Version of Record: https://www.sciencedirect.com/science/article/pii/S0003687020302106
Manuscript_467e36a59a2eec0fd4b0cf01d51f3548

A Forecasting Framework for Predicting Perceived Fatigue: Using Time Series Methods to Forecast Ratings of Perceived Exertion with Features from Wearable Sensors

Sahand Hajifar[a], Hongyue Sun[b], Fadel M. Megahed[c], L. Allison Jones-Farmer[d], Ehsan Rashedi[e], Lora A. Cavuoto[f]

[a]*Department of Industrial and Systems Engineering, University at Buffalo, Buffalo, NY 14260, USA | Email: sahandha@buffalo.edu*
[b]*Department of Industrial and Systems Engineering, University at Buffalo, Buffalo, NY 14260, USA | Email: hongyues@buffalo.edu*
[c]*Farmer School of Business, Miami University, Oxford, OH 45056, USA | Email: fmegahed@miamioh.edu*
[d]*Farmer School of Business, Miami University, Oxford, OH 45056, USA | Email: farmerl2@miamioh.edu*
[e]*Department of Industrial and Systems Engineering, Rochester Institute of Technology, Rochester, NY 14623, USA | Email: exreie@rit.edu*
[f]*Corresponding author. Department of Industrial and Systems Engineering, University at Buffalo, Buffalo, NY 14260, USA | Email: loracavu@buffalo.edu*

Abstract

Advancements in sensing and network technologies have increased the amount of data being collected to monitor the worker conditions. In this study, we consider the use of time series methods to forecast physical fatigue using subjective ratings of perceived exertion (RPE) and gait data from wearable sensors captured during a simulated in-lab manual material handling task (Lab Study 1) and a fatiguing squatting with intermittent walking cycle (Lab Study 2). To determine whether time series models can accurately forecast individual response and for how many time periods ahead, five models were compared: naïve method, autoregression (AR), autoregressive integrated moving average (ARIMA), vector autoregression (VAR), and the vector error correction model (VECM). For forecasts of three or more time periods ahead, the VECM model that incorporates historical RPE and wearable sensor data outperformed the other models with median mean absolute error (MAE) < 1.24 and median MAE < 1.22 across all participants for Lab Study 1 and Lab Study 2, respectively. These results suggest that wearable sensor data can support forecasting a worker's condition and the forecasts obtained are as good as current state-of-the-art models using multiple sensors for current time prediction.

Keywords: Fatigue Forecast, Material Handling, Vector Error Correction Model, Wearable Sensors

1. Introduction

1.1. A Perspective on Modeling Paradigms in Ergonomics

Over the past decade, advancements in sensing, computation, and network technologies have increased the amount of data being collected in the workplace. From an ergonomics perspective, our interest typically lies in novel applications of those technologies, where human performance data is captured for analysis and decision-making [1]. For example, there are an increasing number of articles in the ergonomics literature that examine the utility of wearable sensors [2, 3, 4, 5], machine vision [6], and exoskeletons [7] in assessing certain characteristics of human/worker performance. The data from these approaches generally occur over time in high frequency multivariate streams. Processing and understanding these multivariate time series data requires advanced analytical techniques that are less commonly applied in the ergonomics literature. These emerging technologies combined with the appropriate modeling of the data allow the ergonomics field to move towards personalized assessment at the worker level, continuous monitoring and modeling of human performance, and proactive identification and elimination/control of risk factors.

The realization of these benefits requires the incorporation of different analytical modeling frameworks. In the context of emerging ergonomics applications, we highlight five relevant modeling paradigms:

(A) *Descriptive modeling:* This type of modeling aims to explore, summarize and visualize data. Statistical models (e.g., regression) may be used in this context, however, the focus tends to be on capturing the association between dependent and independent variables rather than making predictions [8].

(B) *Explanatory modeling:* The goals of explanatory modeling are centered around testing causal theories and causal inference [8]. Applications of explanatory modeling in occupational safety and ergonomics are typically focused on epidemiological studies [9, 10].

(C) *Cross-sectional predictive modeling:* In ergonomics research and practice (e.g., [2, 3]), this corresponds to the situations where statistical models and/or machine learning models are used to predict an output/dependent variable (y) based on K independent variables ($K \geq 1$) at a given snapshot in time.

(D) *Statistical surveillance:* Here, the goal is to monitor characteristic(s) of interest over time. Statistical surveillance methodologies can be divided into: (i) *Phase I* applications, where retrospective analysis is performed to establish a baseline sample, ensure its stability and identify a statistical distribution to characterize the observed variability in the baseline sample; and (ii) *Phase II* applications, where real-time monitoring is performed to detect changes in characteristic(s) of interest as quickly as possible while controlling the false signal probability (e.g., [5, 11]).

(E) *Forecasting:* Whereas in cross-sectional models the variables are measured at the same point in time t, in forecasting models, observations until time t (of the input(s), which can combine both independent and dependent variables) are used to forecast future values of the dependent output at time $t+k$, $k>0$ (y_{t+k}). From an ergonomics perspective, forecasting models are useful in predicting a future state of the worker, which can be useful for scheduling interventions and/or estimating their efficacy.

In this article, we will focus on forecasting. In the context of existing ergonomics research and practice, the modeling frameworks used are typically descriptive, explanatory, predictive, and statistical surveillance. The forecasting framework, which is the most proactive and perhaps most suited for injury prevention, is seldom explored in the literature. One example is in the area of sleep- and shift-induced fatigue and the impact on cognitive function. The Fatigue Avoidance Scheduling Tool (FAST™) is a fatigue forecasting software in which work schedules are integrated with an individual's sleep data and circadian rhythm (based on time of day and data collected from a smart wristband), to generate a forecast of cognitive effectiveness for a given time [12, 13]. To the best of our knowledge, similar forecasting studies/software have not been developed for physical fatigue forecasting. Therefore, in this study, we attempt to address this gap. One of our main goals is to demonstrate the utility of combining forecasting techniques with sensor data in determining when an ergonomics intervention may be needed.

1.2. Problem Description and Motivation

In a recent survey of U.S. advanced manufacturing workers [14], approximately 58% of respondents reported being fatigued over the previous week of the survey. These high fatigue rates are not unique to manufacturing occupations. Zhang et al. [15] showed that 59% of construction workers experienced fatigue regularly over the past three months of the study. Physical fatigue at work can increase the risk of traumatic injury and lead to a decrease in work performance. For example, studies of firefighters have shown that physical fatigue leads to movement errors and increased risk of slips, trips, or falls [16, 17].

The effectiveness of work design hinges on the ability to predict fatigue accumulation and recovery as a function of time-on-task and load [18, 19]. Existing models for predicting fatigue and recovery in job design are primarily centered around models of muscle endurance [20, 21, 22]. These models may be suitable for upper extremity-focused manufacturing work, where many jobs contain work components that involve periods of sustained exertions, such as the holding of hand tools [23, 24]. However, similar conditions do not exist in more load-intensive and faster work that involves regular walking. For metabolically demanding tasks, technological progress has paved the way for exploiting sensors to monitor heart rate and estimate energy expenditure. Earlier work has focused on the use of photoplethysmography sensors, which allow for heart rate (HR) measurements that can be used in the estimation of energy expenditure [25, 26]. Physically-demanding work results in an elevated HR, and recovery periods decrease HR [27, 28]. However, the accuracy of HR monitoring is susceptible to sensing delay and noise [29]. More importantly, HR monitoring in an occupational setting is likely to be of concern to workers since it can violate workers' privacy and be used to infer personal medical information [30].

Alternative indicators of fatigue have been recently proposed to overcome the aforementioned limitations of using HR for occupational fatigue monitoring and modeling applications. For walking-intensive tasks, the following changes in gait parameters were observed with the onset of physical fatigue: (a) increased step width [31], (b) increased step width variability [32], (c) decreased step/stride length [31, 17, 32, 33], (d) increased step length variability [31], and (e) decreased stride height [16]. In addition to being indicators of fatigue, these changes in gait parameters are associated with higher risk of slips, trips, and falls [34, 35].

Proper monitoring and detection of physical fatigue is necessary to avoid short-term discomfort and motor control problems, as well as long-term health issues. Thus, an increasing number of studies have been devoted to prediction and classification of human fatigue. These studies often use statistical and data mining techniques based on the informative kinematics of different body locations from wearable sensors and/or subject information [2, 3, 31, 36, 37]. Recent studies have successfully incorporated gait characteristics for fatigue modeling. For example, Zhang et al. [37], Karvekar et al. [38] and Baghdadi et al. [3] have shown that the classification of fatigued and non-fatigued states can be performed with an accuracy of $\geq 90\%$. In a follow-up work, Baghdadi et al. [5] noted two limitations of existing classification approaches to fatigue modeling: (a) these models assume that a worker can only be in either a non-fatigued state or a fatigued state, which is a simplistic representation of the process of fatigue development, and (b) these models are somewhat reactive in practice since an intervention is only possible once a worker is deemed to be fatigued. In an effort to overcome these limitations, they used a multivariate change-point estimation (i.e. a *statistical surveillance*) methodology to jointly monitor stride length, stride height and stride duration [5]. The advantage of their approach is that it focuses on detecting changes from the non-fatigued baseline; practitioners can then be alerted at the earlier stages of fatigue development when a worker's gait parameters start to deviate from their in-control distributions. However, the approach of Baghdadi et al. [5] cannot inform practitioners when a worker will be fatigued in the future (i.e. their approach is not a *forecasting* method and thus, it is not optimal for assigning ergonomic interventions).

This paper introduces a methodology for forecasting worker physical fatigue levels in walking intensive occupational jobs. Examples of such jobs include in warehouses and distribution centers [39] and manufacturing, where Lu et al. [14] reported that workers walk on average 5.7 hours per day on the job. The overarching goal of this study is to develop an individualized method that can be used to forecast exertion and fatigue during physical work (as captured by the rating of perceived exertion (RPE) [40]). To achieve this goal, we have addressed three research questions: (1) Can an individual's RPE be accurately forecasted using classical time series methods? (2) Can (and by how much can) the forecasts of individual RPE be improved by including, not only the participant's prior RPE, but also relevant features from a single inertial measurement unit (IMU) sensor? and (3) How many time periods ahead can we reliably forecast an individual's RPE?

To address these questions, we propose an individualized framework for forecasting RPE, capitalizing on existing univariate and multivariate time-series prediction models. Specifically, we examine:

(A) *Univariate time-series models*. We examine three popular approaches: (a) a naïve method applied to each individual's RPE scores; (b) an autoregressive (AR) model on each individual's RPE scores; and (c) an AR model applied on the first differences of an individual's RPE scores (i.e., the ARIMA model), which relaxes the stationarity assumption made by the AR model. In these cases, we are using only past observations of the RPE for a given participant to predict future RPE values (i.e., we do not include any of the sensor information here).

(B) *Multivariate time-series models*. Similar to the univariate case, here, we consider two popular models: *Vector autoregression (VAR)* and *vector error correction model (VECM)*, which relaxes the stationarity assumptions made by the VAR. Both approaches can capitalize on a vector of historical observations combining both RPE with kinematic features extracted from a wearable sensor (in Lab Study 1 stride height, stride length, and stride duration; and in Lab Study 2 mean acceleration, maximum acceleration and stride duration) in forecasting future RPE values. Here, we consider our second research question to determine whether the VAR and VECM methods that include the kinematic features in the forecast outperform the univariate forecasts from the AR and ARIMA models.

We can examine the first and third research questions via any of the forecasting models introduced above.

The remainder of the paper is organized as follows. We present our proposed forecasting framework in Section 2. We present our results in Section 3 and discuss their relevance in Section 4. We then provide concluding remarks in Section 5. To facilitate reproducing our analysis and encourage future work, we provide links to a GitHub repository containing our data and code, as well as a web page detailing our analysis/results as *Supplemental Material* to this paper.

2. Methods

Our proposed framework for the application of forecasting techniques in ergonomic and occupational safety studies/applications is divided into four phases. We present the framework as applied to data collected from two lab studies that both used a single IMU sensor data to capture gait parameters for a fatiguing task, with RPE as the forecast outcome of interest. Phase 1 is divided into two main sub-phases: (a) designing the data collection procedure, where decisions are made about the experimental design and technology/sensors used to acquire and store the resulting data; and (b) successful implementation of the designed experiment. Phase 2 involves the preprocessing of the acquired data to obtain a clean, synced, and merged dataset that would be used for modeling. In Phase 3, univariate and/or multivariate time-series models are applied to the preprocessed dataset. Phase 4 involves the evaluation of forecasting results and the selection of the "best" model for deployment. We provide a visual summary of the framework in Figure 1. Furthermore, details on each of the four phases are presented in the following subsections.

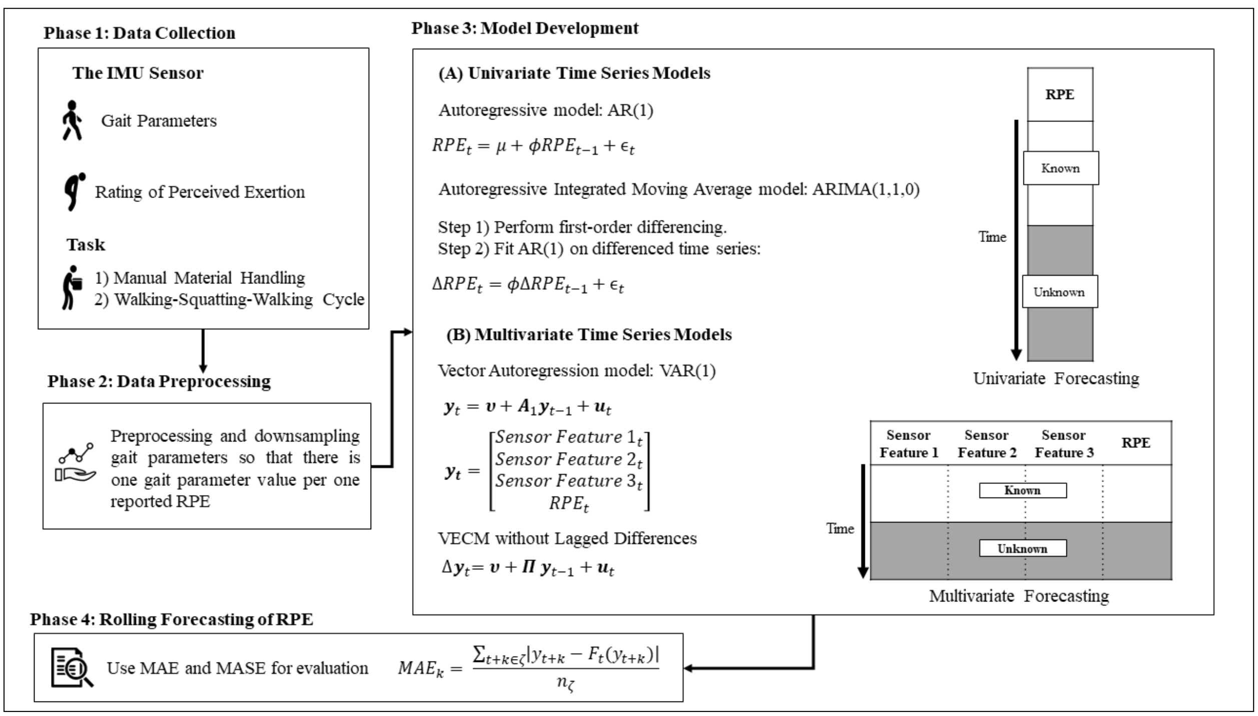

Figure 1: An overview of forecasting framework for acquiring, preprocessing, and analyzing longitudinal ergonomic data.

2.1. *Phase 1 – Data Collection Procedure*

In this study, a secondary analysis of the data collected from two in-lab studies is performed. Lab Study 1 is based on the manual material handling (MMH) study presented in [2, 3, 5], and Lab Study 2 is based on the fatigue inducing squat exercise, with intermittent gait periods, presented in Karvekar et al. [38]. For the reader's convenience, we summarize the relevant details of these studies below. The interested reader is referred to Baghdadi et al. [5] Karvekar et al. [38] for further details on the study design and settings.

2.1.1. Lab Study 1

Fifteen healthy participants (8 males, 7 females) completed a fatiguing three-hour manual materials handling task. During this time, they completed cycles of lifting an assigned carton, loading it on a two-wheeled dolly, pushing the dolly on an approximately 80-meter path, and then lifting the carton off of the dolly and placing it into a stack of cartons. After unloading the carton, this process was repeated for the next carton. Participants spent approximately 20-50 seconds unloading and loading the dolly per cycle, and approximately 1-1.5 minutes pushing the dolly along the 80-meter path. Participants were expected to complete 18 cycles per 30 minutes, for a total target of 108 cycles. The cartons had masses of 10, 18, or 26 kg (6 per mass), which were chosen based on the 50th, 75th, and 90th percentile values of typical lifting levels observed in industry [41]. The order of presentation of the carton masses varied within the three-hour period based on a task instruction sheet, but was kept consistent between participants. Every 10 minutes, when participants were at the loading/unloading location, they were asked to provide their overall rating of perceived exertion and fatigue using the Borg RPE scale. Throughout the session, participants were instrumented with an IMU attached with a strap around their right ankle that continuously collected motion data. The gait cycles were segmented from the data based on the procedure described in Baghdadi et al. [3] to obtain the features of stride length, stride height, and stride duration. Furthermore, per the explanation in Maman et al. [2] and Baghdadi et al. [3, 5], the data used for analysis ignored the first ten minutes of experimental data to allow participants to reach steady-state.

2.1.2. Lab Study 2

Twenty-four participants (12 male, 12 female) were instrumented with a smartphone-based IMU sensor strapped around their left shank. Upon calibrating the participants' perception of exertion using the wall test [38], they were asked to walk approximately 8 meters, followed by performing a set of 16 squats for 2 minutes at the rate of 8 squats/min. For the sake of consistency, the distance was marked on the laboratory's floor to instruct the participants properly. Immediately after the squatting, they reported the RPE scale rating, after which the same cycle was repeated. The IMU data collected at 100Hz during the second walking period was labeled according to the participant's Borg's RPE. This cycle of walking-squatting-walking continued until the participants' RPE was recorded to be ≥ 17, which indicates very hard exertion according to Borg's anchor.

2.2. *Phase 2 – Data Preprocessing*

To preprocess the IMU data for Lab Study 1, we followed the two-stage procedure of Baghdadi et al. [5]. In the first stage, the goal was to standardize the length of tri-variate time series capturing stride length, height and duration across all fifteen participants (i.e., to account for the difference in number of strides among the participants). The experimental data were rescaled from an experimental time scale to a percentile scale. Thus, 2000 strides were retained for each participant; starting at the 0.05% mark and ending with 100% at 0.05% increments [5]. The time domain of the RPE values was also shifted to the percentile scale. In the second stage, a median filter was applied to smooth the tri-variate stride data, using a moving window size (w) of 21 strides and an overlap of $w-1$ [5]. Based on these two pre-processing stages, the resulting (standardized and smoothed/clean) data could be used for forecasting.

For Lab Study 2, gait cycles were segmented based on the method described in Karvekar et al. [38]. From each gait cycle, a variety of features were selected, i.e., mean acceleration, maximum acceleration, and stride duration were extracted. Notably, the first and last steps of each walking period were removed to capture steady-state walking.

For both lab studies, the RPE values were collected at a lower frequency than the stride-related features, due to the self-reported nature of the metric. Thus, we downsampled the stride-related observations so that they aligned with the known RPE values. For Lab Study 1, since the RPE values were collected every 10 minutes, we have chosen to obtain the stride parameter values closest to the time at which the RPE data was collected. Similarly, in Lab Study 2, there was one RPE value reported for each period of squatting exercise. Therefore, we have chosen to extract the stride parameters values related to the last stride for each walking period. Note that we also examined the use of the second stride to evaluate whether the choice of an alternate stride has an impact on the forecasting performance. Our analysis showed that the results were similar. Hence, there is no evidence that the choice of stride selected for down-sampling would significantly impact the prediction results. For the sake of conciseness, we only show the results associated with the last stride following the preprocessing described above.

2.3. *Phase 3 – Application of Univariate and Multivariate Forecasting Models*

The field of *time series analysis* was popularized and unified by the 1970 seminal book of Box and Jenkins [42]. Despite the limited application of time series and forecasting methods in ergonomics, they are widely used in epidemiology, social sciences, engineering and athletics [43]. In the following subsections, we provide general introductions on formulations of five forecasting methods: three univariate (naïve, AR and ARIMA) and two multivariate (VAR and VECM) forecasting techniques. The notations of these methods could be similar (such as the intercept terms in Equations 5 and 7), but their estimated values will be different. Note that the differences among these methods are their underlying assumptions, which form the basis needed for inference/interpretation.

2.3.1. Univariate Forecasting Models

To address our first research question "Can an individual's RPE be accurately forecasted using classical time series methods?", we examine whether the RPE time series alone can be used for forecasting. Since we are only examining one time series (consisting of past values of a participant's RPE), this analysis can be performed using a suitable univariate time series model.

Naïve Forecast. Let us use $\mathrm{F}_t(y_{t+k})$ to denote the forecasted RPE at a future time period $t+k$ obtained at time t. The naïve forecast simply uses the last actual observation (y_t) as the k-step ahead forecast:

$$\mathrm{F}_t(y_{t+k}) = y_t. \tag{1}$$

Despite its simplicity, the naïve method can enable us to quantify the Forecast Value Added (FVA) and assess whether (and by how much) the effort of using advanced forecasting methods are justified [44]. This is discussed in more details in Section 2.4.

Autoregressive Model - AR(1). For the time being, we assume that: (a) the time series of RPE values is stationary to allow us to introduce a simple model to obtain the forecasted RPE value $\mathrm{F}_t(y_{t+k})$; and (b) $\mathrm{F}_t(y_{t+k})$ is regressed solely on its prior value. Note that we relax assumption (a) in our next model, ARIMA; however, we maintain assumption (b) in ARIMA modeling process. Since we have a small number of RPE values in both lab studies, we selected the lowest order (i.e., order 1) for our multivariate time series models and hence, we use an AR model of order $p = 1$ here for a fair comparison among the models. The AR(1) model can be written as:

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t, \tag{2}$$

where μ is a scalar intercept, ϕ is the model coefficient, and ε_t is the error term. Statistically speaking, we assume that $\varepsilon_t \sim N(0, \sigma^2)$. To evaluate the third research question: "How many time periods ahead can we reliably forecast an individual's RPE?", we can generalize the linear equation presented above as follows:

$$\begin{aligned}
\mathrm{F}_t(y_{t+1}) &= \hat{\mu} + \hat{\phi} y_t \\
\mathrm{F}_t(y_{t+2}) &= \hat{\mu} + \hat{\phi} \mathrm{F}_t(y_{t+1}) \\
&\vdots \\
\mathrm{F}_t(y_{t+k}) &= \hat{\mu} + \hat{\phi} \mathrm{F}_t(y_{t+k-1}),
\end{aligned} \tag{3}$$

where k represents the number of time periods ahead for the forecasting. In our solution, we have investigated $k = 1, 2, \ldots, 6$. We refer the reader to Brockwell and Davis [43] for more details on the AR(1) model, and our code (see the **R** Markdown document in *Supplemental Materials*) to reproduce our analysis.

Differenced first-order AR model - ARIMA(1,1,0). To relax the stationarity assumption of the AR(1) model, we apply the AR(1) model on the differenced RPE time series. In the time series analysis literature, the differenced first-order autoregressive model is denoted as ARIMA(1,1,0). Let $\Delta y_t = y_t - y_{t-1}$ represent the first-order differenced RPE values at time t. Here, we no longer include the scalar intercept in ARIMA(1,1,0) since the effect of considering an intercept is equivalent to adding a linear deterministic trend effect in the AR model [45]. Thus, the use of the ARIMA(1,1,0) to predict Δy_{t+k} at time t can be represented as:

$$\begin{aligned} \mathrm{F}_t(\Delta y_{t+1}) &= \hat{\phi}\Delta y_t \\ \mathrm{F}_t(\Delta y_{t+2}) &= \hat{\phi}\mathrm{F}_t(\Delta y_{t+1}) \\ &\vdots \\ \mathrm{F}_t(\Delta y_{t+k}) &= \hat{\phi}\mathrm{F}_t(\Delta y_{t+k-1}). \end{aligned} \tag{4}$$

The estimation/solution procedure described for the AR(1) model is applicable to the ARIMA(1,1,0) model.

From a forecasting perspective, both the AR(1) model and the ARIMA(1,1,0) model are special cases of the general class of *Auto Regressive Integrated Moving Average* models, denoted as *ARIMA*(p,d,q). Parameters p, d, and q are non-negative integers corresponding to the: number of time lags of the AR model, degree of differencing (i.e., the number of first order differences necessary to achieve stationarity), and the order of the *Moving Average* model [42]. Hence, the AR(1) model is also referred to as ARIMA(1,0,0). Note that in our analysis, we have only considered the aforementioned two special cases due to the relatively short length of our time series. In other ergonomic applications, the data may allow for examining more combinations of the *ARIMA*(p,d,q). For more introduction to model selection in the context of ARIMA models, the interested reader is referred to [46], which gives details on a useful **R** package, *forecast* [47]. For a more thorough introduction to ARIMA models, the reader is referred to Brockwell and Davis [43].

2.3.2. Multivariate Forecasting Models

To address our second research question " Can (and by how much can) the forecasts of individual RPE be improved by including, not only the participant's prior RPE, but also relevant features from a single IMU sensor?", we examine the use of two multivariate time series analysis approaches (VAR and VECM) for the purposes of using past observations of stride-related features and RPE to make predictions about $k-$step ahead values of the RPE. Critical to our presentation of both the VAR and VECM approach is the concept of cointegration. Cointegration describes the property of two or more time series that may be nonstationary on their own, but together, some linear combination of the multiple series forms a stationary series [48]. An intuitive explanation of this was given by Murray [49], where they compared the concept of cointegration to a drunk walking her dog. The drunk may wander aimlessly along in the short term, but the dog on a leash never strays too far from its owner over a long walk. The Engle-Granger and Johansen methods are the two most common tests of cointegration. The Engle-Granger test is recommended to be used in VAR systems with two variables, but the Johansen test is applicable for use with systems with two or more variables [50].

Vector Autoregression Model - VAR. To capture the dynamic interrelationships among gait variables and RPE, we will capitalize on the VAR modeling approach that was first proposed by Sims [51]. VAR has become one of the most popular multivariate time series analysis methods due to its simplicity and flexibility [52]. Specifically, the VAR modeling approach uses a linear model containing N equations with N variables where each variable is predicted by its own lagged values and previous values of the other $(N-1)$ variables [53]. Similar to the AR model, the VAR model assumes stationarity of the multiple time series. As mentioned before, since we have a small number of RPE values, we use a VAR model of order $p=1$ here. The VAR(1) model can be written as:

$$\boldsymbol{y}_t = \boldsymbol{\nu} + \boldsymbol{A}\boldsymbol{y}_{t-1} + \boldsymbol{u}_t, \tag{5}$$

where $\boldsymbol{y}_t = (y_{1t}, ..., y_{Nt})'$ is an $N \times 1$ random vector, $\boldsymbol{\nu}$ is an $N \times 1$ vector of intercepts, $\boldsymbol{A}$ is a $N \times N$ coefficient matrix, and $\boldsymbol{u}_t = (u_{1t}, ..., u_{Nt})'$ is an $N \times 1$ white noise process. In this study, the VAR model has four variables $(y_{1t}, y_{2t}, y_{3t}, y_{4t})$ which represent the three features used in each lab study and RPE, respectively. As with the AR model, the estimated VAR(1) equation can be used recursively to determine the k-step ahead forecast:

$$\begin{aligned}
\mathrm{F}_t(\boldsymbol{y}_{t+1}) &= \hat{\boldsymbol{\nu}} + \hat{\boldsymbol{A}}\boldsymbol{y}_t \\
\mathrm{F}_t(\boldsymbol{y}_{t+2}) &= \hat{\boldsymbol{\nu}} + \hat{\boldsymbol{A}}\mathrm{F}_t(\boldsymbol{y}_{t+1}) \\
&\vdots \\
\mathrm{F}_t(\boldsymbol{y}_{t+k}) &= \hat{\boldsymbol{\nu}} + \hat{\boldsymbol{A}}\mathrm{F}_t(\boldsymbol{y}_{t+k-1}).
\end{aligned} \tag{6}$$

The selection of a general VAR model includes two main aspects [54]: model form selection and lag selection. The model form selection procedure requires testing cointegration and stationarity of the multiple time series to determine the appropriate model form [54]. The procedure for lag selection is similar to that of a univariate ARIMA. The interested reader is referred to [55] for more detailed discussions.

Vector Error Correction Model - VECM. The VAR model is appropriate for stationary time series. If the Engle-Granger or Johansen test indicates the cointegration of the multivariate time series, then VECM can be used [56]. The VECM attempts to describe both the short- and long-term movements in a time series. The model is corrected based on prior forecast errors, and is informed by the movement of the other variables in the system. In a sense, the VECM takes advantage of the cointegrating relationships and forecast errors to improve the forecasts. We used the *VECM without lagged differences* [55] which can be written as follows:

$$\Delta \boldsymbol{y}_t = \boldsymbol{\nu} + \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{u}_t, \tag{7}$$

where $\boldsymbol{\nu}$ and $\boldsymbol{\Pi}$ are the $N \times 1$ intercept and the $N \times N$ coefficient matrix respectively. Note that Equation 7 with the intercept $\boldsymbol{\nu}$ corresponds to VAR(1) without drift (i.e., Equation 7's VAR representation is a VAR(1) model as shown in Equation 5) , see Section 6.4 of [55] for detailed derivations of the equations. As a result, the VECM modeling approach can be transformed into a VAR representation ($\boldsymbol{A} = \boldsymbol{\Pi} + \boldsymbol{I}_N$ relates the coefficient matrices in Equations 5 and 7, as shown in Section 6.4 of [55]). This transformation facilitates the forecasting of future observations with VECM, which can be computed in a similar manner to VAR models. Note that due to the limited RPE observations, we use VECM without lagged differences. However, in other ergonomic applications where there are enough observations, the VECM including lagged differences can be used, which corresponds to VAR(p) models with $p \geq 2$. The interested reader is referred to [55] for more details.

2.4. Phase 4 - Evaluation of the Forecasting Models

To systematically assess the forecasting performance of our models for any of the $k-$step ahead predictions, we compare the forecasted and true values of the RPE. In our analysis, we have used the mean absolute error (MAE) and mean absolute scaled error (MASE) metrics. The MAE is a commonly used and intuitive evaluation metric in both forecasting and predictive modeling (when the response is assumed to be continuous) applications. The scale of the MAE depends on the scale of the data itself, while the MASE is a scale-independent metric. For the applications to Lab Studies 1 and 2, the RPE values are forecasted, thus the MAE values are in units of the RPE scale. The MASE removes the scale of the data by "scaling the error based on the in-sample MAE from the naïve forecast method" [57]. This enables us to benchmark the models against the naïve method. In particular, a MASE value lower than one indicates that the method gives a smaller error than that of naïve forecast computed in-sample [58]. Using the notation from above, let y_{t+k} correspond to the actual value of the RPE at time $t+k$, and $\mathrm{F}_t(y_{t+k})$ correspond to the forecasted value of the RPE for time $t+k$ at time t. In such a case, the MAE and MASE in the case of k-steps ahead forecasting can be computed as follows:

$$
\begin{aligned}
MAE_k &= \frac{\sum_{t+k\in\zeta} |y_{t+k} - \mathrm{F}_t(y_{t+k})|}{n_\zeta}, \\
q_{t+k} &= \frac{\mathrm{F}_t(y_{t+k}) - y_{t+k}}{\frac{1}{n_\omega - k}\sum_{i=k+1}^{n_\omega} |y_i - y_{i-k}|}, \\
MASE_k &= \frac{\sum_{t+k\in\zeta} |q_{t+k}|}{n_\zeta},
\end{aligned}
\tag{8}
$$

where ζ is the set of the future observations that can be forecast using a specific method and $k-$steps ahead forecasting target, n_ζ represents the cardinality of ζ, q_{t+k} is a scaled error, ω is the set of in-sample data, and n_ω represents the cardinality of ω.

Let us consider the scenario where we start the forecasting procedure at $t = 8$ with the goal of $k = 2$ steps ahead forecasting, then the possible values of $\omega = \{1,2,..,8\}$ and $\zeta = \{10,11,...,T\}$. In Lab Study 1, the total number of time periods (T) is equal to 16 (since the RPE values were collected at the 6^{th}, 12^{th}, 18^{th}, $\ldots$, 96^{th} time percentile of the experiment). In Lab Study 2, the T ranges from 13 to 38 depending on the participant and we have used $k = 1, 2, \ldots, 6$ for both lab studies. Through the MAE evaluation of the results of the five forecasting approaches for these six step-ahead time periods, we can address the third research question "how many time periods ahead can we reliably forecast an individual's RPE?"

3. Results

3.1. Lab Study 1

We present the individualized forecasting results for the different participants in Figure 2. Figure 2 is interactive, with each panel showing the 1, 2,...6 step ahead forecasts for each participant. Readers viewing through Adobe Acrobat Reader can interact to see all participants. There are three important observations that need to be highlighted from the figure. First, if we compare the true (observed) RPE values across the different subjects (i.e., through the controls or the provided link), we observe that the variations in the RPE profiles are quite large. For example, Participant 1's RPE values start at 10 and peak at 14. On the other hand, Participant 5's RPE values start at 11 and peak at 17. Furthermore, the ranges of the reported RPE across all participants vary significantly (with minimum and maximum ranges of 4 and 12, respectively). Second, on average, the deviations between the predicted RPE values using any of the four forecasting approaches and the observed RPE increase as k increases. This is an expected result, and can be observed clearly by comparing the 1-step ahead and 6-step ahead results. Third, for $k \geq 3$, the AR(1) underestimates the RPE values for all values of t and for all participants with a monotonically increasing time series of true RPE (i.e., all except participants 4 and 9).

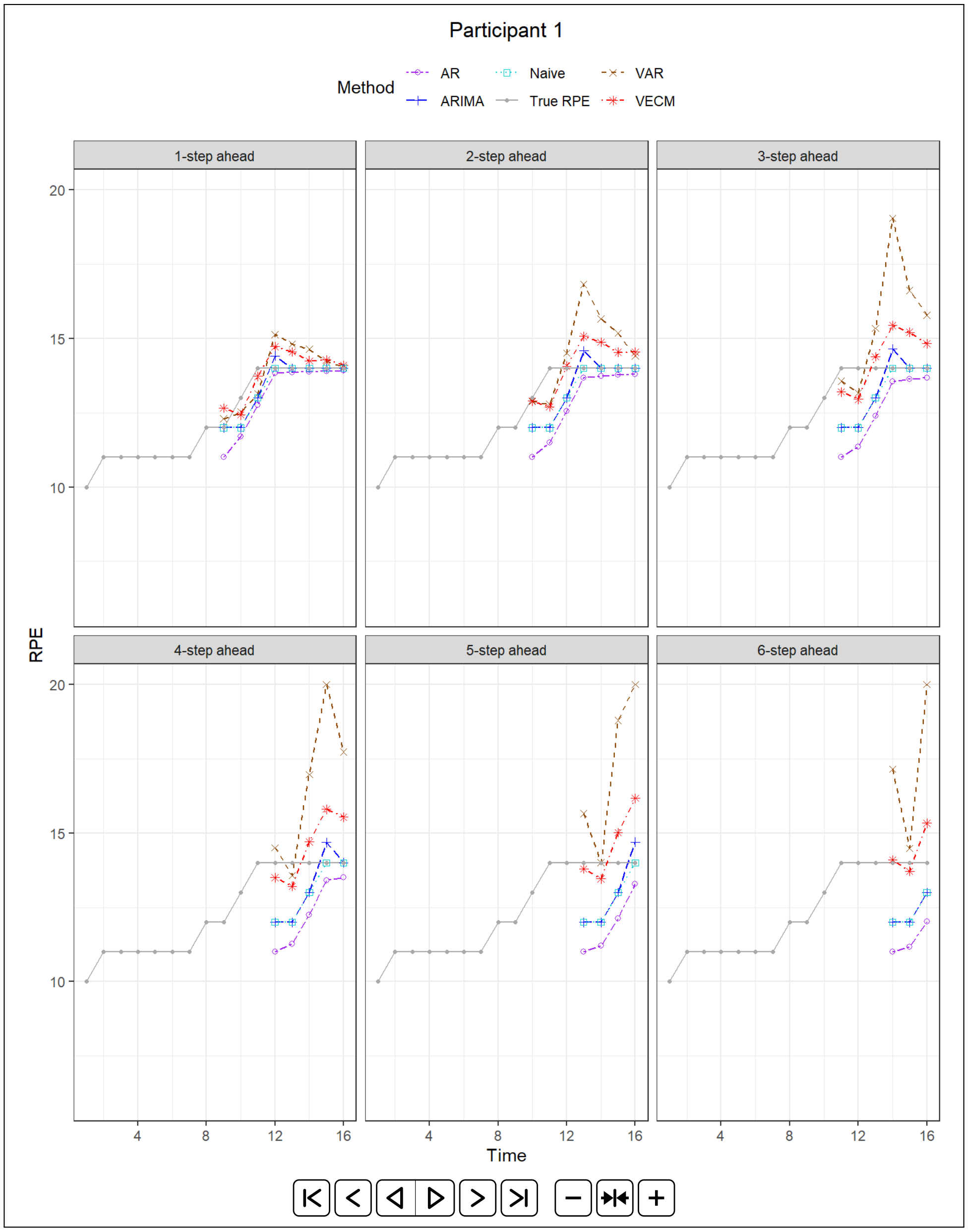

Figure 2: Forecasting results for different participants. The reader is encouraged to interact and engage with this visualization in Adobe Acrobat Reader (not through their web browser). Alternatively, the reader can access a web-based version of this visualization under Section 1.4 at: https://sahand-hajifar.github.io/RPE_Forecasting.html.

The MAE and median MASE results across participants are provided in Figure 3 and Table 1, respectively. Note that the presented results exclude those of participants 2 and 6 since we could not forecast all values of t for these participants using our five models. In particular, the AR(1) model could not forecast all possible values of t for Participant 2, since the autoregressive coefficients could not be reliably estimated due to nonstationarity. Participant 6 was removed from the analysis since its RPE for $t = 1, 2, \ldots, 8$ was invariant ($RPE = 11$).

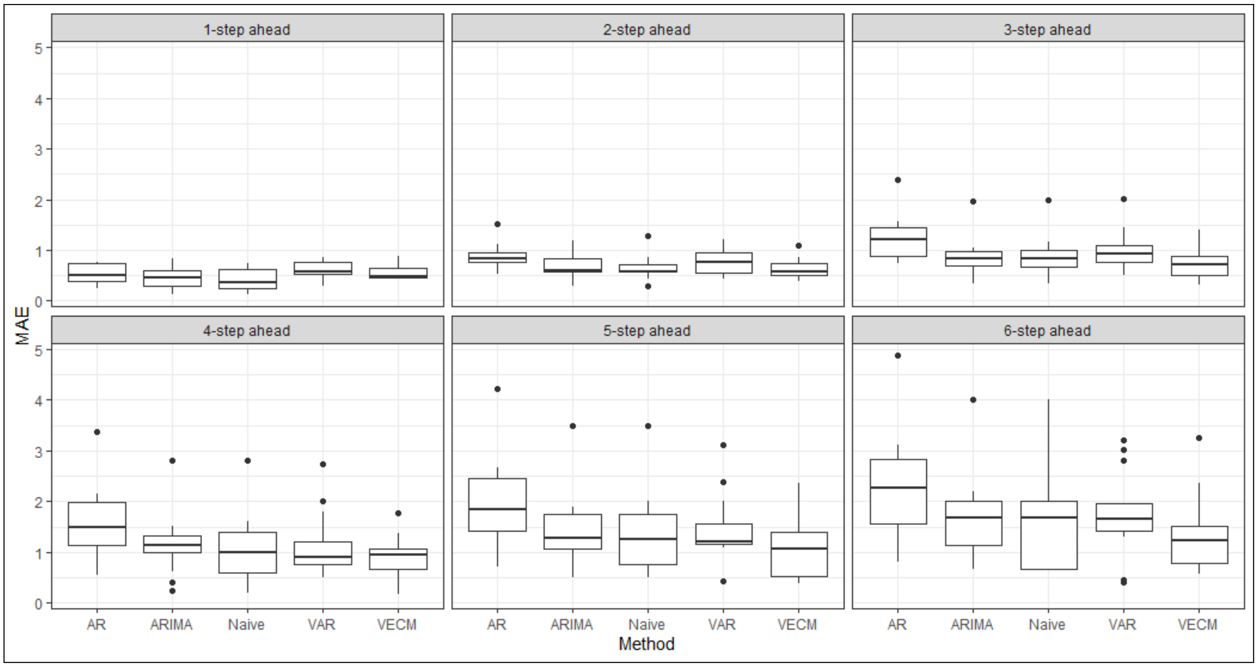

Figure 3: Summary of MAE results for all participants in Lab Study 1

Table 1: Median MASE results for Lab Study 1. The **bolded** values reflect the best model for a given k-step ahead forecast.

Method	Median MASE across all participants					
	1-step ahead	2-step ahead	3-step ahead	4-step ahead	5-step ahead	6-step ahead
Naïve	**1.00**	0.99	0.80	0.67	0.67	0.59
AR	1.63	1.25	1.07	0.99	0.96	0.87
VAR	1.67	1.10	0.97	0.79	0.79	0.72
ARIMA	1.10	1.01	0.83	0.79	0.67	0.59
VECM	1.33	**0.76**	**0.58**	**0.61**	**0.55**	**0.58**

Each panel of Figure 3 summarizes the MAE across all participants and forecasting methods when forecasting RPE $1, 2, \ldots, 6$ steps ahead. For example, the first panel shows the MAE when forecasting 1-step ahead, and all methods perform comparably, with MAE < 1 for most participants. For cases with $k \geq 3$ (which corresponds to forecasting RPE three or more periods in advance), the results in Figure 3 show that the VECM model has consistently lower MAE values across participants, especially for the 5- and 6-step ahead forecasts. For example, when forecasting RPE six periods in advance (corresponding to $k = 6$ or the lower right panel of Figure 3), 75% of the MAE values are below 1.52 units.

In an attempt to summarize the contributions of each of the three stride variables to the forecast/prediction, we have captured the sign of their coefficients in the obtained RPE forecast for the one-step ahead forecast of $t = 9$. The results for the VAR(1) representation of the VECM model (see Equation 5, the transformation was done by $\boldsymbol{A} = \boldsymbol{\Pi} + \boldsymbol{I}_N$) are shown in Table 2. There was a positive relationship for 10 of the 13 participants between lagged stride length and forecasted RPE. The forecasted RPE was negatively related to the lagged stride height and stride duration for the majority of participants, with negative relationships for 9 of the 13 participants (stride height) and 10 out of the 13 participants (stride duration). In all participants, there was a positive relationship between the lagged RPE scores and forecasted RPE (which is an expected result given that the experiment did not involve any rest period).

Table 2: Relationships between the lagged stride parameters and lagged RPE with the forecast RPE in Lab Study 1 (+ denotes a positive relationship, - denotes a negative relationship) when the first 50% of the data points are used (i.e., the training model where the forecast RPE for time period nine was made).

Subject	Lagged SL	Lagged SH	Lagged SD	Lagged RPE
1	-	+	-	+
3	-	-	+	+
4	+	-	-	+
5	+	-	-	+
7	+	-	-	+
8	+	-	-	+
9	+	+	-	+
10	+	-	+	+
11	+	-	-	+
12	-	+	+	+
13	+	+	-	+
14	+	-	-	+
15	+	-	-	+

3.2. *Lab Study 2*

Similar to Lab Study 1, we provide the personalized forecasting results for different participants of Lab Study 2 in the interactive Figure 4. For all participants, the RPE values start at 6 (no exertion at all) and peak at different, but similar, values. The lowest and the highest RPE peaks were 16 and 18.5 for participants 19 and 3, respectively. Although the range of reported RPE values across all participants did not vary significantly (with minimum and maximum ranges of 10 and 12.5), the experimental time and, consequently, the number of reported RPE values varied across participants (with minimum and maximum number of reported RPE values of 13 and 38 for participants 3 and 10, respectively). It should be noted that the observed difference between the forecasted RPE values obtained by any of the five methods and true RPEs increases as k increases, which is an inherent characteristic in any forecasting approach. Typically, the AR(1) method underestimates the response value for higher values of k.

We report the MAE results across all participants (excluding Participant 5) in Figure 5. Participant 5 was excluded since the AR(1) could not forecast all RPE values due to nonstationary data. Thus, we excluded

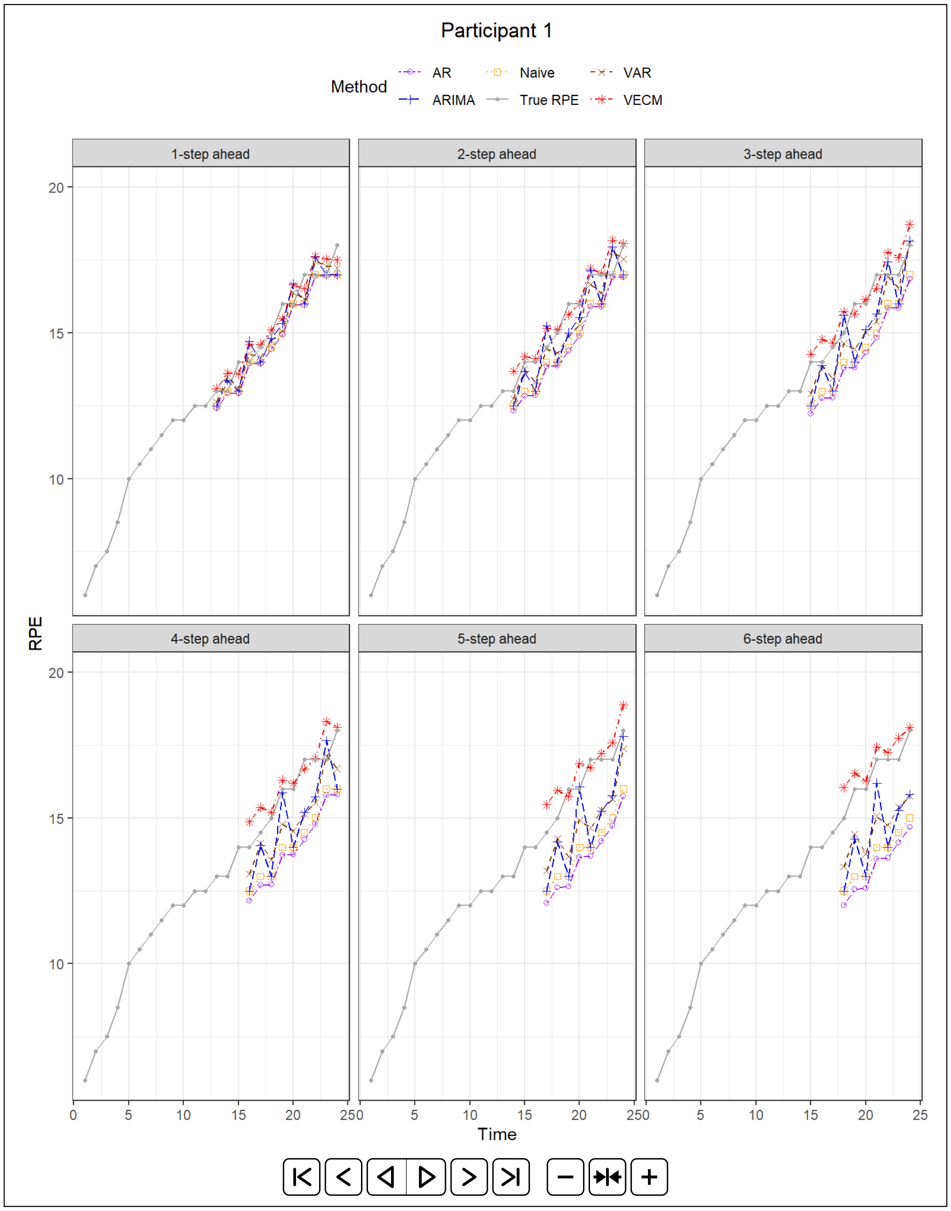

Figure 4: Forecasting results for different participants in Lab Study 2. The reader is encouraged to interact and engage with this visualization in Adobe Acrobat Reader (not through their web browser). Alternatively, the reader can access a web-based version of this visualization under Section 2.4 at: https://sahand-hajifar.github.io/RPE_Forecasting.html.

the participant from all the subsequent reported results to ensure a fair comparison across models. In the first panel of Figure 5, we see that all models performed similarly in the 1-step ahead forecast. While there are minor differences in MAE for the 2-step ahead forecast, we begin to see the performance benefits of the VAR and the VECM methods which include the sensor information in the forecasts in the performance of the 3-step ahead forecast ($k = 3$ periods ahead). Specifically, the median of the MAE values is smaller for the VECM model than for the other models considered for all cases when $k \geq 3$. This result is consistent with our findings from Lab Study 1.

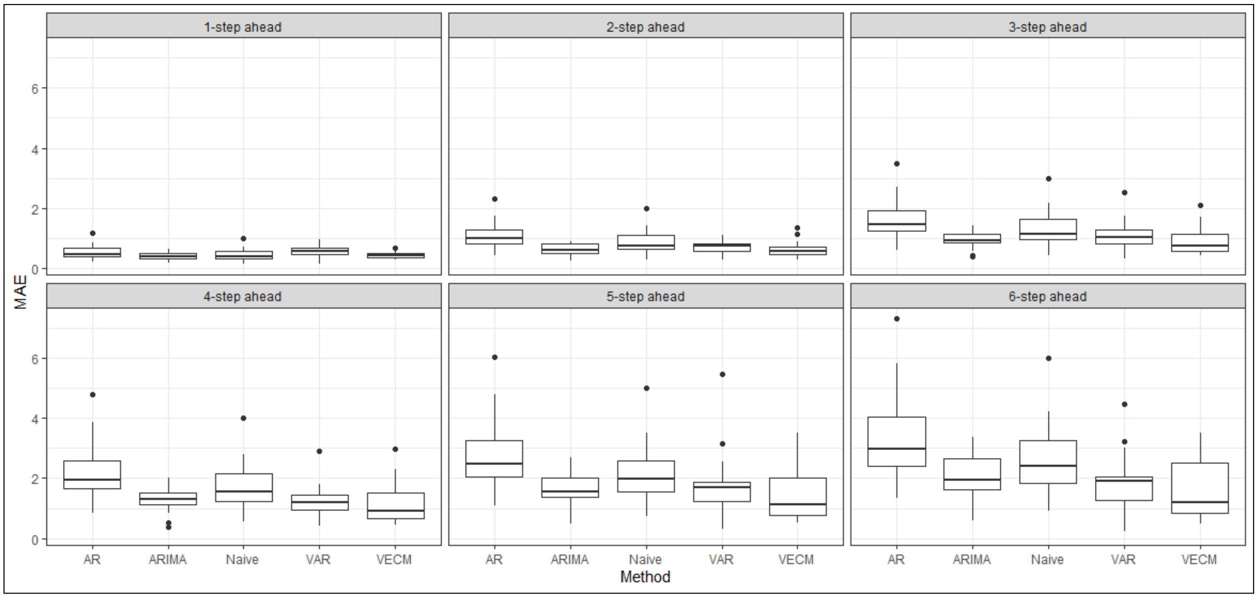

Figure 5: Summary of the MAE results for all participants (with the exception of Participant 5) in Lab Study 2

The median MASE results are provided in Table 3. When compared to Lab Study 1's results in Table 1, all the median MASE values reported here are < 1, i.e., the prediction performance of the naïve method on the in-sample training data is worse than the performance of our four forecasting models on the test data. In our estimation, the lower performance of the naïve model in this study, when compared to Lab Study 1, can be explained by the larger number of forecasts to be made (i.e. the upper value for t) across all participants.

Table 3: Median MASE results for Lab Study 2. The **bolded** values reflect the best model for a given k-step ahead forecast.

Method	Median MASE across all participants					
	1-step ahead	2-step ahead	3-step ahead	4-step ahead	5-step ahead	6-step ahead
Naïve	**0.73**	0.77	0.74	0.74	0.76	0.76
AR	0.90	0.87	0.95	0.91	0.93	0.98
VAR	0.91	0.61	0.57	0.47	0.55	0.45
ARIMA	**0.73**	0.51	0.53	0.49	0.49	0.51
VECM	0.74	**0.48**	**0.46**	**0.46**	**0.46**	**0.43**

Similar to Lab Study 1, the sign of the coefficients in the VAR(1) representation of the VECM based on the first 50% of the data points were captured. These signs help us to summarize the contributions of each of the three stride variables to forecasted RPE (see the results in Table 4). There was a negative relationship for 14 of the 23 participants between the lagged mean acceleration and forecasted RPE. The forecasted RPE was positively related to the lagged maximum acceleration and lagged stride duration, with positive relationships for 14 of the 23 participants. For all participants, there was a positive relationship between the lagged RPE scores and forecasted RPE.

Table 4: Relationships between the lagged stride parameters and lagged RPE with the forecast RPE in Lab Study 2 (+ denotes a positive relationship, - denotes a negative relationship) when the first 50% of the data points are used (i.e., the training model where the forecast RPE for time period nine was made).

Subject	Lagged Mean Acc	Lagged Max Acc	Lagged SD	Lagged RPE
1	+	-	-	+
2	+	-	-	+
3	-	+	+	+
4	-	+	+	+
6	-	+	-	+
7	-	+	+	+
8	-	+	-	+
9	+	-	+	+
10	+	-	+	+
11	+	-	+	+
12	-	+	+	+
13	+	-	-	+
14	-	+	-	+
15	-	+	+	+
16	-	+	+	+
17	-	+	+	+
18	+	-	-	+
19	-	+	+	+
20	-	+	+	+
21	+	-	-	+
22	-	+	+	+
23	+	-	-	+
24	-	+	+	+

4. Discussion

4.1. Assessment of the Results with Regards to the Research Questions

To illustrate the potential utility for a forecasting framework in fatigue modeling, we performed a secondary analysis of the experimental work of Maman et al. [2], Baghdadi et al. [3, 5] and Karvekar et al. [38], using time series models. The first objective was to determine whether an individual's RPE can be accurately forecasted using time series methods. Overall, we were able to show that the RPE predictions were relatively accurate for all four models examined (where the univariate models capitalized on only the RPE data, and the multivariate models used all four variables). As a means of interpreting the performance of the models, we consider two methods of comparison: (1) the MAE of the RPE value with regards to the verbal anchors of the Borg RPE scale, and (2) against the regression approach of [2] for Lab Study 1.

When considering the anchors provided with the Borg 15-grade RPE scale [59], with the exception of the starting and ending values in updated versions, verbal anchors are provided at each of the odd values.

Thus, a rating of 9, for example, is associated with the anchor "Very Light", and a rating of 10 falls between "Very Light" and "Light" (the anchor for 11). From this, we could consider that an error of < 1, would result in the same interpretation of the anchor (e.g., a predicted RPE of 9.9, would still equate to "Very Light"). Thus, we would set a target of obtaining a modeling prediction error < 1. In the current study, we achieved a median prediction error of < 1 for both datasets for the forecasts of 1-4 time periods ahead. This indicates that, on average, we are able to "perfectly" predict the level of exertion for a given participant.

For Lab Study 1, the VECM model, which performed best, had a prediction performance for all examined values of k with median MAE 1.24. This result is excellent compared to the results in Maman et al. (2017), whose real-time (i.e., current RPE value at time t based on information extracted from 5 wearables sensors at time t) had a mean testing set MAE = 2.16. Although the assessment was performed on a separate dataset, for Lab Study 2, we also obtained a median MAE using VECM for all models of $k < 1.22$. Given that the computed predicted errors for the VECM are small for all examined values of k, we conclude that an individual's RPE can be accurately forecasted for up to six time periods ahead. The comparison with Maman et al. [2] is only used to provide context pertaining to the "goodness" of the MAE values depicted in Figure 3. It is important to note, however, that the two studies cannot be directly compared since the models generated by Maman et al. [2]: (a) were not designed to be task-dependent; (b) were intended for real-time analysis (i.e. is the participant currently fatigued, which makes them less proactive than the approaches utilized in the current study); (c) used five sensors; and (d) were not individualized (i.e., they combined data across participants in the training of the models). In addition, in this study we capitalize on the previous RPE values in our prediction, which were ignored by Maman et al. [2].

The second research question focused on the forecast value added by the use of relevant features from a single IMU. For both datasets, owing to the fact that the VECM outperforms its counterpart (ARIMA) and other univariate methods for larger values of k, the overall answer to this question is *yes*, the forecasts of individual RPE can be improved by including the gait features. Model performance for the VECM, compared to the others, also improves for larger values of k. In the case of a univariate fatigue outcome/measure, the generalized $ARIMA(p,d,q)$ is very flexible and can be tuned/optimized to account for the availability and nature of the time series dataset. Perhaps more interestingly, the multivariate approaches discussed here can be applied in scenarios where other measures of fatigue and/or other internet of things (IoT) sensors are used. In fact, a major strength of the discussed multivariate time series approaches is that they are motivated by the reality that predictors simultaneously influence one another. Therefore, these approaches can account for the multidimensional nature of fatigue, which can be translated into several endogenous variables that affect work performance. While we have shown that the VECM model outperforms the others, an important question is whether the improvement gain is of practical relevance. This would help the practitioner determine whether the cost of the added sensors (both the sensor and the time associated with managing the data), is justified by the benefit in performance gain. In considering the comparison to the RPE scale anchors, when forecasting 1 to 4 steps ahead, the VECM model prediction results in the same anchor interpretation as the self-reported measure (MAE < 1). For the ARIMA model, this same threshold was met at 3 steps ahead. This comparison also addresses the third question of how many time periods ahead we can reliably forecast. Across all models for both studies, even at 6 steps ahead, the MAE < 1.25. Organizations need to consider two main decisions: at what threshold would they intervene, and what is the value of earlier prediction. An important note is that, for the two studies presented here, one time step represented a different amount of clock time (10 minutes in Lab Study 1 and $\sim$2 min in Lab Study 2).

One of the advantages of the individualized nature of the presented method is that individual differences are captured and the model performance is not harmed by heterogeneity across participants. This between-subject variability would negatively impact the performance of other population-based methods, such as support vector machine in machine learning (a common approach used in fatigue classification studies in ergonomics and biomechanics [37, 60, 61]). For a running task, Buckley et al. [60] reported 100% classification accuracy using personalized performance measures extracted from an IMU attached to the shank (similar location to the current study), compared to $\sim$70% accuracy for global/population features. The heterogeneity in the current analysis is documented through the relationships between the lagged stride parameters and the forecasted RPE (Tables 2 and 4). As seen in Table 2, most participants had a negative relationship between lagged stride height and forecasted RPE, which is consistent with findings from other studies of decreased stride height with fatigue [16]. Here we show a positive relationship between lagged stride length and forecasted RPE for 10 of 13 participants. In the literature there has been mixed evidence on the change in stride length with fatigue for group-level comparisons. For example, Morris et al. [62] reported reduced stride length when subjective fatigue was higher for individuals with multiple sclerosis, and Boda et al. [63] found smaller normalized stride length for patients with chronic faitgue syndrome versus a control group. However, others have found longer [64] or no significant fatigue-related change [31] in stride length. When considering the stride duration factor, Tables 2 and 4 highlight the opposite relationships identified for the two studies included here. The individualized model allows for assessment of fatigue even if a specific person does not follow the pre-determined expected pattern. In this study, we have advocated for the use of forecasting and time series analysis methods in ergonomics applications. The forecasting approach to physical fatigue modeling is very flexible, and the framework/cases highlighted in this paper is intended to illustrate the benefits of this underutilized modeling paradigm in ergonomics.

4.2. *Study Limitations*

As with any study, there are limitations that need to be highlighted. First, we implemented our forecasting procedure on two experimental studies that used self-reported RPE ratings, which may be biased by non-physical fatigue aspects [2]. Second, the forecasting models are trained based on historical data. Given the statistical nature of these models, an understanding of both the past/training (work tasks performed by the worker) data and future data is needed and should be incorporated in the model. For example, in our two case studies, we did not consider: (a) any periodic patterns in our data (e.g., environments where jobs are rotated one or more times during the work shift), and (b) a possible trend in the data due to training, fatigue, etc. These decisions were primarily guided by the limited number of observations per participant. However, in other applications where the variable(s) are collected more frequently and over a longer period, adjustments may be necessary in our approaches to account for a possible trend and/or periodicity in the data. We refer the reader to Brockwell and Davis [43] for an introduction of such techniques. Third, we have implicitly assumed that the worker will continue to perform the same type of work/tasks. This assumption, which is inherent in any predictive model, forms the basis of "past performance is somewhat indicative of future performance". If there are no regular sets of tasks that a worker routinely performs, our modeling approach would not be as successful. Fourth, the two case studies did not collect multi-visit data for the same subject. Therefore, we could not assess the repeatability of the individualized forecasting models over multiple days. This is an important consideration in practice (e.g., due to training effects and/or changes in job pacing requirements) and a possible limitation due to the collected data in both studies. To overcome this limitation, practitioners can resort to re-training the models (i.e., re-estimating the parameters of the forecasting model) at the beginning of each shift/day, which can be easily coded/programmed using existing information technology infrastructure. Fifth, both case studies show that the multivariate models (VAR and VECM) have consistently resulted in lower forecasting errors than their univariate counterparts that did not incorporate the features extracted from the wearable sensors data. While this supports the possible advantage of using wearables in ergonomic applications, we cannot assess whether the differences would be practically significant in field applications. More research is needed to assess whether our results can be replicated in other settings.

4.3. *Opportunities for Future Research*

We focused on introducing univariate and multivariate forecasting methods, and illustrating their utility through a secondary data analysis. While our two examples were suitable for highlighting the advantages of using forecasting techniques, our work needs to be expanded to consider more practical scenarios. An immediate extension of this work is to assess the stability of model parameters for each individual over a large number of days, since the two case studies examined here only included one session for a given task-participant combination. An improved understanding of the model's stability would provide insight into how the model could be deployed in practice. With the goal of developing a personalized model for forecasting physical fatigue, the impact of a fatigue intervention (e.g., scheduled rest breaks/job rotation) on recovery needs to be quantified. It is unclear whether the effectiveness of assigning these interventions over the course of the workday would decrease or would remain the same. The field of ergonomics can capitalize on the literature on *predictive maintenance* in manufacturing and production systems, where these ideas are developed for non-human systems [see e.g., 65]. In our estimation, this literature would complement some of the recent developments in injury prediction (see e.g., the application of the *Fatigue Failure Theory* in assessing injury risk [66]) and human reliability analysis [67]. From a statistical modeling perspective, we have simplified the analysis by down sampling the IMU data to match the low-frequency nature of the RPE variable. In the future, mixed-frequency time series models should be explored to fully leverage the information from the high frequency sensors.

5. Concluding Remarks

Through the application of the forecasting framework presented, we answered the research questions of whether an individual's historic RPEs can be used for accurate forecasting of future RPEs, whether specific features extracted from an IMU improved the forecasting accuracy, and how far in advance a reliable forecast can be generated. Based on the results, we determined that historic perceived exertion and wearable sensor data can be used to accurately forecast future RPEs up to 6 time points ahead. The ability to make accurate and personalized physical fatigue forecasts can inform the assignment and scheduling of appropriate interventions, which can eliminate or mitigate the negative health outcomes of fatigue on the worker [68, 69, 70], as well as the productivity losses [71, 72], worker compensation costs [73], and high employee turnover experienced by the organization. Furthermore, the use of forecasting strategies can transform fatigue modeling in two major ways: (a) from the somewhat reactive state (i.e., is the worker currently fatigued) to a more proactive paradigm (i.e., intervening based on the forecast of fatigue); and (b) towards personalized modeling, away from standard approaches to injury and risk prediction that are based on population parameters.

Acknowledgments

Funding: The modeling approach, analysis and computational resources were supported in part by the American Society of Safety Professionals Foundation [grant titled "ASSIST: Advancing Safety Surveillance using Individualized Sensor Technology"] and the National Science Foundation [CMMI-1635927]. Dr. Megahed's research was also partially supported by the Neil R. Anderson Endowed Assistant Professorship at Miami University. Dr. Jones-Farmer's research was partially supported by the Van Andel Endowed Professorship at Miami University.

Supplementary Materials

Data, Code and Analysis: The data, code and **R** scripts used in this paper can be accessed at following GitHub repository: https://github.com/sahand-hajifar/RPE_Forecasting. Moreover, our analysis and results are detailed in an **R** Markdown, which we host at: https://sahand-hajifar.github.io/RPE_Forecasting.html.

References

[1] C. G. Drury, Human factors/ergonomics implications of big data analytics: Chartered Institute of Ergonomics and Human Factors annual lecture, Ergonomics 58 (5) (2015) 659–673.

[2] Z. S. Maman, M. A. A. Yazdi, L. A. Cavuoto, F. M. Megahed, A data-driven approach to modeling physical fatigue in the workplace using wearable sensors, Applied Ergonomics 65 (2017) 515–529.

[3] A. Baghdadi, F. M. Megahed, E. T. Esfahani, L. A. Cavuoto, A machine learning approach to detect changes in gait parameters following a fatiguing occupational task, Ergonomics 61 (8) (2018) 1116–1129.

[4] L. Tsao, L. Li, L. Ma, Human work and status evaluation based on wearable sensors in human factors and ergonomics: a review, IEEE Transactions on Human-Machine Systems 49 (1) (2018) 72–84.

[5] A. Baghdadi, L. A. Cavuoto, A. Jones-Farmer, S. E. Rigdon, E. T. Esfahani, F. M. Megahed, Monitoring worker fatigue using wearable devices: A case study to detect changes in gait parameters, Journal of Quality Technology (2019) 1–25.

[6] Y. Yu, H. Li, X. Yang, L. Kong, X. Luo, A. Y. Wong, An automatic and non-invasive physical fatigue assessment method for construction workers, Automation in Construction 103 (2019) 1–12.

[7] S. Kim, A. Moore, D. Srinivasan, A. Akanmu, A. Barr, C. Harris-Adamson, D. M. Rempel, M. A. Nussbaum, Potential of Exoskeleton Technologies to Enhance Safety, Health, and Performance in Construction: Industry Perspectives and Future Research Directions, IISE Transactions on Occupational Ergonomics and Human Factors (2019) 1–7.

[8] G. Shmueli, et al., To explain or to predict?, Statistical Science 25 (3) (2010) 289–310.

[9] M. Cheng, M. S. Thiese, E. M. Wood, J. Kapellusch, J. Foster, D. Drury, A. Merryweather, K. T. Hegmann, B. S. Team, et al., Relationship Between Opioid Use and Pain Severity Ratings in Workers With Low Back Pain, Journal of Occupational and Environmental Medicine 61 (10) (2019) 836–840.

[10] M. Yung, A. M. Dale, J. Kapellusch, S. Bao, C. Harris-Adamson, A. R. Meyers, K. T. Hegmann, D. Rempel, B. A. Evanoff, Modeling the Effect of the 2018 Revised ACGIH® Hand Activity Threshold Limit Value®(TLV) at Reducing Risk for Carpal Tunnel Syndrome, Journal of Occupational and Environmental Hygiene 16 (9) (2019) 628–633.

[11] Z. Sedighi Maman, A. Baghdadi, F. Megahed, L. Cavuoto, Monitoring and change point estimation of normal (in-control) and fatigued (out-of-control) state in workers, in: ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers Digital Collection, DETC2016–60487, 2016.

[12] S. R. Hursh, T. J. Balkin, J. C. Miller, D. R. Eddy, The fatigue avoidance scheduling tool: Modeling to minimize the effects of fatigue on cognitive performance, SAE Transactions (2004) 111–119.

[13] C.-T. Lin, M. Nascimben, J.-T. King, Y.-K. Wang, Task-related EEG and HRV entropy factors under different real-world fatigue scenarios, Neurocomputing 311 (2018) 24–31.

[14] L. Lu, F. M. Megahed, R. F. Sesek, L. A. Cavuoto, A survey of the prevalence of fatigue, its precursors and individual coping mechanisms among US manufacturing workers, Applied Ergonomics 65 (2017) 139–151.

[15] M. Zhang, L. Murphy, D. Fang, A. J. Caban-Martinez, Influence of fatigue on construction workers' physical and cognitive function, Occupational Medicine 65 (3) (2015) 245–250.

[16] M. J. Angelini, R. M. Kesler, M. N. Petrucci, K. S. Rosengren, G. P. Horn, E. T. Hsiao-Wecksler, Effects of simulated firefighting and asymmetric load carriage on firefighter obstacle crossing performance, Applied Ergonomics 70 (2018) 59–67.

[17] K. Park, P. Hur, K. Rosengren, G. P. Horn, E. T. Hsiao-Wecksler, Changes in kinetic and kinematic gait parameters due to firefighting air bottle configuration, in: North American Congress on Biomechanics (NACOB), URL http://www.asbweb.org/conferences/2008/abstracts/579.pdf, 2008.

[18] L. M. Rose, W. P. Neumann, G. M. Hägg, G. Kenttä, Fatigue and recovery during and after static loading, Ergonomics 57 (11) (2014) 1696–1710.

[19] L. M. Rose, C. A. Beauchemin, W. P. Neumann, Modelling endurance and resumption times for repetitive one-hand pushing, Ergonomics 61 (7) (2018) 891–901.

[20] P. Dode, M. Greig, S. Zolfaghari, W. P. Neumann, Integrating human factors into discrete event simulation: a proactive approach to simultaneously design for system performance and employees' well being, International Journal of Production Research 54 (10) (2016) 3105–3117.

[21] M. Y. Jaber, Z. Givi, W. P. Neumann, Incorporating human fatigue and recovery into the learning–forgetting process, Applied Mathematical Modelling 37 (12-13) (2013) 7287–7299.

[22] M. Y. Jaber, W. P. Neumann, Modelling worker fatigue and recovery in dual-resource constrained systems, Computers & Industrial Engineering 59 (1) (2010) 75–84.

[23] J. S. Casey, R. W. McGorry, P. G. Dempsey, Getting a grip on grip force estimates: Avaluable tool for ergonomic evaluations, Professional Safety 47 (10) (2002) 18.

[24] A. Mital, R. R. Bishu, S. Manjunath, Review and evaluation of techniques for determining fatigue allowances, International Journal of Industrial Ergonomics 8 (2) (1991) 165–178.

[25] G. Spurr, A. Prentice, P. Murgatroyd, G. Goldberg, J. Reina, N. Christman, Energy expenditure from minute-by-minute heart-rate recording: comparison with indirect calorimetry, The American Journal of Clinical Nutrition 48 (3) (1988) 552–559.

[26] J. Achten, A. E. Jeukendrup, Heart rate monitoring, Sports Medicine 33 (7) (2003) 517–538.

[27] M. Calzavara, A. Persona, F. Sgarbossa, V. Visentin, A device to monitor fatigue level in order-picking, Industrial Management & Data Systems 118 (4) (2018) 714–727.

[28] B. Daria, C. Martina, P. Alessandro, S. Fabio, V. Valentina, Fatigue and recovery: research opportunities in order picking systems, IFAC-PapersOnLine 50 (1) (2017) 6882–6887.

[29] S. Hwang, J. Seo, H. Jebelli, S. Lee, Feasibility analysis of heart rate monitoring of construction workers using a photoplethysmography (PPG) sensor embedded in a wristband-type activity tracker, Automation in Construction 71 (2016) 372–381.

[30] E. Kristal-Boneh, M. Raifel, P. Froom, J. Ribak, Heart rate variability in health and disease., Scandinavian Journal of Work, Environment & Health 21 (2) (1995) 85–95.

[31] J. L. Helbostad, S. Leirfall, R. Moe-Nilssen, O. Sletvold, Physical fatigue affects gait characteristics in older persons, The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 62 (9) (2007) 1010–1015.

[32] X. Qu, J. C. Yeo, Effects of load carriage and fatigue on gait characteristics, Journal of Biomechanics 44 (7) (2011) 1259–1263.

[33] Y. Lee, S. Ulman, S. Kim, D. Srinivasan, Effects of Mental and Physical Fatigue Inducing Tasks on Balance and Gait Characteristics, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 63, SAGE Publications Sage CA: Los Angeles, CA, 1103–1104, 2019.

[34] K. S. Rosengren, E. T. Hsiao-Wecksler, G. Horn, Fighting fires without falling: Effects of equipment design and fatigue on firefighter's balance and gait, Ecological Psychology 26 (1-2) (2014) 167–175.

[35] K. Park, J. F. Sy, G. P. Horn, R. M. Kesler, M. N. Petrucci, K. S. Rosengren, E. T. Hsiao-Wecksler, Assessing gait changes in firefighters after firefighting activities and while carrying asymmetric loads, Applied Ergonomics 70 (2018) 44–50.

[36] P. Parijat, T. E. Lockhart, Effects of lower extremity muscle fatigue on the outcomes of slip-induced falls, Ergonomics 51 (12) (2008) 1873–1884.

[37] J. Zhang, T. E. Lockhart, R. Soangra, Classifying lower extremity muscle fatigue during walking using machine learning and inertial sensors, Annals of Biomedical Engineering 42 (3) (2014) 600–612.

[38] S. Karvekar, M. Abdollahi, E. Rashedi, A Data-Driven Model to Identify Fatigue Level Based on the Motion Data from a Smartphone, in: 2019 IEEE Western New York Image and Signal Processing Workshop (WNYISPW), IEEE, 1–5, 2019.

[39] W. S. Marras, S. A. Lavender, S. A. Ferguson, R. E. Splittstoesser, G. Yang, Quantitative biomechanical workplace exposure measures: distribution centers, Journal of Electromyography and Kinesiology 20 (5) (2010) 813–822.

[40] G. A. Borg, Psychophysical bases of perceived exertion., Medicine & Science in Sports & Exercise .

[41] P. G. Dempsey, A survey of lifting and lowering tasks, International Journal of Industrial Ergonomics 31 (1) (2003) 11–16.

[42] G. Box, G. Jenkins, Statistical Models for Forecasting and Control, Holden-Day, 1970.

[43] P. J. Brockwell, R. A. Davis, Introduction to time series and forecasting, Springer, 2016.

[44] P. Goodwin, F. Petropoulos, R. J. Hyndman, A note on upper bounds for forecast-value-added relative to naïve forecasts, Journal of the Operational Research Society 68 (9) (2017) 1082–1084.

[45] J. D. Cryer, K.-S. Chan, Models For Nonstationary Time Series, Time Series Analysis: With Applications in R (2008) 87–107.

[46] R. J. Hyndman, G. Athanasopoulos, Forecasting: principles and practice, OTexts, 2018.

[47] R. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, F. Yasmeen, forecast: Forecasting functions for time series and linear models, R package version 8.12. Installation details are available online at https://cran.r-project.org/web/packages/forecast/forecast.pdf. Last updated March 31, 2020., 2018.

[48] H. Shi, K. Worden, E. J. Cross, A cointegration approach for heteroscedastic data based on a time series decomposition: An application to structural health monitoring, Mechanical Systems and Signal Processing 120 (2019) 16–31.

[49] M. P. Murray, A drunk and her dog: an illustration of cointegration and error correction, The American Statistician 48 (1) (1994) 37–39.

[50] L. Drake, Modelling UK house prices using cointegration: an application of the Johansen technique, Applied Economics 25 (9) (1993) 1225–1228.

[51] C. A. Sims, Macroeconomics and reality, Econometrica: Journal of the Econometric Society (1980) 1–48.

[52] R. B. Litterman, Forecasting with Bayesian vector autoregressions—five years of experience, Journal of Business & Economic Statistics 4 (1) (1986) 25–38.

[53] J. H. Stock, M. W. Watson, Vector autoregressions, Journal of Economic Perspectives 15 (4) (2001) 101–115.

[54] G. Adomavicius, J. Bockstedt, A. Gupta, Modeling supply-side dynamics of IT components, products, and infrastructure: An empirical analysis using vector autoregression, Information Systems Research 23 (2) (2012) 397–417.

[55] H. Lütkepohl, New introduction to multiple time series analysis, Springer Science & Business Media, 2005.

[56] W. Enders, Applied Econometric Time Series, 4th Edition, Wiley, 2015.

[57] R. J. Hyndman, A. B. Koehler, Another look at measures of forecast accuracy, International Journal of Forecasting 22 (4) (2006) 679–688.

[58] R. J. Hyndman, Another look at forecast-accuracy metrics for intermittent demand, Foresight: The International Journal of Applied Forecasting 4 (4) (2006) 43–46.

[59] G. Borg, Perceived exertion as an indicator of somatic stress., Scandinavian Journal of Rehabilitation Medicine 2 (2) (1970) 92–98.

[60] C. Buckley, M. A. O'Reilly, D. Whelan, A. V. Farrell, L. Clark, V. Longo, M. Gilchrist, B. Caulfield, Binary classification of running fatigue using a single inertial measurement unit, in: 2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN), IEEE, 197–201, 2017.

[61] A. Aryal, A. Ghahramani, B. Becerik-Gerber, Monitoring fatigue in construction workers using physiological measurements, Automation in Construction 82 (2017) 154–165.

[62] M. E. Morris, C. Cantwell, L. Vowels, K. Dodd, Changes in gait and fatigue from morning to afternoon in people with multiple sclerosis, Journal of Neurology, Neurosurgery & Psychiatry 72 (3) (2002) 361–365.

[63] W. L. Boda, B. H. Natelson, S. A. Sisto, W. N. Tapp, Gait abnormalities in chronic fatigue syndrome, Journal of the Neurological Sciences 131 (2) (1995) 156–161.

[64] H. Nagano, L. James, W. A. Sparrow, R. K. Begg, Effects of walking-induced fatigue on gait function and tripping risks in older adults, Journal of Neuroengineering and Rehabilitation 11 (1) (2014) 1–7.

[65] A. J. Henry, J. A. Nachlas, An equivalent age model for condition-based maintenance, in: 2012 Proceedings Annual Reliability and Maintainability Symposium, IEEE, 1–6, 2012.

[66] S. Gallagher, M. C. Schall Jr, Musculoskeletal disorders as a fatigue failure process: evidence, implications and research needs, Ergonomics 60 (2) (2017) 255–269.

[67] V. Di Pasquale, S. Miranda, W. P. Neumann, A. Setayesh, Human reliability in manual assembly systems: a Systematic Literature Review., IFAC-PapersOnLine 51 (11) (2018) 675–680.

[68] J. N. Côté, D. Raymond, P. A. Mathieu, A. G. Feldman, M. F. Levin, Differences in multi-joint kinematic patterns of repetitive hammering in healthy, fatigued and shoulder-injured individuals, Clinical Biomechanics 20 (6) (2005) 581–590.

[69] M. A. Huysmans, M. J. Hoozemans, A. J. van der Beek, M. P. de Looze, J. H. van Dieën, Position sense acuity of the upper extremity and tracking performance in subjects with non-specific neck and upper extremity pain and healthy controls, Journal of Rehabilitation Medicine 42 (9) (2010) 876–883.

[70] M. Yung, Fatigue at the Workplace: Measurement and Temporal Development, PhD Dissertation, University of Waterloo. Located at: https://uwspace.uwaterloo.ca/handle/10012/10119, 2016.

[71] A. Kolus, R. Wells, P. Neumann, Production quality and human factors engineering: A systematic review and theoretical framework, Applied Ergonomics 73 (2018) 55–89.

[72] M. Yung, A. Kolus, R. Wells, W. P. Neumann, Examining the fatigue-quality relationship in manufacturing, Applied Ergonomics 82 (2020) 102919.

[73] J. A. Ricci, E. Chee, A. L. Lorandeau, J. Berger, Fatigue in the US workforce: prevalence and implications for lost productive work time, Journal of Occupational and Environmental Medicine 49 (1) (2007) 1–10.