Analysis and Prediction of Daily Physical Activity Level Data Using Autoregressive Integrated Moving Average Models

Xi LONG[a,c,1], Steffen PAUWS[a,1], Marten PIJL[a], Joyca LACROIX[a], Annelies H. GORIS[b], and Ronald M. AARTS[a,c]
[a] *User Experiences Group, Philips Research Laboratories, Eindhoven, The Netherlands*
[b] *Philips New Wellness Solutions, Eindhoven, The Netherlands*
[c] *Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*

Abstract. Results are provided on predicting daily physical activity level (PAL) data from past data of participants of a physical activity lifestyle program aimed at promoting a healthier lifestyle consisting of more physical exercise. The PAL data quantifies the level of a person's daily physical activity and reflects the daily energy expenditure of this person. In this wellbeing program, a mobile body-worn activity monitor with a built-in triaxial accelerometer was used to record the PAL data of an individual for a period of 13 weeks. The autoregressive integrated moving average (ARIMA) models were employed to predict future PAL data of every next week. This paper proposes a categorized-ARIMA (CARIMA) prediction method which achieves a large reduction in computation time without significant loss in prediction accuracy compared with the traditional ARIMA. In the current method, PAL data were categorized as being stationary, trend or seasonal via assessing their autocorrelation functions. The most appropriate ARIMA model for these three categories was automatically selected by applying the objective penalty function criterion. The results show that our CARIMA method performed well in terms of PAL prediction accuracy (~9% mean absolute percentage error), model parsimony and robustness.

Keywords. activity monitoring, ARIMA, Kalman recursion, order selection, PAL, prediction, wellbeing.

Introduction

Physical activity has large beneficial effects on people's mental and physical health [1]. Physical inactivity may not only increase risk of chronic disease, but also reduce quality of life resulting in a significant burden to the healthcare system [2]. The increasing number of people with an inactive lifestyle requires the need for highly persuasive physical activity interventions to stimulate a healthier lifestyle. Therefore, research into the development of effective physical activity promotion programs receives much attention nowadays [3, 4]. For participants of these programs, the

[1] Corresponding Authors: Xi LONG, Steffen PAUWS, User Experiences Group, Philips Research Laboratories, HTC 34, 5656 AE, Eindhoven, The Netherlands; E-mail: xi.long@philips.com, steffen.pauws@philips.com

provision of feedback about changes in physical activity behavior has proven to be highly important to stay motivated and to attain their goals [4, 5]. The use of mobile body-worn sensors is a promising solution to realize real-time monitoring and feedback provision of a person's physical activity behaviour [6, 7]. A triaxial accelerometer is an inexpensive, effective and feasible sensor which has been used often in acquiring activity information [6, 8, 9]. Scientific proof has been reported on the feasibility of wearing an accelerometer sensor for assessing reliably daily energy expenditure [6, 10, 11].

In addition to feedback about behaviour, the level of personalization of messages and program content has shown to be important for the effective support of behaviour change. One powerful way to personalize a program is through the incorporation of interaction with a human coach that has insight into the participant's behaviour and can provide feedback and support regarding specific motivational dips or barriers that the participant encounters. Nevertheless, human coaching support can be very labor intensive, which limits the number of participants that a human coach can reach. To increase the coach's reach, we propose to support the coach by models that can predict future activity behaviour of participants. In particular, we introduce a technique that predicts the most likely near-future daily physical activity level[2] (PAL) from past physical activity of a participant in a wellbeing program. Such prediction models can be helpful to determine the most relevant coaching support. For example, when the model predicts that a participant will fail to reach personal activity goals in the near future, a proactive human coaching intervention may be helpful.

Predicting future PAL data using the well-known Box-Jenkins methodology [12] requires time series data only. The Box-Jenkins methodology, commonly known as the autoregressive integrated moving average (ARIMA) model, has already been widely used in a number of related areas such as economic time series forecasting [13, 14], ecological and weather prediction [15, 16], medical monitoring [17, 18], traffic flow prediction [19], and also physical activity recognition [20]. Generally, the application of ARIMA models is mostly focused on predicting a single univariate time series. For multi univariate series, an automatic method was developed [21, 22]. However, the automatic method takes unacceptable computation time due to the need to repetitively choose the most appropriate model from a large number of candidate models. Therefore, we applied a categorized ARIMA (CARIMA) method in which the models were categorized according to an apriori identified data category (i.e., stationarity, trend and seasonality). Then, the most appropriate model for each category was chosen beforehand rather than computed from the pool of candidate models. The parameter estimation of the models was based on the maximum likelihood (ML) method via a Kalman recursion [23].

The remainder of the paper is organized as follows: Section 1 briefly introduces the data collection. The ARIMA method is explained in Section 2. Section 3 discusses the results of the ARIMA modeling and prediction. Conclusions are provided in the final section.

[2] PAL is calculated by dividing the total energy expenditure (TEE) by the resting energy expenditure (REE).

1. Data Collection

In a wellbeing program promoting a healthier lifestyle by being physical active[3], participants were provided with a Philips activity monitor with a built-in triaxial accelerometer to measure the acceleration data of their activities performed throughout the day. The monitor is a light and small-sized portable device (3×3 cm) that can be worn easily and unobtrusively in an arbitrary orientation on the human body in a free-living environment. In order to obtain information of daily energy expenditure, the acceleration data is converted to PAL data because of the found correspondence between acceleration data and energy expenditure [6]. The program lasted 12 weeks for every participant. Before the start of the actual program, participants entered an assessment week, during which they learned to use the provided monitor and completed personal data (e.g., age, gender, weight). The PAL data that are recorded during the assessment week serves as a baseline measurement. In total, there are 91 (13×7) days in which participants wore the device. Each day consisted of a data point containing the PAL value accumulated over an entire 24 hour day of a participant in the program. All data were stored in a database.

In total, 950 participants were recruited for the participation of the wellbeing program. Daily PAL data should lie between 1.2 and 2.5 for adults [25]. The lower bound refers to a sedentary level. PAL values that are lower than this bound reflects not wearing the device. The upper bound refers to a vigorously active level. An extremely high level of physical activity (e.g., endurance) allows a PAL value as high as 4.5, but this level of activity cannot be persevered for a long time [24]. In this study, we treated the extremely low and extremely high PAL values ($PAL<1.2$ or $PAL>5$) as missing data. For the purpose of evaluating ARIMA models, the database was cleaned up by choosing 227 time series out of a total of 950 time series.

Participants in the program were classified into two levels of successful program participation (i.e., success, unsuccess) according to the PAL time series. Success was defined as the achievement of a daily PAL value at the end of the program that is at least 10% higher than the mean daily PAL value during the assessment week. Unsuccess was defined for those participants that were not successful in achieving a 10% increase in physical activity, among which, drop-out was defined as the lack of daily PAL values in the mid of the program until the end of the program.

2. ARIMA Method

In this study, ARIMA models were used to fit the observed daily PAL data and to make a prediction on future PAL data. To this end, an ARIMA method consists of procedure for modeling and for prediction. In addition, the robustness of the fitted ARIMA models on noise and missing data is also important for prediction purposes and will be evaluated accordingly.

[3] URL: http://www.directlife.philips.com/

2.1. ARIMA Modeling

The ARIMA modeling aims at constructing the most appropriate model to fit observed data. A general ARIMA model can be structurally classified as the form of ARIMA(p,d,q)(P,D,Q)$_S$ models [12]. The model can be written as below

$$\phi(B)\Phi(B^S)(1-B)^d(1-B^S)^D y_t = \theta(B)\Theta(B)^S \varepsilon_t \tag{1}$$

$$\phi(B) = 1-\phi_1 B-\phi_2 B^2 - ... -\phi_p B^p \tag{2}$$

$$\theta(B) = 1-\theta_1 B-\theta_2 B^2 - ... -\theta_q B^q \tag{3}$$

$$\Phi(B^S) = 1-\Phi_1 B^S-\Phi_2 B^{2S} - ... -\Phi_P B^{PS} \tag{4}$$

$$\Theta(B^S) = 1-\Theta_1 B^S-\Theta_2 B^{2S} - ... -\Theta_Q B^{QS} \tag{5}$$

where the symbols used are defined as follows:

- y_t: data point at time t
- ε_t: the independent, identical, normally distributed residual at time t
- B: backward shift operator, where $B^n y_t = y_{t-n}$,
- p: order of non-seasonal autoregressive (AR) terms
- d: order of non-seasonal differencing
- q: order of non-seasonal moving average (MA) terms
- P: order of seasonal autoregressive (SAR) terms
- D: order of seasonal differencing
- Q: order of seasonal moving average (SMA) terms
- S: seasonal order.

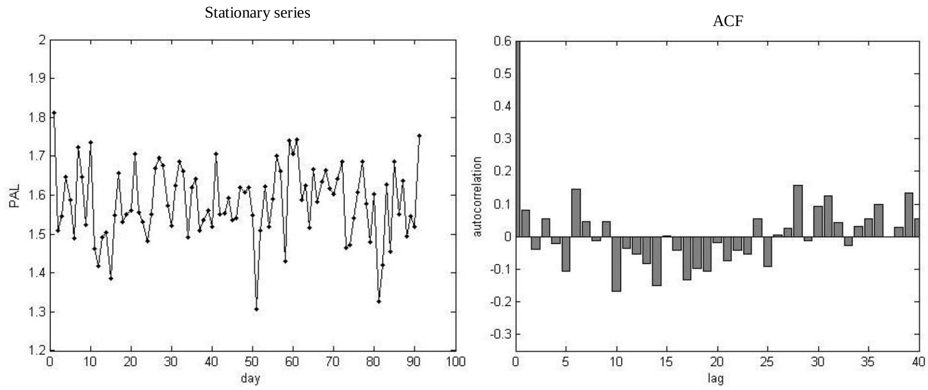
The ARIMA modeling process consists of identification, model estimation and order selection, diagnostic checks and residual analysis.

- *Identification*

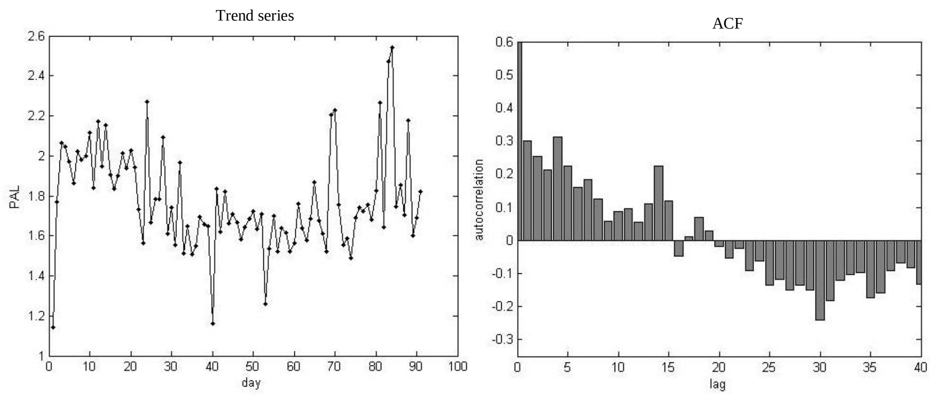
The orders of the ARIMA models need to be determined to ensure that the selected model fits the observed PAL time series best. In the ARIMA models, the non-seasonal differencing and the seasonal differencing are crucial to remove the trend and seasonality of the time series (i.e., to achieve stationarity). The selection of the orders d and D can be made by tentatively identifying the stationarity and seasonality of the observed data. A stationary time series has a mean and a variance that are constant over time. A non-stationary series has either a trend upwards or downwards or fluctuates over different levels. In practice, an approximately stationary assumption is sufficient rather than a strictly stationary one. Seasonality is yet another general pattern that can be observed in the data. Intuitively, a seasonal series means that the data behaves periodically. The periodicity of our daily PAL time series is assumed to be seven days as we expected weekly behaviour in physical activity. First-order non-seasonal differencing (d=1) and seasonal differencing (D=1) can effectively remove the trend and the seasonality of the data [26]. The use of higher differencing orders may result in over-differencing the data. In this study, the analysis of autocorrelation function (ACF) was used to identify the time series as being a stationary, a trend-wise or a seasonal time series. Considering the observations y_1, y_2, …, y_t, the autocorrelation at lag k is

$$\rho_k = \frac{\sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2} \tag{6}$$

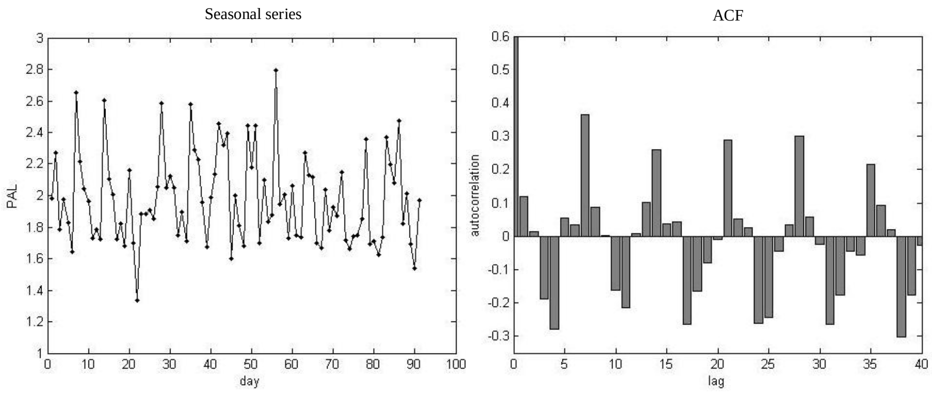
where $\bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t$, n is the sample size. If the ACF magnitude of a time series cuts off or dies down fairly quickly as k increases, then this series is considered to be stationary. Otherwise, if the ACF of a series dies down slowly, it is considered to be non-stationary. Besides, if the ACF of a series has spikes at specific lags, it is considered to be seasonal. As an example, Figure 1 shows the typical examples of a stationary, a trend-wise and a seasonal time series and their ACFs.



(a)



(b)



(c)

Figure 1. The plots of typical examples of a stationary (a), a trend (b), and a seasonal (c) daily PAL time series and their ACFs.

In order to examine the stationarity and the seasonality of a time series quantitatively and automatically rather than to inspect them visually, the t_{ρ_k} statistic can be calculated

$$t_{\rho_k} = \rho_k / s_{\rho_k} \tag{7}$$

where the standard error of ρ_k is

$$s_{\rho_k} = [(1 + 2\sum_{i=1}^{k-1} \rho_i^2) / n]^{1/2} \tag{8}$$

For identifying significant spikes at lags of an ACF automatically, various critical absolute *t*-values are suggested [26]. Table 1 indicates the spikes identification to examine stationarity and seasonality in an ACF. Alternatively, the Augmented Dickey-Fuller (ADF) unit root test [27] can also be utilized to identify the stationarity and seasonality of a given series.

Table 1. Critical absolute *t*-values for identified spikes in the ACF [26]

Lags	Critical Absolute t-values in ACF
Non-seasonal lags	
Low (1,2,3)	1.6
Other non-seasonal	2.0
Seasonal lags	
Exact seasonal (S, 2S, 3S)	1.25
Near & half seasonal	1.6
Other lags	2.0

- *Model Estimation and Order Selection*

In this study, the well-known recursive algorithm, Kalman filtering [23], was applied to the observed PAL time series data for model estimation. In this so-called state space model, the filtered state is the predicted state. Considering the observations $\{y_1, y_2, \dots, y_{t-1}\}$, the Kalman recursion aims at searching for optimal values y_t at point t. The fitting residuals ε_t and their variances F_t are of particular interest. The Kalman recursion process consisted of two groups of equations: the observation equations and the state updating equations. The observation equations are responsible for incorporating a new incoming observation into the estimate and the state updating equations (unobserved) are responsible for updating the current state and residual variances estimates to obtain the fitted values in the next time step. Details of Kalman recursion for ARIMA models are provided elsewhere [28, 29].

The unknown initial elements of the unobserved state of the state space model include the initial disturbances of different components, the initial state mean value and its variance at time $t = 1$. They were defined by using diffuse initialization method [30]. In addition, the ARIMA models normally contain two or more parameters (hyper-parameters) such as the autoregressive and moving average coefficients. The maximum likelihood estimation (MLE) was used to optimize these parameters iteratively. In this case, a log-likelihood function was constructed and the value of it was maximized by minimizing the residuals and their variances simultaneously [30].

In order to select the most appropriate model for a specific series, the orders of the AR, MA, SAR, and SMA processes should be determined. Models with a too high order (i.e., p, q, P and Q) result in over-fitting the data. In practice, p, q, P and Q are equal to or less than 2 [31]. The seasonal order S was assumed to be 7 (i.e., reflected weekly periodicity). The objective penalty function criteria such as Akaike Information Criterion (AIC) [32] and Schwarz's Information Criterion (SIC) [33] are most widely used in the order selection for ARIMA models. These criteria try to find a trade-off between the goodness-of-fit and the parsimony of the models. The SIC is defined as

$$SIC = -2 \cdot \ln(L)/n + k \cdot \ln(n)/n \tag{9}$$

where L is the likelihood function of the ARIMA model, k is the number of parameters to be estimated (i.e., $k=p+q+P+Q+1$), and n is the number of observed data points. For a small or moderate sample size, SIC performs better than AIC in order selection of ARIMA models [34, 35]. Compared to AIC, SIC penalizes the number of parameters more heavily and focuses more on model parsimony, which effectively decreases the possibility of over-fitting. Initially, in this study, all possible 144 candidate models encompassing models until ARIMA(3,d,3)(2,D,2) were involved in model selection. However, the results show that 99% of the time series have the best fit by ARIMA models that encompass orders until (2,d,2)(1,D,1) using SIC. Therefore, only 36 candidate ARIMA models with an maximum order (2,d,2)(1,D,1) are tested under SIC for model selection.

- *Diagnostic Checking and Residual Analysis*

The residuals $r = \{r_1, r_2, \dots, r_n\}$ of the prediction need to be uncorrelated. To check for this, the rescaled residuals (i.e., standardized residuals) are defined as

$$r_t = \varepsilon_t / \sqrt{F_t} \tag{10}$$

where ε_t is the fitting residual and F_t is the residual variance at time t. The standardized residuals from an ARIMA fitting process should be uncorrelated and normally distributed.

Serial correlation in the residuals means that the model is inadequate. There are several statistical tests to check for correlation. The Ljung-Box Q test [36] is a widely-used test based on the ACF of the residuals. The Q statistic is

$$Q = n(n+2)\sum_{l=1}^{h}(n-l)^{-1}\rho_l^2 \tag{11}$$

where n is the data sample size, ρ_l is the autocorrelation of the residuals at lag l, and h is the number of lags being tested. The choice of h is arbitrary and is normally set at 10 [30]. For a significance level of 5% (i.e., the probability of a Type I error), the assumption of uncorrelation is rejected if Q is larger than $\chi^2_{[5\%]}(df)$, the critical value of chi-square distribution [37] with df degrees of freedom.

The fitting performance of the models can also be compared by the criterion of residual measures. We used the mean absolute relative residual (MARR), which is usually used, to measure and quantify the quality of fit,

$$MARR = \tfrac{1}{n}\sum_{t=1}^{n}| r_t / y_t | \tag{12}$$

2.2. Prediction

The ARIMA modeling aims at choosing the most appropriate model for a specific PAL time series based on in-sample fitting. During the modeling process, the parameters and orders were automatically updated with every incoming new observation. In practice, once a model has been chosen, it should be used to predict future data, preferably using out-of-sample data to assess the prediction performance. Take the seasonal PAL time series in Figure 1-(c) as an example. Assume that 70 data points in 10 weeks (including the assessment week) are observed. The modeling period of 70 days and prediction period are indicated in Figure 2. The automatic order selection method based on a penalty function criteria for the ARIMA models has been frequently used [21, 22]. This automatic-order-selection ARIMA (AOS-ARIMA) method searches for the best model for each individual time series evaluating 36 candidate models; this is computational intensive. Another method applies only one single model with specific AR, MA, SAR, and SMA orders for all the time series. However, this single ARIMA (Single-ARIMA) method may choose the model apriori that has the best fit with the majority of categories of stationarity, trend and seasonality. We propose a categorized ARIMA (CARIMA) method by using the most appropriate model that best fit one of the categories. The advantages of the CARIMA include low computational intensity and good quality of fit for the different PAL time series categories.

The out-of-sample one-step-ahead prediction (1-SAP) provides a prediction of the PAL value for the next day. To assess the prediction accuracy, we use the measure of mean absolute absolute percentage error (MAPE),

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} | e_t / y_t | \quad (13)$$

where n is the sample size, e_t is the prediction error and y_t is the actual observation at time t.

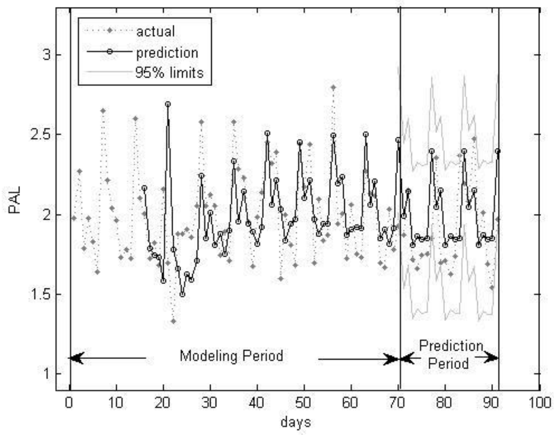


Figure 2. An example of predicting a seasonal PAL time series (with 70 data points observed).

2.3. Model Robustness

The robustness of a prediction model is important for real applications because data in practice contain noise, missing values and outliers. In the areas of economics, ARIMA models have proven to be robust [38]. However, for the PAL data, the robustness of the models on noise, missing values, and outliers need to be assessed.

The noise of the PAL data mainly comes from the background environment and the device itself during the usage. Hence, the noise can be considered as additive white Gaussian noise (AWGN). Since we transformed the outlier problem into a missing value (see Section 1), we will not address outliers. In the database, approximately 11.6% of the PAL data points were missing for each time series on average. The missing data made the model identification and parameter estimation difficult during the ARIMA modeling. In this study, a mean substitution approach was applied to complete incomplete data; this approach simply pads the mean of all non-missing values when calculating the ACF of the data [39]. During the prediction procedure, the missing data were treated as predictions in the same way [30]. In estimating the parameters by computing the log-likelihood function in the Kalman recursion, the value of the residual was simply set to zero when the corresponding observation was missing.

3. Results and Discussion

The ACFs of the PAL data were identified and the results show that the number of stationary series amounts to ~83% (188 out of 227) in the clean dataset, whereas the number of trend and seasonal time series take ~4% (10 out of 227) and ~13% (29 out of 227), respectively. This means that the majority of the series are considered stationary. The 36 ARIMA models encompassed by a (2,*d*,2)(1,*D*,1) model have been applied to fit the data. The ARIMA(1,*d*,1)(0,*D*,0) model was used to compare average SIC statistics over all series. This latter model was chosen because of its best overall fit for the stationary series. In Table 2, the best models are ranked for the stationary, trend and seasonal series according to their average SIC statistics. As shown in Equation 9, SIC tries to establish a trade-off between low residuals and model parsimony.

The CARIMA method proposed that the highest ranked model should be applied for the series in each category. This results in the (1,*d*,1)(0,*D*,0), (0,*d*,1)(0,*D*,0) and (1,*d*,0)(0,*D*,1) models for the stationary, trend and seasonal series, respectively. Table 3 compares the fitting performances of the three methods in terms of MARR. It shows that the AOS-ARIMA method performed the best in fitting accuracy but it is the highest in computational intensity (implemented in Matlab®). The other two methods cost significantly less computation time than the AOS-ARIMA method. In addition, CARIMA achieved lower average MARR than Single-ARIMA for the 227 series in this study.

Table 2. Top 3 ARIMA models ranked by the average SIC over the series of the stationary, trend and seasonal categories

Stationary series		**Trend series**		**Seasonal series**	
Model	Average SIC	Model	Average SIC	Model	Average SIC
(1,d,1)(0,D,0)	-0.3210	(0,d,1)(0,D,0)	-0.5339	(1,d,0)(0,D,1)	-0.0551
(1,d,1)(1,D,0)	-0.2656	(0,d,1)(1,D,0)	-0.4896	(0,d,0)(0,D,1)	-0.0550
(1,d,2)(0,D,0)	-0.2410	(0,d,2)(0,D,0)	-0.4785	(0,d,1)(0,D,1)	-0.0498

Table 3. Overall performance comparison of the in-sample fitting of the three methods used

Method	**Average MARR (±SD)**	**Average Computation Time**
Single-ARIMA	8.9% (±2.8%)	0.5s
AOS-ARIMA	8.5% (±2.7%)	39.3s
CARIMA	8.6% (±2.8%)	0.6s

In order to test for significant effects in the MARRs, we employed the student's *t*-test for each two methods. The null hypothesis states that there is no difference between the two methods. The *t*-test statistic for the 227 MARRs by using AOS-ARIMA and CARIMA methods was calculated to be 2.84, which is larger than $t_{(0.05;\, 452)} = 1.96$, the 5% critical value with the degree of freedom (*df*) of 452. The null hypothesis should be rejected in favour of the alternative hypothesis: there is a significant difference between the means of the two methods. Similarly, the *t*-test statistic of 2.90 calculated from the MARR values by using CARIMA and Single-ARIMA methods also indicates a significant difference. As a conclusion, the use of the CARIMA method largely reduces computation intensity to fit the data with a little loss in its fitting accuracy (~0.1%), in comparison with the AOS-ARIMA method.

Figure 3 shows that for the stationary and trend series, the CARIMA and Single-ARIMA methods performed similar and almost as well as the AOS-ARIMA method. However, for the seasonal series, the fitting performance of Single-ARIMA method becomes worse than the other two methods. In addition, the fitting performance of the trend and seasonal series are better than the fitting performance of the stationary series.

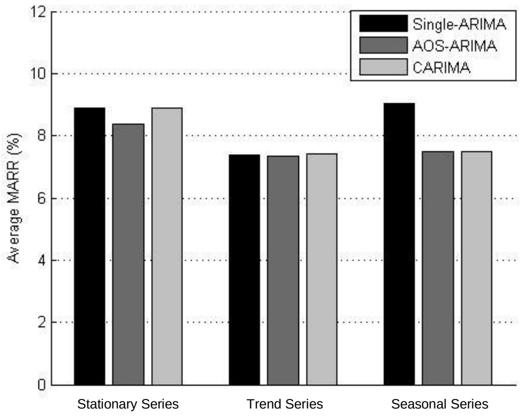


Figure 3. Average MARRs (%) of the data fitting for the stationary, trend and seasonal PAL series by using Single-ARIMA (*black bars*), AOS-ARIMA (*dark gray bars*) and CARIMA (*light gray bars*) methods.

The diagnostic checks, for examining whether or not the applied model accurately represented the underlying process in the observed time series, revealed that residuals were indeed serially uncorrelated in 211, 199 and 176 of the 227 cases when using AOS-ARIMA, CARIMA and Single-ARIMA methods, respectively. Thus, the AOS-ARIMA selected the best fitted model for the series and achieved the highest number of serially uncorrelated residuals. The CARIMA method achieves a slightly lower number of uncorrelated residual cases. The Single-ARIMA performed worse in this respect.

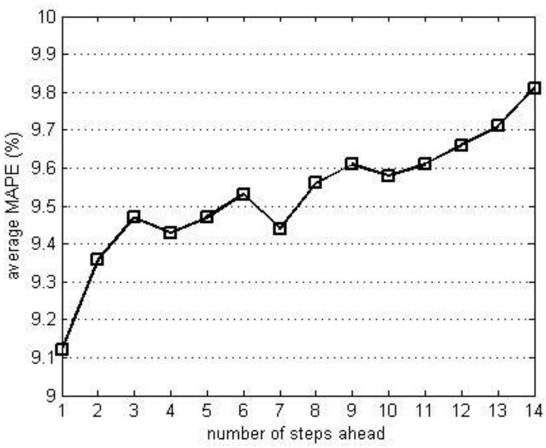


Figure 4. Average MAPEs (%) of multiple-step-ahead predictions by using CARIMA method.

During the prediction process, the data of the assessment week were assumed to be known. The out-of-sample 1-SAP errors show similar results as the in-sample fitting residuals (See Table 4): the CARIMA method is preferred because of to its good prediction performance and computation simplicity. Table 5 shows that the predictions for the trend and seasonal series were more accurate than those for the stationary ones. For the purpose of evaluating the performance of multiple-step-ahead prediction (M-SAP) with the CARIMA method, we plotted the errors (MAPEs) up to the 14-SAP (i.e., the future two weeks) in Figure 4. The models performed well in M-SAP (with an accuracy of <0.7% decline of the 14-SAP compared with that of the 1-SAP).

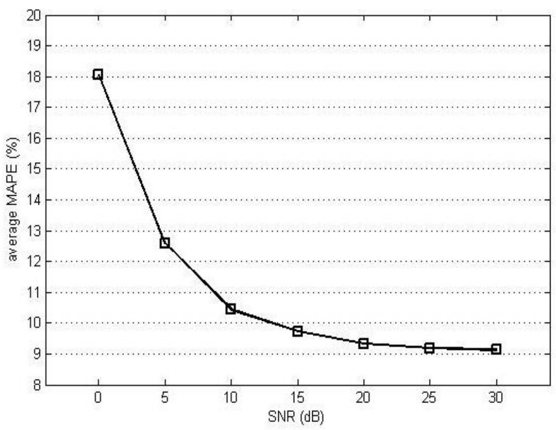
Table 4. Overall performance comparison of the out-of-sample predictions by using the three methods.

Method	Average MAPE (±SD)	Average Computation Time
Single-ARIMA	9.2% (±3.2%)	0.4s
AOS-ARIMA	9.0% (±2.9%)	36.5s
CARIMA	9.1% (±3.1%)	0.5s

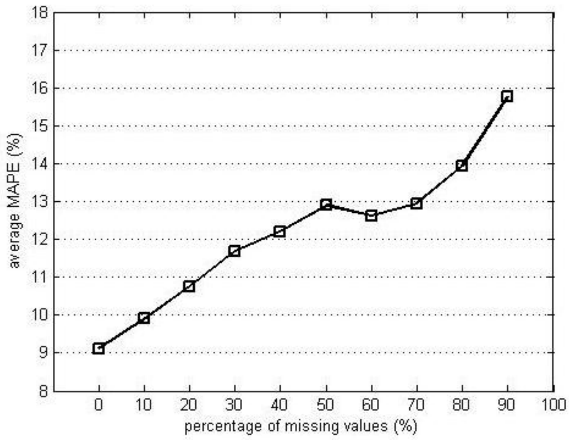
Table 5. Out-of-sample prediction performances of the three categories by using CARIMA method.

Category	Average MAPE (±SD)
Stationary series	9.3% (±3.3%)
Trend series	7.9% (±1.5%)
Seasonal series	8.6% (±2.5%)

To test for robustness of the models, the data were noised by introducing AWGN. As the database has been cleaned, we manually substituted data points with so-called missing values. Figure 5 gives the plots of the signal-to-noise ratio (SNR) (a) and the percentage of missing values (b) versus the prediction performance (average 1-SAF MAPE). The average MAPE of the noisy data is still under 10% when the SNR of the noisy data is 15dB. In addition, when we introduce 12% missing values in our time series, there is still a low mean MAPE of about 10%. Therefore, in this study, the CARIMA models are found to be robust to PAL data contaminated with noise and missing values.



(a)



(b)

Figure 5. Prediction performances of the PAL data with AWGN (a) and with missing values (b).

4. Conclusion

The prediction of human physical activity data in a wellbeing program was studied. The data were collected by a body-worn activity monitor with a built-in single triaxial accelerometer in real-life situations. ARIMA models were employed to predict future physical activity based on past data from previous weeks. The model estimation was done via a Kalman recursion. The CARIMA method was proposed to select the orders of the model and performed effectively in a large reduction in computation time needed with only a little loss in prediction accuracy in comparison to the automatic-order-selection ARIMA method. In average, the MAPEs of 9.3%, 7.9% and 8.6% were achieved for the stationary, trend and seasonal series in the data set. The prediction models were tested to be robust against noisy data (with AWGN at SNR=15dB) and missing values (at a level of 12%). A high precision in predicting next week physical activity of a participant in the study of this paper allows for a well-informed and timely motivation trigger. In addition, it can also aid in deciding whether or not a participant will be successful during the program. In other words, it can be predicted to what extent a participant will achieve next week physical activity targets or the target set at the end of the program and adapt coaching accordingly.

Acknowledgment

The authors would like to thank the New Wellness Solutions Direct Life team for the great and fruitful co-operation and thank Dr. B. Yin (Philips Research Laboratories) for the insightful comments.

References

[1] J.R. Penedo and F.J. Dahn, Exercise and well-being: review of mental and physical health benefits associated with physical activity, *Current Opinion in Psychiatry* 18 (2005), 189–193.

[2] P. Campagna, *Physical Activity Levels and Dietary Intake of Child and Youth in the Province of Nova Scotia*. Nova Scotia Dept. of Education, Health Promotion and Protection, 2005.

[3] L.J. Ware, R. Hurling, O. Bataveljic, W.B. Fairley, L.T. Hurst, P. Murray, L.K. Rennie, E.C. Tomkins, A. Finn, R.M. Cobain, A.D. Pearson, and P.J. Foreyt, Rates and Determinants of uptake and use of an internet physical and weight management program in office and manufacturing work sites in England: Cohort study, *Journal of Medical Internet Research* 10(4) (2008), e56.

[4] A.H.Goris and R. Holmes, The Effect of a Lifestyle Activity intervention program on improving physical activity behavior of employees, *Proceedings of the 3rd International Conference on Persuasive Technology*, PERSUASIVE, Oulu, Finland (2008).

[5] J. Lacroix, P. Saini, and R. Holmes, The Relationship between Goal Difficulty and Performance in the Context of a Physical Activity Intervention Program, *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI)*, Amsterdam, the Netherlands (2008), 415-418.

[6] G. Plasqui, A.M. Joosen, A.D. Kester, A.H. Goris, and K.R. Westerterp, Measuring free-living energy expenditure and physical activity with triaxial accelerometry, *Obesity Research* 13 (2005), 1363-1369.

[7] R. Hurling, M. Catt, M. De Boni, B.W. Fairley, T. Hurst, P. Murray, A. Richardson, and J.S. Sodhi, Using internet and mobile phone technology to deliver an automated physical activity program: Randomized controlled trial, *Journal of Medical Internet Research* 9(2) (2008), e7.

[8] B.P. Clarkson, Life Patterns: Structure from Wearable Sensor. *PhD Thesis*, MIT Media Lab, 2002.

[9] X. Long, B. Yin, and R.M. Aarts, Single-Accelerometer-Based Daily Physical Activity Classification. *31st Annual International IEEE EMBS Conference*, Minneapolis, MN (2009).

[10] C.V.C. Bouten, K.R. Westerterp, M. Verduin, and J.D. Janssen, Assessment of energy expenditure for physical activity using a triaxial acceleromter, *Journal of Medicine & Science in Sports & Exercise* 26(12) (1994), 1516-1523.

[11] M.J. Mathie, A.C. Coster, N.H. Lovell, and B.G. Celler, Detection of daily physical activities using a triaxial accelerometer. *Journal of Medical & Biological Engineering & Computing* 41 (2003), 296-301.

[12] G.E.P. Box, and G.M. Jenkins, *Time Series Analysis Forecasting and Control*, Holden-Day, San Francisco, CA, 1976.

[13] M.P. Clements, and D.F. Hendry, *Forecasting Economic Time Series*, Cambridge University Press, 1998.

[14] P. Pai, and C. Lin, A hybrid ARIMA and support vector machines model in stock price forecasting, *Omega* 33(6) (2005), 497-505.

[15] K. Yurekli, A. Kurunc, and F. Ozturk, Application of linear stochastic models to monthly flow data of Kelkit stream, *Journal of Ecological Modelling* 183(1) (2005), 67-75.

[16] S.D. Campbell, and F.X. Diebold, Weather Forecasting for Weather Derivatives, *Journal of the American Statistical Association*, 100(469) (2005), 6-16.

[17] U. Helfenstein, Box-Jenkins modelling in medical research, *Journal of Statistical Methods in Medical Research* 5(1) (1996), 3-22.

[18] A. Earnest, M.I. Chen, D. Ng, and L.Y. Sin, Using autoregressive integrated Moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore, *BMC Health Services Research* 5:36 (2005).

[19] B.M. Williams, P.K. Durvasula, and D.E. Brown, Urban Freeway Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models, *Journal of the Transportation Research Board*, 1644 (1998), 132-141.

[20] A.M. Khan, Y.K. Lee, and T.-S. Kim, Accelerometer Single-based Human Activity Recognition Using Augmented Autoregressive Model Coefficients and Artificial Neural Nets, *30th Annual International IEEE EMBS Conference*, Vancouver, BC (2008), 5172-5175.

[21] S. Halim, I.N. Bisono, Melissa, and C. Thia, Automatic Seasonal Auto Regressive Moving Average Models and Unit Root Test Detection, *IEEE International Conference on Industrial Engineering and Engineering Management*, Singapore (2007), 1129-1133.

[22] V. Gómez, and A. Maravall, Automatic Modelling Methods for Univariate Series, *Banco de España Working Paper*, no. 9808 (1998).

[23] R.E. Kalman, A new approach to linear filtering and prediction problems, *Journal of Basic Engineering* 82 (1960), 35-45.

[24] World Health Organization, Energy and nutrient requirements. *Report of a joint FAO/WHO/UNI Expert Consultation*, Rome (2001).

[25] P.S. Shetty, C.J. Henry, A.E. Black, and A.M. Prentice, Energy Requirements of Adults: An Update on Basal Metabolic Rates (BMRs) and Physical Activity Levels (PALs). *European Journal of Clinical Nutrition* 50(1) (1996), S1-S23.

[26] B.L. Bowerman and R.T. O'Connell, *Forecasting and Time Series: An Applied Approach*. Duxbury Press, Belmont, CA, 1993.

[27] D.A. Dickey, and W.A. Fuller, Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association* 74 (1979), 427-431.

[28] A.C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, 1989.

[29] J.Y. Peng, and J.A.D. Aston: A MATLAB Software Implementation for Time Series Analysis by State Space Methods, *Proceedings of the American Statistical Association Business and Economics Section*, American Statistical Association (2006).

[30] J. Durbin, S.J. Koopman: *Time Series Analysis by State Space Methods*, Oxford University Press, New York, 2001.

[31] A. Meyler, G. Kenny, T. Quinn: Forecasting Irish Inflation Using ARIMA Models. *Technical Paper*, Central Bank and Financial Services Authority of Ireland 3/RT/98 (1998), 1-48.

[32] H. Akaike: A New look at Statistical Model Identification, *IEEE Transactions on Automatic Control* AC-19 (1974), 716-723.

[33] G. Schwarz: Estimating the Dimension of a Model, *Annual of Statistics* 6 (1978), 461-464.

[34] C.M. Hurvich, C.L. Tsai: Regression and Time Series Model Selection in Small Samples, *Biometrika* 76(2) (1989), 297-307.

[35] Z. Chik: Performance of Order Selection Criteria for Short Time Series, *Pakistan Journal of Applied Sciences* 2(7) (2002), 783-788.

[36] G. Ljung, G. Box: On a Measure of Lack of Fit in Time Series Models, *Biometrika* 66 (1978), 67-72.

[37] M. Merrington: Table of Percentage Points of the t-Distribution, *Biometrika* 32 (1941), 300.

[38] D. Stockton, J. Glassman: An Evaluation of the Forecast Performance of Alternative Models of Inflation, *Review of Economics and Statistics* 69(1) (1987), 108-117.

[39] A.C. Acock: Working With Missing Values, *Journal of Marriage and Family* 67 (2005), 1012-1028.