Behaviour Monitoring and Interpretation – BMI
B. Gottfried and H. Aghajan (Eds.)
IOS Press, 2011

doi:10.3233/978-1-60750-731-4-131

Predicting Daily Physical Activity in a Lifestyle Intervention Program

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

Abstract. The growing number of people adopting a sedentary lifestyle these days creates a serious need for effective physical activity promotion programs. Often, these programs monitor activity, provide feedback about activity and offer coaching to increase activity. Some programs rely on a human coach who creates an activity goal that is tailored to the characteristics of a participant. Throughout the program, the coach motivates the participant to reach his personal goal or adapt the goal, if needed. Both the timing and the content of the coaching are important for the coaching. Insights on the near future state on, for instance, behaviour and motivation of a participant can be helpful to realize an effective proactive coaching style that is personalized in terms of timing and content. As a first step towards providing these insights to a coach, this chapter discusses results of a study on predicting daily physical activity level (PAL) data from past data of participants in a lifestyle intervention program. A mobile body-worn activity monitor with a built-in triaxial accelerometer was used to record PAL data of a participant for a period of 13 weeks. Predicting future PAL data for all days in a given period was done by employing autoregressive integrated moving average (ARIMA) models on the PAL data from days in the period before. By using a newly proposed categorized-ARIMA (CARIMA) prediction method, we achieved a large reduction in computation time without a significant loss in prediction accuracy in comparison with traditional ARIMA models. In CARIMA, PAL data are categorized as stationary, trend or seasonal data by assessing their autocorrelation functions. Then, an ARIMA model that is most appropriate to these three categories is automatically selected based on an objective penalty function criterion. The results show that our CARIMA method performs well in terms of PAL prediction accuracy (~9% mean absolute percentage error), model parsimony and robustness.

Keywords. tailored activity intervention, persuasive technology, activity monitoring, ARIMA, Kalman recursion, order selection, PAL, prediction, wellbeing, health, behaviour change

Introduction

An active lifestyle has large beneficial effects on people's mental and physical health [1]. Both the engagement in intense physical activity (e.g., endurance sports or strength exercises) and the engagement in various types of brief, low to moderately

[1] Corresponding author: Steffen Pauws, Philips Research Laboratories, steffen.pauws@philips.com

intense, everyday physical activities (e.g., walking) contribute to a better mental and physical health [2]. Nevertheless, a growing number of people worldwide fails to meet the recommended levels of physical activity to maintain good health. An inactive lifestyle has been associated with health problems such as obesity, diabetes and cardiovascular disease and hence a reduction in quality of life. These health problems result in a significant burden to the healthcare system [3]. As a consequence of this worrisome situation, the development of effective physical activity intervention programs has become a major focus area [4][5]. For participants of these programs, the provision of feedback on activity level against personal targets has proven to be highly important for staying motivated and attaining goals [5][6]. To track activities of the participants, a mobile body-worn triaxial accelerometer is often used. It is an inexpensive, effective and feasible device with minimal discomfort to the participant and tracks activity by measuring acceleration of body movements [7][8][9][10][11]. The relationship between energy expenditure due to physical activity and body acceleration measured by such a triaxial accelerometer has been evaluated under various controlled and free-living conditions [7][8][12][13]. These studies demonstrated the validity and usefulness of a triaxial accelerometer to measure daily physical activity.

Many physical activity intervention programs have failed to be effective in realizing significant and sustainable changes in activity behaviour [14][15]. They rely on mass intervention by providing a generic solution for the entire participant base that does not accommodate individual differences within variables underlying behaviour change. In contrast to mass interventions, individual one-to-one behaviour change interventions by a human coach are often more effective. The main reason is that a human coach has insight into the participant's performance and characteristics and can provide feedback and support regarding specific motivational dips or barriers that the participant encounters. In this way, the coach can employ a personalized approach to optimize behaviour change by, for instance, setting and adapting individual targets for the participant, providing participant-relevant feedback and adjusting interaction mode to the participant preferences. However, human coaching support can be labour intensive, which limits the number of participants that a human coach can reach. To increase the coach's reach, we propose to provide the coach with insightful information about a participant's performance by predicting his near-future activity levels. In particular, we introduce a modelling technique that predicts the most likely next-period daily physical activity level (PAL) of a participant from his past physical activity patterns. Such information can be helpful for a coach to provide effective coaching in terms of timing and content. For example, when it is likely that a participant will fail to reach a personal target for the next week, a proactive human coaching intervention can reach a participant immediately before the target is actually missed, for instance, by sending an additional motivational trigger or by lowering the target.

For predicting future PAL data, we use autoregressive integrated moving average models (ARIMA) also well-known as the Box-Jenkins methodology for time series data analysis [16]. Unfortunately, this method takes an unacceptably long computation time due to the need of repeatedly choosing the most appropriate model from a large number of candidate models. Therefore, we applied a categorized ARIMA (CARIMA) method in which the times series under study are classified in one of three categories (i.e., stationary data, trend-wise data and seasonal data). Then, the most appropriate model for each category is chosen beforehand instead of computed from a pool of candidate models. The parameter estimation of the models is based on the maximum likelihood (ML) method via a Kalman recursion [17].

The remainder of the chapter is organized as follows: Section 1 presents some related work and possible extensions of the work. Section 2 presents the data collection. The ARIMA method is explained in Section 3. Section 4 discusses the results of the ARIMA modeling and prediction. Conclusions are provided in the final section.

1. Related work

The Box-Jenkins method has already been widely used in a number of other areas such as economic time series forecasting [18][19], ecological and weather prediction [20][21], medical monitoring [22][23], traffic flow prediction [24], and also physical activity recognition [25]. Generally, the application of ARIMA models is mostly focused on predicting a single univariate time series. For multi univariate series, an automatic method has been developed [27][28].

Besides providing insightful information to a coach in a physical activity intervention program, the proposed work can also be applied in telecare or home telemonitoring services. In these healthcare services, elderly patients or those with chronic conditions are helped to maintain their independence and continue living in their own homes by means of communication technologies. For instance, event prediction is a crucial element for providing high-quality telecare. In the case of telemonitoring chronic health failure patients, a timely prediction of a worsening patient condition based on daily measurements of body weight, heart rate and blood pressure is essential to prevent re-hospitalizations or any critical situation. In the case of a senior wandering service, analysis of urban dwelling behaviour of cognitively impaired elderly using GPS data is essential to track outdoor mobility of these elderly.

Lastly, data from a single tri-axial accelerometer measuring human physical activity also allow for automatic classification of typical daily physical activities such as walking, running, cycling or driving [26]. This helps participant and coach to effortlessly link activity patterns with a type of activity.

2. Data Collection

In a lifestyle physical activity intervention program[2], participants were provided with a Philips DirectLife activity monitor containing a built-in triaxial accelerometer to measure the acceleration data of their activities performed throughout the day. This device is a small (3.2 x 3.2 x 0.5 cm), light-weight (12.5 g) instrument (see Figure 1). It is waterproof up to 30 meters depth, and has a battery life of 3 weeks and an internal memory that can store data for up to 22 weeks. The device can be worn on the chest with a key cord, on the waist, or in the trouser pocket in an arbitrary orientation. The features of the device have been designed to enhance unobtrusiveness of wearing and to reduce the interference of the monitoring system with spontaneous activity behaviour. During the monitoring period, the activity monitor was connected several times to a personal computer, using Universal Serial Bus (USB) communication, and the recorded data were uploaded, processed and stored using dedicated software.

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

Figure 1. The Philips DirectLife activity monitor.

The output of the activity monitor is expressed as activity counts per minute, which are the running time summations of absolute output values from the three uniaxial accelerometers in the device. Consecutive counts were summed to arrive at counts per day. By using the correspondence between activity counts and total energy expenditure, daily Physical Activity Level (PAL) data values are calculated using a linear regression model on activity counts and a measure for basal metabolic rate corrected for age, height and body mass [7][8].

The physical activity intervention program was primarily enrolled at different locations in the Netherlands throughout the year with a high participation in the months November, December and January. Each participant in the program, which lasted 13 weeks, took part in one assessment week and 12 intervention weeks. During the assessment week, participants learned to use the activity monitor and completed personal data (e.g., age, gender, weight). The PAL data monitored during the assessment week serve as a baseline on the basis of which a personal activity goal was set to work towards, during the 12 weeks of the program. In total, there were 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 participation in the physical activity intervention program. For the majority of the adult population, daily PAL data lie between 1.2 and 2.5 [30]. A daily PAL around 1.2 corresponds to a sedentary activity level. A PAL below 1.2 indicates that the activity monitor was not been worn that day. A PAL higher than 1.7 corresponds to a healthy activity level. A PAL higher than 2.5 corresponds to a vigorous activity level. An extremely high level of physical activity leads to a PAL value as high as 4.5, but this level of activity is only achieved in extreme situations (e.g., professional sports) [29]. In this study, we treated the extremely low and extremely high PAL values (outliers: $PAL<1.2$ or $PAL>5$) as missing data. Outliers and missing data are generated by non-modelled mechanisms such as not wearing the activity monitor, a flat battery, monitor noise or other disturbances. They can lead to ARIMA model misspecification and bad prediction performance. Therefore, the database was cleaned up by removing all time series contaminated with missing data (and outliers). Of a total of 950 time series, 227 were kept for further study. The impact of noise and missing data on ARIMA prediction accuracy is treated separately (in Section 2.3) in a systematic fashion.

3. 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 a 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.

3.1. 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 [16]. The model can be written as below

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

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

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

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

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

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 [31]. 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 $y_1, y_2, \ldots, y_t$, 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. Figures 2, 3 and 4 show typical examples of a stationary, a trend-wise and a seasonal time series and their ACFs. One would expect that trend and seasonal components coexist in physical activity time series over twelve weeks. For instance, a person can steadily improve his activity levels but still follow a consistent week routine. We did not observe these patterns in the current data set.

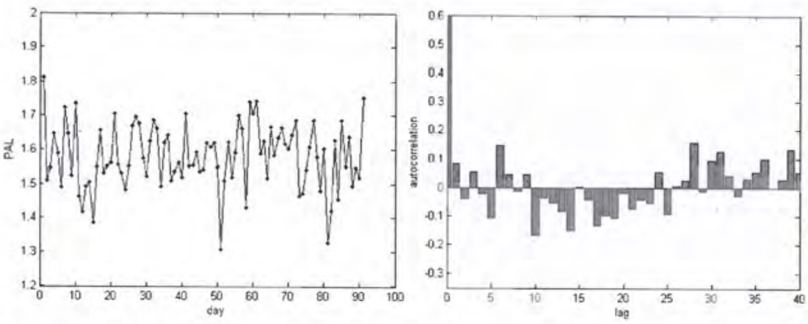

Figure 2. The plots of daily PAL times series that demonstrates stationary behaviour over 91 days (left-hand side) and its corresponding ACF (right-hand side).

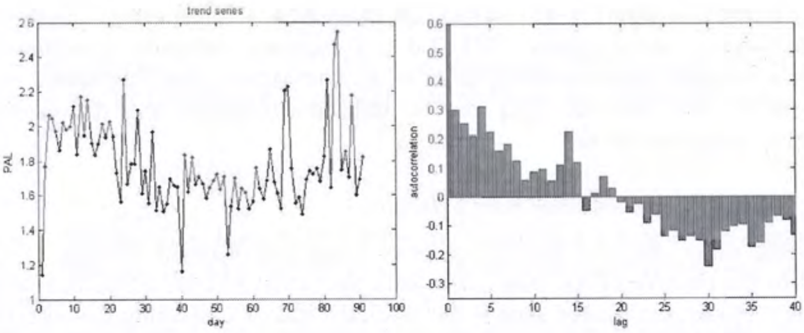

Figure 3. The plots of daily PAL time series that contain a down-ward and up-ward trend over 91 days (left-hand side) and its corresponding ACF (right-hand side).

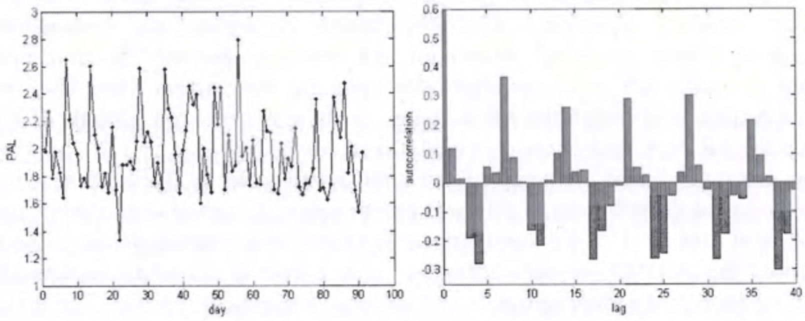

Figure 4. The plots of daily PAL time series containing a seasonal (weekly) trend (left-hand side) and its corresponding ACF (right-hand side).

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

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

Table 1. Critical absolute *t*-values for identified spikes in the ACF [31]

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

For identifying significant spikes at lags of an ACF automatically, various critical absolute t-values are suggested [31]. Table 1 indicates the spike identification to examine stationarity and seasonality in an ACF. Alternatively, the Augmented Dickey-Fuller (ADF) unit root test [32] can be utilized to identify the stationarity and seasonality of a given series.

- *Model Estimation and Order Selection*

In this study, the well-known recursive algorithm, Kalman filtering [17], 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 [33][34].

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 [35]. 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 [35].

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 [36]. 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) [37] and Schwarz's Information Criterion (SIC) [38] 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 [39][40]. 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 a maximum order (2,d,2)(1,D,1) are tested under SIC for model selection.

- *Diagnostic Checking and Residual Analysis*

The residuals $r = \{r_1, r_2, \ldots, r_n\}$ 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 [41] 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 [35]. 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 [42] 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,

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

3.2. 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 4 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 5. The automatic order selection method based on a penalty function criteria for the ARIMA models has been frequently used [27][28]. This automatic-order-selection ARIMA (AOS-ARIMA) method searches for the best model for each individual time series evaluating 36 candidate models; this is computationally 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 percentage error (MAPE),

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

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

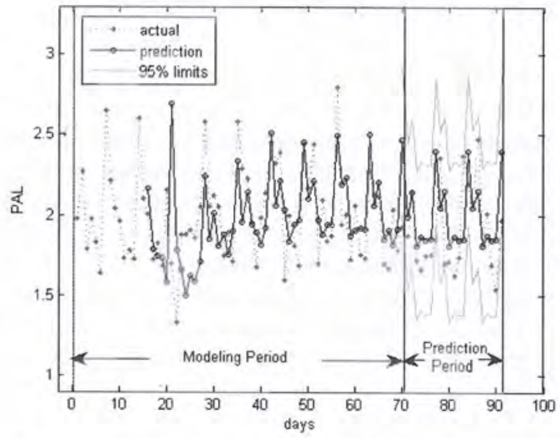

Figure 5. Predicting a (seasonal) PAL time series from 70 observed data points.

3.3. Model Robustness

The robustness of a prediction model is important for real applications because data in practice contains noise, missing values and outliers. In the areas of economics, ARIMA models have proven to be robust [43]. However, for the PAL data, the robustness of the models on noise, missing values, and outliers needs to be assessed.

The noise of the PAL data mainly comes from the background environment and the device itself during 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 deal with incomplete data; this approach simply pads the mean of all non-missing values when calculating the ACF of the data [44]. During the prediction procedure, the missing data were treated as predictions in the same way [35]. 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.

4. Results and Discussion

The ACFs of the PAL data were identified and the results show that the percentage of stationary series amounts to ~83% (188 out of 227) in the clean dataset, whereas the percentages of trend and seasonal time series are ~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 take significantly less computation time than the AOS-ARIMA method. In addition, CARIMA achieved lower average MARR than Single-ARIMA for the 227 PAL time series in this study.

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.

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

Figure 6 shows that for the stationary and trend series, the CARIMA and Single-ARIMA methods performed similarly 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.

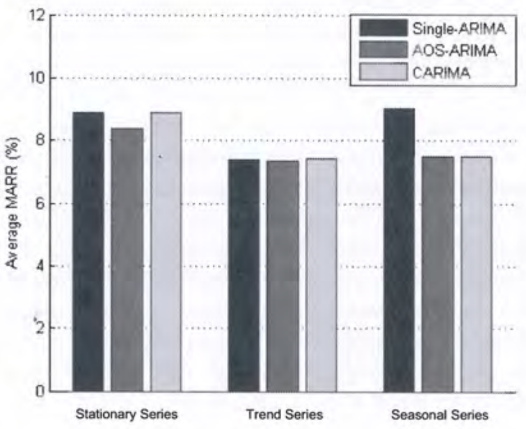

Figure 6. Average MARRs (%) of the data fitting for the stationary, trend and seasonal PAL series by using Single-ARIMA (*black bars*), AOS-ARIMA (*dark gray bars*) and CARIMA (*light gray bars*) methods.

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 fit 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.

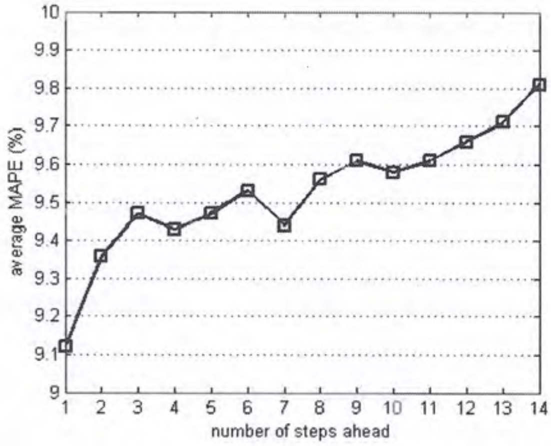

Figure 7. Average MAPEs (%) of multiple-step-ahead predictions by using CARIMA method.

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 its high 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 7. 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 8 gives the plots of various signal-to-noise ratios (SNR) and various percentages of missing values versus prediction performance (average 1-SAP 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.

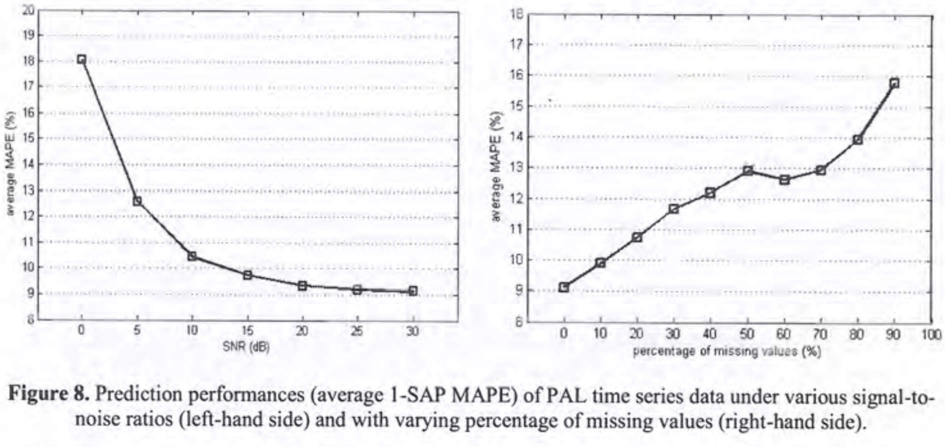

Figure 8. Prediction performances (average 1-SAP MAPE) of PAL time series data under various signal-to-noise ratios (left-hand side) and with varying percentage of missing values (right-hand side).

5. Conclusion

The results presented in this chapter demonstrate the feasibility of rather precise predictability of human physical activity behaviour. The prediction of physical activity data was studied in the context of a lifestyle intervention program. Daily physical activity data of a large number of program participants were collected by a body-worn activity monitor with a built-in single triaxial accelerometer. ARIMA models were employed to predict physical activity in future weeks based on past data from previous weeks. The model estimation was done via a Kalman recursion. Subsequently, the CARIMA method was employed to select the orders of the model. This led to a large reduction in computation time needed with only a minor decline in prediction accuracy in comparison to the traditional, automatic-order-selection ARIMA method. Using CARIMA, MAPEs of 9.3%, 7.9% and 8.6% were achieved for the stationary, trend-wise and seasonal series in the data set, respectively. The CARIMA prediction models were shown to be robust against noisy data (with AWGN at SNR=15dB) and missing values (at a level of 12%). The high precision prediction of near future physical activity of participants in a lifestyle intervention program opens opportunities for timely personalized coaching. It provides insight into future goal achievement and may be indicative of changes in motivation levels. These predictions can be helpful for coaches to optimize the content and timing of their coaching, thereby enabling them to offer more personalized and effective coaching support.

Acknowledgments

The authors would like to thank the Philips DirectLife team for the great and fruitful co-operation and Bin Yin, Wim Stut and John Lamb (Philips Research Laboratories) for their 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] W.L. Haskell, I.M. Lee, R.R. Pate, K.E. Powell, S.N. Blair, B.A. Franklin, C.A. Macera, G.W. Heath, P.D. Thompson and A. Bauman. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Circulation 116*, 9, (2007) 1081-93.

[3] 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.

[4] 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.

[5] 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).

[6] 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.

[7] A.G. Bonomi, G. Plasqui, A.H. Goris and K.R. Westerterp. Estimation of free-living energy expenditure using a novel activity monitor designed to minimize obtrusiveness, *Obesity advance online publication, February 25, 2010*; doi:10.1038/oby.2010.34.

[8] 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.

[9] 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.

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

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

[12] 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.

[13] 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.

[14] M.H. Van den Berg, J.W. Schoones and T.P.M. Vliet-Vlieland. Internet-based physical activity interventions: a systematic review of the literature. *Journal of Medical Internet Research, 9(3):e26* (2007).

[15] W. Zhu. Promoting physical activity through internet: a persuasive technology view. In: *Proceedings of the Second International Conference on Persuasive Technology* (Palo Alto, CA, USA, April 26-27 2007). PERSUASIVE 2007, LNCS 4744, 12-1.

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

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

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

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

[20] 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.

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

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

[23] 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).

[24] 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.

[25] 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.

[26] X. Long, B. Yin and R.M. Aarts, Single-Accelerometer-Based Daily Physical Activity Classification. *Conference Proceedings of 31st Annual International Conference of the IEEE Eng. Medicine and Biology Society (EMBS)*, Minneapolis, Minnesota, USA, Sept 2-6, 2009, 6107-6110.

[27] 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.

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

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

[30] 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.

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

[32] 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.

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

[34] 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).

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

[36] 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.

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

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

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

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

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

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

[43] 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.

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

Ambient Intelligence and Smart Environments 9

Behaviour Monitoring and Interpretation - BMI

Well-Being

Edited by

Björn Gottfried

Hamid Aghajan

IOS Press

Behaviour Monitoring and Interpretation - BMI

The notion of well-being is one which is crucial to many aspects of our daily lives. In addition to providing one of the cornerstones to a healthy lifestyle, the concept of well-being extends to the selection of the type of environment we live in, our interaction with other people and the things we do to realize our plans for the future. Well-being is so intrinsic to our daily lives that it plays a fundamental role at all times and in all places, so it is important that it is taken into account when designing the ubiquitous computing technologies which pervade our lives nowadays.

This book contains the papers presented at the 2009 Behaviour Monitoring and Interpretation (BMI) workshop, third in the annual series launched in 2007, and co-located with the German conference on Artificial Intelligence. The focus of the 2009 workshop on the topic of well-being reflects the significant interest in this area of research, and the book offers state-of-the-art contributions in the application area of well-being from diverse disciplines such as engineering and philosophy.

This overview of a wide range of research projects will be of interest to researchers and developers whose work necessitates a consideration of well-being, either as part of the task of implementing current applications or of designing the systems of the future.

ISBN 978-1-60750-730-7 (print)
ISBN 978-1-60750-731-4 (online)
ISSN 1875-4163 (print)
ISSN 1875-4171 (online)

9 781607 507307