Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 58-15 (2024) 568–573

Understanding "Just-in-Time" States in Behavioral Interventions using System Identification and Data Science Methods ⋆

Mohamed El Mistiri * **Sarasij Banerjee** * **Daniel E. Rivera** *
Predrag Klasnja ** **Junghwan Park** *** **Eric Hekler** ***

* *Control Systems Engineering Lab, Arizona State University, Tempe, AZ 85281 USA (e-mail: {melmisti, sbaner74, daniel.rivera}@asu.edu).*
** *Division of Biomedical and Health Informatics, School of Information, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: klasnja@umich.edu)*
*** *Center for Wireless & Population Health Systems, University of California, San Diego (UCSD), La Jolla, CA 92093 USA (e-mail: {jup014, ehekler}@ucsd.edu)*

Abstract: Insufficient physical activity (PA) is highly prevalent in society, despite its negative impact on personal health. Effective interventions are clearly needed. However, improvements to behavioral interventions face many challenges, from noisy and missing measurements to inadequate understanding of the dynamics of behavior change in context. In this paper, we present a comprehensive, data-driven, system identification approach aimed at overcoming challenges to better understand the dynamics of PA behavior change, which in turn can improve the efficacy of these interventions. The proposed approach consists of an innovative input signal design (aimed at providing informative data sets to study the concept of "just-in-time" dynamics), Singular Spectrum Analysis (SSA; for noise reduction and exploring the separability of the measured output signal), and Model-on-Demand (MoD) estimation, a hybrid data-driven modeling approach, which allows identifying dynamics under changing operating conditions. The proposed approach is evaluated on data for a representative participant from the *JustWalk JITAI* study. The results demonstrate significant potential of the methodology in enhancing the understanding of the dynamics of behavior change in context.

Keywords: eHealth, nonlinear system identification, time series modelling, filtering and smoothing

1. INTRODUCTION

The benefits of physical activity (PA) to health are numerous; an increase from 4,000 to 8,000 steps/day is linked to the reduction of all causes of mortality by 51% (Saint-Maurice et al., 2020). However, the majority of the population does not meet the recommended CDC guidelines for PA; therefore most people do not reap the benefits of healthy levels of PA (Olson et al., 2018). Although digital behavior change interventions (DBCIs) have demonstrated their potential in promoting healthy behaviors including PA, the efficiency of classical DBCIs is hurdled by various obstacles (Schoeppe et al., 2016). Much of traditional behavior medicine focuses on static models on a nomothetic level (i.e., group or population level), that only examine if an intervention component has positive outcomes without explaining the dynamics and drivers behind behavior change. Consequently, traditional studies do not take into account idiosyncrasies within a group or a population, and on a deeper level idiosyncrasies within an individual over time. Hence, there is a lack of understanding of the dynamic and context-varying nature of systems associated with behavior change, especially at an individual level. Therefore, the delivery of behavioral interventions is not personalized or optimized to ensure the effectiveness of DBCIs in promoting sustained healthy behavior.

The availability of temporally dense PA data afforded by advances in and popularity of wearable technologies that track PA levels (e.g., FitBit) has set the stage for the re-emergence of idiographic data analysis methods. Such methods are particularly geared towards understanding behavior change dynamics on an individual level, by analyzing and predicting patterns within time series data corresponding to each individual. These temporally dense longitudinal data sets provide unprecedented opportunities to apply system identification, and control systems engineering principles to implement data-driven solutions to problems faced in behavioral medicine, in terms of optimization and personalization of decision-making in an intervention. However, many challenges are faced in extracting value out of data, both on exploratory and predictive levels (Hekler et al., 2016). Hence, it is particularly important to focus on 1) the design of experiments to provide dynamically informative data sets; and

⋆ This research was funded by the National Institutes of Health (NIH) grants R01LM013107 and R01CA244777.

10.1016/j.ifacol.2024.08.590

2) developing and improving signal processing and model estimation methods that capture system nonlinearities and idiosyncrasies in context.

In behavioral medicine, the concept of "just-in-time" adaptive interventions (JITAIs) has been introduced to improve the understanding of the impact of context and multi-timescale dynamics associated with behavior change, and by extension increase the efficiency of decision-making. The essence of JITAIs is in providing support only when a participant has the need, opportunity, and receptivity to respond positively to that support. This can result in meaningful and sustained adaptations of healthy behavior over time in what is known as "just-in-time" (JIT) states. Therefore, it is imperative to understand JIT context to address JITAIs as an optimization problem, where robust prediction and decision-making are needed to provide support that contributes towards sustainable healthy behavior change. In pursuit of exploring and understanding these concepts, a digital health intervention study (*JustWalk JITAI*; Park et al. (2023)) has been developed and implemented. The results of which are presented in this paper.

JustWalk JITAI is one of the first empirical studies of JIT states based on system identification principles. To facilitate understanding and analysis of such complex dynamic phenomena, innovative input signals are designed and implemented in the design of experiment. Additionally, advanced data-driven modeling techniques are utilized in a comprehensive approach. This paper provides a brief overview of elements comprising the proposed approach and examines its application on a representative *JustWalk JITAI* participant. The results presented demonstrate the significant potential of the proposed system identification approach in improving the understanding of behavior change systems in context.

The paper is organized as follows. Section 2 provides a brief description of the *JustWalk JITAI* study and the input signal design. Section 3 details the implemented estimator, and its advantages in terms of modeling behavior change in context, while in Section 4 a brief summary of Singular Spectrum Analysis (SSA) is provided. In Section 5, results of the proposed approach are presented and discussed for a representative participant. Section 6 provides conclusions along with implications on future work.

2. STUDY DESCRIPTION & INPUT SIGNAL DESIGN

JustWalk JITAI (R01LM013107) is an NIH-funded study aimed at advancing the understanding of multi-timescale dynamics and JIT states context for supporting PA, in the form of daily Step Count. A new and innovative design of experiments is introduced in this unique study. The intervention was designed to provide informative data on short and long timescale dynamics in behavior change systems related to within-day and between-day PA, respectively. This is essential in operationalizing JITAIs as a multi-timescale robust optimization problem, where support is only provided at JIT states in a manner that ensures it contributes towards sustained behavior change. To accomplish this, two intervention components are utilized: 1) adaptive daily goals and 2) walking notifications.

In the adaptive daily goals component, each participant is given daily targets of the number of steps they should meet on a given day, to help them move towards eventually meeting national PA recommendations; these are dictated by a realization of the designed multisine signal for this component. To adapt the signal to the performance of each participant, the maximum and minimum daily step goals in each cycle are adjusted based on the performance in the previous cycle. Consequently, this intervention component is further personalized to each participant, by providing ambitious yet achievable goals. This aspect of the design produces informative data that captures the longer term dynamics at high and low participant performance.

The walking notifications component of the *JustWalk JITAI* focuses on within-day dynamics and JIT states. This component consists of inspiring messages sent to participants within-day on 4 decision points (once every 3 hours starting at 7 am), to invite them to go on a short walk (e.g., 10 minutes). The decision on whether to send a notification or not at every decision point is dictated by a decision rules signal designed for this intervention component. This input signal is designed to test the effectiveness of notifications sent on what is perceived as full or partial JIT states in comparison to fully randomized notifications. To operationalize this decision, three conditions about the current moment are taken into account:

- *Need* (N): If the participant is not on-track to meet the given daily goal.
- *Opportunity* (O): Whether the next 3-hours window is predicted to be an opportune window for the participant to engage in PA, based on a previously developed algorithm .
- *Receptivity* (R): If the participant has received less than six notifications, and responded favorably (i.e., walked) to half of the notifications sent to them in the last 72 hours.

Three different combinations of the JIT conditions to send notifications are chosen as the decision rules in the intervention: 1) N+R, 2) N+O, 3) N+O+R which is labeled as the JIT decision rule. These three combinations are compared to fully randomized decisions.

To design a categorical input signal for the walking notifications component, a unique approach is followed. A pseudo-random binary sequence (PRBS) is generated to serve as the base signal. Then, a three-level uniformly distributed random multi-level sequence (RMLS) is superimposed on one of the binary levels in the PRBS signal. This is done to evaluate the impact of the different JIT state combinations in comparison to fully randomized notifications, as well as to study the dynamic nature of the response to the notifications under different conditions. A detailed description of the input signal design is provided in El Mistiri et al. (2022); details regarding the study design and experimental protocol are published in Park et al. (2023).

3. MODEL-ON-DEMAND ESTIMATION

One of the main objectives of the *JustWalk JITAI* is to explore the multi-timescale dynamics and the dynamic impact of JIT states on PA behavior change. In this paper, the focus is on understanding the dynamic nature of the system in the context of JIT states. To reach this aim, Model-on-Demand (MoD; Braun et al. (2001)) is utilized, which is a sophisticated data-centric modeling approach. MoD is an adaptive modeling approach, which does not rely on a global model, making it a perfect candidate to explore and explain JIT states. In MoD, the weighted least-squares regression problem in (1) is solved to fit the data in a bounded neighborhood around each operating point c, under a specified global regressor structure $[n_a\, n_b\, n_k]$. Therefore, MoD combines both local and global modeling. Hence, in MoD, the estimation data is not discarded, rather a subset of the data is used to estimate a local model at each operating point on demand.

$$\hat{\theta} = \arg\ \min_{\theta} \sum_{k=1}^{N} \ell\left(y_k - \hat{m}(\varphi_k, \theta)\right) K_h\left(\frac{\|\varphi_k - \varphi_c\|_M}{h}\right) \quad (1)$$

where $\hat{\theta} = [\hat{\theta}_0\ \hat{\theta}_{1,1}\ \hat{\theta}_{1,2}\ \cdots\ \hat{\theta}_{1,p_{reg}}]$ represents parameters of the local polynomial model of the regressors, and φ_* is the regressor vector at the point $* \in \{c, k\}$ defined by

$$\begin{aligned}\varphi_* = [&y_{*-1}\ \cdots\ y_{*-n_a}\ u_{1,*-n_{k_1}}\ \cdots\ u_{1,*-n_{b_1}-n_{k_1}+1} \\ &u_{n_u,*-n_{k_{n_u}}}\ \cdots\ u_{1,*-n_{b_{n_u}}-n_{k_{n_u}}+1}]^T\end{aligned} \quad (2)$$

k is the sample instance from the estimation database of length N. The length of the parameter vector p_{reg} is determined by the length of the regressor vector d_{reg}, where $p_{reg} = d_{reg}$ for the linear case, and $p_{reg} = d_{reg} + d_{reg}(d_{reg}+1)/2$ for the quadratic case. Weighting of the data points in the neighborhood of c is performed through a kernel-based function $K_h(\cdot)$, where higher weights are given to points closer to c based on their scaled distance in regressor space and bandwidth h. Distance scaling function defined as $\|\tilde{\varphi}\|_M = \sqrt{\tilde{\varphi}^T M \tilde{\varphi}}$, where $\tilde{\varphi} = \varphi_k - \varphi_c$, and M is the scaling matrix obtained from the inverse covariance of the regressors. The bandwidth h is selected adaptively through an iterative process to optimize over local information criteria (e.g., Akaike Information Criterion (AIC)), within a user-defined range $[k_{min},\, k_{max}]$ for neighborhood size. Other user-defined design variables for MoD include regressor and polynomial orders over which the local model is fitted, the goodness-of-fit (GoF) criteria to select the neighborhood size, and the kernel function utilized in assigning weights to data points.

Five input signals ($n_u = 5$) are utilized to define conditions in regressor space, including designed intervention components and other exogenous signals, as follows:

- *Step Goals* (u_1): the daily step goal given to a participant, as described in Section 2.
- *Decision Rules* (u_2): these define the utilized JIT rule for sending notifications on a specific day, as described in Section 2.
- *Viewed Walking Notifications* (u_3): number of walking notifications viewed by the participant on a day.
- *Temperature* (u_4): the recorded highest daily temperature.
- *Weekend* (u_5): a binary signal representing whether a given day is a weekend (1) or a weekday (0).

The modeled output of interest is PA behavior (y), in terms of daily Step Count, which also serves as part of the regressor φ_c to define context at each point.

4. NOISE REDUCTION & SIGNAL SEPARABILITY: SINGULAR SPECTRUM ANALYSIS (SSA)

There are several challenges faced in analyzing dynamic data related to PA behavior. For instance, the accuracy of the measurements can be heavily influenced by the quality of the sensors used in activity tracking, and the algorithms utilized to filter the raw data (Bender et al., 2017). In addition, the availability of the measurement and its quality are predicated on participants' adherence to wearing the measurement device. As a result, missingness can exist on no wear days and measurement levels can be significantly low on low wear-time days, which does not represent the actual behavior. Moreover, the collected Step Count signal represents the aggregate impact of idiosyncratic forces impacting behavior on different timescales. For example, life rhythms (e.g., work, weekends) can contribute to the daily Step Count in a periodic manner, which is independent of the overall trend seen due to intentionally engaging in PA.

In this work, we rely on Singular Spectrum Analysis (SSA) to reduce noise in the Step Count measurement and study its separability. SSA is a Singular Value Decomposition (SVD) based nonparametric technique for signal processing and time series analysis. SSA is beneficial in analyzing different aspects of the system, like studying separability of the measured signal to its components, filtering measurement noise, and studying causality (Hassani and Zhigljavsky, 2009). In SSA, time series data is decomposed into a sum of its components by performing SVD on the Hankle matrix of the original data. Each component in the sum is then designated into groups of trend, periodic, quasi-periodic, or noise components. Moreover, crosscorrelated components can be grouped and summed to capture their collective characteristics. Finally, the filtered output signal is constructed through the sum of the most relevant system components, excluding noise.

The Hankel matrix H is constructed by transferring the one-dimensional output signal $Y = [y_1\, y_2\, \ldots\, y_N]$ of length N into a series of L-lagged vectors as follows:

$$H = [h_1\, h_2\, \ldots\, h_M] = \begin{bmatrix} y_1 & y_2 & \cdots & y_M \\ y_2 & y_3 & \cdots & y_{M+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & \cdots & y_N \end{bmatrix} \quad (3)$$

where $M = N - L + 1$, and L is an integer such that $2 \leq L < N$ which represents the window length utilized to construct the Hankel matrix. L is a user selected parameter and must be sufficiently large to capture important system dynamics. In this work, the window length L is selected based on the periodicity observed in the autocovariance of the output signal. This is done by obtaining the mean distance between lags at which peaks in the autocovariance occur. L is then selected to include all lags within twice the mean distance between peaks.

SVD is then performed on H to decompose it into a total of L rank-one bi-orthogonal elementary matrices. The eigenvalues of HH^T are represented by λ_i in descending order ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L \geq 0$), while their corresponding orthonormal left and right eigenvectors are denoted by U_i and V_i respectively, where $i = 1, \cdots, L$. Therefore, the Hankel matrix can be re-written as the linear combination of L matrices $H = \sum_{i=1}^{L} H_i$ where $H_i = \sqrt{\lambda_i} U_i V_i^T$.

Each matrix H_i represents a component of the output signal, which is then Hankelized into $\hat{H}_i$ by calculating the antidiagonal average. Consequently, the summation of the Hankel matrices of all output signal components $\hat{H}_i$ represents the Hankel matrix of the original signal H. Time series signals $Y_{SSA,i}$ for each component i are then extracted from each Hankelized component matrix $\hat{H}_i$. As a result, the original signal is the summation of the reconstructed components

$$H = \sum_{i=1}^{L} \hat{H}_i \Rightarrow Y = \sum_{i=1}^{L} Y_{SSA,i} \tag{4}$$

The relevance of each component r_i is determined by its contribution to the sum of all real singular values

$$r_i = 100 \frac{\sqrt{\lambda_i}}{\sum_{i=1}^{L} \sqrt{\lambda_i}} \tag{5}$$

A threshold value (Th) is selected to filter out the least relevant components, where the subset of the selected important components is of size $d = \max(i \mid r_i > Th \ \forall \ i = 1, \cdots, L)$. Because of SSA's formulation, high frequency components (which are often associated with noise) represent the least contribution to the total of the singular values $(\sum_{i=1}^{L} \sqrt{\lambda_i})$. Therefore, this step effectively represents noise reduction in the reconstruction of the filtered output signal $Y_{SSA} = \sum_{i=1}^{d} Y_{SSA,i}$. Subsequently, crosscorrelared signals from the selected reconstructed components $(Y_{SSA,1}, \ldots, Y_{SSA,d})$ are aggregated to represent their collective impact in each group (e.g., trend, seasonality 1) at its respective frequency range.

5. RESULTS & DISCUSSION

5.1 SSA: Noise Reduction and Signal Separability

The autocovariance function of the measured Step Count signal for this participant indicated a periodicity of 7 lags. Consequently, a window length of $L = 15$ days is utilized. As a result, 15 different SSA components are obtained. A threshold value $Th = 4\%$ is utilized to filter out higher frequency components, reducing the number of relevant components to $d = 5$. These components are used to reconstruct the SSA-filtered output signal shown in Fig. 1.

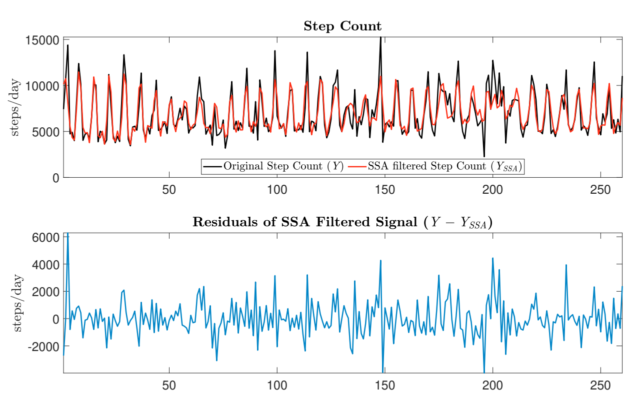

Fig. 1. Original (Y) and SSA-filtered (Y_{SSA}) daily Step Count signals in steps/day, along with residuals.

Results in Fig. 1 demonstrate the effectiveness of SSA in noise reduction. This is evident in how the SSA-filtered signal appears as a smoothed version of the original signal; maintaining important dynamics, while excluding high frequency changes that can be attributed to noise. Moreover, the impact of significantly high and low data points, which can be considered outliers that do not represent the actual system dynamics, is reduced in the SSA-filtered time series.

By analyzing crosscorrelation between the relevant components, they are found to form three distinct groups:

(1) *Trend*: contains the first component, which covers low frequencies and represents the underlying slower dynamics of changes in the Step Count trend over time.
(2) *Seasonality 1*: consistent of the second and third components. This group covers intermediate frequencies, where periodicity is observed in a 7 day pattern.
(3) *Seasonality 2*: covers higher frequencies, with an observed periodicity over a 3-4 day span. The fourth and fifth components are grouped in this set.

The reconstructed signals for each group are presented in Fig. 2. The obtained results illustrate that the collected daily Step Count signal is separable. Moreover, these results demonstrate that the benefit of SSA goes beyond noise reduction; SSA breaks down time series data components into uncorrelated groups that capture different features of the overall behavior. Consequently, each component can be modeled independently to help understand the dynamic nature of the idiosyncratic forces influencing behavior change in context, at different frequencies.

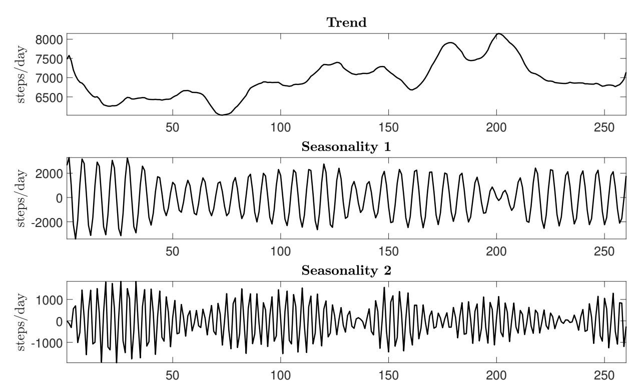

Fig. 2. Reconstructed signals of uncorrelated groups of SSA components for Step Count. The reconstructed groups are arranged from top to bottom based on their increasing covered frequencies.

5.2 Model Estimation and Crossvalidation

For the sake of brevity, we focus in this paper on modeling the filtered output signal Y_{SSA} as a reduced noise representation of the Step Count. The measured data is standardized to ensure well-conditioned matrices in model estimation. The data is then segmented into sub-experiments, where each experiment contains two consecutive goal setting cycles. Consequently, five different sub-experiments are constructed, with 52 days of the intervention in each sub-experiment. The selected regressor order for the local polynomial is $[n_a \, n_b \, n_k] = [4 \ 4 \ 1]$. Data for each sub-experiment is then transferred to regressor space φ, reducing the number of data points available in each sub-experiment to 48. The sub-experiments are then grouped into estimation and validation data groups, where four sub-experiments are designated for estimation, and one is for validation, yielding a total of five possible combinations. Sub-experiments' data in regressor space is then merged based on their group to allow for model estimation and crossvalidation across all possible combinations, where the combination with the highest weighted normalized root mean square error (NRMSE) fit index is selected.

AIC is utilized as a GoF measure for the localized models, with a variance penalty of 3. Because of data scarcity, the neighborhood size has been fixed to the maximum number of available data points in regressor space, which is accomplished by selecting $k_{min} = k_{max} = 192$. Consequently, variations in the local models are only dependent on the kernel-based weighting function of the relevance of points in the neighborhood around c, which is selected as a tricube kernel.

Fig. 3 compares the performance of MoD and ARX-based estimators of the same regressor orders in a simulation setting, over both estimation and validation data. The estimated models illustrate the effectiveness of the proposed input signal design and SSA noise reduction in providing dynamically informative data sets. This is particularly evident in the obtained NRMSE fits, which are sufficiently high for noisy systems associated with behavior change. The effectiveness of this approach is asserted by the performance of the presented estimators in predicting the unfiltered Step Count Y, as summarized in Table 1.

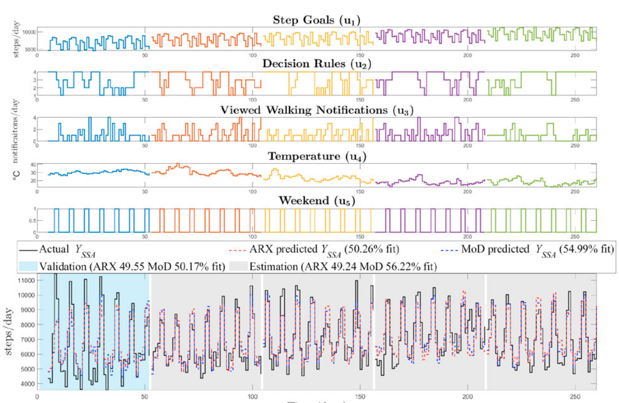

Fig. 3. Simulation results for MoD and ARX-based estimators compared to the SSA-filtered Step Count Y_{SSA}.

MoD achieves better results than the ARX-based global model in predicting the SSA-filtered Step Count based on the presented NRMSE fit indices. The outperformance of the MoD estimator, utilizing the SSA-filtered output signal Y_{SSA} in the estimation database, is further highlighted in comparison with the original unfiltered Step Count Y.

Table 1. NRMSE fit indices for MoD and ARX-based estimators utilizing SSA-filtered Step Count, in comparison to raw Step Count Y and SSA-filtered Step Count Y_{SSA}

Method	Step Count Y		SSA-filtered Step Count Y_{SSA}	
	Validation Fit (%)	Overall fit (%)	Validation Fit (%)	Overall fit (%)
MoD	40.65	29.76	50.17	54.99
ARX	38.36	26.31	49.55	50.26

Additionally, SSA noise reduction significantly enhances the performance of MoD. Because of the reduced noise impact on the localized models, MoD is capable of estimating local models that predict the original signal Y with higher NRMSE fits, in comparison to the same estimator utilizing unfiltered Step Counts Y in the estimation database. While improvements from using MoD might seem modest in terms of NRMSE fit indices, the real added benefit of the MoD estimator is that it optimizes over localized models under a global structure. Therefore, MoD yields improved global fits with local models that vary in dynamics with respect to the operating conditions. This provides the framework needed to analyze idiosyncrasies in a participant's behavior in context.

5.3 MoD: Analyzing Behavior in Context

Having validated the MoD estimator, the benefit of the estimator in explaining PA behavior change in context is now explored. To allow for more data points to be included in estimating localized models, the entirety of the available data is included in the MoD estimation database. Consequently, the total length of the MoD estimator's database is $N = 256$ in regressor space. The neighborhood range is updated accordingly to $k_{min} = k_{max} = 256$. The MoD estimator is then implemented in simulation for hypothetical scenarios, to illustrate the impact of changing operating conditions on the obtained responses.

To showcase the difference between the MoD and ARX-based estimators, the simulation conditions are fixed for the exogenous environmental inputs, while some intervention components are manipulated. In this scenario, Temperature is held constant at $u_4 = 30°C$, and it is assumed that the intervention occurs on a weekday ($u_5 = 0$.) As per intervention components, the JIT decision rule is implemented, while Step Goals start initially at an ambitious level $u_1 = 10,000$ steps/day, then are reduced to a less challenging level $u_1 = 6,000$ steps/day. During periods of both ambitious and non-ambitious goals, a pulse of walking suggestions is sent to the participant and assumed to be viewed, $u_3 = 3$ notifications/day for five consecutive days.

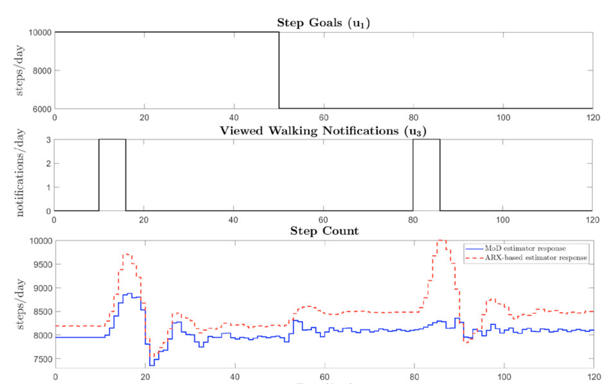

Fig. 4. Pulse responses for MoD and ARX-based estimators in a simulated hypothetical scenario.

As illustrated in Fig. 4 impulse responses for the MoD estimator vary significantly based on operating conditions. During the first pulse of notifications, the Step Goal signal is at an ambitious level relative to the steady state value of the Step Count. Under these operating conditions, the sent walking suggestions have a significant positive impact on the Step Count. The daily Step Count increases by a factor of 930 steps/day at the peak, due to the Viewed Walking Notifications u_3. Meanwhile, the second impulse of notifications, with the same magnitude and duration, has a diminished effect on the system. Viewed Walking Suggestions in this case, contribute to an increase of only 285 steps/day when the given goal is lower than the steady-state value of the Step Count. Moreover, variations in MoD's pulse responses extend beyond the magnitude of the gain, and include the speed and shape of response as seen in Fig. 4. On the other hand, the ARX-based estimated model yields the same pulse responses, regardless of the operating conditions.

From a behavioral science perspective, variations in responses to the same notification pulse can be attributed to the participant's perception of the utility of the received notification with respect to the context at which the notifications are received. For instance, if the support through the notification is provided while the goal is not achieved (i.e., there is a need for support) then it is of high utility. Therefore, the participant acts upon the viewed notifications, and these materialize as efforts toward the goal. Conversely, when support is provided without the need for it (e.g., when the given goals are easily attainable), then its perceived value diminishes and it can be burdensome for the participant. Hence, it does not lead to the intended positive outcomes. Therefore, the obtained MoD simulation results for these hypothetical scenarios, presented in Fig. 4, fit perfectly within JIT concepts and explain idiosyncrasies in the responses of this representative *JustWalk JITAI* participant to walking suggestions based on context. The same analysis can be extended to various scenarios, including at different Decision Rules u_2, to study and predict their impact.

6. CONCLUSION & FUTURE WORK

In this work, a comprehensive system identification approach to studying PA behavior change is presented. This approach includes: informed input signal design, signal processing, and noise reduction, as well as a data-driven model estimator geared towards capturing nonlinearities in the system based on operating conditions. The findings in this paper illustrate the cumulative effect of this approach in analyzing and understating idiosyncrasies in the dynamics of PA related behavior change, based on context.

The proposed input signal design based on *a priori* knowledge, provides dynamically informative data sets at frequencies of interest. Furthermore, SSA utilization in Step Count time series analysis is remarkably effective in reducing measurement noise in an informed manner, and studying the separability of the measured output signal. The analysis shows that the daily Step Count has separable uncorrelated groups of components that can be analyzed independently. Finally, MoD is an excellent alternative approach to global modeling that captures nonlinearities associated with behavior change systems and JIT states. The full potential of this approach will culminate in the development of efficient personalized JITAIs within the control-optimization-trial (COT) framework, which can be disseminated on a large scale to reduce physical inactivity and improve public health.

Future efforts will include further personalization of the MoD estimator for each participant, through the incorporation of algorithms to optimize over input and order selection for the global MoD estimator structure. Moreover, analysis of the reconstructed signals of uncorrelated groups of SSA components for Step Count will be explored. The presented comprehensive system identification approach will be applied to intraday data (i.e., sampled at 3 hours intervals) to help understand the multi-timescale dynamics associated with JIT states and PA behavior change.

ACKNOWLEDGEMENTS

We thank Dr. Guillaume Mercère for introducing our lab to SSA. Research support has been provided by the National Institutes of Health (NIH) through grants R01LM013107 and R01CA244777. The opinions expressed here are the authors' own and do not necessarily reflect the views of NIH.

REFERENCES

Bender, C.G., Hoffstot, J.C., Combs, B.T., Hooshangi, S., and Cappos, J. (2017). Measuring the fitness of fitness trackers. In *2017 IEEE Sensors Applications Symposium (SAS)*, 1–6. IEEE.

Braun, M.W., Rivera, D.E., and Stenman, A. (2001). A Model-on-Demand identification methodology for nonlinear process systems. *Int. J. Control*, 74(18), 1708 – 1717.

El Mistiri, M., Rivera, D.E., Klasnja, P., Park, J., and Hekler, E. (2022). Enhanced Social Cognitive Theory dynamic modeling and simulation towards improving the estimation of 'just-in-time' states. In *2022 American Control Conference (ACC)*, 468–473.

Hassani, H. and Zhigljavsky, A. (2009). Singular spectrum analysis: methodology and application to economics data. *Journal of Systems Science and Complexity*, 22, 372–394.

Hekler, E.B., Klasnja, P., Riley, W.T., Buman, M.P., Huberty, J., Rivera, D.E., and Martin, C.A. (2016). Agile science: creating useful products for behavior change in the real world. *Translational Behavioral Medicine*, 6(2), 317–328.

Olson, R.D., Piercy, K.L., Troiano, R.P., Ballard, R.M., Fulton, J.E., Galuska, D.A., Pfohl, S.Y., Vaux-Bjerke, A., Quam, J.B., George, S.M., Sprow, K., Carlson, S.A., Hyde, E.T., and Olscamp, K. (2018). Physical activity guidelines for americans 2nd edition.

Park, J., Kim, M., El Mistiri, M., Kha, R., Banerjee, S., Gotzian, L., Chevance, G., Rivera, D.E., Klasnja, P., Hekler, E., et al. (2023). Advancing Understanding of Just-in-Time States for Supporting Physical Activity (Project JustWalk JITAI): Protocol for a System ID Study of Just-in-Time Adaptive Interventions. *JMIR Research Protocols*, 12(1), e52161.

Saint-Maurice, P.F., Troiano, R.P., Bassett, David R., J., Graubard, B.I., Carlson, S.A., Shiroma, E.J., Fulton, J.E., and Matthews, C.E. (2020). Association of Daily Step Count and Step Intensity With Mortality Among US Adults. *JAMA*, 323(12), 1151–1160.

Schoeppe, S., Alley, S., Van Lippevelde, W., Bray, N.A., Williams, S.L., Duncan, M.J., and Vandelanotte, C. (2016). Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 13(1), 1–26.