited source plots draw on a linear scale, the equation \( y = mx + c \) is used for the recovery of the final set from any initially established arbitrary scale. Graphical Sets 1-15 are used with the scaling information provided in Table A.1.

Following appropriate rounding per the facts in Section A.2 and exclusion of re-
dundant points (i.e., staircase plots, polygon object), the final recovered sets from
the described high-precision digitization procedure are provided in Tables A.2-A.6.
Finally, estimated accuracy information and confidence statements are presented in
Table A.7.
Table A.2: Final Processed Set (1/3) of Published *Just Walk* Plot from Merecere, 2017 (Participant A) Used in Chapter 4: Intervention Days 0-40 (see Table A.7 for Estimated Accuracy and Confidence)

<table>
<thead>
<tr>
<th>Day</th>
<th>Goals Expected</th>
<th>Expected Points</th>
<th>Granted Points</th>
<th>Predicted Busyness</th>
<th>Predicted Stress</th>
<th>Behavior ((\Delta_k))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>steps/day</td>
<td>100-500</td>
<td>0-500</td>
<td>1-4</td>
<td>1-5</td>
<td>steps/day</td>
</tr>
<tr>
<td>0</td>
<td>5193</td>
<td>100</td>
<td>175</td>
<td>3</td>
<td>2</td>
<td>8075</td>
</tr>
<tr>
<td>1</td>
<td>4818</td>
<td>300</td>
<td>100</td>
<td>2</td>
<td>2</td>
<td>5438</td>
</tr>
<tr>
<td>2</td>
<td>8474</td>
<td>450</td>
<td>300</td>
<td>4</td>
<td>3</td>
<td>6699</td>
</tr>
<tr>
<td>3</td>
<td>8060</td>
<td>200</td>
<td>450</td>
<td>3</td>
<td>3</td>
<td>8861</td>
</tr>
<tr>
<td>4</td>
<td>8608</td>
<td>150</td>
<td>200</td>
<td>4</td>
<td>3</td>
<td>9207</td>
</tr>
<tr>
<td>5</td>
<td>5815</td>
<td>450</td>
<td>150</td>
<td>2</td>
<td>2</td>
<td>12694</td>
</tr>
<tr>
<td>6</td>
<td>4317</td>
<td>475</td>
<td>450</td>
<td>3</td>
<td>3</td>
<td>6554</td>
</tr>
<tr>
<td>7</td>
<td>7961</td>
<td>425</td>
<td>475</td>
<td>2</td>
<td>2</td>
<td>8813</td>
</tr>
<tr>
<td>8</td>
<td>5189</td>
<td>500</td>
<td>425</td>
<td>3</td>
<td>2</td>
<td>8180</td>
</tr>
<tr>
<td>9</td>
<td>4813</td>
<td>300</td>
<td>500</td>
<td>3</td>
<td>2</td>
<td>5264</td>
</tr>
<tr>
<td>10</td>
<td>8492</td>
<td>150</td>
<td>300</td>
<td>3</td>
<td>3</td>
<td>4870</td>
</tr>
<tr>
<td>11</td>
<td>8038</td>
<td>400</td>
<td>150</td>
<td>3</td>
<td>3</td>
<td>9082</td>
</tr>
<tr>
<td>12</td>
<td>8634</td>
<td>450</td>
<td>400</td>
<td>2</td>
<td>3</td>
<td>8486</td>
</tr>
<tr>
<td>13</td>
<td>5789</td>
<td>150</td>
<td>450</td>
<td>1</td>
<td>2</td>
<td>9097</td>
</tr>
<tr>
<td>14</td>
<td>4339</td>
<td>125</td>
<td>150</td>
<td>2</td>
<td>3</td>
<td>7538</td>
</tr>
<tr>
<td>15</td>
<td>7948</td>
<td>175</td>
<td>125</td>
<td>3</td>
<td>2</td>
<td>5276</td>
</tr>
<tr>
<td>16</td>
<td>5193</td>
<td>100</td>
<td>175</td>
<td>3</td>
<td>2</td>
<td>8637</td>
</tr>
<tr>
<td>17</td>
<td>4818</td>
<td>300</td>
<td>100</td>
<td>3</td>
<td>3</td>
<td>5479</td>
</tr>
<tr>
<td>18</td>
<td>8474</td>
<td>450</td>
<td>300</td>
<td>4</td>
<td>4</td>
<td>5120</td>
</tr>
<tr>
<td>19</td>
<td>8060</td>
<td>200</td>
<td>450</td>
<td>4</td>
<td>4</td>
<td>9086</td>
</tr>
<tr>
<td>20</td>
<td>8608</td>
<td>150</td>
<td>200</td>
<td>3</td>
<td>4</td>
<td>8504</td>
</tr>
<tr>
<td>21</td>
<td>5815</td>
<td>450</td>
<td>150</td>
<td>1</td>
<td>2</td>
<td>9819</td>
</tr>
<tr>
<td>22</td>
<td>4317</td>
<td>475</td>
<td>450</td>
<td>3</td>
<td>2</td>
<td>10434</td>
</tr>
<tr>
<td>23</td>
<td>7961</td>
<td>425</td>
<td>475</td>
<td>4</td>
<td>4</td>
<td>7002</td>
</tr>
<tr>
<td>24</td>
<td>5193</td>
<td>500</td>
<td>425</td>
<td>3</td>
<td>2</td>
<td>8040</td>
</tr>
<tr>
<td>25</td>
<td>4818</td>
<td>300</td>
<td>500</td>
<td>3</td>
<td>3</td>
<td>6717</td>
</tr>
<tr>
<td>26</td>
<td>8492</td>
<td>150</td>
<td>300</td>
<td>4</td>
<td>3</td>
<td>6170</td>
</tr>
<tr>
<td>27</td>
<td>8038</td>
<td>400</td>
<td>150</td>
<td>4</td>
<td>4</td>
<td>11126</td>
</tr>
<tr>
<td>28</td>
<td>8634</td>
<td>450</td>
<td>400</td>
<td>2</td>
<td>2</td>
<td>9349</td>
</tr>
<tr>
<td>29</td>
<td>5789</td>
<td>150</td>
<td>450</td>
<td>3</td>
<td>3</td>
<td>10127</td>
</tr>
<tr>
<td>30</td>
<td>4339</td>
<td>125</td>
<td>150</td>
<td>4</td>
<td>4</td>
<td>6421</td>
</tr>
<tr>
<td>31</td>
<td>7948</td>
<td>175</td>
<td>125</td>
<td>4</td>
<td>4</td>
<td>5139</td>
</tr>
<tr>
<td>32</td>
<td>5193</td>
<td>100</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>2669</td>
</tr>
<tr>
<td>33</td>
<td>4818</td>
<td>300</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>2026</td>
</tr>
<tr>
<td>34</td>
<td>8474</td>
<td>450</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>1485</td>
</tr>
<tr>
<td>35</td>
<td>8060</td>
<td>200</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>4187</td>
</tr>
<tr>
<td>36</td>
<td>8608</td>
<td>150</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>3193</td>
</tr>
<tr>
<td>37</td>
<td>5815</td>
<td>450</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5745</td>
</tr>
<tr>
<td>38</td>
<td>4317</td>
<td>475</td>
<td>450</td>
<td>4</td>
<td>3</td>
<td>6695</td>
</tr>
<tr>
<td>39</td>
<td>7961</td>
<td>425</td>
<td>475</td>
<td>4</td>
<td>3</td>
<td>4335</td>
</tr>
<tr>
<td>40</td>
<td>5193</td>
<td>500</td>
<td>425</td>
<td>3</td>
<td>3</td>
<td>8120</td>
</tr>
<tr>
<td>Day</td>
<td>Goals Expected Points 100-500</td>
<td>Granted Points 0-500</td>
<td>Predicted Busyness 1-4</td>
<td>Predicted Stress 1-5</td>
<td>Behavior</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>--------------------------------</td>
<td>---------------------</td>
<td>------------------------</td>
<td>---------------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>4813</td>
<td>300</td>
<td>500</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>8492</td>
<td>150</td>
<td>300</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>8038</td>
<td>400</td>
<td>150</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>8634</td>
<td>450</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>5789</td>
<td>150</td>
<td>450</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>4339</td>
<td>125</td>
<td>150</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>7948</td>
<td>175</td>
<td>125</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>5193</td>
<td>100</td>
<td>175</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>4818</td>
<td>300</td>
<td>100</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>8474</td>
<td>450</td>
<td>300</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>8060</td>
<td>200</td>
<td>450</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>8608</td>
<td>150</td>
<td>200</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>5815</td>
<td>450</td>
<td>150</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>4317</td>
<td>475</td>
<td>450</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>7961</td>
<td>425</td>
<td>475</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>5189</td>
<td>500</td>
<td>425</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>4813</td>
<td>300</td>
<td>500</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>8492</td>
<td>150</td>
<td>300</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>8038</td>
<td>400</td>
<td>150</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>8634</td>
<td>450</td>
<td>400</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>5789</td>
<td>150</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>4339</td>
<td>125</td>
<td>150</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>7948</td>
<td>175</td>
<td>125</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>5193</td>
<td>100</td>
<td>175</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>4818</td>
<td>300</td>
<td>100</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>8474</td>
<td>450</td>
<td>300</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>8060</td>
<td>200</td>
<td>450</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>8608</td>
<td>150</td>
<td>200</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>5815</td>
<td>450</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>4317</td>
<td>475</td>
<td>450</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>7961</td>
<td>425</td>
<td>475</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>5189</td>
<td>500</td>
<td>425</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>4813</td>
<td>300</td>
<td>500</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>8492</td>
<td>150</td>
<td>300</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>8038</td>
<td>400</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>8634</td>
<td>450</td>
<td>400</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>5789</td>
<td>150</td>
<td>450</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>4339</td>
<td>125</td>
<td>150</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>7948</td>
<td>175</td>
<td>125</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>5193</td>
<td>100</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Table A.3: Final Processed Set (2/3) of Published *Just Walk* Plot from Mercere, 2017 (Participant A) Used in Chapter 4: Intervention Days 41-80
(see Table A.7 for Estimated Accuracy and Confidence)
Table A.4: Final Processed Set (3/3) of Published *Just Walk* Plot from Mercere, 2017 (Participant A) Used in Chapter 4: Intervention Days 81-87 (see Table A.7 for Estimated Accuracy and Confidence)

<table>
<thead>
<tr>
<th>Day</th>
<th>Goals steps/day</th>
<th>Expected Points 100-500</th>
<th>Granted Points 0-500</th>
<th>Predicted Busyness 1-5</th>
<th>Predicted Stress 1-5</th>
<th>Behavior steps/day</th>
</tr>
</thead>
<tbody>
<tr>
<td>81</td>
<td>4818</td>
<td>300</td>
<td>100</td>
<td>4</td>
<td>5</td>
<td>5638 (12)</td>
</tr>
<tr>
<td>82</td>
<td>8474</td>
<td>450</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>1944 (18)</td>
</tr>
<tr>
<td>83</td>
<td>8060</td>
<td>200</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>5215 (16)</td>
</tr>
<tr>
<td>84</td>
<td>8608</td>
<td>150</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>3656 (16)</td>
</tr>
<tr>
<td>85</td>
<td>5815</td>
<td>450</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>6528 (18)</td>
</tr>
<tr>
<td>86</td>
<td>4317</td>
<td>475</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>3226 (12)</td>
</tr>
<tr>
<td>87</td>
<td>7961</td>
<td>425</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>4092 (7)</td>
</tr>
<tr>
<td>Day</td>
<td>Expected Points</td>
<td>Granted Points</td>
<td>Predicted Busyness</td>
<td>Predicted Stress</td>
<td>Predicted Typical</td>
<td>Weekday/Weekend</td>
</tr>
<tr>
<td>-----</td>
<td>----------------</td>
<td>----------------</td>
<td>--------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td>steps/day</td>
<td>100-500</td>
<td>0-500</td>
<td>1-4</td>
<td>1-5</td>
<td>{0, 1}</td>
</tr>
<tr>
<td>0</td>
<td>8442</td>
<td>175</td>
<td>200</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>9102</td>
<td>450</td>
<td>175</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>5834</td>
<td>325</td>
<td>450</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>4551</td>
<td>100</td>
<td>325</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>8424</td>
<td>175</td>
<td>100</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5143</td>
<td>150</td>
<td>175</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>5120</td>
<td>125</td>
<td>150</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>8970</td>
<td>400</td>
<td>125</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>8640</td>
<td>425</td>
<td>400</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>9097</td>
<td>150</td>
<td>425</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>5825</td>
<td>275</td>
<td>150</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>4569</td>
<td>500</td>
<td>275</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>8397</td>
<td>425</td>
<td>500</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>5170</td>
<td>450</td>
<td>425</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>5093</td>
<td>475</td>
<td>450</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>8993</td>
<td>200</td>
<td>475</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>8442</td>
<td>175</td>
<td>200</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>9102</td>
<td>450</td>
<td>175</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>5834</td>
<td>325</td>
<td>450</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>4551</td>
<td>100</td>
<td>325</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>8424</td>
<td>175</td>
<td>100</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>5143</td>
<td>150</td>
<td>175</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>5120</td>
<td>125</td>
<td>150</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>23</td>
<td>8970</td>
<td>400</td>
<td>125</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>8640</td>
<td>425</td>
<td>400</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>9097</td>
<td>150</td>
<td>425</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>5825</td>
<td>275</td>
<td>150</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>4569</td>
<td>500</td>
<td>275</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>28</td>
<td>8397</td>
<td>425</td>
<td>500</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>29</td>
<td>5170</td>
<td>450</td>
<td>425</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>5093</td>
<td>475</td>
<td>450</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>31</td>
<td>8993</td>
<td>200</td>
<td>475</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>8442</td>
<td>175</td>
<td>200</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>33</td>
<td>9102</td>
<td>450</td>
<td>175</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>34</td>
<td>5834</td>
<td>325</td>
<td>450</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>35</td>
<td>4551</td>
<td>100</td>
<td>325</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>36</td>
<td>8424</td>
<td>175</td>
<td>100</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>37</td>
<td>5143</td>
<td>150</td>
<td>175</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>38</td>
<td>5120</td>
<td>125</td>
<td>150</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>39</td>
<td>8970</td>
<td>400</td>
<td>125</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>8640</td>
<td>425</td>
<td>400</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>41</td>
<td>9097</td>
<td>150</td>
<td>425</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>42</td>
<td>5825</td>
<td>275</td>
<td>150</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>43</td>
<td>4569</td>
<td>500</td>
<td>275</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>44</td>
<td>8397</td>
<td>425</td>
<td>500</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>45</td>
<td>5170</td>
<td>450</td>
<td>425</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>46</td>
<td>5093</td>
<td>475</td>
<td>450</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>47</td>
<td>8993</td>
<td>200</td>
<td>475</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>48</td>
<td>8442</td>
<td>175</td>
<td>200</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>49</td>
<td>9102</td>
<td>450</td>
<td>175</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>5834</td>
<td>325</td>
<td>450</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

Behavior steps/day
Table A.6: Final Processed Set (2/2) of Published *Just Walk* Plot from Freigoun et al., 2017 (Participant B) Used in Chapters 4 and 5: Intervention Days 51-87 (see Table A.7 for Estimated Accuracy and Confidence)

<table>
<thead>
<tr>
<th>Day</th>
<th>Goals Expected Steps/day</th>
<th>Expected Points</th>
<th>Granted Points</th>
<th>Predicted Busyness</th>
<th>Predicted Stress</th>
<th>Predicted Typical Weekday/Weekend</th>
<th>Behavior Steps/day</th>
</tr>
</thead>
<tbody>
<tr>
<td>51</td>
<td>4551</td>
<td>100</td>
<td>325</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>6162</td>
</tr>
<tr>
<td>52</td>
<td>8424</td>
<td>175</td>
<td>100</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5398</td>
</tr>
<tr>
<td>53</td>
<td>5143</td>
<td>150</td>
<td>175</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>8458</td>
</tr>
<tr>
<td>54</td>
<td>5120</td>
<td>125</td>
<td>150</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5522</td>
</tr>
<tr>
<td>55</td>
<td>8970</td>
<td>400</td>
<td>125</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5381</td>
</tr>
<tr>
<td>56</td>
<td>8460</td>
<td>425</td>
<td>400</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>9194</td>
</tr>
<tr>
<td>57</td>
<td>9097</td>
<td>150</td>
<td>425</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>58</td>
<td>5825</td>
<td>275</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3977</td>
</tr>
<tr>
<td>59</td>
<td>4569</td>
<td>500</td>
<td>275</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>6120</td>
</tr>
<tr>
<td>60</td>
<td>8307</td>
<td>425</td>
<td>500</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4725</td>
</tr>
<tr>
<td>61</td>
<td>5170</td>
<td>450</td>
<td>425</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>8482</td>
</tr>
<tr>
<td>62</td>
<td>5093</td>
<td>475</td>
<td>450</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5385</td>
</tr>
<tr>
<td>63</td>
<td>8993</td>
<td>200</td>
<td>475</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>6395</td>
</tr>
<tr>
<td>64</td>
<td>8442</td>
<td>175</td>
<td>200</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>9235</td>
</tr>
<tr>
<td>65</td>
<td>9102</td>
<td>450</td>
<td>175</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>8900</td>
</tr>
<tr>
<td>66</td>
<td>5834</td>
<td>325</td>
<td>450</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>9359</td>
</tr>
<tr>
<td>67</td>
<td>4551</td>
<td>100</td>
<td>325</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5985</td>
</tr>
<tr>
<td>68</td>
<td>8424</td>
<td>175</td>
<td>100</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4698</td>
</tr>
<tr>
<td>69</td>
<td>5143</td>
<td>150</td>
<td>175</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>8533</td>
</tr>
<tr>
<td>70</td>
<td>5120</td>
<td>125</td>
<td>150</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>6279</td>
</tr>
<tr>
<td>71</td>
<td>8970</td>
<td>400</td>
<td>125</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>5729</td>
</tr>
<tr>
<td>72</td>
<td>8460</td>
<td>425</td>
<td>400</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>9442</td>
</tr>
<tr>
<td>73</td>
<td>9097</td>
<td>150</td>
<td>425</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>8673</td>
</tr>
<tr>
<td>74</td>
<td>5825</td>
<td>275</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2739</td>
</tr>
<tr>
<td>75</td>
<td>4569</td>
<td>500</td>
<td>275</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6043</td>
</tr>
<tr>
<td>76</td>
<td>8307</td>
<td>425</td>
<td>500</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4633</td>
</tr>
<tr>
<td>77</td>
<td>5170</td>
<td>450</td>
<td>425</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>8910</td>
</tr>
<tr>
<td>78</td>
<td>5093</td>
<td>475</td>
<td>450</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>6511</td>
</tr>
<tr>
<td>79</td>
<td>8993</td>
<td>200</td>
<td>475</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5099</td>
</tr>
<tr>
<td>80</td>
<td>8442</td>
<td>175</td>
<td>200</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>9135</td>
</tr>
<tr>
<td>81</td>
<td>9102</td>
<td>450</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2317</td>
</tr>
<tr>
<td>82</td>
<td>5834</td>
<td>325</td>
<td>450</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>10016</td>
</tr>
<tr>
<td>83</td>
<td>4551</td>
<td>100</td>
<td>325</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6047</td>
</tr>
<tr>
<td>84</td>
<td>8424</td>
<td>175</td>
<td>100</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>6396</td>
</tr>
<tr>
<td>85</td>
<td>5143</td>
<td>150</td>
<td>175</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>8755</td>
</tr>
<tr>
<td>86</td>
<td>5120</td>
<td>125</td>
<td>150</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5258</td>
</tr>
<tr>
<td>87</td>
<td>8970</td>
<td>400</td>
<td>125</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5222</td>
</tr>
</tbody>
</table>
Table A.7: Presentation of Round Off Accuracy and Confidence in the Recovery of Published Graphical Data in 1-15 from Cited Sources ($N_A = N_B = 88$ measurements)

<table>
<thead>
<tr>
<th>Just Walk Participant</th>
<th>Graphical Set #</th>
<th>Rounding Accuracy</th>
<th>Confidence</th>
<th>Notes*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Min</td>
<td>Mean</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>99.18%</td>
<td>99.99%</td>
<td>Exact Recovery</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>98.09%</td>
<td>99.97%</td>
<td>Exact Recovery</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>99.81%</td>
<td>99.95%</td>
<td>Exact Recovery</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>99.98%</td>
<td>99.99%</td>
<td>Exact Recovery</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>99.98%</td>
<td>99.99%</td>
<td>Exact Recovery</td>
</tr>
<tr>
<td></td>
<td>6, 7</td>
<td>N/A**</td>
<td>N/A**</td>
<td>$</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>99.9997%</td>
<td>99.9999%</td>
<td>Exact Recovery</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>99.9992%</td>
<td>99.9998%</td>
<td>Exact Recovery</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>99.9995%</td>
<td>99.9997%</td>
<td>Exact Recovery</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>99.9999%</td>
<td>99.9999%</td>
<td>Exact Recovery</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>99.9992%</td>
<td>99.9997%</td>
<td>Exact Recovery</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>99.9996%</td>
<td>99.9997%</td>
<td>Exact Recovery</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>100%</td>
<td>100%</td>
<td>Exact Recovery</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>99.9999%</td>
<td>99.9999%</td>
<td>Exact Recovery†</td>
</tr>
</tbody>
</table>

*All conclusions in this tabulation and below are drawn with the strict use of published *Just Walk* information outlined in Section A.2 (sufficient to the extent of given statements).

**In this case, information from both polyline objects in Sets 6 and 7 (resulting from ‘cutting’ an obtained polygon object) are used jointly; only confidence can be estimated. Unique data points (in time) from both polyline objects are selected with an accuracy of a single polyline (i.e., similar to accuracy in all other sets).

†*Estimated* from $\lceil \max \Delta_k \rceil$, where $\Delta_k \in \mathbb{R}^{88}$ is the vector containing maximum integer gap values (absolute) between both polyline objects from Sets 6 and 7. *Estimated accuracy* is $> 99.07\%$. Minimum gap is 0 steps/day; maximum is 25 steps/day.

‡“Exact” is reported since the recovery/digitization procedure produced a single polyline object provided in Set 15 (with identical properties to Sets 1-5, 8-14), rendering the *estimated accuracy* $> 99.9999\%$ and hence the “exactness” from knowing that steps/day must be a whole number.