

Towards a Generalized Regression Model for On-body Energy Prediction from Treadmill Walking

Harshvardhan Vathsangam, Adar Emken, E. Todd Schroeder, Donna Spruijt-Metz and Gaurav S. Sukhatme

Abstract—Walking is a commonly available activity to maintain a healthy lifestyle. Accurately tracking and measuring calories expended during walking can improve user feedback and intervention measures. Inertial sensors are a promising measurement tool to achieve this purpose. An important aspect in mapping inertial sensor data to energy expenditure is the question of normalizing across physiological parameters. Common approaches such as weight scaling require validation for each new population. An alternative is to use a hierarchical model to model subject-specific parameters at one level and cross-subject parameters connected by physiological variables at a higher level. In this paper, we evaluate an inertial sensor-based hierarchical model to measure energy expenditure across a target population. We first determine the optimal physiological parameter set to represent data. Weight is the most accurate parameter ($p < 0.1$) measured as percentage prediction error. We compare the hierarchical model with a subject-specific regression model and weight exponent scaled models. Subject-specific models perform significantly better ($p < 0.1$ per subject) than weight exponent scaled models at all exponent scales whereas the hierarchical model performed worse than both. We study the effect of personalizing hierarchical models using model results as initial conditions for training subject-specific models with limited training data. Using an informed prior from the hierarchical model produces similar errors to using a subject-specific model with large amounts of training data ($p < 0.1$ per subject). The results provide evidence that hierarchical modeling is a promising technique for generalized prediction energy expenditure prediction across a target population in a clinical setting.

Index Terms—Accelerometer, Bayesian Linear regression, Gyroscope, Hierarchical Linear Model

I. INTRODUCTION

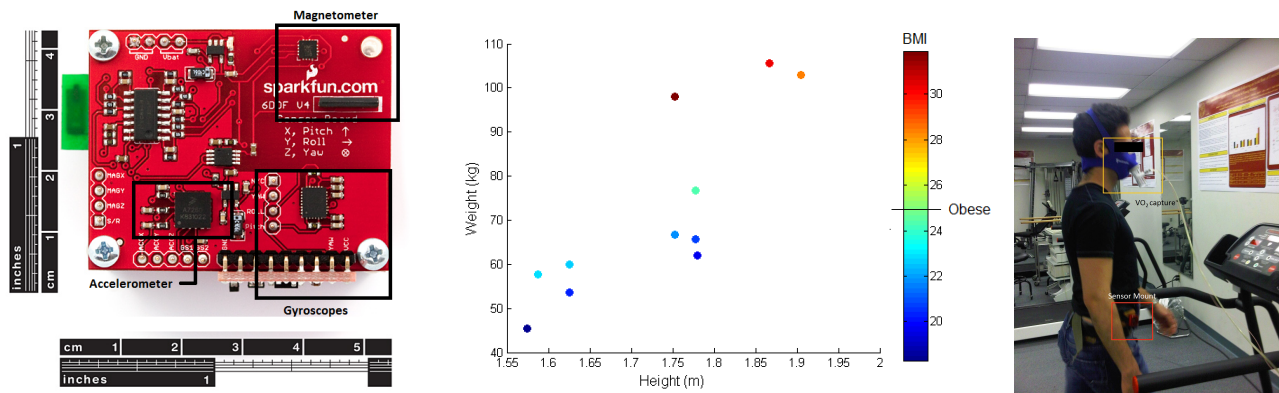
Regular physical activity is important due to its role in weight control, reducing risk of cardiovascular disease, type 2 diabetes, some cancers and in improving mental health and bone strength [1]. An easily available practice to maintain an active lifestyle is to walk regularly [2]. Characterizing energy expenditure from walking would provide a valuable input in characterizing activity intensity and duration and in the assessment of activity-based intervention measures.

Inertial sensors such as accelerometers and gyroscopes have gained traction as promising tools for the ubiquitous detection and quantification of energy expenditure from human movement [3, 4, 5, 6]. With reference to energy expenditure from walking, one common technique is to relate inertial sensor features to ground truth values using a regression technique [7]. An important aspect in mapping movement to energy expenditure across populations is the question of normalizing for differences in physiological parameters such as height, weight, leg length, age, sex etc. One of the most important and well studied physiological parameters for estimating energy expenditure is weight. However, weight is not a good indicator of the participant.

A common technique to account for weight is to scale $\dot{V}O_2$ values by a suitable weight exponent. Most common scaling coefficients include a range from 0.6 – 1.0 [8], the most popular being 0.67 [9], 0.75 [10] and 1.0 [11, 12, 13]. A common issue with weight scaling is determining the appropriate scaling coefficient across a target population. Scaling coefficients vary across age groups and stages of development in individuals [10]. Therefore, a different scaling coefficient has to be validated for each study. Likewise, the effect of other physiological parameters such as sex, stride length and heart rate on energy expenditure during walking also have to be incorporated [11]. The challenge is to fuse all these parameters into a unified framework while maintaining the simplicity of standard regression techniques.

One way to extend the capability of single linear regression is by Hierarchical Linear Modeling (HLM) [14]. Given a target population with a set of participants, HLM based techniques use linear models at levels within and across participants. We adopt this approach in our study by designing a multilevel HLM. At one level we have participant specific models relating inertial sensor features to energy expenditure. At a second level we capture the inter-dependence of different subject-specific models on physiological parameters using a (second) regression model. The advantages of such an approach are many. Using a second level to capture commonalities across subjects allows the separation of the dependence on physiological parameters from participant-specific inertial sensor data. Such an approach also allows flexibility in deciding the right combination of physiological parameters to represent participants. Training this model allows joint modeling of cross-level interactions in a population. Most importantly, HLM allows one to generate informed participant-specific models using only higher level information. This is an advantage when limited or no data is available for a new participant. Thus we retain all the benefits of subject-specific monitoring using linear regression while capturing the generalizability across populations.

This paper describes an experimental evaluation of a Hierarchical Linear Model to predict energy expenditure using on-body, wireless inertial sensors for level treadmill walking. We describe sensor hardware and method of capturing human movement in Sec. II. We present the details of a subject-specific Bayesian Linear Regression model and an HLM to map movement information to energy expenditure in Sec. III. We present experimental results identifying the best representative physiological parameters to predict energy expenditure in Sec. IV-A. The model with the optimal physiological parameters is compared with subject-specific models in Sec. IV-B. We examine the effect of personalizing of HLMs using them as informative initial conditions for subject-specific models with



(a) Sensor board used to collect data. Data streams consist of triaxial accelerometer and gyroscopic information. Source:www.sparkfun.com (b) Illustration of population statistics plotted as Weight (vs) Height for each subject. Each point is color coded by BMI. Average weight = 72 ± 21 kgs. Average height = 1.73 ± 0.11 m. (c) An example recording procedure for a single participant.

Figure 1: Illustration of hardware, ground truth collection and population statistics

limited training data in Sec. IV-C Both models are in turn compared with weight exponent scaled models in Sec. IV-D. In doing so, we provide a technique for the normalization of energy expenditure prediction using inertial sensors across a population for treadmill walking in a clinical setting. Our approach builds on earlier work in developing data-driven subject-specific regression models [15].

II. CAPTURING HUMAN MOVEMENT

A. Hardware

Human movement was captured using a modified Sparkfun 6DoF Inertial Measurement Unit (IMU) [16, 17] worn on the right iliac crest. The sensor provided 6 sensor streams conveying triaxial acceleration (Freescale MMA7260Q tri-axial accelerometer) and triaxial rotational rates (2 Invensense IDG300 gyroscopes) allowing translational and rotational motion capture in all three planes – sagittal, frontal and transverse. Data were sampled at 100 Hz and transmitted via Bluetooth (RN-41 Bluetooth module) to a nearby PC. Figure 1a illustrates the sensor board used for collecting data.

B. Representation of human walk

We use the principle that steady state human walk is cyclic in nature [18] to represent walking. Inherent periodicity was captured using a single inertial sensor worn above the iliac crest on the right hip. Sensor data corresponded directly to the accelerations and rotational rates of the hip in the sensor’s local frame of reference. A natural candidate to represent the periodicity of walk is the Fourier transform of sensor streams [19].

C. Data collection

Nine healthy adults (Five men, four women) participated in the study. Height and weight of each participant were recorded using a Healthometer balance beam scale. Figure 1b illustrates participant statistics using a Weight (vs) Height graph. Each point is color coded by Body Mass Index (BMI). Informed written consent was obtained from participants and the study was approved by the Institutional Review Board, University of Southern California. Participants had average weight = 72 ± 21 kgs and average height = 1.73 ± 0.11 m.

Rate of oxygen consumption ($\dot{V}O_2$, mL/min) as measured using the MedGraphics Cardio II metabolic system with BreezeSuite v6.1B (Medical Graphics Corporation) was the representation of energy expenditure. Before each test, the flow meter was calibrated against a 3.0 L syringe and the system was calibrated against O_2 and CO_2 gases of known concentration. This system outputs data at the frequency of every breath. Figure 1c illustrates a typical recording procedure. Each participant was asked to walk at 5 speeds (2.5, 2.8, 3.0, 3.3, and 3.5 mph) on a motorized treadmill for 7 minutes of recording time per speed. At each speed transition, $\dot{V}O_2$ readings were allowed to stabilize for 2 minutes prior to the start of data collection. Speeds were chosen based on the Compendium of Physical Activities [20].

Each sensor stream from the IMU was passed through a lowpass filter with 3dB cutoff at 6 Hz. The cutoff frequency was chosen based on manual inspection of the Fourier transforms and keeping in mind that everyday activities fall in the frequency range of 0.1-10 Hz [21]. Each of the 7 minute streams from the IMU were divided into 10 second intervals or epochs. The 10 second interval was chosen based on previous successful implementations [4] on this time scale. Thus each subject’s data consists of roughly 210 data points. Within each epoch, feature vectors were extracted from each sensor stream by computing the 1024 point FFT for the signal stream in each epoch. The FFTs for all sensor streams within an epoch were then concatenated to obtain the complete feature vector. Data for each user was thus a set of points consisting of epochs, each containing concatenated FFT features from the IMU and the average rate of oxygen consumption ($\dot{V}O_2$) for that epoch. These represent per-user data while walking at five different speeds.

III. ENERGY ESTIMATION ACROSS POPULATIONS

Our approach to energy expenditure estimation involves learning data-driven user models and using these models to predict energy expenditure for unseen data points. One candidate techniques is to train an individualized parametric model Bayesian Linear Regression (BLR) model for each participant and energy consumption for unseen data points relevant to that participant alone. Such models are not likely to

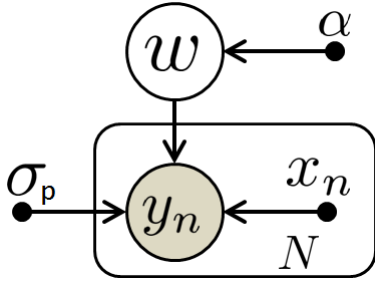


Figure 2: Illustration of individualized energy prediction model. Training the model is equivalent to finding optimal w to maximize data likelihood. A distribution is assumed over $w : w \sim \mathcal{N}(w : 0, \alpha^{-1}\mathbf{I})$. α is a hyperparameter.

be successful for unseen data points of another participant. To model cross-participant dependencies, one has to incorporate joint modeling of participant data based on physiological parameters (e.g., weight and height). This is done through a Hierarchical Linear Model (HLM).

A. Subject-specific energy estimation models with Bayesian Linear Regression

We first describe the individualized regression model. A linear relationship between the feature vector and target energy expenditure is assumed. We adopt a Bayesian framework through Bayesian Linear Regression (BLR). The approach is similar to our earlier work in predicting energy expenditure [4]. Fig. 2 represents a graphical model based plate representation of the conventional BLR technique.

In the context of energy prediction, for a given person p , with input feature dataset $\mathbf{X}_p = \{x_{n_p}\}_{n=1}^{N_p}$ and target energy values $\mathbf{Y}_p = \{y_{n_p}\}_{n=1}^{N_p}$, assuming a linear model, we have for each data pair:

$$y_{n_p} = \mathbf{w}_p^T x_{n_p} + \epsilon, \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma_p^{-2}) \quad (1)$$

where ϵ is a noise parameter and $\mathbf{w}_p = (w_{0_p}, \dots, w_{M-1_p})^T$ are the model weights. Training a model amounts to learning the weights \mathbf{w}_p and the noise parameters σ_p . Using the properties of Gaussians, we have

$$p(y_{n_p} | x_{n_p}, \mathbf{w}_p, \sigma_p) = \mathcal{N}(y_{n_p}; \mathbf{w}_p^T x_{n_p}, \sigma_p^{-2}) \quad (2)$$

Bayesian Linear Regression (BLR) [22] adopts a Bayesian approach to linear regression problem by introducing a prior probability distribution, specifically a Gaussian prior, $p(\mathbf{w}_p) = \mathcal{N}(\mathbf{w}_p; \mathbf{0}, \alpha^{-1}\mathbf{I})$ over the model parameters \mathbf{w}_p in Eq. 1 where α is a hyperparameter. The optimal prediction for a new data point is given by the predictive distribution:

$$p(y_{*p} | x_{*p}, \mathbf{Y}_p, \alpha, \sigma_p) = \mathcal{N}(\mathbf{m}_N^T x_{*p}, \sigma_N^2(x_{*p})) \quad (3)$$

$$\text{and } \sigma_N^2(x_{*p}) = \sigma_p^{-2} + x_{*p}^T \mathbf{S}_{N_p} x_{*p} \quad (4)$$

$$\mathbf{m}_{N_p} = \sigma_p^2 \mathbf{S}_N \mathbf{X}_p^T \mathbf{Y}_p \quad (5)$$

$$\text{and } \mathbf{S}_{N_p}^{-1} = \alpha \mathbf{I} + \sigma_p^2 \mathbf{X}_p^T \mathbf{X}_p \quad (6)$$

Model parameters are estimated using an iterative approach by finding the best α and σ_p to maximize the evidence function and finding the best parameters $\hat{\mathbf{w}}$ to maximize the likelihood given a fixed α and σ_p alternately until convergence. This technique provides a subject-specific BLR model that can be used on one participant alone.

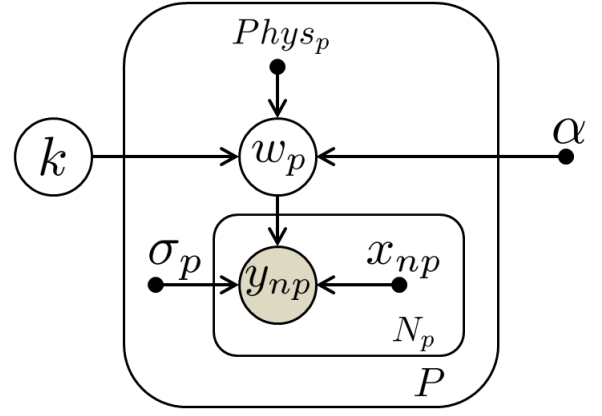


Figure 3: Illustration of generalized model for energy expenditure prediction for P people and N_p training points per person. Each person has a subject-specific weight w_p that is influenced by physiological parameters such as height, weight and age through k and in turn influences energy prediction given an input point x_{np} . The task is to learn w_p and k given instances of each person's physiological parameters and energy training data per person.

B. Generalized Energy Estimation Models with Hierarchical Regression

1) *Model description:* We build on the individualized model to obtain an HLM incorporating physiological parameters. Consider a test population consisting of P participants. For each participant p , Let there be N_p data points collected, consisting of the energy values $\mathbf{Y}_p = \{y_{1_p}, y_{2_p}, \dots, y_{N_p}\}$ and D -dimensional Fourier transform inputs $\mathbf{X}_p = \{x_{1_p}, x_{2_p}, \dots, x_{N_p}\}$ as detailed in Sec. II-C. We denote $\mathbf{Y} = \{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_P\}$ and $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_P\}$ to be complete set of training data for all participants. Let each participant have a physiological parameters determined by \mathbf{Phys}_p and the complete set be $\mathbf{PHY} = \{\mathbf{Phys}_p\}_{p=1}^P$. These include parameters such as height, weight and age (with a constant term for bias). We model top-down dependence of \mathbf{w}_p on each participant's physiological parameters \mathbf{Phys}_p through a universal weight parameter k . Each \mathbf{w}_p in turn influences energy predictions y_{n_p} for an input x_{n_p} . Fig. 3 illustrates the plate representation of this model. Similar to Sec. III-A, each output energy value, y_{n_p} is linearly dependent on input x_{n_p} . This can be expressed as:

$$\mathbf{Y}_p \sim \mathcal{N}(\mathbf{Y}_p; \mathbf{w}_p^T \mathbf{X}_p, \sigma_p^{-2}\mathbf{I}) \quad (7)$$

$$\text{or } y_{n_p} \sim \mathcal{N}(y_{n_p}; \mathbf{w}_p^T x_{n_p}, \sigma_p^{-2}) \quad (8)$$

We assume that \mathbf{w}_p is not a point estimate but a distribution with linear dependence of each participant's parameter \mathbf{w}_p on k and \mathbf{Phys}_p :

$$\mathbf{w}_p \sim \mathcal{N}(\mathbf{w}_p; \mathbf{k}^T \mathbf{Phys}_p, \alpha^{-1}\mathbf{I}) \quad (9)$$

It is worth noting that both \mathbf{w}_p and k are hidden variables which need to be estimated from data. Variable k is also not a point estimate but has a prior distribution $k \sim \mathcal{N}(k; \mathbf{0}, \sigma^{-2}\mathbf{I})$. Each \mathbf{w}_p is now an informative prior dependent on the person's physiological parameters \mathbf{Phys}_p through k . We denote $\mathbf{W} = \{\mathbf{w}_p\}_{p=1}^P$

Training the multilevel model is equivalent to learning individual \mathbf{w}_p 's, the overall parameter \mathbf{k} as well as the noise parameters $\{\sigma_p\}_{p=1}^P, \alpha$ and σ . The HLM combines P local regression models in two ways. First, the local regression coefficients \mathbf{w}_p determine energy values for each person. Second, the different coefficients are connected through the group-level model parameter \mathbf{k} . Intuitively, the HLM captures the inherent similarity in walking across different people while accounting for individual walking styles and energy consumption.

2) *Likelihood evaluation*: We aim to find the optimal parameters that maximize the likelihood of each energy prediction y_{n_p} for each person p , given the input data points x_{n_p} and the physiological parameters \mathbf{Phys}_p . This likelihood can be written as:

$$l = p(\mathbf{Y}|\mathbf{X}, \mathbf{PHY}) \quad (10)$$

$$= \int p(\mathbf{Y}, \mathbf{W}|\mathbf{X}, \mathbf{PHY}) d\mathbf{W} \quad (11)$$

$$= \int p(\mathbf{Y}|\mathbf{W}, \mathbf{X}, \mathbf{PHY}) p(\mathbf{W}|\mathbf{X}, \mathbf{PHY}) d\mathbf{W} \quad (12)$$

$p(\mathbf{W}|\mathbf{X}, \mathbf{PHY})$ can be expressed in terms of hidden variable \mathbf{k} as:

$$\begin{aligned} p(\mathbf{W}|\mathbf{X}, \mathbf{PHY}) &= \int p(\mathbf{W}, \mathbf{k}|\mathbf{X}, \mathbf{PHY}) d\mathbf{k} \\ &= \int p(\mathbf{W}|\mathbf{X}, \mathbf{k}, \mathbf{PHY}) p(\mathbf{k}|\mathbf{X}, \mathbf{PHY}) d\mathbf{k} \end{aligned} \quad (13)$$

From the graphical model in Fig. 3, the probabilities $p(\mathbf{Y}|\mathbf{W}, \mathbf{X}, \mathbf{PHY})$ and $p(\mathbf{W}|\mathbf{X}, \mathbf{k}, \mathbf{PHY})$ can be broken down into individual distributions as:

$$p(\mathbf{W}|\mathbf{X}, \mathbf{k}, \mathbf{PHY}) = \prod_{p=1}^P p(\mathbf{w}_p|\mathbf{X}_p, \mathbf{k}, \mathbf{Phys}_p)$$

$$p(\mathbf{Y}|\mathbf{W}, \mathbf{X}, \mathbf{PHY}) = \prod_{p=1}^P \prod_{n_p=1}^{N_p} p(y_{n_p}|x_{n_p}, \mathbf{w}_p, \mathbf{Phys}_p)$$

From the model graph, we can infer $\mathbf{w}_p \perp\!\!\!\perp \mathbf{X}_p|\mathbf{k}, \mathbf{Phys}_p$, $y_{n_p} \perp\!\!\!\perp \mathbf{Phys}_p|x_{n_p}, \mathbf{w}_p$ and $\mathbf{k} \perp\!\!\!\perp \mathbf{Phys}_p, \mathbf{X}_p$. Thus:

$$p(\mathbf{W}|\mathbf{X}, \mathbf{k}, \mathbf{PHY}) = \prod_{p=1}^P p(\mathbf{w}_p|\mathbf{k}, \mathbf{Phys}_p) \quad (14)$$

$$p(\mathbf{Y}|\mathbf{W}, \mathbf{X}, \mathbf{PHY}) = \prod_{p=1}^P \prod_{n_p=1}^{N_p} p(y_{n_p}|x_{n_p}, \mathbf{w}_p) \quad (15)$$

$$p(\mathbf{k}|\mathbf{X}, \mathbf{PHY}) = p(\mathbf{k}) \quad (16)$$

Substituting Eqs. 14, 15 and 16 into Eqs. 12 and 13, we have:

$$\begin{aligned} l &= p(\mathbf{Y}|\mathbf{X}, \mathbf{PHY}) = \\ &= \int \prod_{p=1}^P \prod_{n_p=1}^{N_p} p(y_{n_p}|x_{n_p}, \mathbf{w}_p, \mathbf{Phys}_p) p(\mathbf{W}|\mathbf{X}, \mathbf{PHY}) d\mathbf{W} \\ p(\mathbf{W}|\mathbf{X}, \mathbf{PHY}) &= \int \prod_{p=1}^P p(\mathbf{w}_p|\mathbf{k}, \mathbf{Phys}_p) d\mathbf{k} \end{aligned} \quad (17)$$

The pair of equations represented by [17] represent the likelihood of the observations \mathbf{Y} given physiological parameters \mathbf{PHY} and inputs \mathbf{X} . Maximizing the log-likelihood, $\mathbb{L} = \log l$ is equivalent to finding the optimal \mathbf{w}_p, \mathbf{k} and respective

noise parameters that maximize these equations. Of particular interest is parameter $\mathbf{k} \in \mathbb{R}^D$ which helps generate a person dependent weight \mathbf{w}_p given only the physiological parameters. The probabilities $p(y_{n_p}|x_{n_p}, \mathbf{w}_p)$ and $p(\mathbf{w}_p|\mathbf{k}, \mathbf{Phys}_p)$ are defined by Eqs. 8 and 9 respectively. For the class of exponential distributions in the absence of prior information, there is no closed form solution for $\{\mathbf{w}_p\}_{p=1}^P$ and \mathbf{k} . Hence approximation techniques are required.

3) *Algorithm description*: We propose an EM like algorithm [23] to learn the parameters $\{\mathbf{w}_p\}_{p=1}^P$ and \mathbf{k} . Our original aim was to maximize the likelihood $\mathbb{L} = \log p(\mathbf{Y}|\mathbf{X}, \mathbf{PHY})$. We approximate the likelihood term to incorporate the Maximum A posteriori estimates (MAP estimates) of individual weights \mathbf{w}_p denoted by $\hat{\mathbf{w}}_p$. Each $\hat{\mathbf{w}}_p$ is now a point estimate assumed to be known and can be interpreted as a parameter that has to be optimized. From Eq. 9, the MAP estimate corresponds to the mean of each \mathbf{w}_p . The modified algorithm maximizes the incomplete log likelihood $\log p(\mathbf{Y}|\hat{\mathbf{W}}, \mathbf{X}, \mathbf{PHY})$ by maximizing the expected complete log likelihood $\langle \log p(\mathbf{Y}, \mathbf{Z}|\hat{\mathbf{W}}, \mathbf{X}, \mathbf{PHY}) \rangle$. This expectation is written as:

$$\begin{aligned} Q(\hat{\mathbf{W}}, \hat{\mathbf{W}}^{old}) &= \\ &= \sum_{\mathbf{k}} \log p(\mathbf{Y}, \mathbf{Z}|\hat{\mathbf{W}}, \mathbf{X}, \mathbf{PHY}) p(\mathbf{Y}, \mathbf{Z}|\hat{\mathbf{W}}, \mathbf{X}, \mathbf{PHY}) \end{aligned}$$

We treat \mathbf{k} as a hidden variable that needs to be estimated and hence $\mathbf{Z} = \mathbf{k}$. Correspondingly, we have the following steps:

a) *Initialization*: Initialize $\hat{\mathbf{w}}_p, \sigma_p, \alpha$ to appropriate values. A good initialization condition is one obtained using least squares estimates.

b) *E step*: Evaluate $p(\mathbf{k}|\{\hat{\mathbf{w}}_p^{old}\}_{p=1}^P, \{\mathbf{Phys}_p\}_{p=1}^P)$. For this, we take advantage of the linear dependence of each person's weight vector $\hat{\mathbf{w}}_p$ given their physiological parameters \mathbf{Phys}_p through \mathbf{k} . This means we have a dataset of P target $\hat{\mathbf{w}}_p$'s and their corresponding input variables \mathbf{Phys}_p 's. We can thus frame a regression problem from a person's physiological parameter variable \mathbf{Phys} to a weight variable $\hat{\mathbf{w}}$ (of which there are P examples) through \mathbf{k} . Given that $\mathbf{k} \in \mathbb{R}^D$ and there are no cross-correlations, we frame D separate linear regression problems. The m^{th} regression problem, maps the physiological parameters to the m^{th} component of $\hat{\mathbf{w}}$ through the m^{th} component of \mathbf{k} . Thus we have D separate linear regression problems mapping physiological parameters \mathbf{Phys}_p to \mathbf{w}_{EM} through \mathbf{k} in a component-wise manner. We solve each of these regression problems in a BLR framework similar to that described in Sec. III-A to obtain a mean and variance measure for \mathbf{k} .

c) *M step*: Maximize likelihood of the dataset given \mathbf{k} . This corresponds to re-estimating parameters using the current value of \mathbf{k} . The learned \mathbf{k} from the E-step is used to estimate individual $\hat{\mathbf{w}}_p$'s given their physiological parameters \mathbf{Phys}_p . From the linear relationship as defined by Eq. 9, we have:

$$\hat{\mathbf{w}}_p = \mathbf{k}^T \mathbf{Phys}_p \quad (18)$$

We use this estimate as initial conditions and maximize the likelihood of the dataset. Given \mathbf{k} (which is fixed after the E-step), this is equivalent to maximizing individual likelihoods of

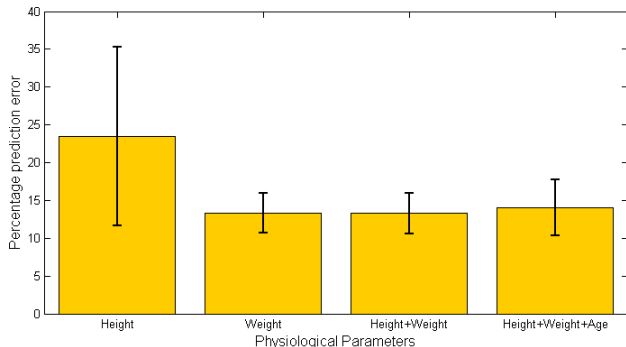


Figure 4: Illustration of prediction errors (expressed as percentage of ground truth) with different combinations of physical parameters. Maximum errors were obtained when only height was used at feature vector. Among features chosen weight is the single best physiological parameters to estimate k . Combining weight and height marginally improved performance and adding age degraded performance.

each of the participants. Maximizing the individual likelihoods is the same as finding the optimal w_p given N_p data pairs $\{x_{n_p}, y_{n_p}\}_{n_p=1}^{N_p}$. This is equivalent to solving P individual Bayesian Linear Regression problems with the initial conditions of w_p 's as defined above and finding the optimal w_p 's.

d) *Evaluate log likelihood:* The total log likelihood is the sum of the P individual log likelihoods found from the M step. We check for convergence of log likelihood and if not, repeat the E and M steps again. Using this algorithm we learn a generalized energy prediction model that maps physiological parameters to subject specific weights w_p and uses these weights to predict energy expenditure for each subject.

IV. EXPERIMENTAL VALIDATION

Given P participants, the subject-specific model simply trains P separate regression models using the technique outlined in Sec. III-A. Whenever the generalized and subject-specific models were compared for participant p , a percentage of subject p 's data were partitioned into testing and training data. The generalized model was trained using all data from the remaining $P - 1$ participants and energy predictions were made for participant p 's test data. Making a prediction corresponded to learning the optimal k , using this k to predict individual w_p 's for and using this w_p to predict energy expenditure for unseen data points. Simultaneously, a subject-specific BLR model was trained using participant p 's training data. This model was then used to predict energy expenditure for the same test data as the generalized model. Errors for both models were measured as percentage deviation from the ground truth value output from the metabolic cart. For each participant, this was repeated 20 times with different randomly sampled training and testing data and at different percentages of training data. This was repeated for all participants and errors were averaged across participants.

A. Choosing Optimal Physiological Parameters

Given the HLM, our first study examined the best set of physiological parameters that minimized energy expenditure

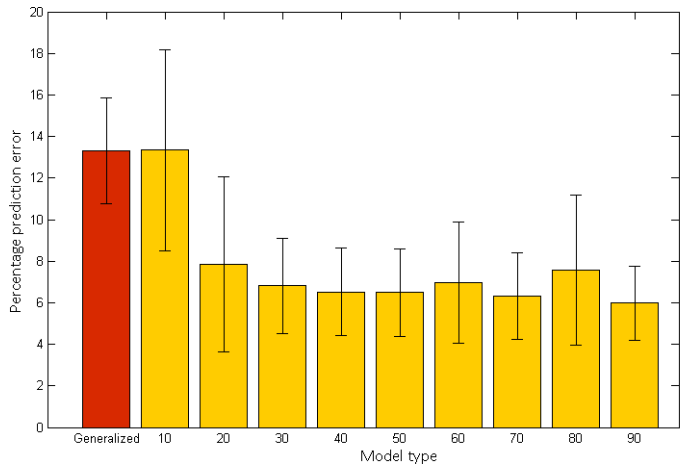


Figure 5: Comparison of errors obtained from the HLM (red) with individualized subject-specific models averaged across all participants (yellow). The generalized model showed similar errors to subject-specific models with 10% of training data. With more training data available, subject-specific models outperformed the generalized model.

prediction errors. The combinations of parameters used were Height, Weight, Weight + Height, Weight + Height + Age. Maximum errors were obtained when only height was used at feature vector. The best individual physiological input for this population was weight. Combining weight and height only marginally improved performance and adding age degraded performance. Based on the above results, it was decided to use weight as the only physiological parameter (with a constant term for bias). Fig. 4 illustrates the errors obtained.

B. Comparison with Subject-specific Modeling

An important question is how the HLM compares with subject-specific models. The HLM showed comparable performance for all percentages of training data and hence only one bar is shown. The HLM showed comparable errors to subject specific models with 10% of training data used. With more training data available, subject specific models outperformed the HLM. As is expected, the availability of more training data allows stronger modeling capability of subject-specific models. Using an HLM however is still equivalent to training a subject-specific with small amounts of data and might be a preferable trade-off in cases where subject-specific models are not available. It can be shown that predictions from a BLR model (used in estimating k 's) approach the ideal value with increasing large amounts of training data. Hence it is expected that as the target population is expanded, the HLM will perform competitively with the subject-specific approach. Fig. 5 illustrates the relative errors obtained when an HLM (shown in red) trained using data from $P - 1$ people is compared with a subject-specific model for the p^{th} person with varying training data (shown in yellow).

C. Personalization of HLM with Limited Training Data

Given the superior performance of subject-specific models, this section of the study explores whether weights obtained using the generalized model could be combined with limited training data to obtain an equivalent (if not better) model

in terms of prediction accuracy. Such an approach would be beneficial because of the potential in using less training data for training subject-specific models. Given the learned \mathbf{k} , for an unknown subject, one can predict the subject specific weight $\hat{\mathbf{w}}_p$ using Eq. 18. One can then train a subject-specific model with this $\hat{\mathbf{w}}_p$ as an initial condition and limited subject-specific data. Fig. 6 illustrates percentage errors obtained with varying amounts of training data. Training a model with an initial condition (red) shows smaller errors than when no such condition is used (yellow). This effect is most pronounced when small amounts of training data are used ($p < 0.1$ per subject). It can also be seen from the figure that using an initial condition with small amounts of training data shows comparable errors to training with no initial condition with substantial training data. The use of this technique suggests that while the HLM by itself produced higher errors than a subject-specific model, the lowest errors can be obtained by using the \mathbf{k} from the model to obtain a subject specific weight $\hat{\mathbf{w}}_p$ and using this $\hat{\mathbf{w}}_p$ as an informative prior with small amounts of training data.

Fig. 7 illustrates the relative predictive capacities of the HLM and subject-specific model as applied to one participant. It can be seen that while the generalized model makes as good $\dot{V}O_2$ predictions as the subject-specific model in the middle energy range, the model breaks down when predicting lower or higher energy ranges. This can be understood as follows, the HLM has to simultaneously fit model parameters across participants and within each participant. Most participants' walking styles differ the most at their lowest and highest walking speeds. Typically, the higher the speed, the larger is the energy consumed. Intuitively, given that the generalized model has to tradeoff between overall accuracy and subject-specific accuracy, the parameters are optimized over the most similar looking input points which occur in the middle ranges of speeds of each participant. This is typical of the bias-variance tradeoff in learning such models.

D. Comparison with Weight-dependent Models

Current normalizing approaches model weight dependence by using power law scaling. $\dot{V}O_2$ values are scaled by a weight exponent of the form W^s . This section of the study focused on how the HLM and subject-specific models compare against such scaling based approaches. Here, all participants were considered to represent a single dataset. Instead of modeling dependence on physiological parameters, each $\dot{V}O_2$ value was scaled by a suitable weight exponent W^s . This represented the scaled $\dot{V}O_2$ value, $\dot{V}O_{2,scaled} = \frac{\dot{V}O_2}{W^s}$ for that feature-energy data pair. The input features were the same as previous models. This unified dataset was divided into training and testing data and regression models were trained. This was repeated for different randomly sampled data and percentages of training and testing data. Various exponent coefficients in the range $0.6 < s < 1.6$ were used in increments of 0.1 and percentage errors (as defined in previous section) were recorded. Thus for different combinations of training data and exponent coefficients, we have an error surface. Fig. 8a illustrates such a surface. From the figure, it can be seen that there exists an optimal exponent-percentage combination basin that produces

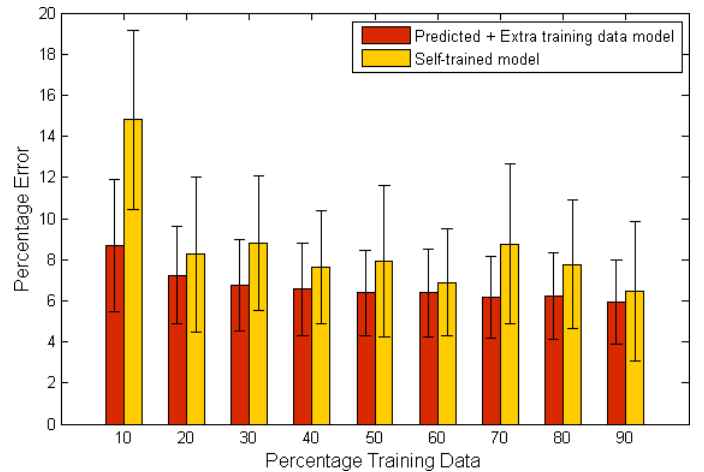


Figure 6: Illustration of the effect of adding a small amount of training data to the HLM. An informative initial condition from physiological parameters is used to train a subject-specific model, shown in red. For comparison, a subject-specific model (shown in yellow) with no such initial condition is also trained. Using an HLM with initial conditions along with a small percentage of training data produces similar errors to using a subject-specific model with large amounts of training data ($p < 0.1$ per subject).

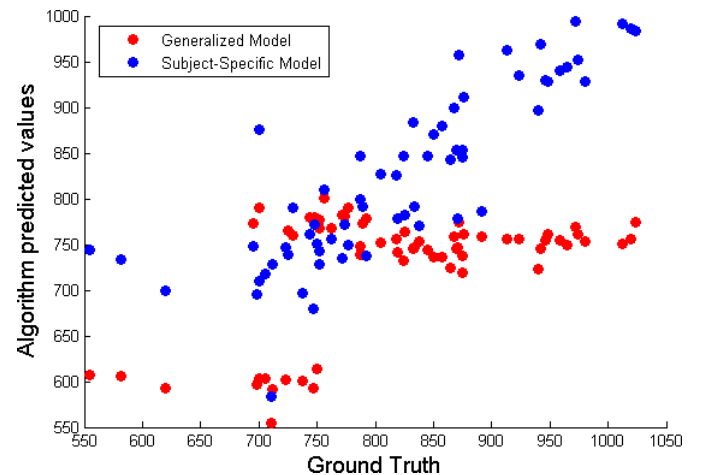
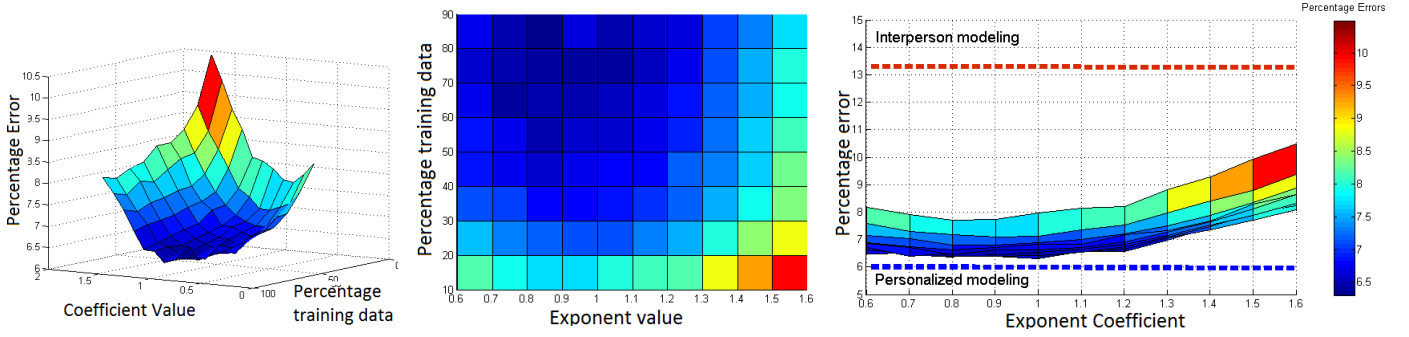


Figure 7: Illustration of the predicted values versus ground truth using both generalized algorithms (red) and subject-specific algorithms (blue) for a single participant. Similar plots exist for other participants. HLMs perform poorly in the end regions. This could be because the parameters are optimized over the most similar looking input points. These occur in the middle ranges of speeds of each participant.

the least error. Fig. 8b shows the same surface as seen from above. This represents a color plot of errors for various training percentage-exponent coefficient combinations. Lowest errors ($\approx 6.5\%$) were seen with an exponent coefficient of $0.7 - 1.0$ and a large percentage of training data ($> 50\%$). This exponent coefficient value corresponded to previous research indicating that the optimal value is approximately $0.65 - 1.0$ [10].

Fig. 8c illustrates a comparative performance between power law scaled approaches, generalized model and subject-specific model based approaches for different power coef-



(a) Illustration of percentage errors obtained as a function of weight exponent used and training data available. There exists an optimal exponent-percentage basin that produces the least error. (b) A colormap illustration of percentage errors as a function of training data and weight exponent extracted from the previous figure. Lowest errors of $\approx 6.5\%$ are seen when an exponent coefficient of $0.7 - 1.0$ and a large percentage of training data ($> 50\%$) are used. (c) Illustration of the training errors as a function of different exponents as compared with subject-specific models and HLM. Subject-specific models outperform weight exponent scaled models regardless of exponent. Generalized models perform worse than subject-specific or weight exponent scaled models.

Figure 8: Comparison of HLM and subject-specific models with weight-scaled linear approaches. $\dot{V}O_2$ values are scaled by a weight exponent W^s where $0.6 < s < 1.6$ and probabilistic linear models are trained, these are compared against the subject-specific and generalized model described in this study.

ficients at all training percentages. Subject-specific models outperformed weight exponent scaled models irrespective of exponent. This is because each subject-specific model used training data only from that particular participant. This would naturally result in a better fit than any form of weight scaling across all participants. Generalized models performed worse than subject-specific or weight exponent scaled models.

V. CONCLUSION

Regular physical activity has a number of health benefits and walking is a commonly available activity to maintain an active lifestyle. Accurately tracking and measuring calories burned from walking is important to help improve user feedback and designing of effective intervention measures. The last decade has seen the emergence of inertial sensors along with pattern recognition techniques to detect, characterize and quantify physical activity in general and walking specifically. An issue with using inertial sensor data to estimate energy expenditure is how data can be normalized across varying physiological parameters such as height, weight, age etc. Common approaches such as weight scaling require validation across each new target population. An alternative is to extend the capability of standard linear regression through Hierarchical Linear Modeling (HLM). Using an HLM, at one level we have participant specific models relating inertial sensor features to energy expenditure. At a second level we capture the interdependence of different subject-specific models on physiological parameters using a second regression model. This paper explored the validation of an inertial sensor based HLM to measure energy expenditure while accounting for variations in physiological parameters, specifically weight. The model was described and an EM-like approach to learn parameters was detailed. Our chief contributions are summarized next:

Flexibility in modeling physiological parameters: The use of HLMs allows flexibility in incorporating new physiological parameters. By placing these parameters at a higher level, one can easily switch, add or remove various combinations

of parameters and examine their effects on prediction accuracy. Examining the relative weight coefficients also helps determine which parameter dominates predictions. In our study, weight was the single best physiological parameter. Maximum errors were obtained when only height was used at feature vector. Combining weight and height only marginally improved performance. Adding age degraded performance.

Comparison with subject-specific and weight-scaled modeling: Using a second level to capture commonalities across subjects allows generalized modeling from participant-specific inertial sensor data. We compared the modeling capability of a hierarchical model with a subject-specific linear regression model and weight exponent scaled models. The generalized model showed similar errors to subject-specific models with 10% of training data used. Subject-specific models performed better than weight exponent scaled models for all exponent scales. An important insight from our approach was that generalized models were as good as the subject-specific model in the middle energy range for each subject but broke down when predicting lower or higher energy ranges. This is most likely because most subjects exhibit similar walking patterns in the mid-speed ranges.

Accurate models with sparse data: In many data collection studies, researchers often have to deal with inadequate or unequal amounts of data from a subsection of participants. A subject-specific model trained with sparse data might not be as statistically significant as a model trained with copious data. Capturing inter-participant dependencies through a higher level of modeling allows researchers to effectively “transfer” model information from those participants for whom extensive data are available to those where only limited data are available. In our study, an informative initial condition initialized from physiological parameters was used as an initial condition in training a subject-specific model with limited data. For comparison, a subject-specific model with no such initial condition was also trained with the same training data. Using an informative initial condition produced similar errors to a subject-specific model with large training data. The use of a

generalized model as an initial condition for subject-specific models allows for accurate modeling with smaller amount of training data. This is also useful for obtaining more accurate models for participants with sparse data sets.

VI. FUTURE WORK

We plan to expand our work in a number of directions. The most important extension is to test our algorithm across a much larger population to examine whether a larger dataset improves the performance of the HLM measured against the weight-scaled and subject-specific models. Testing across a larger population will also allow us to test the effects of other physiological parameters such as height, stride length and sex on prediction accuracies. The advantage of our approach is that one can simply incorporate these parameters as additional features in the top level of the hierarchical. We also aim to test the algorithm in free-living conditions across common activities such as walking, sitting and standing. Finally, we plan to test algorithms that learn the parameters \mathbf{k} and \mathbf{w}_p simultaneously. These include Gibbs-sampling algorithms, variational approximations and belief-propagation based approaches.

VII. ACKNOWLEDGEMENTS

This work was supported in part by Qualcomm, Nokia, NSF (CCR-0120778) as part of the Center for Embedded Networked Sensing (CENS), and the USC Comprehensive NCMHD Research Center of Excellence (P60 MD 002254). Support for H. Vathsangam was provided by the USC Annenberg Doctoral fellowship program.

The authors would like to thank David Erceg of the Division of Biokinesiology and Physical Therapy, USC, for his invaluable support and guidance.

APPENDIX

Table I: Glossary of Terms

Term	Description	Dimension
x_{n_p}	n^{th} input data point for person p	$1 \times D$
y_{n_p}	n^{th} target $\dot{V}O_2$ for person p	1×1
\mathbf{w}_p	Weight parameter for person p	$1 \times D$
\mathbf{k}	Universal weight parameter	$1 \times D$
$Phys_p$	Physiological parameters	$1 \times (M + 1)$
σ_p	Subject-dependent noise parameter	1×1
σ	Prior variance for \mathbf{k}	1×1
α	Mapping variance for w_p	1×1
N_p	Number of data points for person p	–
\mathbf{X}_p	Collection of input data for person p	$N_p \times D$
\mathbf{Y}_p	Collection of $\dot{V}O_2$ values for person p	$N_p \times 1$
\mathbf{X}	Collection of all input data points	$P \times N_p \times D$
\mathbf{Y}	Collection of all $\dot{V}O_2$ values	$P \times N_p \times 1$
PHY	Collection of all physiological values	$P \times (M + 1) \times 1$

REFERENCES

- [1] D. Warburton, C. Nicol, and S. Bredin, "Health benefits of physical activity: the evidence," *CMAJ*, vol. 174(6), pp. 801–809, 2006.
- [2] A. Dunn, B. Marcus, J. Kampert, M. Garcia, H. K. III, and S. Blair, "Comparison of lifestyle and structured interventions to increase physical activity and cardiorespiratory fitness: A randomized trial," *JAMA*, vol. 281, pp. 327–334, 1999.
- [3] K. Aminian and B. Najafi, "Capturing human motion using body-fixed sensors: outdoor measurement and clinical applications," *Computer Animation and Virtual Worlds*, vol. 15, no. 2, pp. 79–94, 2004.

- [4] H. Vathsangam, A. Emken, D. Spruijt-Metz, T. E. Schroeder, and G. S. Sukhatme, "Energy Estimation of Treadmill Walking using On-body Accelerometers and Gyroscopes," in *EMBC*, 2010.
- [5] R. P. Troiano, "Translating accelerometer counts into energy expenditure: advancing the quest.," *J Appl Physiol*, vol. 100, pp. 1107–1108, Apr 2006.
- [6] F. Albinali, S. Intille, W. Haskell, and M. Rosenberger, "Using wearable activity type detection to improve physical activity energy expenditure estimation," in *12th conference on Ubiquitous Computing*, 2010.
- [7] A. Godfrey, R. Conway, D. Meagher, and G. O'Laughlin, "Direct measurement of human movement by accelerometry," *Medical Engineering & Physics*, vol. 30, no. 10, pp. 1364 – 1386, 2008.
- [8] I. Zakeri, M. Puyau, A. Adolph, F. Vohra, and N. Butte, "Normalization of Energy Expenditure Data for Differences in Body Mass or Composition in Children and Adolescents," *J of Nutrition*, vol. 136, pp. 1371–1376, 2006.
- [9] A. Neville, R. Ramsbottom, and C. Williams, "Scaling physiological measurements for individuals of different body size," *Eur J Appl Physiol Occup Physiol*, vol. 65(2), pp. 110–7, 1992.
- [10] D. Rogers, B. Olson, and J. Wilmore, "Scaling for the vo₂-to-body size relationship among children and adults," *J Appl. Physiol.*, vol. 79(3), pp. 958–967, 1995.
- [11] R. L. Waters and S. Mulroy, "The energy expenditure of normal and pathologic gait," *Gait & Posture*, vol. 9, no. 3, pp. 207 – 231, 1999.
- [12] C. Wyndham, W. van der Walt, A. van Rensburg, G. Rogers, and N. Strydom, "The influence of body weight on energy expenditure during walking on a road and on a treadmill," *European Journal of Applied Physiology and Occupational Physiology*, vol. 29, pp. 285–292, 1971.
- [13] M. Pearce, D. Cunningham, A. Donner, P. Rechnitzer, G. Fullerton, and J. Howard, "Energy cost of treadmill and floor walking at self-selected paces," *European Journal of Applied Physiology and Occupational Physiology*, vol. 52, pp. 115–119, 1983.
- [14] A. Gelman and J. Hill, *Data Analysis Using Regression and Multi-level/Hierarchical Models*. Cambridge University Press, 2007.
- [15] H. Vathsangam, A. Emken, T. E. Schroeder, D. Spruijt-Metz, and G. S. Sukhatme, "Determining Energy Expenditure from Treadmill Walking using Hip-Worn Inertial Sensors: An Experimental Study," *IEEE Transactions on Biomedical Engineering*, 2011. Under Review.
- [16] "6DoF Inertial Measurement Unit Datasheet," tech. rep., Sparkfun Electronics, 2009.
- [17] H. Vathsangam, A. Emken, G. S. Sukhatme, and D. Spruijt-Metz, "Evaluation of A Tri-axial, Gyroscope-enhanced, And Bluetooth-enabled Accelerometer In Estimating Walking Speeds," in *ACSM Annual Meeting*, vol. 42 of *Supplement*, p. 479, Medicine & Science in Sports & Exercise, May 2010.
- [18] C. Chang, R. Ansari, and A. Khokhar, "Efficient tracking of cyclic human motion by component motion," pp. 941–944, December 2004.
- [19] H. Vathsangam, A. Emken, D. Spruijt-Metz, and G. S. Sukhatme, "Toward Free-Living Walking Speed Estimation Using Gaussian Process-based Regression with On-Body Accelerometers and Gyroscopes," in *Pervasive Health*, Mar 2010.
- [20] B. Ainsworth, W. Haskell, M. Whitt, A. Swartz, S. Strath, W. O'Brien, D. B. Jr., K. Schmitz, P. Emplaincourt, D. Jacobs, and A. Leon, "Compendium of physical activities: classification of energy costs of human physical activities.," *Med. & Sci. in Sports & Exer.*, vol. 32, pp. S498–S516, 2000.
- [21] L. Bao and S. S. Intille, "Activity Recognition from User-Annotated Acceleration Data," in *PERVASIVE* (Alois, ed.), pp. 1–17, Massachusetts Institute of Technology, Springer-Verlag GmbH, April 2004.
- [22] C. M. Bishop, *Pattern Recognition and Machine Learning*, ch. 3.3, pp. 152–156. Springer, 2006.
- [23] C. M. Bishop, *Pattern Recognition and Machine Learning*, ch. 9.4, pp. 450–454. Springer, 2006.