Taylor & Francis
Taylor & Francis Group

Journal of Quality Technology
A Quarterly Journal of Methods, Applications and Related Topics

ISSN: 0022-4065 (Print) 2575-6230 (Online) Journal homepage: https://www.tandfonline.com/loi/ujqt20

Monitoring worker fatigue using wearable devices: A case study to detect changes in gait parameters

Amir Baghdadi, Lora A. Cavuoto, Allison Jones-Farmer, Steven E. Rigdon, Ehsan T. Esfahani & Fadel M. Megahed

JOURNAL OF QUALITY TECHNOLOGY
https://doi.org/10.1080/00224065.2019.1640097

CASE REPORT

Monitoring worker fatigue using wearable devices: A case study to detect changes in gait parameters

Amir Baghdadi[a], Lora A. Cavuoto[b], Allison Jones-Farmer[c], Steven E. Rigdon[d], Ehsan T. Esfahani[a], and Fadel M. Megahed[c]

[a]Department of Mechanical & Aerospace Engineering, University at Buffalo, Buffalo, New York; [b]Department of Industrial & Systems Engineering, University at Buffalo, Buffalo, New York; [c]Farmer School of Business, Miami University, Oxford, Ohio; [d]College for Public Health and Social Justice, Saint Louis University, Saint Louis, Missouri

ABSTRACT
The goal of this case study is to answer four research questions related to fatigue through features derived from wearable sensors to measure patterns in steps: (1) How do important gait parameters change over time? (2) How do these sensor-based changes relate to the participant's subjective fatigue ratings over time? (3) Are there consistent patterns in performance across different individuals over time? and (4) Do these patterns vary systematically based on specific demographic characteristics? To answer these questions, we have combined multivariate changepoint methods with hierarchical time-series clustering and exploratory data analysis. The results improve our understanding of fatigue development.

KEYWORDS
clustering; internet of people; median filter; nonparametric multivariate inference; signal processing

1. Motivation

Many authors have noted how recent technological advancements have led to innovative statistical methods to help understand, monitor, and/or control complex data sets. For example, in 2018, the *Journal of Quality Technology* published three special issues: (a) "Statistical process control on big data streams," (b) "Quality engineering in advanced manufacturing," and (c) "Reliability and maintenance modeling with big data." In all, 19 articles were published in these special issues, and we applaud the focus on process improvement and reliability related to technological advancement. However, we want to emphasize how the advances in technology can affect our ability to improve the performance of human workers who are central to productivity and quality in many production systems.

Our motivation behind presenting this article/case study is to encourage the quality community to investigate relevant research problems that pertain to human operators. Specifically, we posit that the three research streams highlighted by the special issues can extend to "human systems." For example, let us consider the domain of *advanced manufacturing* (often referred to as *Industry 4.0*), which is defined by The (U.S.) President's Council of Advisors on Science and Technology (2011) as a subset of manufacturing tasks that is built upon the utilization of automation, computing, and sensing technologies. *Advanced manufacturing* is different from the computer-integrated manufacturing of the 1990s. Specifically, "the end goal of computer-integrated manufacturing was a workless manufacturing environment (i.e., lights-out manufacturing facilities); however, advanced manufacturing aims to integrate workers into the cyber–physical infrastructure to maximize the impact of their skills . . ." (Lu et al. 2017, 139). From that perspective, each operator can be seen as an integral component of modern production systems. We hope to demonstrate how statistical methods that have traditionally been applied to problems of process quality monitoring (e.g., filtering, change point methods, and CUSUMs) can be applied to monitor worker fatigue.

The quality community has traditionally ignored the impact of human operators on process/product quality in manufacturing environments. While it is outside the scope of this article to investigate potential reasons for this observation, it is due time to encompass human performance monitoring as an integral component of quality monitoring activities. Our view is supported by the recent review of Kolus, Wells, and

CONTACT Fadel M. Megahed fmegahed@miamioh.edu Farmer School of Business, Miami University, Oxford, OH 45056, USA.
Color versions of one or more figures in the article can be found online at www.tandfonline.com/ujqt

Neumann (2018), who identified 73 published empirical studies discussing the impact of human factors (HF) in manufacturing operations on system quality performance. The authors clearly stated that: "quality deficits were associated with undesirable human effects of workload like fatigue and injury-related risk factors. Forty-six percent of the studies reported on efforts to improve HF in the [operations systems] with effect sizes for quality improvements reaching up to 86%." Moreover, we would like to mention two examples that show why our community should be more amenable for monitoring human performance as a part of quality control. First, it is already a standard practice in [lean] Six Sigma to collect "human performance related data" as a part of the *Safety* and *Morale* key indicators in the SQDCM (safety, quality, delivery, cost, and morale) framework. Second, the research community has been indirectly measuring human performance in surgical quality performance monitoring for about two decades (see the review of Woodall, Fogel, and Steiner 2015 for more details). We believe that the aforementioned discussion should encourage and justify why our community needs to be more involved in human performance monitoring.

With recent developments leading to numerous sensors that can be used to capture human performance (e.g., wearables, motion capture devices, sensors embedded in work stations), we believe that the three research streams can be extended to human performance modeling applications. First, the use of sensors and the need to analyze the generated data using automated methods is a key feature of *advanced manufacturing*. Second, because the sensors often have multiple frequencies and generate different data types, their analysis fits the "big data streams" paradigm. Third, a degradation in a worker's performance can require an ergonomic intervention (i.e., maintenance). Thus, the ability to estimate when the performance will deteriorate and what type of intervention optimizes recovery can improve quality and productivity and reduce injury risk. This application is conceptualized using a "big data" predictive maintenance paradigm.

2. Problem description

In this case study, we consider data gathered from 15 subjects participating in an experimental study performing an occupational task. Data were gathered on each participant using wearable sensors over a 3-hour period for the purposes of understanding physical fatigue development over time. Therefore, the main goals of our analysis are to (a) detect the onset of fatigue during the task, (b) characterize the stages of fatigue development, and (c) identify how different participants fatigue over time.

From an occupational perspective, physical fatigue is an adverse condition that leads to "a lower level of strength, physical capacity and job performance" (Lu et al. 2017, 140). Antecedents include (a) high workloads, (b) unergonomic postures, (c) long shifts, and/or (d) poor physical environments (Yung 2016). Interestingly, the recent technological advancements have burdened highly skilled workers, increasing their workload and fatigue rates (Brocal and Sebastián 2015; Gust et al. 2017; Romero et al. 2016).

High fatigue rates are being reported in many U.S. industries. For example, 58 percent of U.S. advanced manufacturing workers have reported being fatigued during the past work week (Lu et al. 2017). Similar fatigue rates have also been reported among construction (Zhang et al. 2015) and distribution center workers (Schneider, Copsey, and Irastorza 2010). It is important to highlight that high workplace fatigue rates are not unique to U.S. workplaces because they have also been reported in Japan (Kajimoto 2008), Sweden (Evengård 2008), Canada (Yung 2016), and the European Union (Loriol 2017).

Physical fatigue results in numerous adverse outcomes. These include (a) increases in sickness absence (Janssen et al. 2003), (b) higher likelihood of being injured in an occupational accident (Swaen et al. 2003), (c) decreased physical and cognitive function (Bláfoss et al. 2019; Zhang et al. 2015), (d) reduced work capacity/performance (Barker and Nussbaum 2011; Macintosh, Svedahl, and Kim 2004), (e) diminished quality (Kolus, Wells, and Neumann 2018; MacLeod 1994; Yeow and Sen 2003) and (f) higher rates of health complaints (Sluiter et al. 2003). The consequences of worker fatigue are not limited to the worker alone. For example, Ricci et al. 2007 estimated that U.S. employers lose $136 billion annually in fatigue-related costs (i.e., diminished production and worker compensation).

Due to the damaging effects of physical fatigue on both the individual and firm, it is important to quickly detect the onset of fatigue in order to minimize its consequences. Currently, the measurement of physical fatigue can be divided into three approaches: (a) exact methods, which involve the application of blood sampling techniques to detect cellular and metabolic changes (Garde, Hansen, and Jensen 2003); (b) electromyography (EMG)-based methods, which capture the electrical activity produced by muscles through the use of skin recording electrodes; and (c) feature-based methods that attempt to detect/monitor fatigue symptoms including changes in posture control, walking patterns, and increased sway (Davidson, Madigan, and Nussbaum 2004), which are typically recorded via visual inspection of work tasks or through the use of self-reported questionnaires. Approaches (a) and (b) are unlikely to be used in practice because they are intrusive and will likely be resisted by individuals/unions. Additionally, the use of EMG is only suitable for stationary tasks, which makes it unfit for many occupational settings. Thus, the first two approaches are primarily limited to laboratory studies (Cavuoto and Megahed 2017).

Although there are several physiological and physical indicators of fatigue onset, we limit our analysis to detecting changes in gait (i.e., walking pattern) parameters. Fatigue is a multifaceted phenomenon, which heavily depends on task, a person's individual characteristics, medical history, and environmental conditions (Cavuoto and Megahed 2017). Fatigue indicators will differ according to these conditions; thus, the corresponding feature-based methods to measure fatigue must be tailored to specific tasks. It is important to develop a methodology that can potentially translate to a large subset of occupations; we have to identify a component that is common to many tasks. Walking (often to pick items) is a pervasive component of many occupational tasks in manufacturing, construction, mining, and nursing. For example, U.S. advanced manufacturing workers have reported performing tasks that involve an average of 5.7 hours of walking per shift (Lu et al. 2017). Similarly, distribution center workers typically spend 75 percent of their time picking orders and walk up to 6 miles per shift (Fiveash 2016). The reader is referred to Hernandes 2016 for similar statistics involving other occupations.

In the fields of industrial statistics, quality, and reliability, fatigue generally refers to the weakening of materials after prolonged use or load (King et al. 2016). In this study, we consider physical fatigue of humans in an occupational setting. In both material fatigue testing and human fatigue, the stress can accumulate to the point that a failure occurs; for the case of a material, this might mean that the item breaks and, in the case of a worker, it could mean that job performance is negatively affected. Human fatigue is also related to degradation testing of equipment, where items are placed on test and the important characteristic is monitored across time so that a failure can be predicted. For example, a lamp may be tested and the luminosity monitored across time. When the luminosity drops below a specified level, the unit is deemed to have failed (Tseng, Hamada, and Chiao 1995). Lu and Meeker (1993) developed methodology to monitor the crack length in materials in order to predict when a crack will lead to a failure. Degradation testing and human fatigue do, however, differ in a few aspects. First, the profile from a degradation test is almost always monotonic because luminosity does not get higher and cracks don't get shorter. Also, in degradation testing, the general form of the degradation profile is known; for example, cracks will grow in size until a failure but, for humans, it is not known *a priori* what type of profile will suggest fatigue. Third, there is usually a single output variable in degradation testing, whereas human fatigue can be assessed through many output variables that are measured using wearable sensors.

With the continued advancements in wearable technologies and the associated decrease in their costs, wearable sensors are widely considered to be an integral component of next-generation fatigue/gait monitoring systems (Dempsey et al. 2018; Schall, Sesek, and Cavuoto 2018; Tsao, Ma, and Papp 2018; Zhao and Obonyo 2018). The advantages for using wearables technology for feature-based prediction/monitoring methods include portability, noninvasiveness, affordability, multifunctionality (which makes them translatable between different applications and/or occupational tasks (Tongen and Wunderlich 2010; Yang and Li 2012), and reliability (Tao et al. 2012). These advantages make them more appealing than video-based motion capture systems (which are static) and questionnaires (which are less timely).

The existing literature on wearable sensors for fatigue detection has been limited to classification methodologies, where regression and/or machine learning approaches are used to predict whether a subject is fatigued or not (Baghdadi, Cavuoto, and Crassidis 2018; Baghdadi et al. 2018; Pantelopoulos and Bourbakis 2008, 2009, 2010; Maman et al. 2017, n.d.). In these approaches, the response variable is typically dichotomous, corresponding to fatigued and nonfatigued states based on a subjective user/participant rating. The predictor variables are typically features extracted from the wearable sensors capturing statistical, kinematic, and physiological changes. There are three main limitations for the methodologies currently published in the literature. First, temporal variation in the predictors are not explicitly accounted for in the existing models, which are meant to be applied at regular intervals (e.g., every 15 minutes) instead of accumulating information over time. Second, machine learning approaches are black-box and often provide no diagnostic information for why a person is deemed fatigued at a given time point. Finally, the published sensitivity and specificity rates for individual subjects are typically between 75 percent and 85 percent. When applying these classification methods to multiple subjects in practice, it is likely that the misclassification rates would be very high.

As a first step toward addressing these limitations, we examine how wearable sensor data can be better explored and analyzed. Specifically, we combine exploratory data analysis (EDA) techniques, multivariate nonparametric change point methodology, and time series clustering to address four research questions:

1. How do important gait parameters (e.g., stride length, height, and duration) change over time?
2. How do these sensor-based changes relate to the participants' subjective fatigue ratings?
3. Are there consistent patterns in performance across different individuals over time?
4. If so, do these patterns vary systematically based on specific demographic characteristics?

Addressing these questions can set the foundation for prospective fatigue monitoring.

The remainder of the article is structured as follows. Section 3 details the experimental setup and feature engineering (i.e., how stride length, height, and duration were generated from the sensors' data). In Section 4, we present how the data were preprocessed. Section 5 details the application of the nonparametric change point approach to detect the different stages of fatigue development, and address research questions (1) and (2). In Section 6, we discuss how a hierarchical time series clustering approach was used to help us address research questions (3) and (4). Finally, our concluding remarks are presented in Section 7.

3. Experimental setup and feature engineering

The goal of this article is to examine the use of EDA techniques, multivariate change point analysis techniques, and time series clustering to describe the different stages of fatigue development based on data collected from wearable sensors. This case study presents a secondary analysis of the experimental data of Baghdadi et al. (2018), where fatigue was classified using a machine learning approach (i.e., not accounting for the temporal nature of wearable sensors data). To improve the readability of our article, in this section, we present a brief description of (a) the experimental procedure, (b) sensor data collection and recording, and (c) the feature engineering, explaining how the gait cycles were extracted from the raw data and how the stride length, height, and duration were computed based on the segmented gait cycles.

3.1. Experimental procedure

The case presented in this manuscript is part of the broader study published by Maman et al. (2017) and further reported on in Baghdadi et al. (2018). The study was designed as a cross-sectional laboratory study using a one-factor within-subjects design. The designed factor was the physical level of the task at three levels (low, medium, and high) based on postural, biomechanical, and physiological demand. The low-level task included an assembly task completed in a standing position at a workstation, the medium-level task involved supply pickup and delivery with sustained back flexion at the delivery point, and the high-level task involved manual materials handling. These tasks represent the range of tasks performed regularly and repeatedly in complex manufacturing environments (Lu et al. 2017). Each task level was performed in a separate session and the session involved 3 hours of continuous work. Due to the nature of the tasks, only the high-level manual materials handling (MMH) task involved a significant period of continuous walking that would allow for analysis of changes in gait over the duration of the session. Similar to Baghdadi et al. (2018), we only used the data corresponding to the high-level MMH task in our case study. The 3-hour period was selected to represent a typical period of continuous manufacturing work.

A sample of 15 participants completed the study. The participants had an average (standard deviation): (a) age of 37.6 (16.7) years, (b) body mass of 74.2 (14.4) kilograms, and (c) height of 170.8 (9.2) centimeters. Eight participants were men and seven were women. These subjects included a combination of students with some experience of performing manual tasks and local workers, which is typical of many ergonomic studies (Evstigneeva et al. 2012; Rashedi et al. 2014; Tanoue et al. 2016). Participants were excluded based on the *Physical Activity Readiness Questionnaire* (Thomas, Reading, and Shephard 1992).

At the start of the experimental session, participants were provided with instructions for performing the task and had a period to become acquainted with the task and ask any questions as needed. This period also served as a warm-up for the physical demand of the task. Participants were instrumented with a small inertial measurement unit (IMU of size 51 mm × 34 mm × 14 mm) attached around their right ankle (see Figure 1). Note that an IMU is a device that encompasses three different sensors: (a) accelerometer, measuring a body's specific force, (b) gyroscope to measure the angular rate, and (c) magnetometer for measuring the magnetic field, which is useful for determining directions in a global field of reference (e.g., one can think of the role of a compass). The IMU, used in Baghdadi et al. (2018), recorded data continuously at 51.2 Hz (i.e., 51.2 measurements per sensor per second) throughout the task. Each sensor recorded three time series at 51.2 Hz, one for each axis of the Cartesian coordinate system.

Once instrumented, baseline data were collected with the participant in a stationary position. Once the MMH task commenced, the participants completed a set of deliveries of weighted cartons using a two-wheeled dolly. At 10-minute intervals during the session, participants provided a rating of their perceived exertion (RPE) (Borg 1982). A visual summary of the experimental procedure is shown in Figure 2 (see Baghdadi et al. 2018 for more details).

3.2. Feature engineering

In this subsection, we describe the approach needed to preprocess the accelerometer data to generate the required features. The challenge is to transform the three-dimensional accelerometer data from the reference of an individual body (which we denote with x, y, and z) to a global frame of reference (which we distinguish through the use of the uppercase X, Y, and Z). To achieve this transformation, we first estimate the angular orientation of the ankle/sensor. This is achieved by combining the angular orientation from the accelerometer and gyroscope through the application of a Kalman Filter (KF). Through a series of preprocessing steps, we translate the raw data into the stride length, height, and duration for each participant over time. An overview of the preprocessing steps involved in feature generation are depicted in Figure 3. While the major components of the IMU preprocessing procedure are the same across multiple studies (Baghdadi, Cavuoto, and Crassidis 2018; Baghdadi et al. 2018; Rebula et al. 2013), the sequence/segmentation approach can differ according to the location of the IMU sensor on the body, the task being monitored, and/or the goal of application (real-time versus retrospective monitoring). For this reason, the segmentation approach presented here is different than that used by Baghdadi et al. (2018).

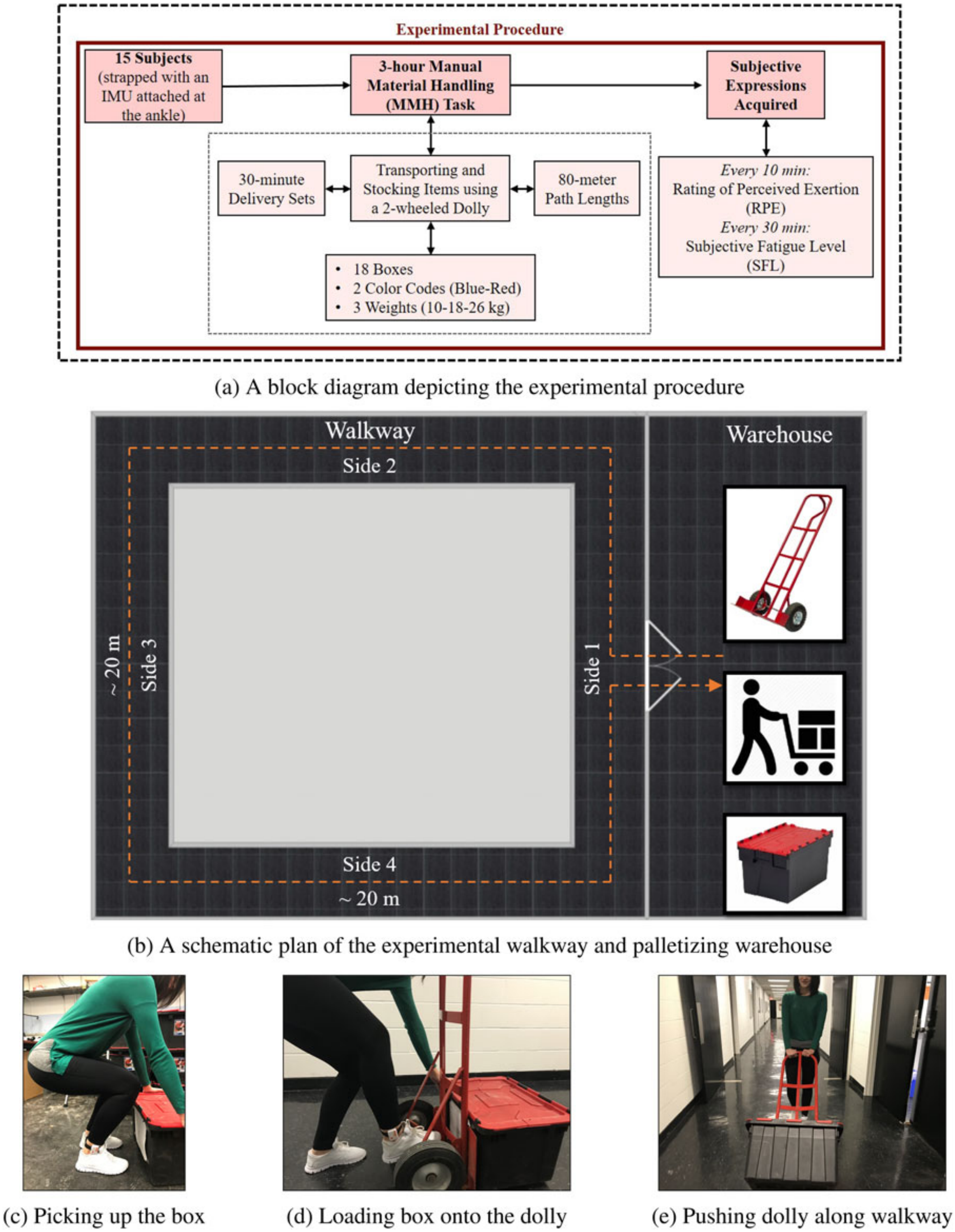
3.2.1. Fusing the accelerometer and gyroscope data using a Kalman filter

From Figure 3, the first step in preprocessing the IMU data is to estimate the angular orientation of the ankle. In this case study, we implement the procedure used by Baghdadi et al. (2018) and developed in the seminal work of Luinge and Veltink (2005), where a Kalman filter (KF) was designed to fuse accelerometer and gyroscope signals for ambulatory recording of a body part's orientation. For the sake of conciseness, we do not repeat the description of the process here. The interested reader is referred to Luinge and Veltink (2005) for a detailed description of the statistical modeling and the assumptions pertaining to both movement mechanism and sensor error behavior. In addition, we provide the MATLAB code used to implement that procedure in the GitHub repository (see the folder titled "feature-engineering") introduced in the *Supplementary Materials Section*.

The KF implementation results in an optimal/fused estimate of the angular orientation, which accounts for the different accuracies and sources of error for each of the accelerometer and gyroscope (Luinge and Veltink 2005). To demonstrate the utility of the KF implementation, consider the sample experimental

(a) The IMU sensor (b) The relative size of the IMU when placed on the right ankle

Figure 1. Visualizing the IMU sensor and its placement on the right ankle of one of the participants. Figure 1b Source: Baghdadi, Megahed et al. (2018), Figue 2. © 2018. Amir Baghdadi, Fadel M. Megahed, Ehsan T. Esfahani & Lora A. Cavuoto. All Rights Reserved. Reproduced with permission.



(a) A block diagram depicting the experimental procedure

(b) A schematic plan of the experimental walkway and palletizing warehouse

(c) Picking up the box (d) Loading box onto the dolly (e) Pushing dolly along walkway

Figure 2. A visual summary of the experimental procedure. Adapted from Figures 2 and 3 in Baghdadi, Megahed et al. (2018).

data from one participant shown in Figure 4. In this sample, the participant was standing for the first ~10 seconds and then walking for the remaining ~20 seconds. Figure 4(a) highlights the effect of movement on the angular orientation estimate obtained from the accelerometer. In particular, one can see the large variation in the angular position of the ankle with movement. This variation is much larger than the variation expected with movement on a flat surface and is attributed to the inability of the accelerometer to accurately predict the angular position on its own (Jimenez et al. 2009; Rebula et al. 2013).

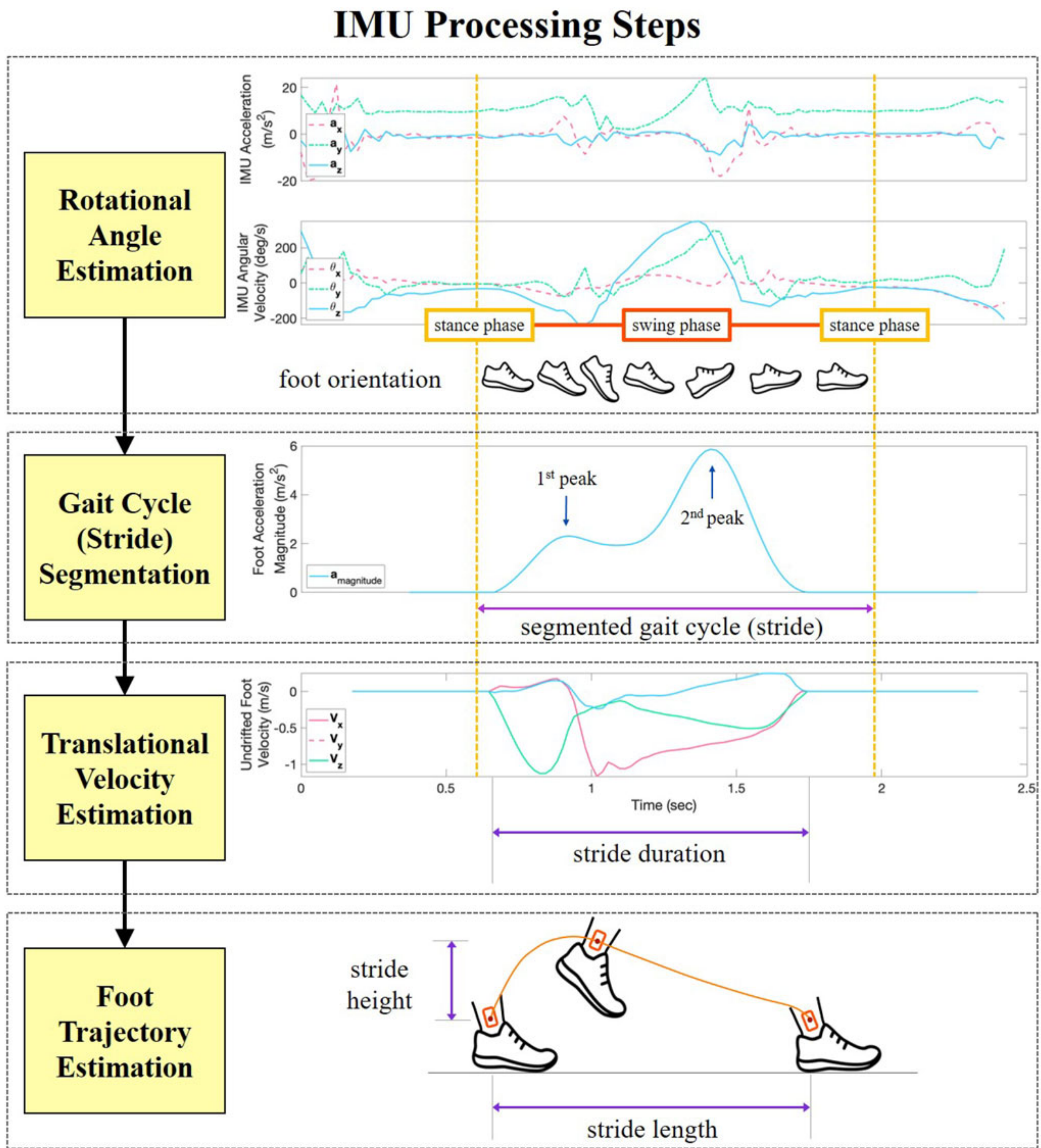


Figure 3. An overview of the main IMU processing tasks (shown through the flowchart on the left), with their corresponding data trajectories (plotted at right). Figure is inspired by Rebula et al. (2013).

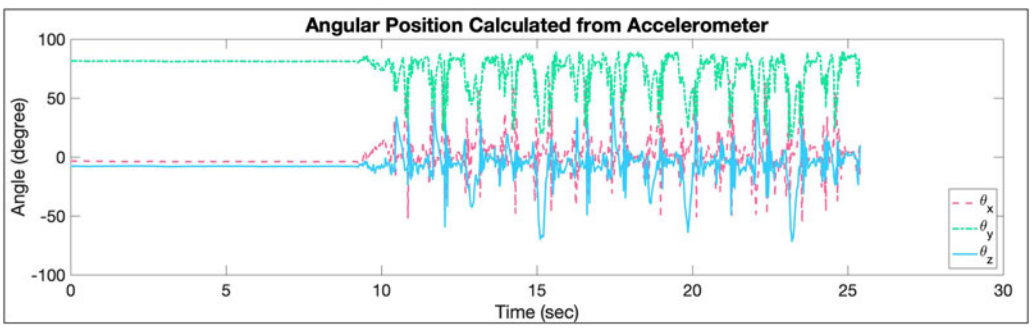
Moreover, Figure 4(b) demonstrates the drift in angular position due to the accumulated error of gyroscope biases (Chang, Li, and Chen 2015). Figure 4(c) shows how the KF procedure of Luinge and Veltink (2005) resulted in a realistic estimate of the changes (within $\sim 25^{\circ}$) in the ankle's angular orientation due to walking.

Estimating the angular orientation is pivotal in the usage and understanding of the acceleration data captured by the IMU's triaxial accelerometer. Similar to the approach of Baghdadi et al. (2018) (and detailed in Ghobadi and Esfahani 2017), we express the estimated angular positions, from the KF, in the form of quaternions, which are then used for translation of the accelerometer signal from the body coordinate system to the global frame of reference. Consequently, the gravitational effects will be limited to the vertical axis and will not affect/influence the accelerations along the horizontal plane. The reader is referred to Ghobadi and Esfahani (2017) for a detailed description of the mathematical formulations/ derivations.

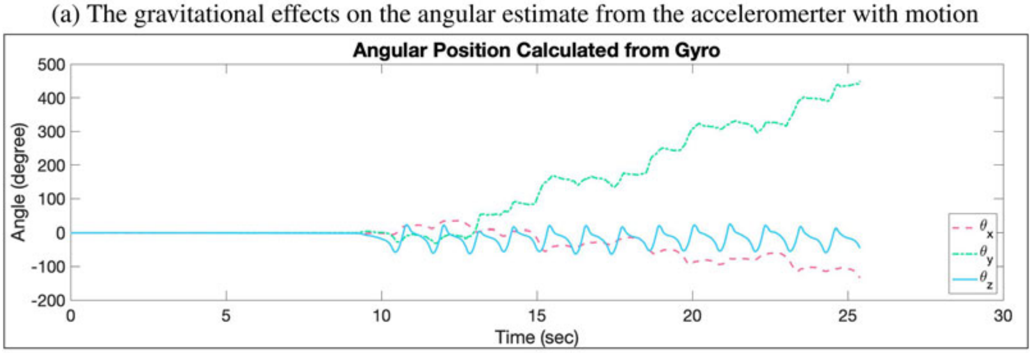
Once the accelerations are expressed in the global reference frame, high-frequency noise in the acceleration signals (generated from external sources, e.g., electromagnetic interferences and small shifts in the sensor positioning with movement) were removed through a smoothing procedure. A low-pass Butterworth filter, with a cutoff frequency of 4 HZ, was used to smooth the transformed acceleration signals in the ***X***, ***Y***, and ***Z*** global frame of reference directions. The reader should note that the Butterworth filter is commonly applied in smoothing biophysical sensor signals (see, e.g., Baghdadi et al. 2018; Bastos et al. 2001; Merletti and Di Torino 1999). After smoothing, the overall resultant acceleration was calculated as the magnitude of the accelerometer signal across the three axes (i.e., $\mathbf{a_m} = \sqrt{\mathbf{a_X^2} + \mathbf{a_Y^2} + \mathbf{a_Z^2}}$). The $\mathbf{a_m}$ signal is then further smoothed using the Butterworth filter, with the same cutoff frequency of 4 HZ, to ensure that the sensor signal is smoothed so that it is suitable for usage in the subsequent gait cycle segmentation procedure. This was done to mimic the analysis performed in Baghdadi et al. (2018).

3.2.2. Gait cycle segmentation

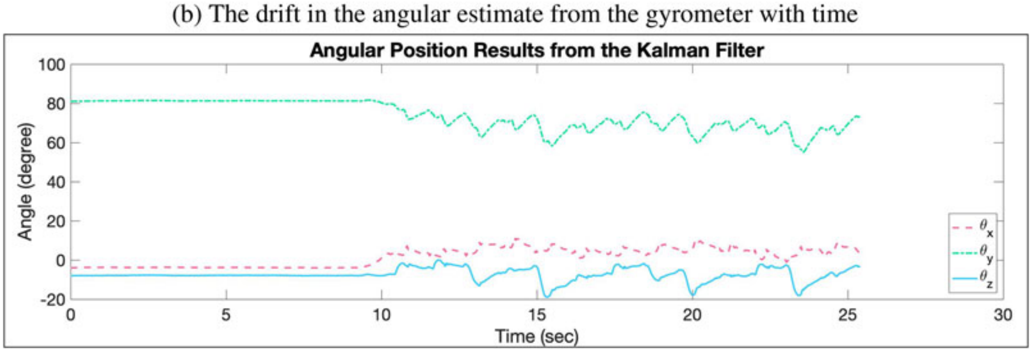
In order to extract the gait features of stride length, stride height, and stride duration, the time-series magnitude of the acceleration (i.e., $\mathbf{a_m}$) was segmented using a modified version of the algorithm introduced in Baghdadi et al. (2018). This method isolates each stride based on locating the two peaks of acceleration



(a) The gravitational effects on the angular estimate from the accelerometer with motion



(b) The drift in the angular estimate from the gyrometer with time



(c) A more reliable estimate of the angular orientation, as obtained from the KF procedure

Figure 4. The utility and effectiveness of the KF in estimating the angular orientation over time. The three signals correspond to the angular orientation in the ***x***, ***y***, and ***z*** directions.

indicating the heel strike and toe-off portions of the gait cycle. This assumption was justified based on the results of Tongen and Wunderlich (2010). As each stride was expected to last around 1 second, the locations of the two peaks are first searched within a time window containing 50 data points (recall our IMU has a frequency of around 50 HZ). From a biomechanics perspective, true gait segments were expected to fall within a 50–65 data point window. A moving-window approach was used to search the entire time series. After finding a gait segment, the search continued from the data point immediately following the end of that segment. However, if no segment was found within the time window, the ending point of the window was shifted by one data point. The procedure continued until all gait segments were extracted. Algorithm 1 summarizes our segmentation procedure. The algorithm contains several empirically determined thresholds. These thresholds are data dependent and may differ for other applications.

Algorithm 1 Gait Cycle Segmentation Approach Consisting of Two Main Procedures/Functions

procedure PARAMETERS DETERMINATION(*accMgnFiltTrain*)
Apply the SEGMENTATION PROCEDURE on *accMgnFiltTrain* with time window data points = 50 (*segLenAvg* = 50), and *peakAccThr* threshold $\tau = 5$. ▷ *Values are for initialization purposes only*
Determine the average number of data points per gait cycle in the training dataset.
Define the threshold τ for identifying acceleration peaks as 80 percent of smallest peak.
Output: [*segLenAvg, peakAccThr*]
procedure SEGMENTATION(*accMgnFilt, segLenAvg, peakAccThr*)
while {data length allows} **do**
Identify the 1*st* peak of the gait cycle acceleration.
Identify the 2*nd* peak of the gait cycle acceleration.
if {2*nd* peak > 1*st* peak by 20 percent && 2*nd* peak > τ} **then**
Determine *start point* for the gait segment by searching for the minimum point within a window 20 points before and 10 points after the current window *start point*.
Determine *end point* for the gait segment by searching for the minimum point within a window of 20 points past the 2*nd* peak.
if {determined segment length is > 1.3 × average segment length} **then**
Discard the segment as an outlier and start from the next point after the original time window *start point*.
else
Update the time window to start from the determined next data point after the *end point* to search for the next segment.
Update the array of segments to include the determined *start* and *end points*.
Start from the next point after the original time window *start point*.
return
Output: [*segStartPntArr, segEndPntArr, peakTwoArr*]

List of abbreviations:
accMgnFiltTrain: pure gait sample of $\mathbf{a}_m$ of 2,000 points (~ 40 sec) for training the segmentation parameters across the 3 hours.
accMgnFilt: filtered acceleration magnitude calculated from the transformed acceleration signal ($\mathbf{a}_m$).
segLenAvg: average length of segmentations from parameters determination phase.
peakAccThr: threshold τ from parameters determination phase.
segStartPntArr: array of refined start point for all segments in the input data set.
segEndPntArr: array of refined end point for all segments in the input data set.

3.2.3. Kinematics estimation and feature generation

From the segmented gait cycles, one can directly compute the stride duration. For a given gait segment, the stride duration equals the time difference between the ending and starting points for that segment. For generating stride length and height, we followed the procedure of Baghdadi et al. (2018), which entailed (a) numerically integrating the smoothed and transformed acceleration signals in the $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ global reference frame to obtain the corresponding velocity signal; (b) utilizing the *Zero Velocity Update* algorithm of Skog et al. (2010) to remove the drift in the velocity calculations and force the segmentation end point to zero (because this is a stationary period of gait); (c) numerically integrating the corrected velocity signal to get the position trajectories in the global $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ directions; and (d) from the position trajectories, estimating the stride length (i.e., the distance covered when a subject takes two steps, one with each foot) and height for each gait segment. Because one of our main goals is to understand fatigue development across participants, we rescaled the stride length and height, dividing each feature with the participant's height. This allows us to better interpret the features and facilitates the comparison of gait changes over time and across participants (Öberg, Karsznia, and Öberg 1993). The three time series containing the scaled stride height, length, and duration for each participant are depicted in Figure 5.

From Figure 5, there are three characteristics of our data that can complicate any subsequent analysis. First, the length of the data (which can be observed from the x-axis of the graph) is different for each participant. Note that the data length is a function of both the total time of walking per participant and number of strides. This can be problematic when we attempt to perform time-series clustering on the participants' data because these approaches require that the length of each series is identical. Second, there are several outliers within each series that will impact the performance of almost any statistical approach that will be used for detecting the onset of fatigue. Third, the time series for each feature exhibits high levels of autocorrelation and an underlying cyclical pattern that captures typical walking behavior. This is problematic because the existing literature on change point detection for time-series data assumes independent vector of observations over time (e.g., see Cabrieto et al. 2017; Capizzi and Masarotto 2017; De Oca et al. 2010; Matteson and James 2014). In the next section, we describe our attempts to tackle each of these issues.

4. Signal processing

4.1. Standardizing the length of each participant's time series

To account for the different lengths of the time series across participants, we rescaled the data from nominal time spent walking to a 0–100 percent scale, reflecting the percentage completion time for the experiment. To be specific, for each participant, the three features were sampled in increments of 0.05 percent time from start (starting with 0.05 percent and ending with 100 percent time from start). We used this sampling strategy because it allows us to (a) have an equal number of observations, $n = 2{,}000$, among the 15 participants; (b) the 2,000 observations present sufficient granularity to capture any changes in kinematic patterns associated with fatigue development (as the observations are spaced at approximately 3–5 seconds depending on the

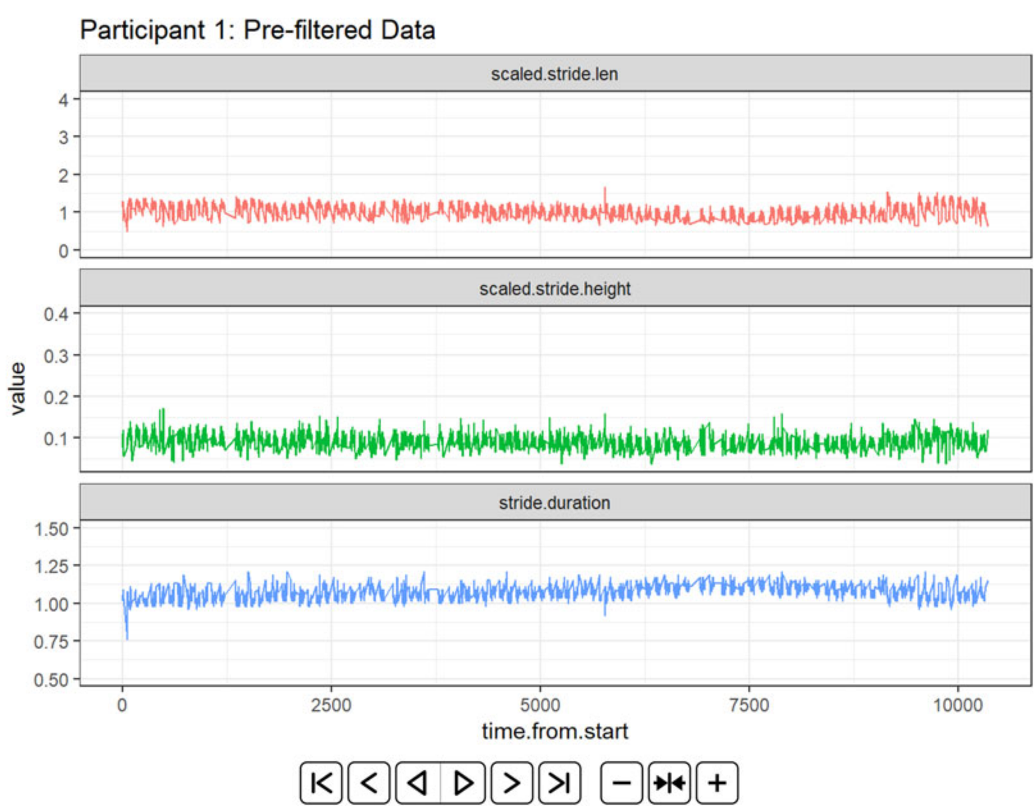


Figure 5. The three time series for stride length, height, and duration for Participant 1. The reader is encouraged to visit https://fmegahed.github.io/fatigue_case_jqt.html to examine the visualizations for the other participants. The time series data are stored in a "FeatureGeneration.RData" file, which can be accessed through our GitHub repository.

participant's walking behavior); and (c) maintain the time order of observation, which is needed for any subsequent statistical analysis. The links provided in the *Supplementary Materials* allow the reader to access our code and results from this standardizing stage.

4.2. Using the median filter to smooth the data

The sampling/standardization approach of Subsection 4.1 resulted in a trivariate vector (scaled stride length, scaled stride height, and stride duration) of 2,000 observations for each participant. Thus, it solved the unequal length problem highlighted in Subsection 3.2.3; however, the sampling did not remove/correct the outliers observed in Figure 5. This subsection describes our approach to outlier correction for the data obtained at the end of Subsection 4.1.

Based on plotting the standardized data (not shown here for conciseness; refer to Figure 5 for the nonstandardized data or see https://fmegahed.github.io/fatigue_case_jqt.html for the standardized plots), we observed the following:

A. From a kinematic perspective, the spikes in the data reflect errors from the segmentation methodology and/or from the sensor signal. For example, a stride length ≥ 2 is not realistic because this means that the person's stride length (refer to Figure 3 for a visual interpretation) is twice their height. The challenge here, however, is that there is no hard limit that we can use for outlier detection because the kinematics literature reports mean values across the population instead of attempting to provide a physical threshold of what is possible.

B. There are several instances of the data, where the outliers occurred in short succession (typically 2–3 observations). These situations should be corrected because a participant is unlikely to be fatigued for a few strides only. Specifically, the application of the standardization step, in the previous section, means that the maximum time difference between two consecutive observations ranges between 3 to 5 seconds, depending on the participant.

C. As noted earlier in Subsection 3.2.3, the data for each participant seems to be nonstationary and autocorrelated. This means that commonly used outlier detection methods, which are based on normal theory (e.g., box plots for each sensor signal or the Mahalanobis-based distance measures for our trivariate vector) are not appropriate because their distributional assumptions will be violated.

From these observations, the outliers can be corrected through a robust filtering technique. This allows the data to retain its statistical characteristics and simplifies the signal processing and imputation to one step.

We used the *median filter* to smooth the data. The use of median filtering to smooth data is reportedly first suggested in Tukey (1974) and has been widely adopted ever since (see, e.g., Dougherty and Astola 1994; Gallagher and Wise 1981; Pitas and Venetsanopoulos 2013).

From an algorithmic perspective, our implementation of the median filter consisted of three steps. First, we had to select an appropriate window size (w). We chose $w = 21$ because it (a) captures 60–100 seconds of continuous walking, depending on the participant, which allows us to extend our approach to fatigue prediction applications, where 2-minute windows are often used (see e.g., Maman et al. n.d.); and (b) is an odd number, which makes the median one of the actual sensor values from the window. Second, we had to determine the magnitude of overlap of the moving window. The three most common overlap types are no overlap, an overlap of approximately $\frac{w}{2}$, and an overlap of $w - 1$ (Pitas and Venetsanopoulos 2013). We chose an overlap of $w - 1$ (i.e., we moved the window by one observation at a time) because this allowed us to preserve the changes due to fatigue (see the following subsection for a detailed discussion). Third, once the window size and overlap magnitude were selected, we applied the median filter to the three features and all participants.

The application of the median filter smoothed the data and reduced the length of each participant's time series by $w = 21$, from $n = 2{,}000$ to $n = 1{,}979$ observations (our first observation corresponds to 0.5 percent percent from the start instead of 0.05 percent). The results after the standardization and median filtering signal processing steps are depicted in Figure 6.

4.3. On the tradeoff between autocorrelation and change point detection

As mentioned in the previous subsection, the selection of the degree of window overlap for the median filter affects the amount of data available for analysis, the ability to visualize patterns in the data, as well as the degree of autocorrelation. These two effects can have a profound influence on the results of multivariate change point methodologies, most of which are developed to detect changes in observation vectors that are independent over time.

Figure 7 shows how the window overlap affects the scaled stride length for participant 10. The top panels give the filtered series for windows of size $w = 21$ with no overlap, an overlap of $\frac{w}{2}$, and an overlap of $w - 1$. The lower panels give the autocorrelation functions (acf) up to lag 30 for the same overlap values. Due to the nature of the biometric process we are observing, one would expect the cyclical autocorrelation pattern as we observe in the series with windows that overlap by $w - 1$. As expected, when the overlap is reduced to $\frac{w}{2}$ and no overlap, the autocorrelation reduces, but so does our ability to visualize the patterns in the data over time.

Because the goal of the analysis is to detect fatigue, we determined that it is best to retain as many of the patterns in the data as possible by using an overlap of $w - 1$. By doing this, we are left with highly autocorrelated data, which violates the assumptions of most multivariate change point methods. This decision reflects a realistic decision that practitioners often must make in practice. Transforming or manipulating data to meet statistical assumptions necessary for certain methods often renders the features and patterns of the data indiscernible, but failing to transform the data means that the conclusions drawn from the methods are often limited or even invalid. For the purposes of this analysis, we will proceed with the data filtered using a window size of $w = 21$ with an overlap of 20, noting that the data are significantly autocorrelated, and that our change point analysis must be interpreted with this limitation in mind.

5. Change point analysis

5.1. Rationale for the use and selection of change point methodology

Let $\mathbf{z}_{i,1}, \mathbf{z}_{i,2}, \ldots, \mathbf{z}_{i,1979}$ be the sequence of standardized and median filtered kinematic features for participant i, where $i = 1, 2, \ldots, 15$. With this notation, the vector $\mathbf{z}$ is composed of the median filtered kinematic features (scaled stride length, scaled stride height, and stride duration) and the second subscript denotes the index for the percent time from start. In this section, we describe the statistical approaches used to analyze $\mathbf{z}_{i,1}, \mathbf{z}_{i,2}, \ldots, \mathbf{z}_{i,1979}, (i = 1, 2, \ldots, 15)$ to answer the following research questions: (1) How do important gait parameters (e.g., stride length, height, and duration) change over time? (2) How do these sensor-based

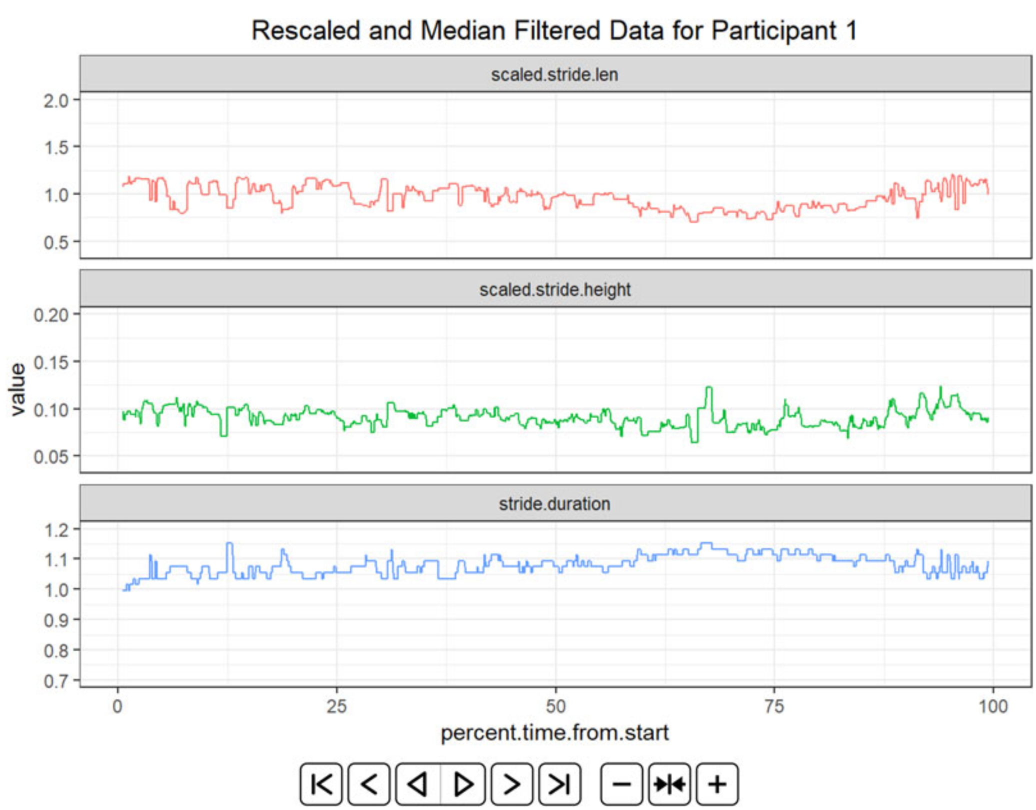


Figure 6. The standardized and median filtered time series for stride length, height, and duration for Participant 1. The reader is encouraged to visit https://fmegahed.github.io/fatigue_case_jqt.html to examine the visualizations for the other participants. The standardized and median filtered data for each participant are stored in a "NormMedianFilteredData.RData" file, which can be accessed through our GitHub repository (see Supplementary Materials).

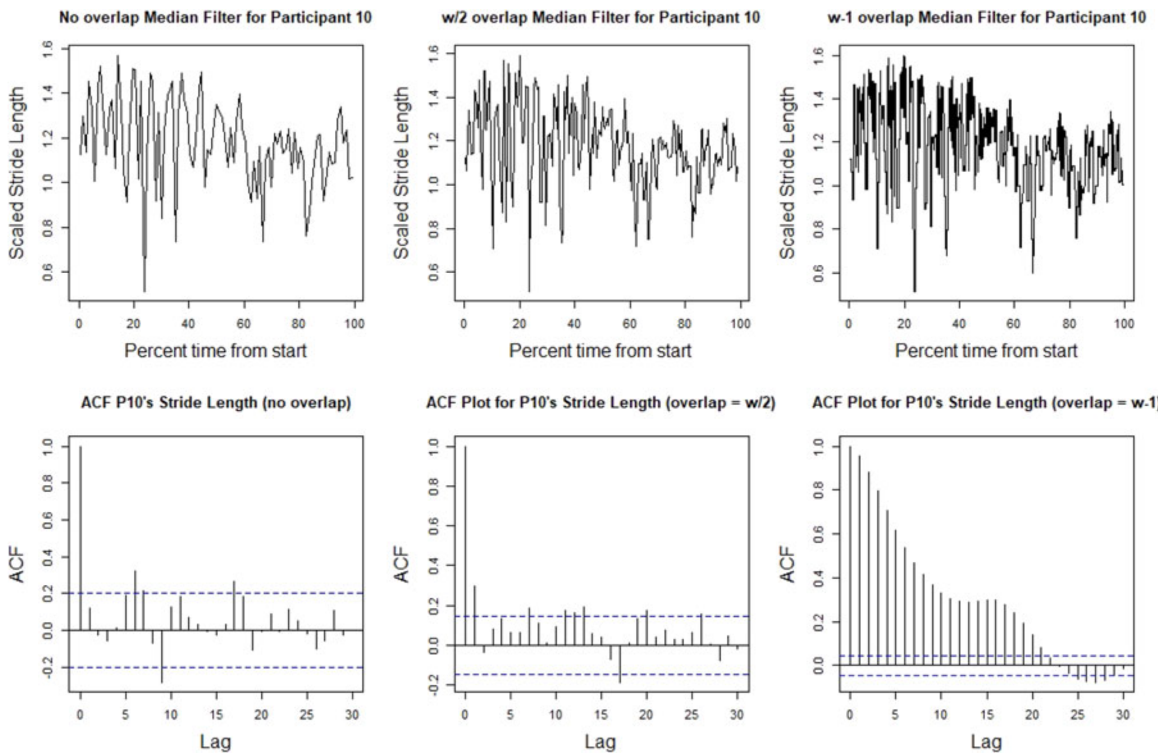


Figure 7. An illustration of the effect of overlap window on the independence of successive observations for the scaled stride length for Participant 10.

changes relate to the participants' subjective fatigue ratings? To answer these questions, we used a multivariate change point methodology supplemented with a graphical analysis for diagnosis.

From Figure 6, for a given participant i, one can make the following observations about the data $\mathbf{z}_{i,1}, \mathbf{z}_{i,2}, \ldots, \mathbf{z}_{i,1979}$: (a) the data are trivariate and are sampled from a continuous, but unknown, distribution; (b) change point locations are unknown, with the possibility of multiple unknown change points occurring in the data; (c) each vector of observations is nonstationary; (d) the variance of each feature is not constant; and (e) the vectors are autocorrelated, that is, the typical independence assumption between $\mathbf{z}_{i,j}$ and $\mathbf{z}_{i,j+k}$ cannot be met. Statements (a)–(d) require the use of a nonparametric, multivariate change point methodology that allows for detecting multiple change points. Examples of such methods include those of Capizzi and Masarotto (2017) and Matteson and James (2014); however, our data violate the assumption of independence between observation vectors for both of these methods. At this time, there is little published research on change point methods suitable for this data that are designed for a time-dependent multivariate and multiple change point scenario. For this reason and given the scope of this case study article, we have elected to use an existing approach and then evaluate whether the detected change points are reasonable based on combining the information from the experimental procedure and the participant's subjective ratings of fatigue.

In addition to the statistical requirements, there are important practical requirements that should inform method selection. The method should be reasonably easy to implement and replicate, easy to understand, and able to be automated. To ensure that the implementation is straightforward and easy to replicate, we limited our selection of methods to those that have an R package. Both Matteson and James (2014) and Capizzi and Masarotto (2017) met the requirements (a)–(d) above and had complementary R packages: ECP and dfphase1. In addition, both Matteson and James (2014) and Capizzi and Masarotto (2017) were straightforward to implement, providing results that were easy to understand. Finally, because we do not expect practitioners to inspect multiple graphs to visually determine the location of change points, it is important that the methods could be automated. The methods of Matteson and James (2014) and Capizzi and Masarotto (2017) both met this need.

5.2. Application of the change point methods

Capizzi and Masarotto (2017) proposed a multivariate Phase I method based on the signed rank statistic for detection of process shifts of various types (e.g., transient, sustained, step, etc.). The method uses a permutation test to determine the location of change points and is complemented with a LASSO-based post-signal diagnosis tool to identify which variable shifted. We implemented this method, which has a strict assumption of independence of the observation vectors over time, on our filtered series, $\mathbf{z}_{i,1}, \mathbf{z}_{i,2}, \ldots, \mathbf{z}_{i,1979}$, and we observed numerous change points that we believe to be false alarms. For this reason, we did not include this analysis in this case study.

Matteson and James (2014) introduces a retrospective, nonparametric change point method based on a multivariate distance measure given in Szekely and Rizzo (2005) known as *e*-distance. The *e*-distance is an extension of Ward's minimum variance method (Ward 1963), which is used in hierarchical clustering to minimize the total within-cluster sum of squared error. The *e*-distance between two vectors of observations $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ measures jointly the between and within distances between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. Szekely and Rizzo (2005) also developed a hierarchical clustering algorithm to merge pairs of clusters with minimum *e*-distances.

Matteson and James (2014) adapted the clustering method of Szekely and Rizzo (2005) for use as a change point method for multivariate observations over time, presenting two possible methods. One method, referred to as the *e*-divisive, estimates multiple change points by iteratively applying a single change point method and using a permutation test. The second method, referred to as the *e-Agglo* method, is an agglomerative clustering method. This method requires an initial segmentation of the data and observations are grouped together in sequence as in hierarchical clustering. The algorithm is designed to preserve the time order of the observations within and between clusters. In the absence of prior information about the location of change points to create an initial segmentation, the initial segments can be set to the individual observations or pairs of adjacent observations.

Because of its computational efficiency and good performance, we used the *e-Agglo* method of Matteson and James (2014) for our analysis. Details on the implementation of the *e-Agglo* method can be found in James and Matteson (2013). Implementation requires an initial segmentation of the data as well as a penalty function to discourage over fitting and a parameter, $0 < \alpha \leq 2$. For our analysis, we used an initial segmentation of subsequent pairs of observations because this segmentation reduced the computational burden over the choice of individual observations as segments, yet did not substantially affect our results. We chose the *penalty2* function (see James and Matteson 2013, 6), which discourages the identification of small change points and/or change points in close proximity, because we are interested in detecting large and sustained shifts in the features indicating fatigue. In certain special cases of clustering, the parameter α can be thought of as the power of the Euclidean distance between clusters. For example, $\alpha = 2$ corresponds to the use of Ward's method in standard hierarchical clustering. For our analysis, we chose $\alpha = 1$ as recommended by Matteson and James (2014). The reader is referred to Szekely and Rizzo (2005) and Matteson and James (2014) for more details on selecting α.

Plots of the filtered data with the identified change points overlaid in black are given in Figure 8. From this interactive figure, there are three trends from the graphs. We describe each trend in a paragraph below.

First, the number of change points varied between 1 and 4. In most cases, the single change point separated an initial interval (equating to 1–1.5 hours of length) of no signs of fatigue to a secondary interval where fatigue started to develop. When two change points were detected (e.g., see participants 5, 6, and 10), the third interval captured the participant's deteriorated performance near exhaustion. When three change points were found (single case of participant one), the initial interval was short, indicating that the subject was still in his warm-up/learning phase. When four change points were found (i.e., participant 15), the time difference between the second and third phase was short, indicating that the participant had a significant and large change in his walking patterns. We do not have a kinematic/physiological justification for this change. However, from a psychological perspective, this short change can indicate distraction and/or a decrease in motivation.

Second, the change points captured changes in mean, variance, and correlation structure among the three features. From a statistical perspective, the choice of $\alpha = 1$ allows for detecting different types of shifts. From a kinematic perspective, we wanted to detect these patterns because they correspond to different fatigue modes, which may require different interventions. For example, changes in mean can reflect the onset of fatigue, while changes in variance of an acceleration signal is a symptom of tremor, which can happen in extreme fatigue cases (Cavuoto and Megahed 2017).

Third, participant 3 had faulty sensor readings, which resulted in a large amount of missing and/or inconsistent data. This can be observed from the

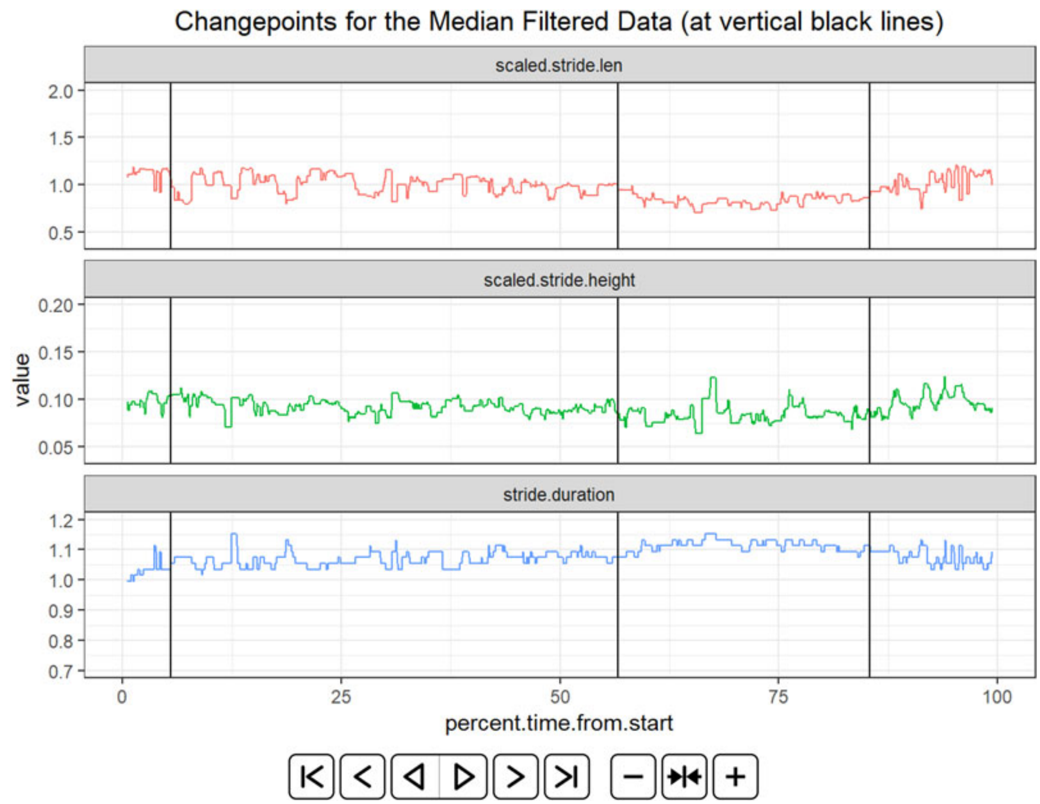


Figure 8. The *e-Agglo*-based change points are overlaid on the CUSUMs of the standardized and median filtered time series for stride length, height, and duration for Participant 1. The reader is encouraged to visit https://fmegahed.github.io/fatigue_case_jqt.html to examine the visualizations for the other participants.

sensor signal data (as well as Figures 5 and 6). These inconsistencies, unsurprisingly, led to a large number of change points. Hereafter, we will exclude participant 3 from our analysis.

5.3. Using cumulative sums to explain the changes to a non-statistical audience

Due to the nonstationarity and the cyclical patterns observed in our data, we hypothesized that the cumulative sums (CUSUMs) of the mean-corrected series would contain useful diagnostic information. Because we would be implementing the CUSUMs retrospectively (i.e., a Phase I application), we used the overall mean of each series for a given participant to estimate the mean parameter within the CUSUM. The top three panels in Figure 9 show cumulative changes above and below the average level for each participant during the experimental period for each kinematic feature. For a given participant, the black vertical lines denote the *e-Agglo* change points first shown in Figure 8 (i.e., the change points were obtained on the filtered series, $\mathbf{z}_{i,1}, \mathbf{z}_{i,2}, \ldots, \mathbf{z}_{i,1979}$, and not the CUSUMs). Additionally, the lower panel denotes the participant's subjective ratings of fatigue (RPE) using the Borg 6–20 scale. The blue vertical line in the panels corresponds to the first instance when the $RPE \geq 13$, which can be used to estimate the onset of fatigue according to Maman et al. (2017).

There are several noteworthy observations to be made from Figure 9. First, the reader should note that all top three panels, that is, the CUSUM plots, always end at zero for all participants. This is not a coincidence, and this is an expected mathematical result given that we have used the CUSUMs retrospectively, with the overall mean of each series as an estimate for the mean parameter used in the recursive CUSUM calculation. Second, the change points did not coincide with the RPE threshold of $\geq$ 13. In some cases, the proximity of a change point and the RPE threshold were close (e.g., participants 1 and 2). However, in most cases, the perceived exertion threshold did not coincide with the identified change points in the processes. We estimate that this observation is attributed to the differences between the psychological and physiological reactions. In other terms, while the $RPE \geq 13$ can be used to denote the onset of fatigue, the effects of fatigue on performance can precede or follow the occurrence of this threshold. Third, and perhaps the most interesting observation, large changes in the

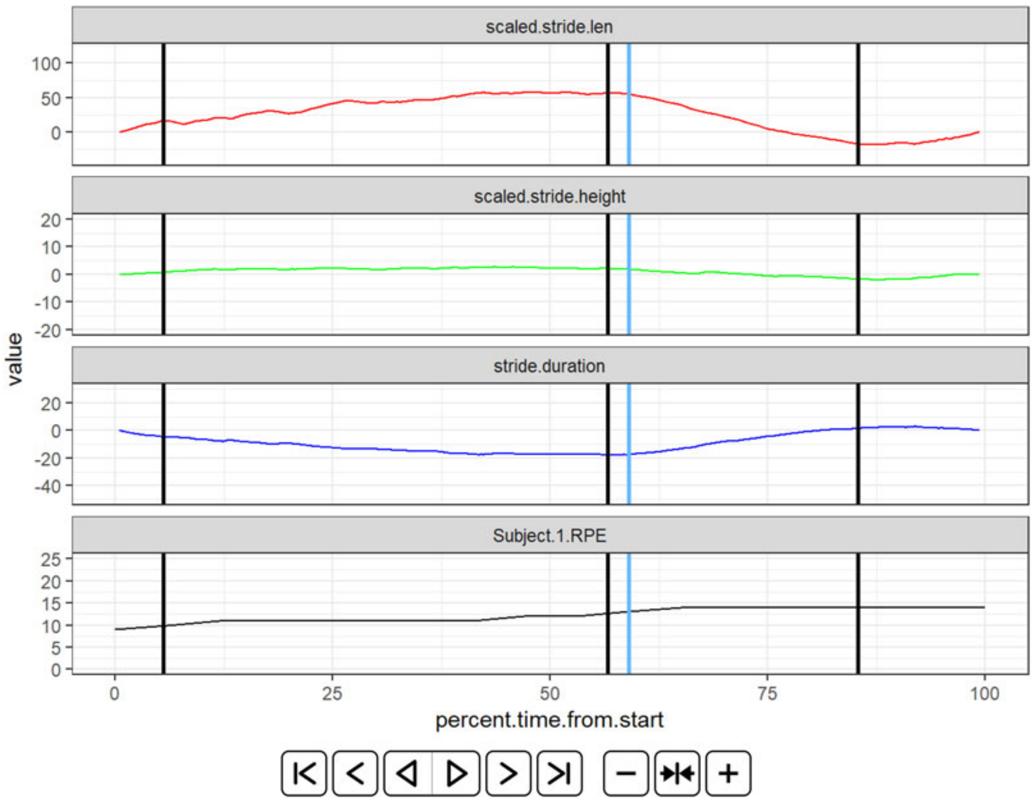


Figure 9. A visualization of the *e-Agglo*-based change points overlaid on (a) the CUSUMs of the standardized and median filtered time series for stride length, height, and duration for Participant 1 (top three panels) and (b) the participant's subjective ratings of fatigue (bottom panel). The blue vertical line captures the first instance when the subjective rating of fatigue is $\geq$ 13. The reader is encouraged to visit https://fmegahed.github.io/fatigue_case_jqt.html to examine the visualizations for the other participants.

CUSUM slopes were often accompanied by change points (e.g., see participants 1, 2, 4, 5, 7, 8, 9 10, 12, 13, and 15). Although we interpret the analysis retrospectively, this seems to indicate that changes in direction above and below the within-person average are being identified as possible changes in exertion or performance over time. We have no definitive conclusions from this analysis; we simply find this an interesting observation worthy of further investigation.

5.4. Discussion of change point modeling results

In Section 5, we performed a retrospective statistical analysis of the trivariate series of observations from wearable sensors for the purposes of attempting to better understand/answer the characteristics of fatigue development over time. More specifically, the analysis performed in this section attempted to answer two main research questions.

For addressing the first question, we used a multivariate nonparametric change point method based on agglomerative clustering on our data, which allowed us to automate the detection of changes in walking patterns over time for a given participant. From a kinematic point of view, we have learned that, over time, participants tended to have systematic changes in their walking pattern. To summarize the results depicted in Figure 8, we expect that a person's walking behavior in a manual material-handling task involves the following four stages (based on the occurrence of three change points). The first stage can be defined as a warm-up/ramp-up stage. This stage is often highlighted in the literature; it is customary to use an arbitrary 10-minute cutoff in the data preprocessing to remove its effects (see, e.g., Baghdadi et al. 2018; Maman et al. 2017). Our results indicate that this arbitrary cutoff may not be suitable for all participants, as indicated by the early change for participants 1 and 10. Future studies may benefit from using data from the initial stages of the experiment and determining the appropriate cutoff for the warm-up period on a participant-by-participant basis using a suitable change point method. The second stage involves the participant reaching a steady-state walking pattern (or more generally work performance). From an occupational safety and ergonomics perspective, we hope that job tasks are designed such that this steady-state pattern can be maintained for the shift duration. If the job is not well designed and/or the worker is not well rested, the third stage occurs when the steady-state pattern changes. Given the nature of the experiment and our analysis, we cannot conclusively state that a detrimental change in performance is observed in the third stage. However, we can state that there is a change in the patterns observed by the sensor, which can reflect either a worker's attempt to adjust to the onset of fatigue (by changing work posture) and/or a different level of performance. In the fourth stage, the magnitude of change in pattern is larger, reflecting a higher level of fatigue effects.

For the second question, we combined the RPE values with the change points obtained from the first analysis. Our results seemed to indicate that there is a difference between the psychological feeling of fatigue and its transcendence to changes in job performance. Future research in this area should investigate our observation further.

6. Time series clustering

In Section 2, we posed four research questions, the last two of which were

(3) Are there consistent patterns in performance across different individuals over time?
(4) If so, do these patterns vary systematically based on specific demographic characteristics?

To answer these research questions, we applied a multivariate time series clustering method followed by a descriptive analysis of demographic and anthropometric characteristics of the resultant clusters. Note that these questions can be answered using either the scaled and median filtered trivariate series or the CUSUM of the aforementioned series. In this section, for the sake of conciseness, we limit our discussion to when the clustering was performed on the CUSUM of the trivariate series. The reader is referred to Section 4.1 in https://fmegahed.github.io/fatigue_case_jqt.html for the clustering results on the median filtered trivariate series.

6.1. Application of clustering methodology

Clustering methods are unsupervised methods that are generally used in large data situations. Although we have a very small number of participants in our study, we found ourselves trying to heuristically group participants according to patterns in their performance. As such, we decided to try using a more algorithmic approach and apply a clustering method for purely exploratory purposes. With only 14 observations in our study (recall that we dropped participant 3 because of a large number of suspected faulty sensor readings), we realize that the conclusions we can draw from this analysis are limited to a purely exploratory analysis. In addition, this analysis makes an assumption that the participants would exhibit the same or similar pattern for repeated performances.

Time series clustering methods are extensions of cross-sectional clustering algorithms, modified to handle time series data. As with all clustering algorithms, the distance or dissimilarity measure between the clusters is one of the most important decisions that affect the results of the analysis. Because we are interested in understanding whether there are consistent patterns among individuals and not necessarily whether those patterns of change happen at exactly the same point within the experiment, we chose an elastic method to measure dissimilarity known as dynamic time warping (DTW). The idea behind DTW is to compare two series by algorithmically stretching or compressing two series in time in order align the patterns. After doing so, the distances between the two series are computed. It is, perhaps, easiest to understand the intuition behind DTW graphically. Figure 10 shows the DTW alignment between two time series using the aami3a data from the R package dtw. Aghabozorgi, Shirkhorshidi, and Wah (2015) note that the DTW is preferred over Euclidean distance measures for data that has a temporal drift. In addition, the DTW

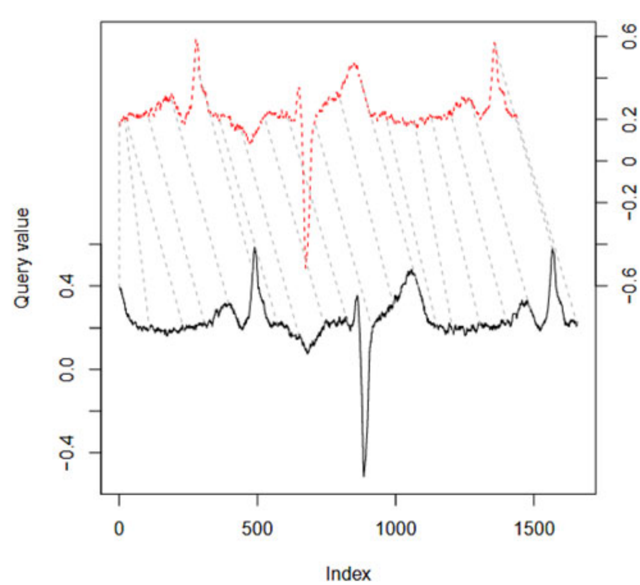
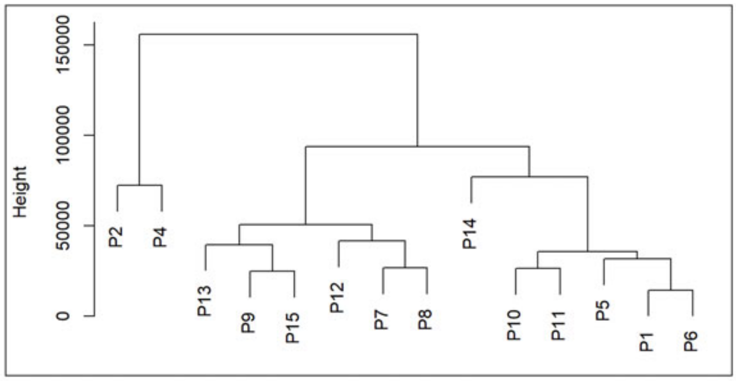


Figure 10. An example of applying DTW to align and compute the distances between two series. Figure is from Giorgino (2009).

method can cluster multivariate time series, which makes it suitable for our case study.

We used the R dtwclust package (Sarda-Espinosa 2018) to perform the multivariate time series clustering of the participants. Specifically, we used the function tsclust() and chose hierarchical clustering with the classic DTW measure computed according to Sakoe and Chiba (1978). Because this is a purely

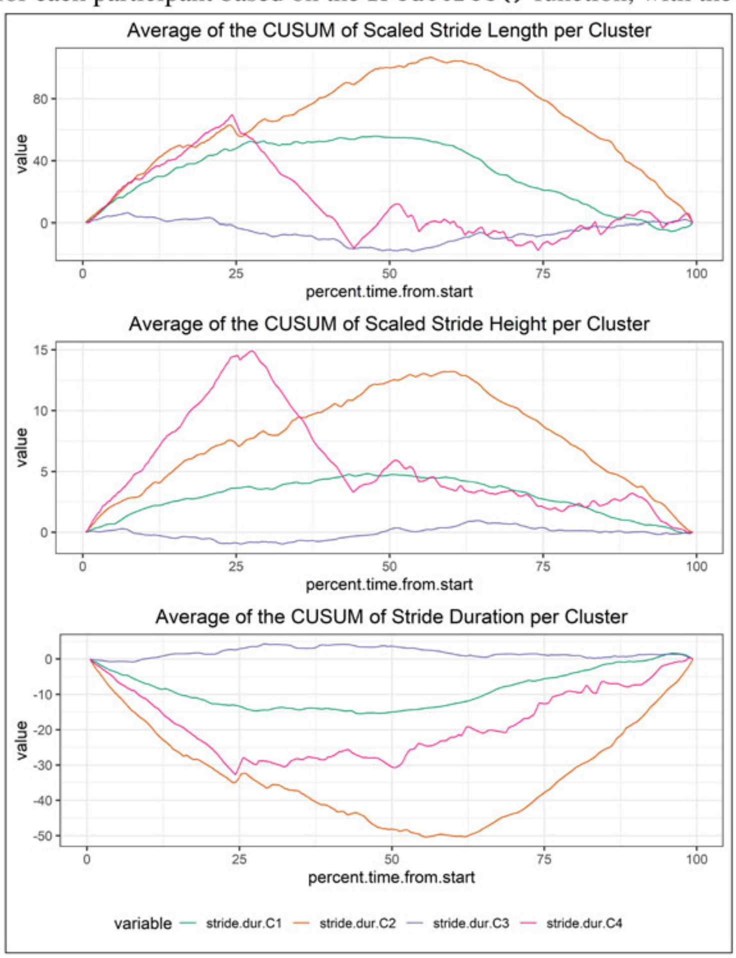


(a) Dendogram based on multivariate heirarchical time series clustering using DTW as a distance measure.

Cluster Assignment for each Participant

P1	P2	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
1	2	2	1	1	3	3	3	1	1	3	3	4	3

(b) Cluster assignment for each participant based on the **R** `cuttree()` function, with the number of clusters = 4.



(c) A summary of cluster membership based on the means of the cumulative sum of each feature.

Figure 11. Results of applying the hierarchical clustering, with the DTW distance measure, on the CUSUMs of the three scaled and median filtered feature data.

exploratory analysis, we used the dendogram shown in Figure 11a to identify an appropriate number of clusters. Based on the dendogram, we selected four clusters. Accordingly, the participants (P) were divided as shown in Figure 11b. To help in understanding why the participants were clustered into different groups, we plot the average "profile" of each feature per cluster in Figure 11c.

Figure 11 (especially Figure 11c) presents several insights that are helpful in addressing our third research question. First, there is a large variation in fatigue development patterns across the participants. This is why the number of clusters is somewhat large considering the sample size. Second, the profiles depicted in Figure 11c show that there are varying patterns of fatigue development across participants. For example, the participants within clusters 1 (green) and 2 (orange) exhibit similar patterns, but with different magnitudes. The CUSUMs of stride length and height increase until the onset of fatigue, when they then decrease. This means that those participants generally take shorter and lower steps when they are fatigued. The reciprocal effect is shown on stride duration, where the time taken for each stride increases with fatigue. The six participants within cluster three exhibit the opposite behavior, where they generally take longer and higher strides in a shorter time span when they are fatigued. For cluster 4 (i.e., participant 14), the pattern is somewhat similar to clusters 1–4; however, the onset of fatigue (at least from a change of performance perspective) is earlier and more abrupt (at least for stride height and length).

From an occupational health and safety perspective, the changes detected from the sensors would not have been observed otherwise. Specifically, the participants had to pick the weighted cartons at a given rate. When fatigued, the participants maintained the prescribed order-picking rate; however, they adjusted their walking pattern (either by having longer and slower strides or shorter and faster strides) to mitigate the effects of their physical fatigue. This is a novel and interesting finding, which requires further investigation to examine whether such a pattern holds in other occupational settings.

6.2. Demographic and anthropometric analysis of cluster solution

In Table 1, we present the recorded summary statistics of the participants within each cluster. A close examination, involving both the demographic/anthropometric measures and number of subjects within the cluster, reveals that there does not seem to be a systematic difference between clusters. Given the secondary nature of the analysis and small sample sizes, we cannot determine whether the lack of systematic differences is attributed to (a) not capturing other anthropometric, demographic, and/or medical factors that are driving the differences or (b) the effects could not be determined due to the small sample sizes.

From an occupational safety and ergonomics perspective, our conclusion that there are no differences due to demographic and anthropometric measures is potentially interesting if this conclusion can be replicated for larger sample sizes. In such a case, occupational safety professionals should examine their workers on an individual basis and not by using some general "rule of thumb." Current injury prevention practices often recommend designing work tasks, for example, 75 percent of the female population, will not get injured. However, these calculations are only valid for simple tasks such as stationary lifting (Salvendy 2012). In more dynamic and complex tasks, we may need to resort to real-time and individualized monitoring, where we account for a worker's baseline. The insights gained from our analysis would support the need for a larger study to examine this observation (personal and body size/shape factors not accounting for differences within the clusters) further.

7. Conclusions

7.1. Summary of impacts and contributions of the case study

7.1.1. An ergonomics and occupational safety perspective

In this case study, we examined the use of wearable sensor data to understand how fatigue develops over time. More specifically, we attempted to address the following research questions: (1) How do important

Table 1. The mean (SD) demographic, anthropometric, and perceived fatigue ratings for participants within each cluster.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
# Subjects within cluster	5	2	6	1
Gender (M= 0 & F= 1)	0.4 (0.55)	0.5 (0.71)	0.5 (0.55)	1 (0)
Age	38.8 (15.1)	24.5 (6.4)	32.8 (15.4)	62 (0)
Weight in kg	78.0 (17.3)	74.7 (6.3)	69.0 (2.2)	92.7 (0)
Height in cm	168.6 (10.7)	177.5 (0.7)	169.7 (6.9)	183 (0)
Final Borg ratings	15.6 (2.4)	15.5 (0.7)	15.2 (2.1)	13 (0)

gait parameters (e.g., stride length, height, and duration) change over time? (2) How do these sensor-based changes relate with the participant's subjective fatigue ratings over time? (3) Are there consistent patterns in performance across different individuals over time? (4) If so, do these patterns vary systematically based on specific demographic characteristics?

After engineering the features and deploying signal-processing techniques to smooth the time series of each feature, we combined exploratory data analysis and the *e-Agglo* multivariate change point methodology to address the first two questions. From our analysis, we have learned that the number of change points for each participant typically varied between one and three change points. Their location (i.e., percent time from start of the experiment) indicated that there are generally four stages of performance changes. These stages correspond to (a) warm-up/learning stage, (b) steady-state performance stage, (c) deviation from the steady-state performance due to the onset of fatigue, and (d) larger deviation from the steady-state performance, where participants start to be exhausted. The identification of these changes represents a novel and significant contribution to the biomechanics and ergonomics literatures. Perhaps, more importantly to the *statistics* community, these were only possible due to using the multivariate change point methodology and the CUSUM-based visualizations.

For the second question, we have learned that the performance changes are not typically aligned with the participant's perception of fatigue. This is an interesting finding because it indicates that there is a difference between the perception and physical changes in performance due to fatigue.

To address questions (3) and (4), we combined the DTW multivariate time series clustering approach with EDA. Our analyses indicated that the participants should be clustered into four groups, which reflect changes in both the magnitude and pattern of fatigue development across clusters. Because participants had to maintain a certain order-picking pace, they adjusted their stride (shorter and faster strides for some participants and longer and slower for others) to mitigate the effects of fatigue. From an operational perspective, this finding supports the use of wearable sensors in the workplace because typical measures, such as cycle time and time on task, would not have been able to capture this change. For question (4), our analysis does not support the hypothesis that the changes in patterns (i.e., the clusters) are driven by anthropometric and demographic factors.

7.1.2. A statistical (process control) perspective

Over the past several years, several prominent industrial statisticians (Nair 2007; Woodall 2017; Woodall and Montgomery 2014) have indicated that there is a gap between the theory and practice of statistical process control (SPC). For example, (Nair 2007, 7) stated that:

> One can identify many reasons for this gap: real problems tend to be messy and do not readily translate into research topics that can be solved and published; there are no incentives for researchers in academia to get close to the problem and develop complete solutions, software and transfer technology (especially in statistics departments that tend to be located in liberal arts colleges) . . .

Additionally, most of the research in SPC has remained centered around manufacturing applications (we acknowledge that there are streams of excellent research in other domains, e.g., public-health surveillance and network applications). Thus, there has been an emphasis, by renowned authors such as Box and Woodall (2012) and Bisgaard (2012), to examine innovative applications of our methods in other domains. Recently, Jensen et al. (2018, 10-11) stated that "Researchers must focus on real, applied problems, which require collaboration with subject matter experts . . . True collaboration with subject matter experts on real projects is the key for critical future research."

In the spirit of the above discussion, there are three aspects that we hope to achieve by publishing this case study. First, we would like to inspire research in human performance modeling/monitoring because it is an important area of work with significant worker's safety and economic implications. By making our data and code available, we hope to reduce the "cost of entry" for statisticians and encourage future work in this area.

Second, we would like to highlight the importance of an *open research culture* (Nosek et al. 2015). We see no reason why publishing the code does not become the norm in our research community. We applaud researchers who put their code into R packages that help the translation of our methodologies into other domains. While we have highlighted only two R packages in this article, there are several other packages developed by members of our community (see, e.g., Bui and Apley 2018; Höhle 2007; Knoth 2018). In our estimation, publishing the code brings us closer to Prof. Nair's call for developing complete solutions that include software development, and this process has become simpler over the last few years with continued developments in both Python and R.

The third, and final observation, that we would like to highlight is that we agree with Prof. Nair's comment about how "real problems tend to be messy." We would like to add that they often require a multidisciplinary and interdisciplinary team. To tackle the questions of this case study, our team was comprised of the following backgrounds/expertise: (a) fatigue and human performance modeling, (b) biomechanics and kinematics, (c) biostatistics, (d) industrial statistics, (e) safety engineering, and (f) data analytics. The team worked together to understand how to best tackle each stage of the analysis. From a research perspective, those different backgrounds can also lead to competing goals.

7.2. Limitations and future work

As we discussed in Section 3.1, the data analyzed in this case stems from Baghdadi et al. (2018), which is part of a larger study examined in Maman et al. (2017, n.d.). One limitation of this analysis is the somewhat small sample size associated with the work. It is important to mention that the cost of the experimental study exceeded $22,000 (advertising, participant incentives, material and equipment, and the 6-month stipend for the graduate student). The reader is referred to the referenced articles for a detailed explanation of the factors that influenced the sample size (and how a small number of participants is not uncommon in ergonomic studies).

In our analysis, we considered a retrospective analysis of a trivariate series of observations from wearable sensors. The observation frequency and the nature of the process rendered the data autocorrelated and cross-correlated. For our application, we could not transform the data to make successive observations independent because we needed to retain the patterns to detect the desired changes in the process. We chose the *e-Agglo* approach for change point analysis because it resulted in reasonable results despite the fact that our data violated the assumption of independence of the observations within the clusters.

Our study and analysis reveals a gap in the literature for multivariate change point methods that can retrospectively detect multiple, unknown change points, are free from distributional assumptions, and are computationally efficient. We realize this is a tall order. We hope that, by publishing this case study, code, and data, we encourage researchers to investigate this problem and produce practical methodologies that can be used in this and other scenarios like this.

We also realize the limitation of our work in terms of signal diagnosis. We have simply detected the change points, but more work needs to be done in terms of understanding what has happened in terms of automating methods for diagnosing what aspect of the kinematic process changed and how to work with the participants to extend the period prior to the onset of fatigue. The change point method of Capizzi and Masarotto (2017) combines detection with a LASSO-based diagnosis method, but did not work in our study (likely because we could not meet the assumption of independence). While there has been some work in the joint detection and diagnosis of changes, much of this work has been done in advanced manufacturing, not in human performance modeling applications.

In future experiments, it would be helpful to apply designed experiments in order to determine effects due to gender, age, and other variables. One possible design would be a factorial design where n subjects in each gender/age category are assigned to perform a task similar to the one described here. With sufficient subjects, it would then be possible to see whether the manifestation of fatigue is different across these various categories.

As we continue to see humans monitored in the workplace using sensor technology, we expect an increased need for monitoring and control systems that can be used to automatically detect and diagnose worker fatigue and safety concerns.

About the authors

Amir Baghdadi received the B.Sc. degree from the Sharif University of Technology, Tehran, Iran, in 2014, and the M.Sc. and Ph.D. degrees in mechanical engineering from the University at Buffalo, State University of New York, NY, USA, in 2017 and 2019, respectively. He currently holds a postdoctoral associate position at Project neuroArm at the Department of Clinical Neurosciences, University of Calgary. Amir's work is toward the development of novel navigation algorithms utilizing machine learning for the neuroArmPLUS robotic system for neurosurgery. Amir was an Adjunct Faculty and an Instructor at The State University of New York College, Buffalo, and a Teaching Fellow with the Department of Mechanical and Aerospace Engineering, University at Buffalo. He was also a Research Affiliate with Roswell Park Cancer Institute, Buffalo, NY, USA, where he has utilized computer vision methods to model automated assessment of surgical performance during robot-assisted surgery; and developed an automated tool for benign and malignant kidney tumor classification from computed tomography (CT) images using convolutional neural network (CNN). His research interests include the intersection of health informatics, data science, sensor fusion, computer vision, and biomechanics. Email: amirbagh@buffalo.edu

Lora A. Cavuoto received her M.S. and Ph.D. degrees in Industrial and Systems Engineering from Virginia Tech, Blacksburg, VA, USA, in 2009 and 2012, respectively. She received her M.S. degree in occupational ergonomics and safety from the University of Miami, Coral Gables, FL, USA, in 2008. She is currently an Associate Professor in the Department of Industrial and Systems Engineering, where she directs the Ergonomics and Biomechanics Laboratory. Her current research interests include quantifying physical exposures and physiological responses in the workplace to identify indicators of fatigue development. Her research work also aims to understand and model the effects of health conditions, particularly obesity and aging, on physical capacity, specifically strength, fatigue, and motor performance. She is an Associate Editor for *Human Factors and Ergonomics in Manufacturing & Service Industries* and a Scientific Editor for *Applied Ergonomics*. Email: loracavu@buffalo.edu.

L. Allison Jones-Farmer is the Van Andel Professor of Business Analytics at Miami University in Oxford, Ohio. Her research focuses on developing practical methods for analyzing data in industrial and business settings. She is on the editorial review board of *Journal of Quality Technology* and *Quality Engineering* and is the current Editor for the Case Study section of JQT. Her email address is farmerl2@miamioh.edu.

Steven E. Rigdon is a Professor of Biostatistics in the College for Public Health and Social Justice at Saint Louis University. He is also Distinguished Research Professor Emeritus at Southern Illinois University Edwardsville. In 2017 he was elected Fellow of the American Statistical Association. From 1996 to 2006 he served on the editorial board of the *Journal of Quality Technology*. He was a guest editor for *Quality Engineering* for a special issue containing papers from the 2018 Hunter Conference. Currently, he is the editor of the *Journal of Quantitative Analysis in Sports*. Email: srigdon@slu.edu.

Ehsan T. Esfahani received the M.S. degree in electrical engineering and the Ph.D. degree in mechanical engineering from the University of California Riverside, Riverside, CA, USA, in 2012. He is currently an Associate Professor with the Department of Mechanical and Aerospace Engineering, University at Buffalo-SUNY, Buffalo, NY, USA. His main research interests include human in the loop systems, human-robot interactions, human activity monitoring, and biorobotics. Email: ehsanesf@buffalo.edu.

Fadel M. Megahed received his M.S. and Ph.D. degrees in Industrial and Systems Engineering from Virginia Tech, Blacksburg, VA, USA, in 2009 and 2012, respectively. He is currently the Neil R. Anderson Assistant Professor in the Farmer School of Business at Miami University located in Oxford, OH. His current research focuses on creating new tools to store, organize, analyze, model, and visualize the large heterogeneous data sets associated with modern manufacturing, healthcare and service environments. He serves on the Editorial Board for the *Journal of Quality Technology* and the *Journal of Financial Economic Policy*. He is a member of ASQ. Email: fmegahed@miamioh.edu.

Supplemental Materials (Data, Code and R Markdown)

To facilitate the replication of our work and encourage future work in this area, we provide the raw data and code in the following GitHub repository: https://github.com/fmegahed/fatigue-changepoint. The repository is divided into three main folders: (a) feature-engineering, which primarily consists of the MATLAB code that we have to perform the analysis of Section 3; (b) fatigue-changepoint, which contains the data and R code used for analysis in Sections 4–6; and (c) fmegahed.github.io, where we host the HTML generated by the R Markdown documenting the code and results obtained from Sections 4–6. In addition, we have created an HTML file based on our R Markdown, which we make available at https://fmegahed.github.io/fatigue_case_jqt.html. The HTML combines our code, analysis, and results. Thus, we consider it as an important part of our work that should be examined by the readers of this case study.

ORCID

Fadel M. Megahed http://orcid.org/0000-0003-2194-5110

References

Aghabozorgi, S., A. S. Shirkhorshidi, and T. Y. Wah. 2015. Time-series clustering – A decade review. *Information Systems* 53:16–38. doi: 10.1016/j.is.2015.04.007.

Baghdadi, A., L. A. Cavuoto, and J. L. Crassidis. 2018. Hip and trunk kinematics estimation in gait through Kalman filter using IMU data at the ankle. *IEEE Sensors Journal* 18 (10):4253–60. doi: 10.1109/JSEN.2018.2817228.

Baghdadi, A., F. M. Megahed, E. T. Esfahani, and L. A. Cavuoto. 2018. A machine learning approach to detect changes in gait parameters following a fatiguing occupational task. *Ergonomics* 61 (8):1116–29. doi: 10.1080/00140139.2018.1442936.

Barker, L. M., and M. A. Nussbaum. 2011. The effects of fatigue on performance in simulated nursing work. *Ergonomics* 54 (9):815–29. doi: 10.1080/00140139.2011.597878.

Bastos, L., P. Tomé, T. Cunha, M. Fernandes, and S. Cunha. 2001. Gravity anomalies from airborne measurements – Experiments using a low cost IMU device. In *Gravity, geoid and geodynamics 2000*, ed. M. G. Sideris, 253–58. Berlin, Heidelberg: Springer.

Bisgaard, S. 2012. The future of quality technology: From a manufacturing to a knowledge economy & from defects to innovations. *Quality Engineering* 24 (1):30–6. doi: 10.1080/08982112.2011.627010.

Bláfoss, R., J. K. Micheletti, E. Sundstrup, M. D. Jakobsen, H. Bay, and L. L. Andersen. 2019. Is fatigue after work a barrier for leisure-time physical activity? Cross-sectional study among 10,000 adults from the general working population. *Scandinavian Journal of Public Health* 47 (3):383–91. doi: 10.1177/1403494818765894.

Borg, G. A. 1982. Psychophysical bases of perceived exertion. *Medicine and Science in Sports and Exercise* 14 (5):377–81.

Box, G. E., and W. H. Woodall. 2012. Innovation, quality engineering, and statistics. *Quality Engineering* 24 (1):20–9. doi: 10.1080/08982112.2012.627003.

Brocal, F., and M. A. Sebastián. 2015. Identification and analysis of advanced manufacturing processes susceptible of generating new and emerging occupational risks. *Procedia Engineering* 132:887–94. doi: 10.1016/j.proeng.2015.12.574.

Bui, A. T., and D. W. Apley. 2018. Monitoring for changes in the nature of stochastic textured surfaces. *Journal of Quality Technology* 50 (4):363–78. doi: 10.1080/00224065.2018.1507559.

Cabrieto, J., F. Tuerlinckx, P. Kuppens, M. Grassmann, and E. Ceulemans. 2017. Detecting correlation changes in multivariate time series: A comparison of four non-parametric change point detection methods. *Behavior Research Methods* 49 (3):988–1005. doi: 10.3758/s13428-016-0754-9.

Capizzi, G., and G. Masarotto. 2017. Phase I distribution-free analysis of multivariate data. *Technometrics* 59 (4): 484–95. doi: 10.1080/00401706.2016.1272494.

Cavuoto, L., and F. Megahed. 2017. Understanding fatigue: Implications for worker safety. *Professional Safety* 62 (12): 16–9.

Chang, L., J. Li, and S. Chen. 2015. Initial alignment by attitude estimation for strapdown inertial navigation systems. *IEEE Transactions on Instrumentation and Measurement* 64 (3):784–94.

Davidson, B. S., M. L. Madigan, and M. A. Nussbaum. 2004. Effects of lumbar extensor fatigue and fatigue rate on postural sway. *European Journal of Applied Physiology* 93 (1–2):183–9. doi: 10.1007/s00421-004-1195-1.

De Oca, V. M., D. R. Jeske, Q. Zhang, C. Rendon, and M. Marvasti. 2010. A CUSUM change-point detection algorithm for non-stationary sequences with application to data network surveillance. *Journal of Systems and Software* 83 (7):1288–97. doi: 10.1016/j.jss.2010.02.006.

Dempsey, P., L. Kocher, M. Nasarwanji, J. Pollard, and A. Whitson. 2018. Emerging ergonomics issues and opportunities in mining. *International Journal of Environmental Research and Public Health* 15 (11):2449–60. doi: 10.3390/ijerph15112449.

Dougherty, E. R., and J. Astola. 1994. *An introduction to nonlinear image processing*. Vol. 16. Bellingham, WA: SPIE Press.

Evengård, B. 2008. *Fatigue: Epidemiology and social/industrial aspects*, 17–31. Tokyo: Springer Japan.

Evstigneeva, M., A. Aleksandrov, S. E. Mathiassen, and E. Lyskov. 2012. Concurrent cognitive task may improve motor work performance and reduce muscle fatigue. *Work* 41 (Supplement 1):2893–6.

Fiveash, C. 2016. Warehouse automation: The next generation. Accessed November 24, 2018. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/pcast-advanced-manufacturing-june2011.pdf.

Gallagher, N., and G. Wise. 1981. A theoretical analysis of the properties of median filters. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 29 (6):1136–41. doi: 10.1109/TASSP.1981.1163708.

Garde, A. H., Å. M. Hansen, and B. R. Jensen. 2003. Physiological responses to four hours of low-level repetitive work. *Scandinavian Journal of Work, Environment & Health* 29 (6):452–60. doi: 10.5271/sjweh.753.

Ghobadi, M., and E. T. Esfahani. 2017. A robust automatic gait monitoring approach using a single IMU for home-based applications. *Journal of Mechanics in Medicine and Biology* 17 (05):1750077. doi: 10.1142/S0219519417500774.

Giorgino, T. 2009. Computing and visualizing dynamic time warping alignments in r: The DTW package. *Journal of Statistical Software* 31 (7):1–24. doi: 10.18637/jss.v031.i07.

Gust, P., U. Müller, N. Feller, and M. Schiffmann. 2017. *Field study on the application of a simulation-based software tool for the strain-based staffing in industrial manufacturing*, 3–12. Cham: Springer International Publishing.

Hernandes, V. 2016. Study: Waiters take average of 23,000 steps a day. Accessed November 24, 2018. http://www.ibtimes.com.au/study-waiters-take-average-23000-steps-day-1295370.

Höhle, M. 2007. Surveillance: An R package for the monitoring of infectious diseases. *Computational Statistics* 22 (4):571–82. doi: 10.1007/s00180-007-0074-8.

James, N. A., and D. S. Matteson. 2013. ECP: An R package for nonparametric multiple change point analysis of multivariate data. arXiv preprint arXiv:1309.3295.

Janssen, N., I. Kant, G. Swaen, P. Janssen, and C. Schröer. 2003. Fatigue as a predictor of sickness absence: Results from the Maastricht cohort study on fatigue at work. *Occupational and Environmental Medicine* 60 (>90001): 71i–i76. doi: 10.1136/oem.60.suppl_1.i71.

Jensen, W. A., D. C. Montgomery, F. Tsung, and G. G. Vining. 2018. 50 years of the Journal of Quality Technology. *Journal of Quality Technology* 50 (1):2–16. doi: 10.1080/00224065.2018.1404881.

Jimenez, A. R., F. Seco, C. Prieto, and J. Guevara. 2009. A comparison of pedestrian dead-reckoning algorithms using a low-cost MEMS IMU. Paper presented at the IEEE Intelligent Signal Processing, 2009. WISP 2009. IEEE International Symposium on, Budapest, Hungary, August 26–28, 2009, 37–42.

Kajimoto, O. 2008. *Development of a method of evaluation of fatigue and its economic impacts*, 33–46. Tokyo: Springer Japan.

King, C. B., Y. Hong, S. P. Dehart, P. A. Defeo, and R. Pan. 2016. Planning fatigue tests for polymer composites. *Journal of Quality Technology* 48 (3):227–45. doi: 10.1080/00224065.2016.11918163.

Knoth, S. 2018. Package 'SPC.' Accessed December 20, 2018. https://cran.r-project.org/web/packages/spc/spc.pdf.

Kolus, A., R. Wells, and P. Neumann. 2018. Production quality and human factors engineering: A systematic review and theoretical framework. *Applied Ergonomics* 73: 55–89. doi: 10.1016/j.apergo.2018.05.010.

Loriol, M. 2017. A sociological stance on fatigue and tiredness: Social inequalities, norms and representations. *Neurophysiologie Clinique/Clinical Neurophysiology* 47 (2): 87–94. doi: 10.1016/j.neucli.2016.12.001.

Luinge, H. J., and P. H. Veltink. 2005. Measuring orientation of human body segments using miniature gyroscopes and accelerometers. *Medical & Biological Engineering & Computing* 43 (2):273–82. doi: 10.1007/BF02345966.

Lu, C. J., and W. Q. Meeker. 1993. Using degradation measures to estimate a time-to-failure distribution. *Technometrics* 35 (2):161–74. doi: 10.2307/1269661.

Lu, L., F. M. Megahed, R. F. Sesek, and L. A. Cavuoto. 2017. A survey of the prevalence of fatigue, its precursors and individual coping mechanisms among US manufacturing workers. *Applied Ergonomics* 65:139–51. doi: 10.1016/j.apergo.2017.06.004.

Macintosh, B. R., K. Svedahl, and M. Kim. 2004. Fatigue and optimal conditions for short-term work capacity. *European Journal of Applied Physiology* 92 (4–5):369–75. doi: 10.1007/s00421-004-1177-3.

MacLeod, D. 1994. *The ergonomics edge: Improving safety, quality, and productivity*. New York, NY: John Wiley & Sons.

Maman, Z. S., Y.-J. Chen, A. Baghdadi, S. Lombardo, L. A. Cavuoto, and F. Megahed. n.d. A data analytic framework for physical fatigue management using wearable sensors. *International Journal of Production Economics*. Submitted for publication.

Maman, Z. S., M. A. A. Yazdi, L. A. Cavuoto, and F. M. Megahed. 2017. A data-driven approach to modeling physical fatigue in the workplace using wearable sensors. *Applied Ergonomics* 65:515–29. doi: 10.1016/j.apergo.2017.02.001.

Matteson, D. S., and N. A. James. 2014. A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association* 109 (505):334–45. doi: 10.1080/01621459.2013.849605.

Merletti, R., and P. Di Torino. 1999. Standards for reporting EMG data. *Journal of Electromyography and Kinesiology* 9 (1):3–4.

Nair, V. 2007. Industrial statistics: The gap between research and practice. Youden Address. Fall Technical Conference, Jacksonville, FL. Accessed December 20, 2018. http://207.67.83.164/statistics/2010/02/statistical-process-control-(spc)/2007-youden-address.pdf.

Nosek, B. A., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen et al. 2015. Promoting an open research culture. *Science* 348 (6242):1422–5. doi: 10.1126/science.aab2374.

Öberg, T., A. Karsznia, and K. Öberg. 1993. Basic gait parameters: Reference data for normal subjects, 10–79 years of age. *Journal of Rehabilitation Research and Development* 30 (2):210–23.

Pantelopoulos, A., and N. Bourbakis. 2008. A formal language approach for multi-sensor wearable health-monitoring systems. Paper presented at the 2008 8th IEEE International Conference on BioInformatics and BioEngineering, 1–7, Athens, Greece, December 8.

Pantelopoulos, A., and N. Bourbakis. 2009. A health prognosis wearable system with learning capabilities using NNS. Paper presented at the IEEE Tools with Artificial Intelligence, 2009. ICTAI '09. 21st International Conference, 243–7.

Pantelopoulos, A., and N. Bourbakis. 2010. Design of the new prognosis wearable system-prototype for health monitoring of people at risk. In *Advances in biomedical sensing, measurements, instrumentation and systems*, 29–42. Berlin, Heidelberg, Germany: Springer.

Pitas, I., and A. N. Venetsanopoulos. 2013. *Nonlinear digital filters: Principles and applications*. Vol. 84. New York, NY: Springer Science & Business Media.

Rashedi, E., S. Kim, M. A. Nussbaum, and M. J. Agnew. 2014. Ergonomic evaluation of a wearable assistive device for overhead work. *Ergonomics* 57 (12):1864–74. doi: 10.1080/00140139.2014.952682.

Rebula, J. R., L. V. Ojeda, P. G. Adamczyk, and A. D. Kuo. 2013. Measurement of foot placement and its variability with inertial sensors. *Gait & Posture* 38 (4):974–80. doi: 10.1016/j.gaitpost.2013.05.012.

Ricci, J. A., E. Chee, A. L. Lorandeau, and J. Berger. 2007. Fatigue in the US workforce: Prevalence and implications for lost productive work time. *Journal of Occupational and Environmental Medicine* 49 (1):1–10. doi: 10.1097/01.jom.0000249782.60321.2a.

Romero, D., P. Bernus, O. Noran, J. Stahre, Å. Fast-Berglund. 2016. The operator 4.0: Human cyber-physical systems & adaptive automation towards human–automation symbiosis work systems. In *APMS* (*Advances in production management systems*), ed. I. Nääs, 677–686. Cham: Springer.

Sakoe, H., and S. Chiba. 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26 (1):43–9. doi: 10.1109/TASSP.1978.1163055.

Salvendy, G. 2012. *Handbook of human factors and ergonomics*. New Jersey: John Wiley & Sons.

Sarda-Espinosa, A. 2018. Package 'dtwclust.' Accessed December 20, 2018. http://cran.ma.imperial.ac.uk/web/packages/dtwclust/dtwclust.pdf.

Schall Jr., M. C., R. F. Sesek, and L. A. Cavuoto. 2018. Barriers to the adoption of wearable sensors in the workplace: A survey of occupational safety and health professionals. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 60 (3):351–62. doi: 10.1177/0018720817753907.

Schneider, E., S. Copsey, and X. Irastorza. 2010. OSH [Occupational Safety and Health] in figures: Work-related musculoskeletal disorders in the EU – Facts and figures. Office for Official Publications of the European Communities.

Skog, I., P. Handel, J.-O. Nilsson, and J. Rantakokko. 2010. Zero-velocity detection – An algorithm evaluation. *IEEE Transactions on Biomedical Engineering* 57 (11):2657–66. doi: 10.1109/TBME.2010.2060723.

Sluiter, J., E. De Croon, T. Meijman, and M. Frings-Dresen. 2003. Need for recovery from work related fatigue and its role in the development and prediction of subjective health complaints. *Occupational and Environmental Medicine* 60 (Suppl. 1):i62–i70. doi: 10.1136/oem.60.suppl_1.i62.

Swaen, G., L. Van Amelsvoort, U. Bültmann, and I. Kant. 2003. Fatigue as a risk factor for being injured in an occupational accident: Results from the Maastricht cohort study. *Occupational and Environmental Medicine* 60 (Suppl. 1):i88–i92. doi: 10.1136/oem.60.suppl_1.i88.

Szekely, G. J., and M. L. Rizzo. 2005. Hierarchical clustering via joint between–within distances: Extending Ward's minimum variance method. *Journal of Classification* 22 (2):151–83. doi: 10.1007/s00357-005-0012-9.

Tanoue, H., T. Mitsuhashi, S. Sako, R. Goto, T. Nakai, and R. Inaba. 2016. Effects of a dynamic chair on pelvic mobility, fatigue, and work efficiency during work performed while sitting: A comparison of dynamic sitting and static sitting. *Journal of Physical Therapy Science* 28 (6):1759–63. doi: 10.1589/jpts.28.1759.

Tao, W., T. Liu, R. Zheng, and H. Feng. 2012. Gait analysis using wearable sensors. *Sensors* 12 (2):2255–83. doi: 10.3390/s120202255.

The (U.S.) President's Council of Advisors on Science and Technology. 2011. Report to the President on ensuring American leadership in advanced manufacturing. President's Council of Advisors on Science and Technology, Executive Office of the President of the United States. Accessed November 23, 2018. https://

obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/pcast-advanced-manufacturing-june2011.pdf.

Thomas, S., J. Reading, and R. J. Shephard. 1992. Revision of the physical activity readiness questionnaire (par-q). *Canadian Journal of Sport Sciences* 17 (4):338–45.

Tongen, A., and R. E. Wunderlich. 2010. Biomechanics of running and walking. In *Mathematics and Sports: Dolciani Mathematical Expositions #43*, ed. J. A. Gallian, 315-328. Washington DC: Mathematical Association of America.

Tsao, L., L. Ma, and C.-T. Papp. 2018. Using non-invasive wearable sensors to estimate perceived fatigue level in manual material handling tasks. Paper presented at the International Conference on Applied Human Factors and Ergonomics, 65–74, Springer.

Tseng, S.-T., M. Hamada, and C.-H. Chiao. 1995. Using degradation data to improve fluorescent lamp reliability. *Journal of Quality Technology* 27 (4):363–9. doi: 10.1080/00224065.1995.11979618.

Tukey, J. 1974. Nonlinear (nonsuperposable) methods for smoothing data. Paper presented at the Proceedings of Congress Record EASCON, Washington, DC, October 7–9, 1974, 673–681.

Ward, Jr., J. H. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58 (301):236–44. doi: 10.2307/2282967.

Woodall, W. H. 2017. Bridging the gap between theory and practice in basic statistical process monitoring. *Quality Engineering* 29 (1):2–15. doi: 10.1080/08982112.2016.1210449.

Woodall, W. H., S. L. Fogel, and S. H. Steiner. 2015. The monitoring and improvement of surgical-outcome quality. *Journal of Quality Technology* 47 (4):383–99. doi: 10.1080/00224065.2015.11918141.

Woodall, W. H., and D. C. Montgomery. 2014. Some current directions in the theory and application of statistical process monitoring. *Journal of Quality Technology* 46 (1):78–94. doi: 10.1080/00224065.2014.11917955.

Yang, S., and Q. Li. 2012. Inertial sensor-based methods in walking speed estimation: A systematic review. *Sensors* 12 (5):6102–16. doi: 10.3390/s120506102.

Yeow, P. H., and R. N. Sen. 2003. Quality, productivity, occupational health and safety and cost effectiveness of ergonomic improvements in the test workstations of an electronic factory. *International Journal of Industrial Ergonomics* 32 (3):147–63. doi: 10.1016/S0169-8141(03)00051-9.

Yung, M. 2016. Fatigue at the workplace: Measurement and temporal development. PhD. diss., Department of Kinesiology, University of Waterloo. Accessed November 23, 2018. http://hdl.handle.net/10012/10119.

Zhang, M., L. Murphy, D. Fang, and A. Caban-Martinez. 2015. Influence of fatigue on construction workers' physical and cognitive function. *Occupational Medicine* 65 (3):245–50. doi: 10.1093/occmed/kqu215.

Zhao, J., and E. Obonyo. 2018. Towards a data-driven approach to injury prevention in construction. Paper presented at the Workshop of the European Group for Intelligent Computing in Engineering, 385–411, Springer.