CHAPMAN UNIVERSITY

Chapman University Digital Commons

Psychology Faculty Articles and Research | Psychology

7-31-2018

A Markov Approach for Increasing Precision in the Assessment of Data-Intensive Behavioral Interventions

Vincent Berardi
Chapman University, berardi@chapman.edu

Ricardo Carretero-González
San Diego State University

John Belletierre
San Diego State University

Marc A. Adams
Arizona State University

Suzanne C. Hughes
San Diego State University

See next page for additional authors

Follow this and additional works at: https://digitalcommons.chapman.edu/psychology_articles

 Part of the Behavior and Behavior Mechanisms Commons, Community Health Commons, Health Information Technology Commons, Health Services Research Commons, Other Psychiatry and Psychology Commons, and the Substance Abuse and Addiction Commons

Recommended Citation

Berardi, V., Carretero-González, R., Bellettiere, J., Adams, M.A., Hughes, S., & Hovell, M. (2018). A Markov approach for increasing precision in the assessment of data-intensive behavioral interventions. *Journal of Biomedical Informatics*. doi: https://doi.org/10.1016/j.jbi.2018.07.023

This Article is brought to you for free and open access by the Psychology at Chapman University Digital Commons. It has been accepted for inclusion in Psychology Faculty Articles and Research by an authorized administrator of Chapman University Digital Commons. For more information, please contact laughtin@chapman.edu.

A Markov Approach for Increasing Precision in the Assessment of Data-Intensive Behavioral Interventions

Comments

NOTICE: this is the author's version of a work that was accepted for publication in *Journal of Biomedical Informatics*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version will be subsequently published in *Journal of Biomedical Informatics* in 2018. DOI: 10.1016/j.jbi.2018.07.023

The Creative Commons license below applies only to this version of the article.

Creative Commons License

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

Copyright

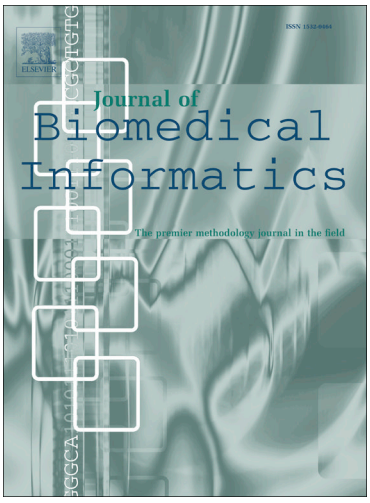
Elsevier

Authors

Vincent Berardi, Ricardo Carretero-González, John Belletierre, Marc A. Adams, Suzanne C. Hughes, and Melbourne Hovell

This article is available at Chapman University Digital Commons: https://digitalcommons.chapman.edu/psychology_articles/123

Accepted Manuscript



A Markov Approach for Increasing Precision in the Assessment of Data-Intensive Behavioral Interventions

Vincent Berardi, Ricardo Carretero-González, John Bellettiere, Marc A. Adams, Suzanne Hughes, Melbourne Hovell

PII: S1532-0464(18)30153-9
DOI: https://doi.org/10.1016/j.jbi.2018.07.023
Reference: YJBIN 3028

To appear in: *Journal of Biomedical Informatics*

Received Date: 1 March 2018
Revised Date: 20 June 2018
Accepted Date: 30 July 2018

Please cite this article as: Berardi, V., Carretero-González, R., Bellettiere, J., Adams, M.A., Hughes, S., Hovell, M., A Markov Approach for Increasing Precision in the Assessment of Data-Intensive Behavioral Interventions, *Journal of Biomedical Informatics* (2018), doi: https://doi.org/10.1016/j.jbi.2018.07.023

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Markov Methodology for More Precise Assessments of Data-Intensive Behavioral Interventions

Vincent Berardi[a,*], Ricardo Carretero-González[b], John Bellettiere[c,d], Marc A. Adams[e], Suzanne Hughes[c], Melbourne Hovell[c]

[a] *Department of Psychology, Chapman University, Orange, CA, USA*
[b] *Department of Mathematics and Statistics, San Diego State University, San Diego, CA, USA*
[c] *Center for Behavioral Epidemiology and Community Health, San Diego State University, San Diego, CA, USA*
[d] *Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA, USA*
[e] *College of Health Solutions, Arizona State University, Phoenix, AZ, USA*

Abstract

Health interventions using real-time sensing technology are characterized by intensive longitudinal data, which has the potential to enable nuanced evaluations of individuals' responses to treatment. Existing analytic tools were not developed to capitalize on this opportunity as they typically focus on first-order findings such as changes in the level and/or slope of outcome variables over different intervention phases. This paper introduces an exploratory, Markov-based empirical transition method that offers a more comprehensive assessment of behavioral responses when intensive longitudinal data are available. The procedure projects a univariate time-series into discrete states and empirically determines the probability of transitioning from one state to another. State transition probabilities are summarized separately in phase-specific transition matrices. Comparing transition matrices illuminates intricate, quantifiable differences in behavior between intervention phases. Statistical significance is estimated via bootstrapping techniques. This paper introduces the methodology via three

*Corresponding Author
Email addresses: berardi@chapman.edu (Vincent Berardi), rcarretero@sdsu.edu (Ricardo Carretero-González), jbellettiere@ucsd.edu (John Bellettiere), marc.adams@asu.edu (Marc A. Adams), shughes@cbeachsdsu.org (Suzanne Hughes), mhovell@cbeachsdsu.org (Melbourne Hovell)

ACCEPTED MANUSCRIPT

case studies from a secondhand smoke reduction trial utilizing real-time air particle sensors. Analysis enabled the identification of complex phenomena such as avoidance and escape behavior in response to punitive contingencies for tobacco use. Additionally, the largest changes in behavior dynamics were associated with the introduction of behavioral feedback. The Markov approach's ability to elucidate subtle behavioral details has not typically been feasible with standard methodologies, mainly due to historical limitations associated with infrequent repeated measures. These results suggest that the evaluation of intervention effects in data-intensive single-case designs can be enhanced, providing rich information that can ultimately be used to develop interventions uniquely tailored to specific individuals.

Keywords: behavioral interventions, longitudinal data, mobile health, e-health, Markov analysis, secondhand smoke

1. Introduction

Interventions aimed at changing behavior are often implemented on an individual level in studies known as single case designs(SCDs) [1]. Within an SCD, a treatment approach is typically evaluated over time by observing an individual within two or more distinct phases (intervals of time), both with and without an active treatment in place. The basic framework includes a baseline phase (A), comprised of several repeated observations of the dependent variable without an active treatment. This baseline phase is usually followed by an intervention phase (B), defined by a discrete point in time where the independent variable begins to be experimentally manipulated for the duration of the phase, concurrent with continued observations of the dependent variable. The frequency of observations within SCDs makes them an attractive option for researchers aiming to observe the precise nuances of how people interact with treatment protocols [2], specifically those that incorporate highly-individualized shaping procedures.

Mobile sensing instruments such as fitness trackers, wearable glucose mon-

ACCEPTED MANUSCRIPT

itors, and devices within the Internet of Things are beginning to enable behavioral responses and the physiological/environmental contexts in which they occur to be assessed continually, in near real time [3]. The longitudinal data produced by this technology offers the possibility of deploying behavior interventions within SCDs that are characterized by an unprecedented amount of data. Analytic approaches capable of navigating such data-rich longitudinal studies have not yet been fully developed. For example, in the popular text *Models for Intensive Longitudinal Data* [4], nearly all data sets were generated by ecological momentary assessments (EMAs), diaries, or a reviews of historical data. Generally, the most intense observation frequencies were associated with EMAs, which generated data points episodically a few times per day. Similar data intensity is noted in the text *Intensive Longitudinal Methods* [5], where the measurement frequency ranged in duration from 10 through 29 observations. The data sets used within these texts are representative of the current status quo, as demonstrated by a meta-analysis of 409 SCDs, which reported that the average number of observations in the baseline phase was 10.22 [6]. In contrast, real-time sensors deployed in studies routinely assess data at frequencies on the order of several times per minute over the course of weeks or months. This increased data volume should enable the effects of interventions on behavioral dynamics to be detailed at a much finer resolution, provided that analytic tools are developed for this purpose.

In addition to increasing the precision of behavioral assessments, technological advancement has the potential to fundamentally change the nature of interventions away from static procedures towards just-in-time adaptive behavioral interventions (JITAIs) [7]. JITAIs enable treatments to be provided on an ongoing basis and to automatically adapt in response to participants' varying behaviors, environmental contexts, and past history. In contrast to the historic paradigm, where a small number of study phases demarcate time intervals of interest, JITAIs do not have well-defined intervention on/off time intervals and instead are hypothesized as an ongoing interaction between patients and providers. In static interventions, the delineation between intervention on/off

ACCEPTED MANUSCRIPT

phases has resulted in analytic evaluations that are necessarily focused on identifying global differences between study phases, often by quantifying changes in level, trend, variability, overlap, and/or immediacy of effect [8, 9, 10, 11, 12]. These analyses were not developed to elucidate the subtleties of behavioral responses to continuously-adapting interventions, making them insufficient for evaluating JITAIs.

To fully realize mobile-sensing technology's potential to increase the resolution of outcomes, it is necessary to develop analytic techniques that i.) capture the nuances of individual responses to treatment and ii.) are capable of assessing ongoing interventions that are frequently encountered by participants throughout the course of a trial. This manuscript describes the development of a Markov-based, transition matrix methodology that has the potential to meet these challenges. For a given case, this approach evaluates the intervention by comparing each observation, throughout the entire course of the trial, to a second observation located within close temporal proximity. The timescales considered are on the order of seconds, allowing detailed profiles of individual responses to the intervention to be created. Additionally, because the frequency of assessment is much higher than in typical behavioral interventions, there are sufficient observations of the dependent variable to evaluate continual exposure to intervention stimuli. The proposed analytic approach is non-parametric and exploratory, characteristics that have been suggested to be valuable for revealing behavioral dynamics [13].

The use of the techniques developed herein requires a very high sampling frequency, meaning it is not appropriate for most current studies. However, intense sampling frequencies are becoming increasingly more common as real-time sensing technology becomes more ubiquitous in multiple contexts [14]. Health promotion interventions characterized by streaming technology and intensive measurement frequency have already begun to be implemented [15, 16, 17]. It is these types of studies, likely representing an increasing proportion of behavioral interventions in the future, that the procedures described in this manuscript were developed to analyze.

ACCEPTED MANUSCRIPT

2. Background

2.1. Project Fresh Air

The methodologies in this paper are generalizable to any study with a sufficient number of observations. Project Fresh Air (PFA), a secondhand smoke (SHS) reduction trial characterized by streaming data and intensive data measurements, was used as a prototype throughout this manuscript (see Ref. [17] for the full details of this study). This trial aimed to ameliorate SHS exposure by leveraging punitive contingencies, which are defined as aversive stimuli contingent on a behavior(s) that makes the behavior(s) less likely to be emitted in the future. Approximately 300 homes were enrolled in this trial, each of which contained at least one adult who generated SHS (typically via indoor cigarette smoking) and at least one child under 14 years old living in the home. To monitor indoor air quality, two Dylos DC1700 air particle quality monitors were installed inside of each home. The monitors were calibrated to detect particles with sizes ranging from 0.5 to 2.5 microns, which is consistent with SHS as well as non-tobacco aerosol sources [18]. One monitor was installed in the room nearest to where most smoking took place and another was placed in the child's bedroom; measurements from only the main smoking room monitor are included in the current analysis. The monitors measured the air particle concentration every ten seconds. In approximately half of the homes, the air particle monitors were fitted with devices that were programmed to deliver aversive visual and auditory feedback (yellow/red lights and tones) when air particle concentrations exceeded 60 $\frac{\mu g}{m^3}$, which previous research indicated was consistent with indoor cigarette smoking [18]. The intensity of the aversive feedback increased [19] if a second 120 $\frac{\mu g}{m^3}$ threshold was breached. For these homes, the trial was stratified into two phases: 1.) *Baseline* – a period during which feedback was not active and 2.) *Treatment* – the period during which the feedback was activated, representing an AB logic. Previously, linear mixed-effects analyses demonstrated that the intervention, on average, significantly reduced particle-related and tobacco-related outcomes between the Baseline and Treatment phases [17], but did

ACCEPTED MANUSCRIPT

not separately examine precise outcomes for individual homes. The aim of the analyses presented below is to investigate the individualized effect of the first $\left(60\frac{\mu g}{m^3}\right)$ threshold on a small subset of homes from PFA.

3. Methods

3.1. Markov Chains and Transition Matrices

The methodology underlying this analysis is based on Markov chains. Markov chains (or processes) are discrete systems that, at any given time, can be characterized as being in a particular state, where the states are mutually exclusive and exhaustive. They are also memoryless, meaning that the probability distribution of the system's next state is determined entirely by the current state. If there are n states, the transition matrix T is an n-by-n structure summarizing these probabilities over all states. $T_{i,j}$, the element of matrix T corresponding to the i^{th} row and j^{th} column, is the probability that a system in State i at a given time will be in State j at the next time step. For instance, $T_{1,1}$ is the probability that a system in State 1 remains in State 1 at the next time. $T_{1,2}$ is the probability that a system in State 1 moves to State 2, and so on. One row for each state is constructed and, by the law of total probability, the sum of each row must be 1. As an example, consider the matrix

$$T = \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.25 & 0.25 \end{pmatrix}.$$

In this case, when in State 1 at time t_i, the system will move to State 2 at time t_{i+1} with probability 1. When in State 2 at time t_i, the system has a 0.5 probability of moving to State 1 and a 0.5 probability of moving to State 3 at time t_{i+1}. At time t_i, if the system is in State 3, then at time t_{i+1}, the system will move to State 1 with probability 0.5, to State 2 with probability 0.25, and remain in State 3 with probability 0.25.

Markov models are extremely dexterous and have been widely applied to longitudinal data for many purposes including the transition of a system among

ACCEPTED MANUSCRIPT

several states [20], the clustering of multivariable time series [21], and to assess pre/post treatment results [22]. The versatility of Markov modeling allows it to be used for the present task of assessing an ongoing behavioral intervention.

3.2. Empirical Transition Matrices

The approach detailed herein proceeds by empirically calculating transition matrices for each phase (A and B) of the intervention. In terms of PFA, for each home one transition matrix was calculated for both the Baseline phase and Treatment as follows. Each data point in the air particle time series was assigned to one of eight states corresponding to the air particle concentration ranges summarized in Table 1. While these states do not necessarily correspond to distinct system characteristics, as is often the case with Markov modeling, this procedure is useful for characterizing system behavior at various levels of particle generation. States were denoted as S_j for $j = 1 \dots 6$; S_4 represents a particle concentration that has breached the 60 $\frac{\mu g}{m^3}$ threshold, thereby activating the aversive feedback. The selection of the boundaries that delineate the states must be specifically determined for each study. Section 3.4 details metrics to aid with this selection and to evaluate the effect that state boundary selection has on overall results.

To populate the transition matrix, a *lag* l was selected, which defines the time interval at which state transitions are evaluated. s_i was defined as the state of the system at time i and s_{i+l} was defined as the state of the system l units later at time $i + l$. (The last $l - 1$ observations were eliminated from this analysis since the system's state l units later was not observed.) We call the s_i's *source* states and the s_{i+l}'s *destination* states. A schematic of this process is depicted in Fig. 1 for a single point where $s_i = S_2$ and $s_{i+l} = S_3$. For each S_j, consider all of the s_i's such that $s_i = S_j$, i.e., all of the observations in S_j. We then determined s_{i+l}, the destination state, for each of these observations. The raw counts were divided by $|S_j|$, the total number of observations contained in state j, to convert them into probabilities describing the transition from each state into every other state. This information was summarized in an empirical

ACCEPTED MANUSCRIPT

State	**Conc. Range** $\left(\frac{\mu g}{m^3}\right)$	**# Observations**		
		Home 1	*Home 2*	*Home 3*
$< S_1$	< 30	715,806	405,519	809,962
S_1	30 – 40	6,881	10,837	2,644
S_2	40 – 50	1,790	3,488	589
S_3	50 – 60	685	1,509	224
S_4^*	60 – 70	396	868	158
S_5	70 – 80	486	544	112
S_6	80 – 90	277	276	173
$> S_6$	> 90	654	1,478	1,399

Table 1: Range of particle concentrations for each of the states used in the Markov analysis along with the number of observations in each state for each of the three homes under consideration (Home 1, Home 2, Home 3). S_4 has an asterisk since it represents the first state where a particle concentration has breached the first threshold and activated aversive feedback.

transition matrix, which described the probability of moving from one state to another after a lag l has elapsed.

T_B was defined as the empirically-determined transition matrix for the Baseline phase of the intervention when the visual and audio feedback was not yet activated and T_T was defined as the empirically-determined transition matrix for the Treatment phase once the feedback had been activated. If present, intervention effects should manifest themselves as differences between these two matrices, which can be summarized by $T_\Delta \equiv T_T - T_B$. T_B, T_T, and T_Δ were calculated for Home 1, Home 2, and Home 3, three households in the PFA study. As will be demonstrated in Section 4.1.1, these homes were chosen since they exemplify different analytic results of interest. In each case, a lag of $l = 6$ measurements, or one minute, was used. The rationale for this choice of lag will is discussed in Section 3.4.

ACCEPTED MANUSCRIPT

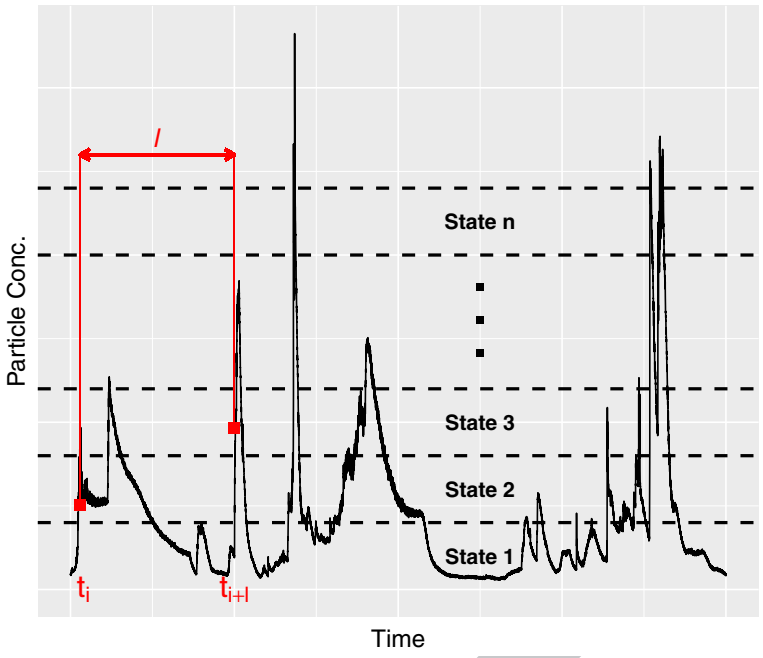


Figure 1: Schematic of the mechanism used to populate Markov process transition matrices based on an air particle time series. Time is shown on the x-axis and particle concentration is shown on the y-axis. At time t_i, the system is in state S_2. After a lag of l time units, denoted by the red double arrow, the system is in S_3. This process is performed for every observation so that the probability of moving from every state to every other state can be summarized.

3.3. Significance of the Differences between T_B and T_T

It is important to determine whether differences summarized in T_Δ matrices were statistically significant and, therefore, likely reflective of distinctive dynamics in different study phases. Bootstrapped confidence intervals for each element of T_Δ were calculated to make this assessment. This procedure took advantage of the fact that each row of the transition matrix defines a multinomial distribution with n categories (each of the n destination states). The probability of the system moving into a given category (typically called a successful trial in a multinomial distribution) was given by the empirically-calculated, discrete distribution summarized in the row. Statistical software can easily generate any number of values at random from a multinomial distribution. For the multinomial distribution defined by the row associated with a given S_j, $|S_j|$ values were

ACCEPTED MANUSCRIPT

randomly generated. These random values were then grouped according to the destination state into which they fell. Dividing the number of elements in each group by $|S_j|$ turns the randomly-generated values into a randomly-generated row of a transition matrix. This procedure was repeated $B \equiv 10,000$ times for both the Baseline and Treatment phase matrices and, for each run, the difference between the distributions was calculated, resulting in B T_Δ-like difference matrices. Each element of the T_Δ matrices was considered separately and the B values were sorted from lowest to highest. The $0.025B^{\text{th}}$ and $0.975B^{\text{th}}$ value define a 95% confidence interval. If the resulting confidence interval did not contain zero, then the change associated with this element in T_Δ was considered significant.

3.4. Selection of Analytic Parameters

Several components of the Markov transition matrix are free parameters that must be chosen. These include the use of {30,40,...,90} as the boundaries defining the states, S_j, and the use of a lag, l, of one minute between measurements. The selection of these parameters can be informed by theoretical and/or empirical criteria. In Appendix Appendix B, methodologies from both of these perspectives that aided in choosing parameters are detailed. The effect of parameter selection on overall results is also investigated. It should be noted that the stratification point between the Baseline and Treatment matrices can also be considered a parameter. Given the strong rationale for selecting the onset of the intervention as this boundary, we discuss this choice as a validity check.

3.5. Validity Investigation

The methodologies outlined above reflect the stratification of measurements based on whether they are in the Baseline or Treatment phase of the study followed by an analysis of the difference between transition matrices associated with each phase. As a validity check, this section details analyses that explore whether this delineation choice leads to larger effects than delineating the data

ACCEPTED MANUSCRIPT

by some other criterion. If the largest changes are indeed associated with stratifying by intervention phase, this will increase the evidence for interpreting the results as a definitive demonstration of the effectiveness of the PFA intervention.

3.5.1. Convergence of the Markov System

Before investigating validity, it is necessary to first determine the number of measurements required for the system to converge to 'mean' dynamics. Our strategy is to consider increasingly larger subsets of the data and determine how many observations are required for results to be consistent. For a given home, this process begins by defining the first time point at which the intervention has been activated, which is designated as the ${t_I}^{\text{th}}$ observation. Those t_i values with $i < I$ are in the Baseline phase and those t_i values with $i \geq I$ are in the Treatment phase. The empirically-calculated probabilities described in Section 3.2 will be calculated using all of the measurements in the Baseline and Treatment phases, but the convergence process seeks to identify a subset of these observations that yields similar dynamics.

The first subset of Baseline phase observations considered consisted of the 10% of all Baseline observations that were temporally closest to t_I. Similarly, the 10% of the Treatment phase observations that were temporally closest to the onset of the intervention were also selected. This process is illustrated in the bottom row of Fig. 2. The process described in Section 3.2 was then repeated for this subset of data by forming the empirical transition matrices T_B and T_T and calculating the difference matrix T_Δ^{10}, where the superscript indicates that 10% of the Baseline/Intervention data was used. To determine the concordance of T_Δ^{10} with T_Δ^{100}, i.e., the results when using all observations, the Frobenius norm $||T_\Delta^{100} - T_\Delta^{10}||_F$ was calculated. The Frobenius norm ($||.||_F$) is the L^2 norm of a vectorized version of the matrix. When performing this calculation, all non-significant T_Δ values were set to zero. To determine the number of observations required for the system to converge, the process described above was repeated when considering 20%, 30%,...100% of the observations from each phase's time series that were closest to the ${t_I}^{\text{th}}$ observation.

ACCEPTED MANUSCRIPT

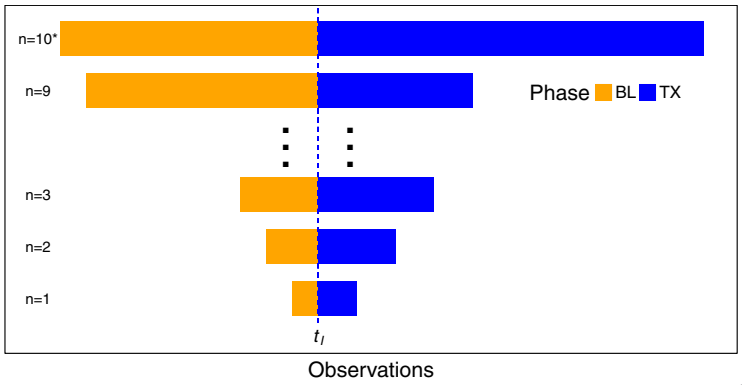


Figure 2: Schematic of the procedure used to evaluate the number of measurements for the Markov analysis to converge. The dashed line t_I indicates the onset of the intervention characterized by the availability of aversive monitor feedback. For $n = 1$, T_B was formed using only a small subset of the data adjacent and just prior to t_I. This is denoted by the small orange rectangle on the lowest row. The small blue rectangle on the lowest row illustrates the small subset of data just after t_I that was used to construct T_T. For $n = 2$, the amount of data used to populate the transition matrices was expanded by 10% of the total number of measurements in each phase. This process continued until $n = 10$, when all of the data in each phase of the intervention were used to construct the transition matrices, which is the case that is illustrated in Fig. 4.

3.5.2. Optimal Boundary between T_B and T_T

Once the number of measurements required for the dynamics to converge has been established, a validity check that considers the effect of alternate boundaries between the T_B and T_T matrices can proceed. Define m_B and m_T, respectively, as the number of measurements required for convergence in the Baseline and Treatment phases. An iterative procedure was implemented where the number of observations used to form the matrices T_B and T_T was held constant at m_B and m_T over all iterations, while the boundary defining the two matrices varied. For the first iteration, the transition matrix T_B was generated from the 1^{st} through the m_B^{th} observations in the study and T_T was generated from the $(m_B + 1)^{\text{th}}$ through the $(m_B + m_T)^{\text{th}}$ observations. This process is illustrated in the bottom row of Fig. 3. The associated T_Δ matrix summarizing the difference between these two transition matrices was then calculated. In a sliding

ACCEPTED MANUSCRIPT

window-type procedure, the observation indices were then shifted by some value δ such that a new T_B and T_T were defined. (See the second from bottom row of Fig. 3.) T_Δ was again calculated for these two matrices. This procedure was repeated (observations were shifted by δ) as long as there were a sufficient number of observations to accommodate shifting the data window. δ was chosen such that after five shifts, the boundary between T_B and T_T exactly matched t_I, the boundary between the Baseline and Treatment phases.

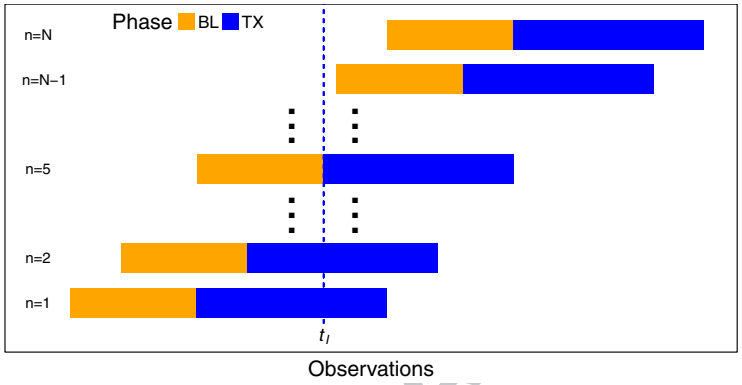


Figure 3: Schematic of the procedure used to evaluate optimal boundary between the Baseline and Treatment transition matrices. For the first iteration, the 1st through $(m_B + m_T)^{\text{th}}$ observations were used, with a boundary at m_B. For each subsequent iteration, the window of observations considered is shifted by δ, which is chosen so that at the fifth iteration, the boundary between the two phases aligns with t_I. For each iteration, transition matrices were empirically calculated and T_Δ, the difference between these matrices, was also determined. n is the iteration number.

4. Results

4.1. Empirical Transition Matrices

Figure 4 illustrates the empirical transition matrices calculated for Homes 1 through 3. The first and second column of panels depict T_B and T_T, respectively, while the third column of panels depicts T_Δ. Only significant changes (as determined via the methodology described in Section 3.3) with an effect size (i.e. difference between matrix probabilities) greater in absolute value than 0.05

ACCEPTED MANUSCRIPT

are denoted in the T_Δ matrices; these probabilities are outlined in blue. Generally speaking, the transition matrices have the largest probabilities along the diagonal, meaning that the source state s_i and the destination state s_{i+l} are the same. In other words, one minute after any given measurement, the air particle concentration is most likely to have not changed by an amount large enough for it to have transitioned into different state. For Home 1 and Home 2, the intervention had the effect of increasing the probabilities in the subdiagonal entries, which represent, on average, a decrease in particle concentrations after l time units have elapsed. These increased subdiagonal values came at the expense of decreasing probabilities on the diagonal and superdiagonal. This result can be seen by examining the transition matrices T_B and T_T themselves or by observing the sign and location of significant differences highlighted in T_Δ. For these homes, the intervention had the effect of creating downward pressure on states, where, once a state had been reached, air particle concentrations were more likely to decrease in the Treatment phase compared to the Baseline phase. In Home 3, there was no systematic pattern and only four of the source-destination pairs in T_Δ were significant.

4.1.1. Behavioral Interpretation of Results

From the standpoint of behavior science, detailed conclusions about the dynamics of a household can be extracted from the transition matrices. The treatment included lights and tones emanating from the monitor once air particle concentrations exceeded a threshold. This behavior-stimulus association is an example of an aversive/punishing contingency. By definition, an aversive/punishing contingency occurs when the presentation of a stimulus made contingent on a specific behavior results in a reduction of the behavior that led to the generation of the stimulus. This reduction can occur in two ways, *escape behavior* or *avoidance behavior*. In the escape paradigm, an individual performs behaviors to immediately alleviate the aversive stimulus associated with its action. For example, in PFA, once the aversive alarm has sounded, an individual may respond by extinguishing a cigarette or by moving outside in order reduce

ACCEPTED MANUSCRIPT

the air particle concentration triggering the sound. Avoidance behavior, on the other hand, is when an individual has discriminated the condition(s) that led to the aversive stimulus and avoids the behavior or environment altogether. In PFA, after several instances of being exposed to aversive monitor feedback as a result of particle-generating behavior, an individual may move outdoors before or just after lighting a cigarette so that particle concentrations do not trigger the monitor's feedback.

The T_Δ matrices appear to indicate that Home 1 exhibited escape behavior while Home 2 exhibited avoidance behavior. For Home 1, the reduction in the value of the diagonal probabilities and associated increase in subdiagonal probabilities only occurred for states S_4 through S_6, precisely those states that triggered the monitor feedback. That is, household members in Home 1 appear to be seeking relief from the aversive stimuli, once it has been activated. In Home 2, though, the effect was present along diagonal and subdiagonal entries for all states, including those prior to the activation of feedback. This can be interpreted inferentially as the household adjusting their behavior in order to avoid triggering the alarm rather than reacting to the alarm once it has become engaged.

4.2. Validity Investigation

The results of the convergence analysis described in Section 3.5.1 are shown in Fig. 5 for each of the three homes under consideration. In these figures, convergence is represented by the asymptoting of $||T_\Delta^{100} - T_\Delta^n||_F$ values; the dip for $n = 100$ is expected as there is an exact match between matrices. For Home 1, the results converged for all $n \geq 70$, which corresponds to using 125,864 measurements in the Baseline phase and 382,635 measurements in the Treatment Phase. For Homes 2 and 3, the system begins to converge at $n = 40$ and $n = 60$, respectively. For consistency and in an effort to be conservative, we consider each of these systems to have converged when using 70% of the data. Table 2 summarizes the number of measurements required to converge for each home.

For the three homes under consideration, Fig. 6 illustrates $||T_\Delta||_F$ for the

ACCEPTED MANUSCRIPT

Variable	*Definition*	*Home 1*	*Home 2*	*Home 3*
N	# observations total	726,428	424,519	815,260
t_I	Baseline/Treatment Phase Boundary	179,807	251,462	330,650
m_B	# obs. for Baseline convergence	125,864	176,022	231,453
m_T	# obs. for Treatment convergence	382,635	121,141	339,228
δ	Shift size for validity check	13,486	18,860	24,799

Table 2: Variable definitions for validity analysis and values for representative homes. See text for details.

various stratification points between the two transition matrices, as detailed in Section 3.5.2. Larger norms indicate a greater difference between T_B and T_T and, therefore, a greater difference between the dynamics summarized by the matrices. t_I, the boundary between the Baseline and Treatment phases, is denoted as iteration 0 and all other windows are reported in terms of their iteration offset from t_I. For Home 2, the maximum of the norm is exactly at t_I, while for Home 1, the maximum occurs when the breakpoint is slightly offset from t_I. For Home 3, the pattern was slightly different and the norm was relatively low until it spiked when the breakpoint was offset by two iterations past t_I. In these cases, the norms are larger for stratification points that, generally speaking, most closely align with the breakpoint between the Baseline and Treatment phases. This indicates that stratifying by the intervention phase led to larger difference in dynamics than alternate boundaries, demonstrating the influence of the PFA intervention in affecting household dynamics and adding validity to the behavioral findings.

While the Frobenius norm provides a measure of the total action of a matrix, it yields no information about the structure of the values within a matrix. It is possible for two matrices to have the same norm and completely different structures. (As a trivial example, the identity matrix and a matrix of the same

ACCEPTED MANUSCRIPT

dimensions with 1's on the anti-diagonal and 0's elsewhere have the same norm, but, at most, one element in common.) For the above sliding-window analysis, it only makes sense to compare the norms of transition matrices across different boundaries if the structure of the underlying matrices are similar. The T_Δ matrices that have been discussed thus far have been characterized by negative values on the diagonal, positive values on the sub-diagonal, and negative values on the superdiagonal. Therefore, to assess the structure of the matrices calculated during the iterative, sliding-window procedure, the mean subdiagonal, diagonal, and superdiagonal values were recorded for each of the boundary iterations. These results are shown in the bottom panels of Fig. 6. For Home 2, the characteristic signs for the 3 interior diagonals are present across all boundaries. For Home 1, the expected structure of a negative diagonal and positive superdiagonal does not emerge until the boundary between the two matrices aligns with t_I. This is additional evidence that the intervention was critical in changing the dynamics of the home in a manner that is face valid. The results for Home 3 do not follow the same pattern. Its largest norm occurs at an offset of 2; however the expected pattern does not emerge until offsets of 3 and 4 from t_I. This is likely a function of the ineffectiveness of the intervention for this home. Overall, this analysis provides more validity to the conclusions concerning the importance of the PFA intervention in generating behavioral responses.

5. Discussion

This paper describes an exploratory Markov procedure that empirically evaluated the individualized effects of repeated exposure to an intervention with a level of detail not possible with standard methodologies. While in the traditional paradigm differences in the level, slope, or variation of a dependent variable are often used to arrive at a binary determination of whether an intervention was efficacious or not, this methodology offers the ability to identify more nuanced, informative effects. In other words, whereas traditional methodologies focus on "if" an intervention affects behavior, the empirical Markov methodology pro-

ACCEPTED MANUSCRIPT

vides a more detailed assessment of the nature of the behavior responses to an intervention. Our example resulted in outcome measures that can be inferred to correspond to avoidance and escape behaviors, a characterization that likely would have been missed by other methodologies. It is probable that different behavioral phenomena will be observed when applying the empirical Markov procedure to other studies. For all cases, though, the precise level of detail offered by this approach provides an opportunity to assess interventions in a manner that is much more consistent with established behavioral theory. This could ultimately lead to more effective, tailored behavioral interventions.

For each case considered herein, the results required over 100,000 observations per phase to converge. In general, this methodology requires a large volume of repeated measures for a single individual with the actual number being dependent on the variance of the observed data. Such intensive longitudinal data have not been the norm in behavioral interventions thus far. This trend is changing and soon more studies will have the requisite data measurement intensity [7, 23]. Wearable devices such as the activity trackers, smart watches, and a myriad of similar technologies enable a large number of physiological variables to be assessed continuously in near real-time. Big data from smart homes, networked cars, and the Internet of Things, which uploads data from sensors on physical devices (thermostats, washers/dryers, etc.) to networks, allow for the measurement of additional behaviors and even the context in which they occur [3]. Real-time data generated by such devices enables a more comprehensive assessment of individuals than has ever been possible. The current trend of quantified self tracking, where individuals record certain aspects of their daily life with great precision, will only add to this ability [24]. As this technology becomes more ubiquitous, there will be a greater number of opportunities to provide the type of personalized, data-intensive health interventions that are amenable to methods such as this empirical Markov approach [25]. This process has already begun to take hold in studies concerning, for example, physical activity [26], dietary intake [27], cigarette smoking cessation [28], and drug abuse [29]. It has been suggested that current analytic approaches are

ACCEPTED MANUSCRIPT

not compatible with the intensive data streams generated by these studies and, therefore, new methodologies are needed [30]. The empirical Markov model described herein has the potential to help fill this methodological gap.

A preponderance of data-centric studies in conjunction with analytic methodologies that are capable of elucidating detailed accounts of behavior (such as the empirical Markov methodology) might have the effect of increasing the prevalence of studies implementing SCDs as opposed to traditional between-subject designs. In such studies, an intervention effect may be sufficiently strong as to leave little doubt as to the efficacy of a treatment, especially if replicated in multiple individuals. This development would have the effect of moving the evaluation of treatment away from the descriptions of "average" change in behavior, which may conceal important functional relations, towards inductive assessments of individual level outcomes [2]. This development is likely to lead to more easily interpretable results that can better inform treatment decisions. Furthermore, analytic results from our methodology can also be used to inform the design and implementation of clinical trials aiming to gain a clearer picture of the variance in patient responses to treatment.

Limitations in this approach are now outlined. Our methodology does not allow for the inclusion of time-variant predictor and/or mediator variables. As with all intensive longitudinal studies, procedures for the management of missing/corrupt data must be developed. Furthermore, this analysis was intended for demonstration purposes and was only performed for three homes. Therefore, the generalizability of the results to other homes has not yet been demonstrated, but will be explored in future work. Applying the empirical Markov methodology to different types of behavior should also be addressed in the future. With a large number of SCDs, it might be possible to assess high-level predictor/mediator variables. A key step in this process is the development of succinct metric to summarize the T_Δ matrices as opposed to the ad-hoc descriptions of behavior that were used here. Possibilities under consideration are a principal components decomposition of the matrix as well as pattern recognition approaches. Such a metric can be used to aggregate results in a way that would

ACCEPTED MANUSCRIPT

allow the overall intervention efficacy to be assessed with greater precision at the individual level of behavior.

6. Software and Data Availability

In accordance with the Peer Reviewers' Openness Initiative, software scripts used for implementing our analysis, along with the necessary data, has been made publicly available. An `R` package called `MarkovSCD` was written explicitly to implement our methodology and is hosted on GitHub. An example script demonstrating how to load the package and use it to produce the figures and tables within this manuscript is provided in Appendix A. All data included within the `R` package is in a de-identified format that complies with the Institutional Review Board that oversaw the trial.

7. Declaration of Interests

Declarations of interest: none

8. Funding Sources

Research reported in this publication was supported by NHLBI of the National Institutes of Health under award number R01HL103684. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. John Bellettiere was funded in part by training grants provided by the National Institutes of Health (T32HL079891-11 and TL1TR001443).

[1] A. E. Kazdin, Single-case research designs: Methods for clinical and applied settings, Oxford University Press, 2011.

[2] M. P. Normand, Less is more: Psychologists can learn more by studying fewer people, Frontiers in psychology 7 (2016) 934.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

[3] M. Chui, M. Löffler, R. Roberts, The internet of things, McKinsey Quarterly 2 (2010) (2010) 1–9.

[4] T. A. Walls, J. L. Schafer, Models for intensive longitudinal data, Oxford University Press, 2006.

[5] N. Bolger, J. Laurenceau, Intensive longitudinal methods: an introduction to diary and experience sampling research, Guilford, 2013.

[6] J. D. Smith, Single-case experimental designs: A systematic review of published research and current standards., Psychological Methods 17 (4) (2012) 510.

[7] I. Nahum-Shani, E. B. Hekler, D. Spruijt-Metz, Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework., Health Psychology 34 (S) (2015) 1209.

[8] N. A. Gage, T. J. Lewis, Analysis of effect for single-case design research, Journal of Applied Sport Psychology 25 (1) (2013) 46–60.

[9] J. Cohen, A power primer., Psychological bulletin 112 (1) (1992) 155.

[10] W. R. Shadish, E. N. Kyse, D. M. Rindskopf, Analyzing data from single-case designs using multilevel models: New applications and some agenda items for future research., Psychological methods 18 (3) (2013) 385.

[11] R. M. de Vries, R. D. Morey, Bayesian hypothesis testing for single-subject designs., Psychological methods 18 (2) (2013) 165.

[12] M. R. Nourbakhsh, K. J. Ottenbacher, The statistical analysis of single-subject data: a comparative examination, Physical therapy 74 (8) (1994) 768–776.

[13] E. Ferrer, Exploratory approaches for studying social interactions, dynamics, and multivariate processes in psychological science, Multivariate behavioral research 51 (2-3) (2016) 240–256.

ACCEPTED MANUSCRIPT

[14] W. T. Riley, K. J. Serrano, W. Nilsen, A. A. Atienza, Mobile and wireless technologies in health behavior and the potential for intensively adaptive interventions, Current opinion in psychology 5 (2015) 67–71.

[15] S. C. Mukhopadhyay, Wearable sensors for human activity monitoring: A review, IEEE sensors journal 15 (3) (2015) 1321–1330.

[16] M. A. Adams, J. C. Hurley, M. Todd, N. Bhuiyan, C. L. Jarrett, W. J. Tucker, K. E. Hollingshead, S. S. Angadi, Adaptive goal setting and financial incentives: a 2× 2 factorial randomized controlled trial to increase adults' physical activity, BMC public health 17 (1) (2017) 286.

[17] S. C. Hughes, J. Bellettiere, B. Nguyen, S. Liles, N. E. Klepeis, P. J. Quintana, V. Berardi, S. Obayashi, S. Bradley, C. R. Hofstetter, et al., Randomized trial to reduce air particle levels in homes of smokers and children, American journal of preventive medicine 54 (3) 359–367.

[18] N. E. Klepeis, S. C. Hughes, R. D. Edwards, T. Allen, M. Johnson, Z. Chowdhury, K. R. Smith, M. Boman-Davis, J. Bellettiere, M. F. Hovell, Promoting smoke-free homes: a novel behavioral intervention using real-time audio-visual feedback on airborne particle levels, PloS one 8 (8) (2013) e73251.

[19] J. Bellettiere, S. C. Hughes, S. Liles, M. Boman-Davis, N. Klepeis, E. Blumberg, J. Mills, V. Berardi, S. Obayashi, T. T. Allen, et al., Developing and selecting auditory warnings for a real-time behavioral intervention, American journal of public health research 2 (6) (2014) 232.

[20] C. A. Stifter, M. Rovine, Modeling dyadic processes using hidden markov models: A time series approach to mother–infant interactions during infant immunization, Infant and child development 24 (3) (2015) 298–321.

[21] S. Ghassempour, F. Girosi, A. Maeder, Clustering multivariate time series using hidden markov models, International journal of environmental research and public health 11 (3) (2014) 2741–2763.

ACCEPTED MANUSCRIPT

[22] E. Holmes, M. Bonsall, S. Hales, H. Mitchell, F. Renner, S. Blackwell, P. Watson, G. Goodwin, M. Di Simplicio, Applications of time-series analysis to mood fluctuations in bipolar disorder to promote treatment innovation: a case series, Translational psychiatry 6 (1) (2016) e720.

[23] D. Spruijt-Metz, E. Hekler, N. Saranummi, S. Intille, I. Korhonen, W. Nilsen, D. E. Rivera, B. Spring, S. Michie, D. A. Asch, et al., Building new computational models to support health behavior change and maintenance: new opportunities in behavioral research, Translational behavioral medicine 5 (3) (2015) 335–346.

[24] M. Swan, The quantified self: Fundamental disruption in big data science and biological discovery, Big Data 1 (2) (2013) 85–99.

[25] M. J. Khoury, J. P. Ioannidis, Big data meets public health, Science 346 (6213) (2014) 1054–1055.

[26] M. A. Adams, J. F. Sallis, G. J. Norman, M. F. Hovell, E. B. Hekler, E. Perata, An adaptive physical activity intervention for overweight adults: a randomized controlled trial, PloS one 8 (12).

[27] G. Shroff, A. Smailagic, D. P. Siewiorek, Wearable context-aware food recognition for calorie monitoring, in: 2008 12th IEEE International Symposium on Wearable Computers, IEEE, 2008, pp. 119–120.

[28] A. Rodgers, T. Corbett, D. Bramley, T. Riddell, M. Wills, R.-B. Lin, M. Jones, Do u smoke after txt? results of a randomised trial of smoking cessation using mobile phone text messaging, Tobacco control 14 (4) (2005) 255–261.

[29] S. F. Butler, S. H. Budman, A. Licari, T. A. Cassidy, K. Lioy, J. Dickinson, J. S. Brownstein, J. C. Benneyan, T. C. Green, N. Katz, National addictions vigilance intervention and prevention program (navippro^{TM}): a real-time, product-specific, public health surveillance system for monitor-

ACCEPTED MANUSCRIPT

ing prescription drug abuse, Pharmacoepidemiology and drug safety 17 (12) (2008) 1142–1154.

[30] W. T. Riley, D. E. Rivera, A. A. Atienza, W. Nilsen, S. M. Allison, R. Mermelstein, Health behavior models in the age of mobile interventions: are our theories up to the task?, Translational behavioral medicine 1 (1) (2011) 53–71.

[31] E.-J. Wagenmakers, P. C. Molenaar, R. P. Grasman, P. A. Hartelman, H. L. van der Maas, Transformation invariant stochastic catastrophe theory, Physica D: Nonlinear Phenomena 211 (3) (2005) 263–276.

Appendix A. Example Implementation of R Package

Prior to running this script, please check the `MarkovSCD` ***package documentation for updates.***

```
#Install package from Github
library(devtools)
install_github("vancebee/MarkovSCD")
library(MarkovSCD)


#Load Baseline and Treatment Phase data for one home
BL = HM2$MassAve[HM2$Phase == "BL"]
TX = HM2$MassAve[HM2$Phase == "TX"]


#Define state boundaries
sb = seq(30,90,10)


#Calculate empirical transition matrices
A = transmat(tseries = BL,statebounds = sb,lag = 6)
B = transmat(tseries = TX,statebounds = sb,lag = 6)
```

ACCEPTED MANUSCRIPT

```
#Mirror left two columns, center row of Fig. 2
A$prob; B$prob

#Calculate delta matrix
dd = deltatrans(A,B)
#Mirror right column,center row of Fig. 2
dd$prettydelta

#Calcualte the mean first passage times for Home 2 BL
#shown in Table 2
mm = mfpt(A$prob,4)

#Calculate level change function
data = HM2$MassAve[HM2$MassAve>30 & HM2$MassAve<90]
lc = levelcross(tseries = data,npts = 20, lag=1)
#Plot 10sec level-change function from right panel of  Fig. 3
plot(lc$evalpts,lc$lvlcrs,type = "o")

#Evaluate different lags
le = lageval(tseries = TX,statebounds = sb,+
  lagrange = c(1,2,seq(3,60,3)))
#Plot State 3 time series in center panel of  Fig. 4
plot(le$lagrange,le$diagbylag[[3]],type = "o")

#Prepare range of state boundaries for sensitivity analysis
sbrng = list()
w = c(5,10,20)
for(ii in 1:length(w)){
sbrng[[ii]] = seq(30,90,w[ii])
}
#Perform sensitivity analysis
```

ACCEPTED MANUSCRIPT

```
ss = sensitivity(tseries1 = BL,tseries2 = TX,stbdyrange = sbrng,+
  lagrange = c(3,6, 30, 60, 180))
#Reproduce matrices in Fig. 6
ss$deltamats

#Determine when dynamics have converged
cv = dynamicsconv(tseries1 = BL, tseries2 = TX, nitvl = 10,+
  statebounds = sb ,lag = 6)
#Reproduce center panel of Fig. 8
plot(cv$normdiff)

#Define convergence windows to search for optimal boundary between matrices
il1 = cv$ilength1[7]
il2 = cv$ilength2[7]
#Perform the vailidity check
vv = validitycheck(tseries1 = BL, tseries2 = TX, ilength1 = il1,+
  ilength2 = il2, statebounds = sb ,lag = 6)
#Reproduce center column of top row of Fig.10
plot(vv$norm,type="o")
#Reproduce center column of bottom row of Fig.10
matplot(1:ncol(vv$diagconfig),t(vv$diagconfig),type = "l")
```

Appendix B. Selection of Analytic Parameters

Appendix B.1. State Boundaries

There are two components that must be taken into account when deciding on state boundaries, 1.) the range of values to be considered and 2.) the discretization of this range. For PFA, theoretical considerations can be used when selecting the first component. The focus of the trial is particle generating behavior which, by definition, result in elevated air particle concentrations. Our expe-

ACCEPTED MANUSCRIPT

rience with the study data indicate that low-level measurements $<\sim 30\frac{\mu g}{m^3}$ are likely associated with background particle concentrations rather than particle-generating events. Therefore, they can be safely ignored by setting $30\frac{\mu g}{m^3}$ as the minimum of values to be considered. Our analysis focuses on the lower, $60\frac{\mu g}{m^3}$ aversive feedback threshold. To avoid confounding from the activation and/or anticipation of the upper, $120\frac{\mu g}{m^3}$ threshold, the maximum value considered was selected as the midpoint between the lower and upper feedback, $90\frac{\mu g}{m^3}$.

Once the range of values to be encompassed by the states was selected, a theoretical rationale was not available to determine how to stratify the states within this range; therefore, a graphical tool was used. An empirical level-crossing function, $\hat{\ell}(x)$, can be used to estimate a longitudinal time series' invariant function, and therefore the presence and number of stable equilibria, which are associated with distinct dynamics [31]. $\hat{\ell}(x)$ is defined as the proportion of all observations where consecutive measurements cross x. It was calculated by considering equally-spaced x values between 30 and 90 and counting the number of instances in which consecutive measurements crossed these values, and then dividing by the total number of measurements. This process was repeated while counting whether lagged measurements separated by 1, 5, and 10 min crossed x. The results are illustrated in Fig. B.7 for Homes 1 and 2. When interpreting a level-crossing function, the focus is on identifying multiple modalities, which is indicative of regions with distinct dynamics. When present, boundaries should be selected to avoid mixing these regions. This was not the case for the PFA data, where $\hat{\ell}(x)$ was monotonically decreasing for lags of 10 sec and 1 min and was unimodal for lags of 5 min and 10 min. The mode for the larger lags likely reflects a greater probability of air particle levels returning to baseline after some time has passed. In either case, the PFA level-change function does not offer insight into the specific delineation of states. However, the results indicate that choosing parsimonious state boundaries that are evenly-spacing between 30 and $90\frac{\mu g}{m^3}$ does not present undue complications from mixing dynamical regions.

ACCEPTED MANUSCRIPT

Appendix B.2. Evaluation of Lag

The values along the diagonal of the transition matrices provide a metric by which to gauge the optimal value for the lag l. For smaller values of l, the probabilities on the diagonal are expected to be larger. Taken to the extreme, if l was chosen to be one measurement (10 sec for PFA), the system will be highly autocorrelated since there has likely been an insufficient amount of time for air particle concentrations to change; therefore, changes to a new state will be rare. As a result, transitions between states could be drowned out by a large number of source-destination observations remaining in the same class. With larger values of l, the effect of the autocorrelation will decrease and for sufficiently large values there will be an essentially random relationship between the concentration at t_i and t_{i+l}. These two time points would be separated by such a long period as to not capture the same particle-generating dynamics (i.e. causal independent variables). The relationship would instead reflect an overall distribution of the different states. The choice of l must strike a balance between the extremes of the maximally autocorrelated and the overall distribution scenarios.

Figure B.8 illustrates the values of each diagonal element of T_T for lags ranging from 1 to 60 measurements (10 sec to 10 min). T_T was used for this calculation since this phase was associated with the greatest number of observations. The smallest variations occurred for S_1 and S_2, since these states were most likely to be associated with background measurements that have little variance. In each home, for S_3 through S_6, beginning at a lag of approximately 30 sec, the diagonal probability value was relatively small (< 0.5 in most cases) compared with the probabilities for the $l = 10$ sec case (> 0.9 in most cases). This indicates that the degree of autocorrelation between source and destination observations was reduced to the point that there were substantial numbers of transitions among states, possibly enabling patterns to emerge. Lags less than $l \approx 5$ min for Home 1 and $l \approx 3$ min for Homes 2 and 3 were was also higher than the horizontal asymptote, which represents the mean, global distribution of probabilities associated with large lags. This is indicative of a meaningful correlation between source and destination states and the capturing of more

ACCEPTED MANUSCRIPT

than just the overall distribution of the states. The results outlined in Fig. B.8 indicate that a lag between $l \approx 0.5$ and $l \approx 5$ min will likely allow meaningful household dynamics to be observed. We concluded that $l = 1$ min best balanced the competing interests of the asymptotic and pure autocorrelation scenarios, so it was selected for all subsequent analyses. Other researchers reviewing Fig. B.8, though, could plausibly justify an alternative selection of l. The effects that differing choices for l have on outcome analyses was explored in the sensitivity analysis discussed below.

Appendix B.3. Sensitivity Analysis

The previous two sections outlined ambiguity in selecting the state boundary delineation and lag for the transition matrix methodology. The level-change analysis did not yield an instinctive choice for state boundaries, although it did indicate that equispaced states were non-problematic. The diagonal lag analysis provided evidence to select a lag of $l = 1$ min, but alternative interpretations of the optimal lag were reasonable. To address this uncertainty in parameter selection, a sensitivity analysis was conducted to investigate the effects of different combinations of state boundaries and lag. The results of this analysis are summarized in Figs. B.9 and B.10 for Homes 1 and 2, respectively. Lags of 0.5, 1, 5, 10, and 40 min were considered and are illustrated as the rows (from top to bottom) in these figures. The columns represent three different stratifications of particle ranges to serve as state boundaries. From left to right, they are: {30,50,70,90}, {30,40,50,60,70,80,90}, and {30,35,40,45,50,55,60,65,70,75,80,85,90}. Below, these cases are referred to by specifying the width of each state, i.e., $\nu = 20$, 10, and 5, respectively.

For each lag-stratification combination, Figs. B.9 and B.10 illustrate the T_Δ matrix resulting from the analysis described in the previous section. The matrix in the second column of the second row corresponds to a lag of 1 min and $\nu = 10$; these are the parameters used throughout the manuscript. For Home 1, the aversive behavior pattern, characterized by significant negative (red) values on the diagonal and significant positive (green) values on the subdiagonal only

ACCEPTED MANUSCRIPT

for states greater than 60 $\frac{\mu g}{m^3}$, is present for many lag-state boundary combinations. The same is true of the avoidance behavior pattern in Home 2. For both homes, the use of a greater number of states ($\nu = 5$) reveals the same patterns. However, the findings are somewhat obfuscated as the inclusion of additional states results in smaller transition values and more transitions being considered non-significant. When using only three states ($\nu = 20$), the results are too blunt and avoidance and escape behaviors within the two homes cannot be differentiated.

In Figs. B.9 and B.10, the first three rows correspond with reasonable lags, as described in the previous section. The results are qualitatively consistent, albeit with larger lags being associated with a greater number of significant transition cells and a shifting of cells with significant values towards destination states associated with attenuated concentrations. This is likely because the larger lags are beginning to approach global differences between the two phases, which will be seen most in lower concentrations, since they represent the largest proportion of observations. For $l = 10$ min the avoidance/escape patterns largely remain intact, with the shift mentioned above being more pronounced. For $l = 40$ min, as expected, the patterns are no longer visible.

The results of this sensitivity analysis indicate that the aversive/avoidance results outlined in the previous section are robust over reasonable choices for modeling parameters. But there is no guarantee that this will be true for other studies. While the parameterization described above is specific to the PFA study, the tools used within this section provide an outline for inferring parameter values. The diagonal lag methodology summarized in Fig. B.8 provided a reasonable basis for selecting the lag. While the optimal lag value was not able to be absolutely determined, the sensitivity analysis indicated that any reasonable interpretation of Fig. B.8 yields a lag that would lead to consistent results. When determining the state boundaries, the level crossing function can be used to assess critical features of the data, such as the presence of stable equilibria that need to be accounted for during stratification. Even if the multiple equilibria are not found, as was the case with PFA, the level crossing function provides

ACCEPTED MANUSCRIPT

a motivation for using parsimonious, equispaced states. The sensitivity analysis indicates that there should be a proclivity for using more, rather than less, states. This ensure that nuanced patterns can be observed. This inclination must be balanced by a need to retain the ability to easily interpret results.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

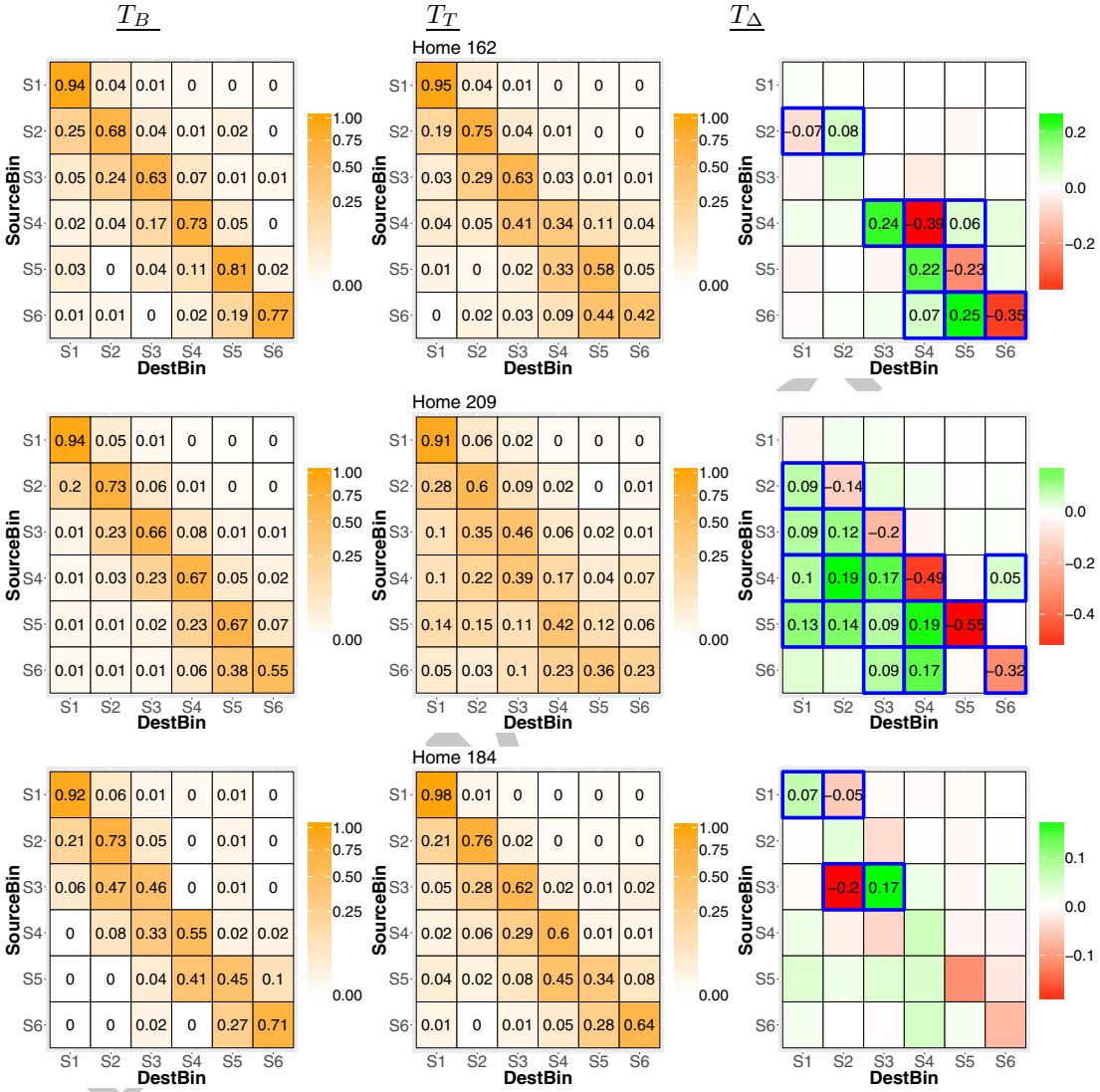


Figure 4: Transition matrices for three homes in PFA. The rows of each matrix represent the source states and the columns represent the destination states. The value of $T_{i,j}$ represents the empirically-calculated probability of the system being in State S_j when it was in S_i l units earlier. The left column of the figure contains baseline transition matrices T_B, the center column contains treatment transition matrices T_T, and the right column contains T_Δ, the difference between these two matrices. The rows of panels correspond, from top to bottom, to Home 1, Home 2, and Home 3. For T_Δ, only those values determined to be statistically significant with an effect greater than $|0.05|$ are shown. These cells are highlighted in blue.

ACCEPTED MANUSCRIPT

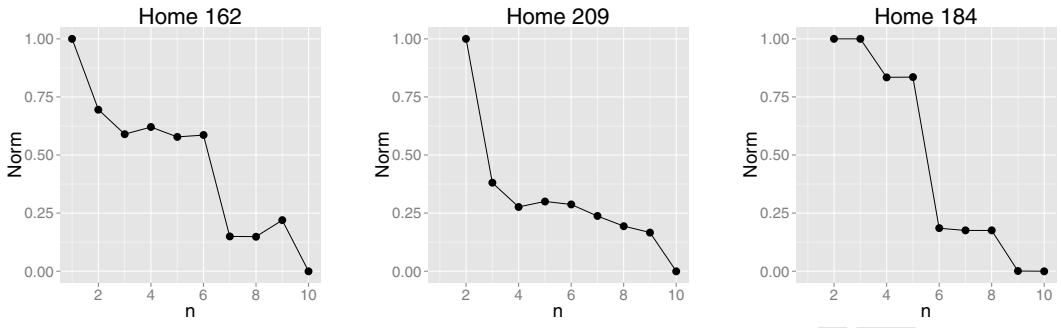


Figure 5: $||T^*_\Delta - T^n_\Delta||_F$ for $n = 1 \ldots 10$ for the three representative homes. In the two left-most cases, note the convergence to the results associated with the full amount of data as larger and larger subsets of data are used. Results were normalized by the largest value to make a visual comparison more tractable.

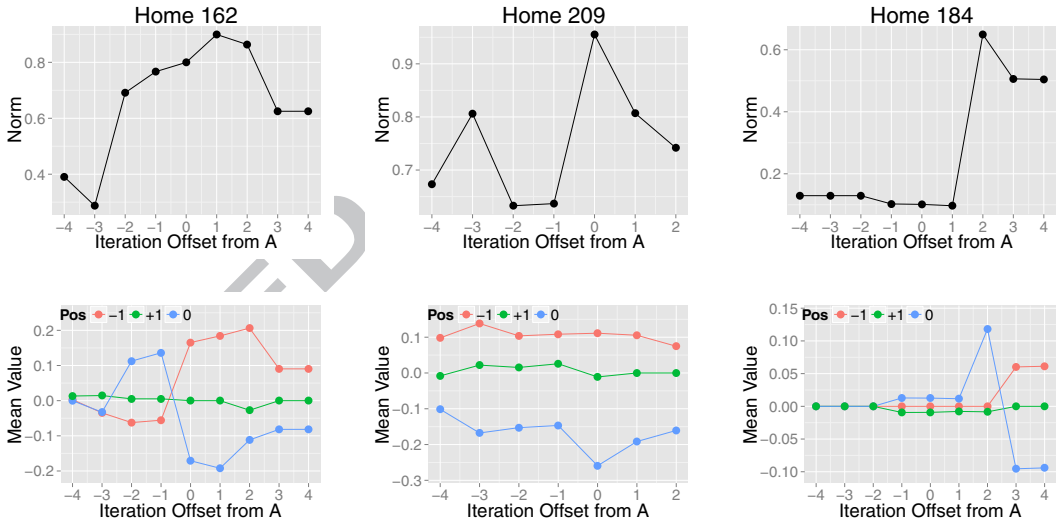


Figure 6: The top panel illustrates $||T_\Delta||_F$, which summarizes the difference between the two transition matrices that are being compared for various boundaries separating the data. The x-axis represents the number of shifts away from t_I, the boundary between intervention phases. The bottom panel illustrates the mean subdiagonal, diagonal, and superdiagonal values for each of the boundaries under consideration.

ACCEPTED MANUSCRIPT

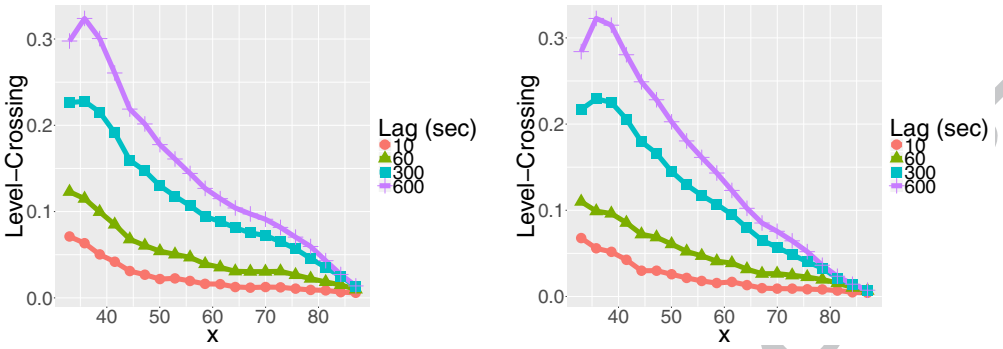


Figure B.7: Empirical level change function $\hat{\ell}(x)$ for various lags for Homes 1 and 2.

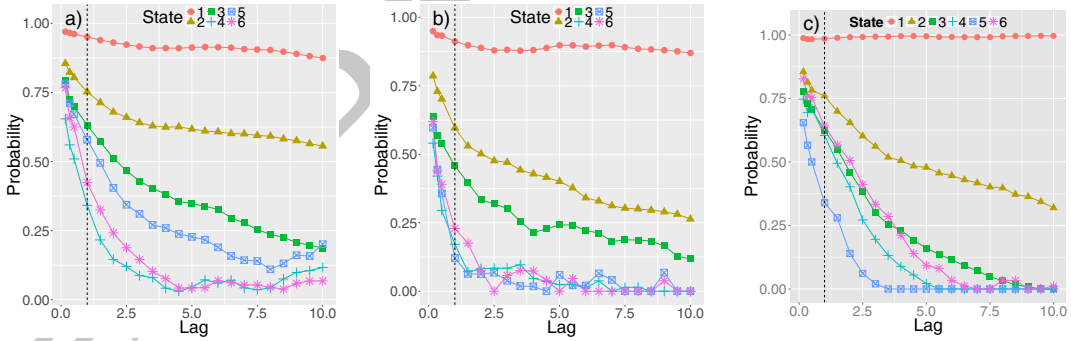


Figure B.8: The value of the diagonal element corresponding to each of the 6 states for values of l ranging from 10 sec–10 min for Homes 1, 2, and 3, from left to right. The vertical dashed line represents a lag of 1 min.

ACCEPTED MANUSCRIPT

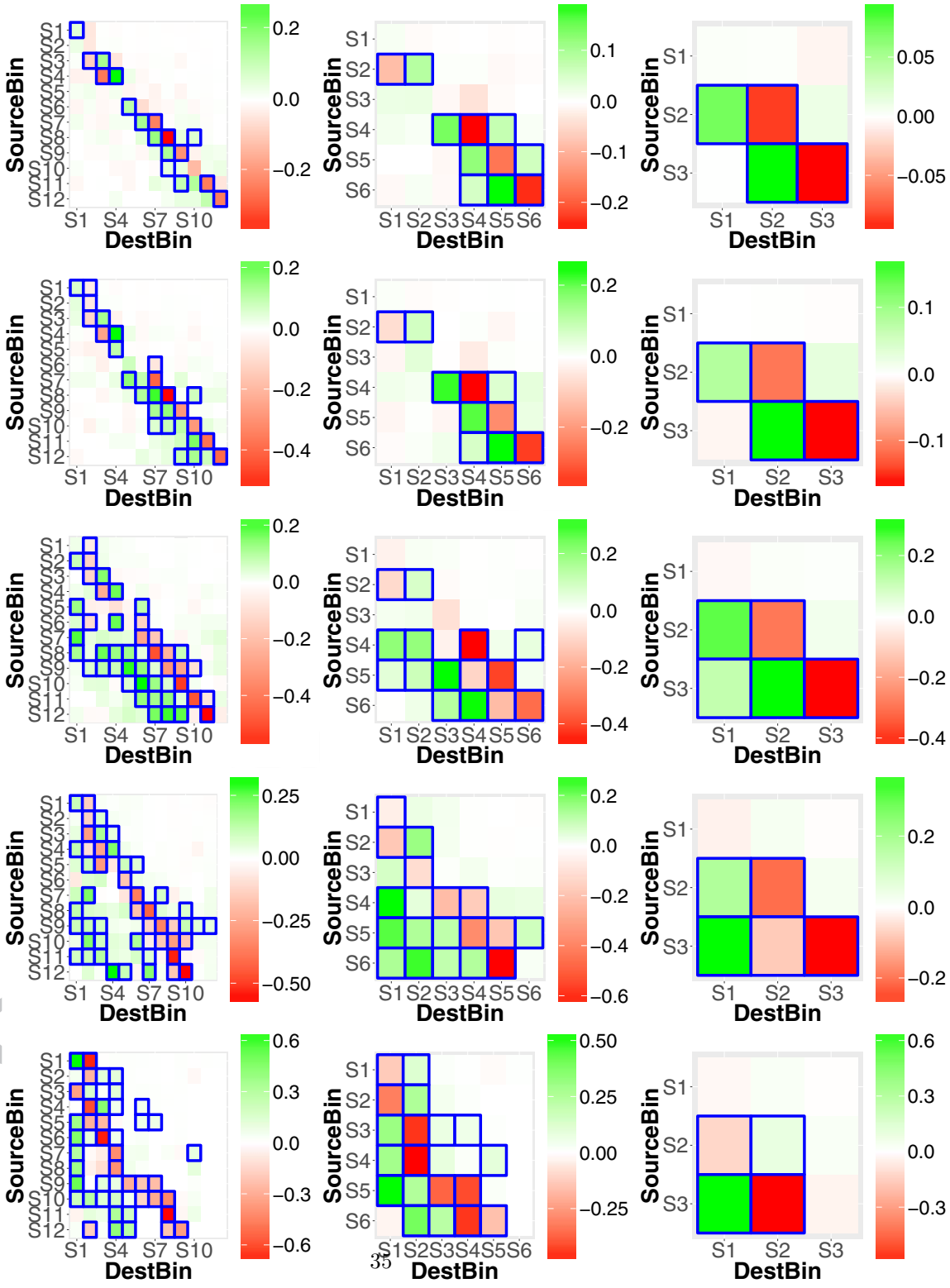


Figure B.9: T_Δ matrices for the sensitivity analysis performed on Home 1. The rows represent, from top to bottom, lags of 0.5 min, 1 min, 5 min, 10 min, and 40 min, respectively. The columns, from left to right, represent states with width ν of 20, 10, and 5.

ACCEPTED MANUSCRIPT

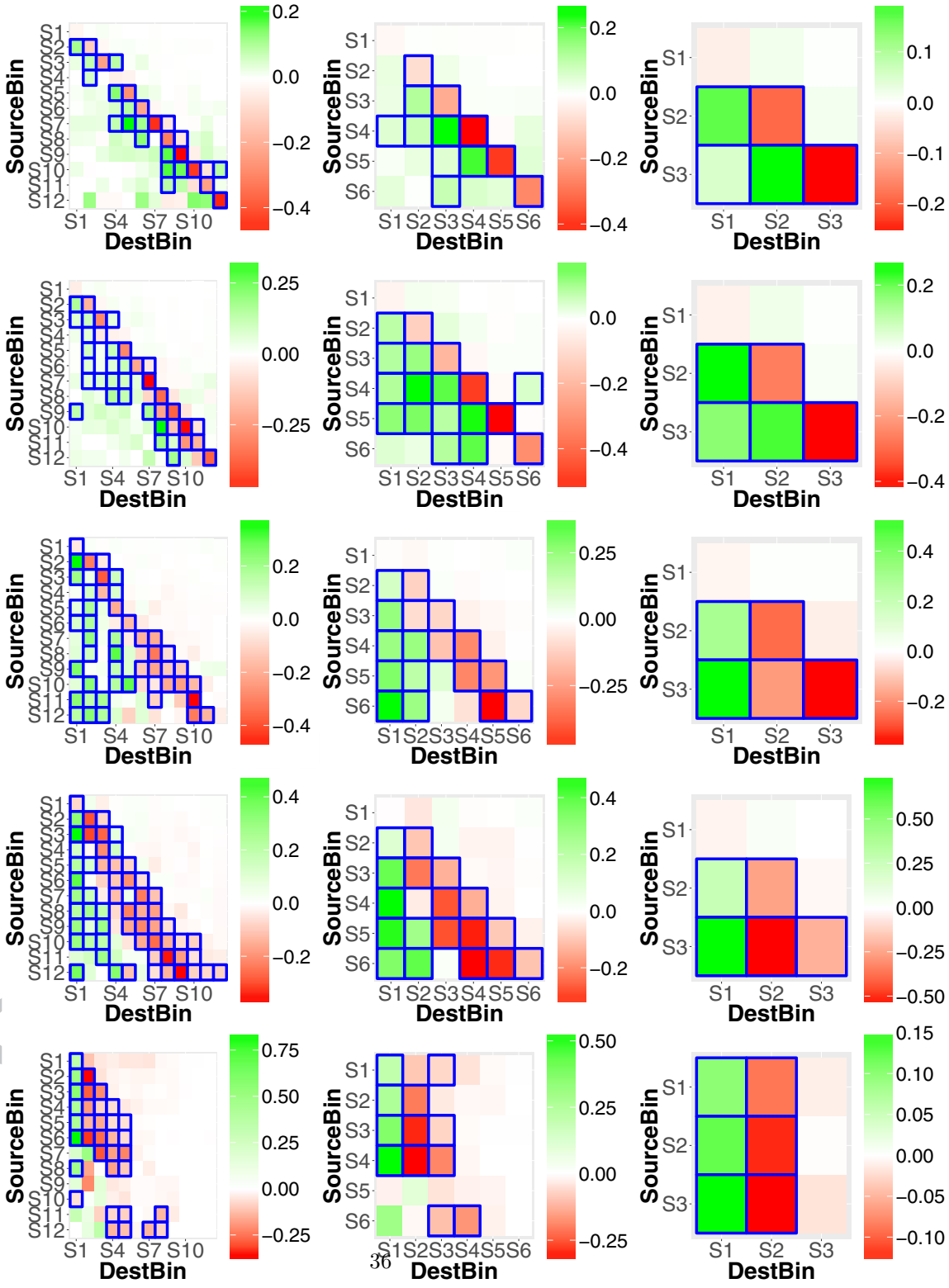


Figure B.10: Results of the sensitivity analysis performed on Home 2. See Fig. B.9 for an explanation.

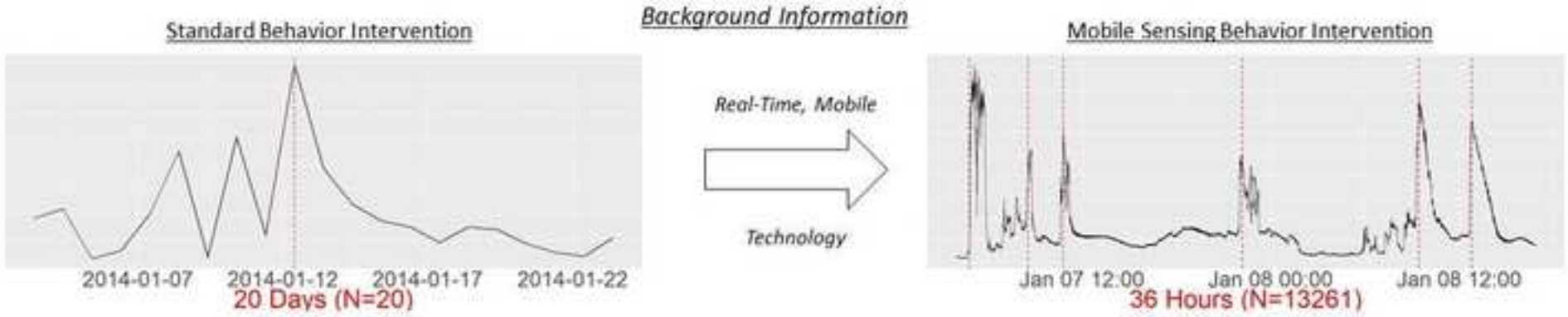
ACCEPTED MANUSCRIPT

- Intensive longitudinal data leads to more precise behavioral interventions
- Avoidance and escape behavior were generated in response to an aversive stimulus
- Providing real-time feedback led to the largest changes in behavior dynamics

ACCEPTED MANUSCRIPT

Graphical Abstract

ACCEPTED MANUSCRIPT



ACCEPTED