Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 56-2 (2023) 6504–6509

Experiment Design Considerations for Estimating Energy Expenditure during Wheelchair Propulsion

Roya Doshmanziari, Håkon Strand Aandahl
Marius Lyng Danielsson, Julia Kathrin Baumgart, Damiano Varagnolo

Abstract
We propose a methodology for estimating energy expenditure (EE) during wheelchair propulsion. The method is based on measured physiological and kinematic signals from wearable sensor devices in an experimental setup design. More specifically, we have developed regression models based on features extracted from heart rate, acceleration and gyroscope data collected during nine experiment stages with twenty participants. Support Vector regression and Gaussian process regression methods were implemented to provide an estimate of EE for each participant during the experiment. Extensive cross validation techniques were applied to evaluate the performance of the proposed models and investigate the necessity of personalizing the algorithms based on personal characteristics.

Keywords: Energy Expenditure, Manual Wheelchair Users, Inertial Measurement Units, Regression

1. INTRODUCTION

Compared to the general population, wheelchair users (WCU) have a more inactive lifestyle, which is related to an increased risk of lifestyle-related diseases such as obesity (Weil et al., 2002). Obesity is caused by an energy intake that consistently exceeds Energy Expenditure (EE). Total daily EE is made up of three components: 65-70 percent attributed to resting EE (REE), 10 percent to diet-induced dissipation of heat, and 15-30 percent to Physical Activity EE (PAEE) (Nightingale et al., 2017b). While both REE and PAEE are dependent on factors such as sex, body-mass and on the daily activity levels (da Rocha et al., 2005; Klausen et al., 1997), PAEE is the most malleable of the three components, and therefore particularly useful for obtaining a balance between EE and energy intake.

WCUs may use objective feedback on daily EE to achieve this balance, and also during lifestyle counselling with their healthcare professionals. However, gold standard methods for measuring EE (including PAEE) - such as doubly labelled water, direct or indirect calorimetry - are expensive and impractical to use over time during daily life situations (Nightingale et al., 2017a). Therefore, promising alternatives are commercial, wearable multi-sensor devices such as Apple watch, Garmin and Fitbit, which estimate EE through data-driven techniques. These wearable devices are now able to estimate EE in the general population during exercise in the most common modalities (ie. running and cycling) with moderate to high accuracy (Dannecker et al., 2013; O'Driscoll et al., 2020). While the Apple watch is one of few wearables specifically tailored to WCU (employing wheelchair propulsion as exercise modality), it was found to underestimate EE in this population with large variations in the error (Moreno et al., 2020).

These variations in the error may be related to the fact that the EE estimation algorithms do not successfully adjust for differences in WCUs' personal characteristics and the ones related to the disability, e.g., in wheelchair propulsion movement patterns and the amount of active muscle mass they are able to recruit (Glasheen et al., 2021; Nightingale et al., 2017a). There is thus the need to better individualize the algorithms for these factors, as a first step in the standardized laboratory setting, where one has good control over the experimental setup. Specific system excitation patterns may then be taken into consideration when developing the EE estimation algorithms. In other words, for training the specific algorithms data need to be collected in a way that allows ad-hoc considerations on how many different aerobic / anaerobic tests the WCUs need to perform, how much recovery is needed between tests within a single session but also between repeated sessions.

The current study aims at providing an estimate of EE based on multi-sensor data and personal characteristics. We address this by tailored regression analyses which estimate EE on the specific case of a group of able-bodied control group participants performing wheelchair propulsion on a treadmill in the standardized laboratory setting. The proposed analyses employ three main sub-models - a feature extraction from inertial measurment unit (IMU) and heart rate (HR) data based on temporal

⋆ R. Doshmanziari, M.l. Danielsson, J.k. Baumgart and D. Varagnolo are with the Norwegian University of Science and Technology in Trondheim, Norway. Hakon.S. Andahl is with the Nord University, Norway. D. Varagnolo is also with the University of Padova, Italy. Emails: {roya.doshmanziari |Marius.l.Danielsson |Julia.k.Baumgart |damiano.varagnolo} @ntnu.no, Hakon.S.Andahl@nord.no

⋆⋆For this study, ethical approval was obtained from the Norwegian Centre for Research Data (NSD). Volunteers participated in the experiments, signed a consent form and experiments were performed at the Elite Sports Science Lab (NeXtMove Core facilities), Norwegian University of Science and Technology

2405-8963 Copyright © 2023 The Authors. This is an open access article under the CC BY-NC-ND license.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2023.10.297

and spectral analysis, a dimensionality reduction model and a regression model to provide EE estimations. We evaluate generalizability of the proposed models, and assess how specific the algorithms must be tailored to personal characteristics of the participants. To this end, we investigate the dependency of the models performances on selected factors like sex, body-mass, age and self-reported fitness index (IPAQ).

Given the inherent limitations concerning time constraints and the practicalities involving data collection with WCU, the present study addresses the data collected from a control group (able-bodied participants). The contributions of the paper may thus be summarized in a set of data-driven considerations about: *a)* how many and which levels of individualization are needed when designing EE estimation algorithms for wheelchair propulsion sessions, and *b)* how to design experiments for training such EE estimation algorithms.

To convey such messages the paper is structured as follows: Section 2 describing the designed experiment setup and measured data, Section 3 presenting data treatment, feature engineering and regression models, Section 4 detailing some numerical results for the EE estimations, and Section 5 concluding the paper with some closing remarks and future research questions.

2. THE EXPERIMENTAL SETUP

The data used in this paper is based on 20 healthy, able-bodied participants performing three 4-minute stages on three separate inclines of wheelchair propulsion on a motorized treadmill (Forcelink Technology, Culemborg, the Netherlands).

More precisely, each participant performed three stages at each incline on a separate of in total three testing days (in total nine recordings per participant). Each day consisted of treadmill wheelchair propulsion at a fixed treadmill incline (0.5, 2.5 or 5 percent), and the order of the inclines for each participant was counterbalanced. Within each session three stages were performed at increasing speed, which were selected to cover a range of submaximal intensities and separated by $2-3$ minutes of rest in between, described in Table 1. Per participant, all three testing days occurred within an 14-day period; the time of testing remained consistent to minimize the effect diurnal rhythm variability. Prior to arriving at the laboratory, all participants were instructed to refrain from performing high-intensity exercise or heavy strength training 24 hours prior to testing. Additional restrictions included consumption of alcohol 24 hours prior, caffeine on the day of, or food 2 hours before testing. On the first test day and before the testing, participants provided information on their sex and age, had their body height and body mass measured, and completed the self-reported International Physical Activity Questionnaire (IPAQ) (Bauman et al., 2009). The IPAQ was used to assess habitual physical activity among participants. The IPAQ score indicated that our participants either had moderate or high physical activity levels, while no participants were included with low physical activity levels.

Table 1. Experimental setup for three testing days, different incline and speed of the treadmill (km/h).

Incline	Day 1 (0.5)		Day 2 (2.5)		Day 3 (5.0)	
	Men	Women	Men	Women	Men	Women
Stage 1	4	3	3	2	2	1
Stage 2	6	5	4	3	3	2
Stage 3	8	7	5	4	4	3

Five IMU (Gait Up Physilog®5 inertial sensor, Gait Up; Lausanne, Switzerland) were attached to the participants and the wheelchair (chest, back, forearm, seat and wheel) capturing movement data during the wheelchair propulsion stages. In addition, the participants' HR was measured by a Polar chest strap connected to a Polar M400 HR monitor watch (Polar Electro Oy, Finland). The participants were also fitted with a face-mask attached to a Vyntus ergospirometer (Vyntus CPX, Vyaire, Medical GmbH, Germany), which was calibrated for volume and against a gas mixture of 80% N_2, 15% O_2 and 5% CO_2 prior to testing each participant. Oxygen uptake and carbon dioxide production were measured continuously. EE (kcal/min) was calculated from oxygen uptake $\dot{V}O_2$ (measured in L/min) and carbon dioxide production $\dot{V}CO_2$ (L/min) based on Weir's formula as in (Weir, 1949).

$$\text{EE} = 3.94\ \dot{V}O_2 + 1.106\ \dot{V}CO_2 \tag{1}$$

Information on the personal characteristics and fitness level of the twenty included participants is summarized in Table 2 and schematized in Figure 1.

Table 2. Summary of the characteristics of the participants

Gender	Number	Age	Body Mass (kg)	Height (cm)	Body mass index (BMI) (kg/m^2)
Male	11	33 ± 11	81.9 ± 11.2	183.5 ± 8.2	24.3 ± 2.3
Female	9	34 ± 11	67.0 ± 7.9	167.3 ± 5.3	24.0 ± 2.6
Total	20	33 ± 11	75.2 ± 11.4	176.2 ± 9.9	24.2 ± 2.4

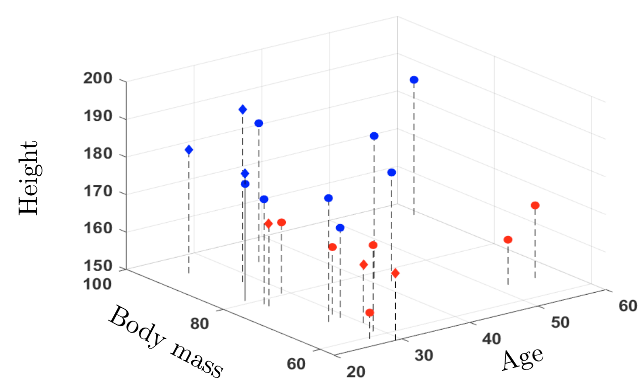

Figure 1. Distribution of personal characteristics of the participants including age, body mass, height, sex and IPAQ score (Sex: Female = red , Male = blue | IPAQ: High = diamond, Moderate = circle).

3. METHODOLOGY

We seek to find a set of data-driven considerations about: *a)* how many and which levels of individualization are needed when designing EE estimation algorithms for wheelchair propulsion sessions, and *b)* how to design experiments for training such EE tracking algorithms. To do this, we ladder on a series of regression models of EE as a function of signals such as HR, acceleration and gyroscope. The discussion below focuses on how these models were produced.

3.1 Preprocessing of Data

We consider synchronized and filtered data collected from five different IMU sensors placed on locations that from intuitive perspectives capture the motion of the torso, the arms and the wheelchair. The current study focused on datastreams from the wrist IMU sensor which were synchronized and filtered to compensate for scaling and offset errors.

We note that acceleration signals (X, Y and Z directions), sampled at a 256 Hz rate, included fast- and slow-varying components. The slow component is due to gravitational forces acting on the body, while the fast components are the result of body movement. This calls for an opportune data processing pipeline automatically separating such components. Geographically speaking, the start and stop location of the participant on the treadmill is in the same. Therefore, the average value of norm of the acceleration signals, i.e., $\|Acc\| = \sqrt{Acc_x^2 + Acc_y^2 + Acc_z^2}$ during the entire session, is assumed to be due to the gravity component and is subtracted from the norm of the acceleration signal ($\|Acc\|$). Furthermore, we consider measured EE values as our target to be estimated. Measured EE values for each record (note: with a sampling rate of 0.1 Hz) were trimmed to only include steady state EE values. More specifically, the first 70 and last 10 seconds of each EE time series were removed to exclude the non-steady state parts in the beginning and the end of each exercise stage. The mean value of the remaining samples was then used as the measured steady state EE of that specific participant performing that specific exercise stage.

3.2 Feature extraction

Existing literature on EE tracking studies have employed various classification/regression techniques based on potential time and frequency domain features extracted from measurement devices. Analyzing the power spectral density of the acceleration signals processed in the current study shows the potential of using frequency domain features in capturing the intensity of the movement trends. Thus, we consider HR, norm of the accelerometer signal, and gyroscope measurements to extract physiological and movement features. We moreover applied the so-called Welch method (Barbé et al., 2009) to analyze the spectral power of the IMU signals, as suggested in the literature for the inspection of such type of information. The list of the 64 features that we extracted from the measurements and used in our consequent analyses is summarized in Table 3.

[1] Median absolute deviation.
[2] Interquartile range of the signal.
[3] Highest recorded HR value throughout all the experiments.
[4] The percentage of the highest HR recorded per session divided by HR peak.

Table 3. *Features extracted from the signals*

Domain	Sources	Features
Time	HR, $\|Acc\|$ GyroX,GyroY,GyroZ	Mean, Max, Min, SD, MAD [1], Energy Entropy, IQR [2], Kurtosis, Skewness
Frequency	$\|Acc\|$ GyroX,GyroY,GyroZ	Max PSD, Min PSD, Mean PSD Freq of Max PSD
Personal Characteristics	Self Reported: Measured:	Sex, IPAQ, Age, Height Body Mass, HR peak [3], HR peak percentage [4]

3.3 Dimensionality Reduction

In our specific domain the results shall be interpretable by medical personnel, and by the participants themselves. Given that a high number of features may limit the interpretability and ease of visualization of the data. Thus, after scaling the feature vectors we employ a Principle Component Analysis (PCA) dimensionality reduction approach (Daffertshofer et al., 2004; Van Der Maaten et al., 2009). To minimize the risk of under-fitting phenomena, we consider a relatively high number of principal components, i.e., enough to explain 95% of the variance of the whole dataset [5].

3.4 Model structure selection

We here report the results obtained through Support Vector Regression (SVR, more specifically with a standard Radial Basis Function kernel), and Gaussian Processes Regression (GPR) defined by the kernel

$$K_{GPR} = k_N + k_C \; k_{Ma}, \tag{2}$$

with k_C a constant kernel, k_N a white noise kernel parameterized by the noise level σ, and the Matern kernel k_{Ma} (a general form of RBF kernel parameterized with a length-scale parameter $l > 0$ as defined in (3)).

$$\begin{aligned} k_{Ma} &= \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2}\right) \exp\left(-\frac{\sqrt{5}r}{l}\right), \\ r &= \|x - x'\| \,. \end{aligned} \tag{3}$$

The choice in (2) is inspired by the ones made by Abdessalem et al. (2017); Awad and Khanna (2015); Schulz et al. (2018); Smola and Schölkopf (2004); Williams and Rasmussen (2006). Note that the various coefficients defining our choice (i.e., the constant associated to the constant kernel, σ, and l) were considered hyperparameters that were also estimated through an opportune training process.

The results obtained with these kernels / radial basis functions, and the ones we obtained using less sophisticated choices (e.g., linear kernels) were not dramatically different. Therefore, the final results were only mildly depending on the selected model structure (starting obviously from rather standard choices). For the sake of brevity we only report the results obtained based on the choices described in the above in the remainder of the paper.

[5] Doing a PCA on the whole dataset is typically a statistically suboptimal choice. In our application we noted that the loadings and scores plots associated to different CV folds tend to be quite stable though – we don't report the more detailed results here due to the lack of space. Given the stability of the PCA results, we decided to opt for a whole-dataset decomposition.

3.5 Cross validation

To estimate the performance of the regression models on unseen data, to investigate different generalization properties of the found models, and to avoid selection bias and over-fitting, we employ shuffle split, k-fold, and repeated k-fold cross validation (CV) strategies. To do so we follow four specific approaches partitioning data into approximately 70-30% train and test (holdout). More specifically, we consider:

- *Record-wise* splitting of the dataset into training and test sets by applying repeated k-fold (RKF) on the total number of records (180 collected datasets). This means creating both training and test sets that each blend in statistical information from every participant and every exercise stage. This dataset creation strategy is expected to provide optimistic results of the performance of the trained models on unseen data - the more optimistic the smaller the original dataset. This strategy helps to understand the general performance of the trained models.
- *Participant-wise* splitting of the dataset into training and test sets by applying shuffle split on the IDs of the participants. In this case the entire data of each participant is included either in the test set or the training set. This strategy is useful to assess the generalization capabilities for unseen persons.
- *Stage-wise* splitting of the dataset by applying a shuffle split on the IDs of the recorded exercise stages per participant (i.e., 1 : 9). This strategy is useful to check the generalization capabilities on unseen exercises.
- *Intensity-wise* splitting of the dataset by first categorizing the various EE signals into three "exercise intensity zones" (low, moderate, and high) containing equal proportions of the data, then creating three separate datasets with the so-created data, and finally performing a 3-fold CV on the new datasets. This strategy is used to check the generalization capabilities of the models in different exercise intensity zones.

Note that RKF and shuffle split CV are performed with ten repeats. We present the corresponding results and conclusions in the sections below.

4. RESULTS

This section summarizes the capabilities of the proposed regression models in generalizing trained models on unseen data, and in doing so we investigate to which extent it is necessary to employ individualized experiment designs and personalized models for the specific case of WCU. To do so we analyse how the model prediction accuracy is affected by sex and the IPAQ factor. More specifically, we consider the following three categories of performance indexes:

- performance of the *overall* prediction capabilities of models built on all the participants data,
- performance of *sex-based* models, i.e., prediction results for the case where the data is split not only as described in Section 3.5, but also based sex,
- performance of *IPAQ-based* models, i.e., prediction results for the case where the data is split based on reported self reported fitness index (IPAQ) factors (in our case, either *moderate* or *high*).

The quality of each model prediction is assessed by the coefficient of determination (R^2 score) and root mean squared error (RMSE). The summary of the results in terms of generalizability and personalized factors are reported in Table 4 and discussed in the following subsections.

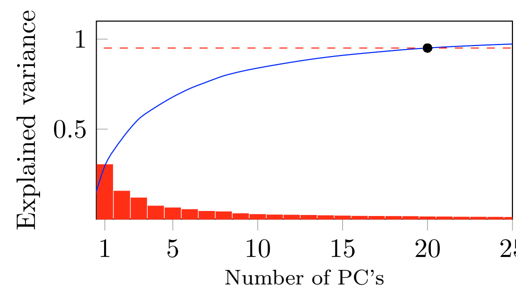

Figure 2. Explained variance versus number of principal components for the dataset. In this case the threshold of 95% of explained variance leads to a number of PCs equal to 20.

4.1 General prediction capabilities of the regression models

The reported performance indexes (high fit values and small residuals) in the overall category (record-/participant-wise CV scenarios) in Table 4 prove the validity of the proposed approaches. Comparing R^2 and RMSE (kcal/min) values shows that GPR outperforms SVR in estimating EE values and succeeds in capturing the trend of the data. More precisely, the trained model in record-/participant-wise CV analysis may predict EE values on unseen records and unseen participant data with an acceptable level of precision. In contrast, the high prediction error and negative R^2 values in session- and intensity-wise CV indicate the need for including a wide range of EE values or in other words different exercise intensity zones in the training datasets.

Table 5 and Figure 3 present then more details on the prediction capabilities of SVR and GPR for the intensity-wise CV based on splitting the dataset into low, moderate and high exercise intensities. The accuracy of the predictions are insufficient for the test data in all intensity zones. It is worth mentioning that high intensity zone predictions for test data show lower fit values and more biased estimations (higher residuals). This could be explained by higher measurement error of the Vyntus ergospirometer and an increasing anaerobic component, which is not captured in the way that we currently measure and calculate EE values.

4.2 On the importance of considering personalized factors

To evaluate the impact of the personal characteristics on the capabilities of predicting EE we assessed the models quality separately for sex and IPAQ scores (see Table 4). In addition, Figure 4 compares RMSE values in overall and personalized scenarios employing participants-wise CV for both regression models.

Table 4. Regression models and cross validation results

Model	CV	Overall		Sex/Male		Sex/Female		IPAQ/MOD		IPAQ/HIGH	
		R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
SVR	Record-wise	0.86	0.72	0.76	0.91	0.80	0.58	0.82	0.77	0.83	0.86
	Participant-wise	0.86	0.70	0.78	0.87	0.80	0.57	0.81	0.79	0.82	0.90
	Stage-wise	0.27	1.11	0.08	1.23	0.15	0.75	0.10	1.16	0.04	1.38
	Intensity-wise	-1.66	1.31	-2.24	1.44	-6.35	1.11	-2.16	1.35	-3.67	1.66
GPR	Record-wise	0.87	0.68	0.80	0.81	0.83	0.53	0.82	0.74	0.91	0.61
	Participant-wise	0.87	0.66	0.83	0.77	0.85	0.49	0.80	0.80	0.88	0.72
	Stage-wise	0.52	0.86	0.32	0.95	0.50	0.59	0.32	0.95	0.73	0.76
	Intensity-wise	-0.42	0.96	-0.86	1.06	-1.06	0.62	-1.10	1.04	-0.12	0.91

As can be seen from Figure 1, the distribution of personal characteristics for female participants is slightly more homogeneous than in male participants. This can explain an apparent difference in RMSE values in terms of mean and standard deviation and better fit in predicting unseen data for females (compared to male) in sex divided datasets. We note that in the current experimental setup, to compensate for biological differences between male and female participants (in regards to body mass, height, and muscle distribution), a lower treadmill speed has been chosen for female participants (as detailed in Table 1). This, however, leads to different excitation levels of the systems. While this needs to be further investigated, the speeds chosen for female group were likely more achievable in comparison with those for the male group.

Figure 4 highlights also the existence of considerable differences in prediction error when diving data into high and moderate IPAQ factors. However, in this study, only 6 out of 20 participants scored within high IPAQ factor range, thus the dataset for this group is relatively small compared to the group of moderate ones. This small dataset size can be a cause of high standard deviations in RMSE values. Moreover, a considerable part of the general population have a low IPAQ factor, a factor that is not included in this study. Therefore, the simulation results highlight the necessity of including a wider range of participants with a inactive to highly active life style in the training dataset.

Table 5. Comparison between test and train model predictions in intensity-wise CV.

Model		Low Intensity		Moderate Intensity		High Intensity	
		R^2	RMSE	R^2	RMSE	R^2	RMSE
SVR	Train	0.96	0.33	0.98	0.36	0.96	0.22
	Test	-2.89	1.20	-0.3	0.58	-1.79	2.14
GPR	Train	0.99	0.20	0.98	0.37	0.89	0.35
	Test	0.11	0.57	-0.76	0.68	-0.61	1.62

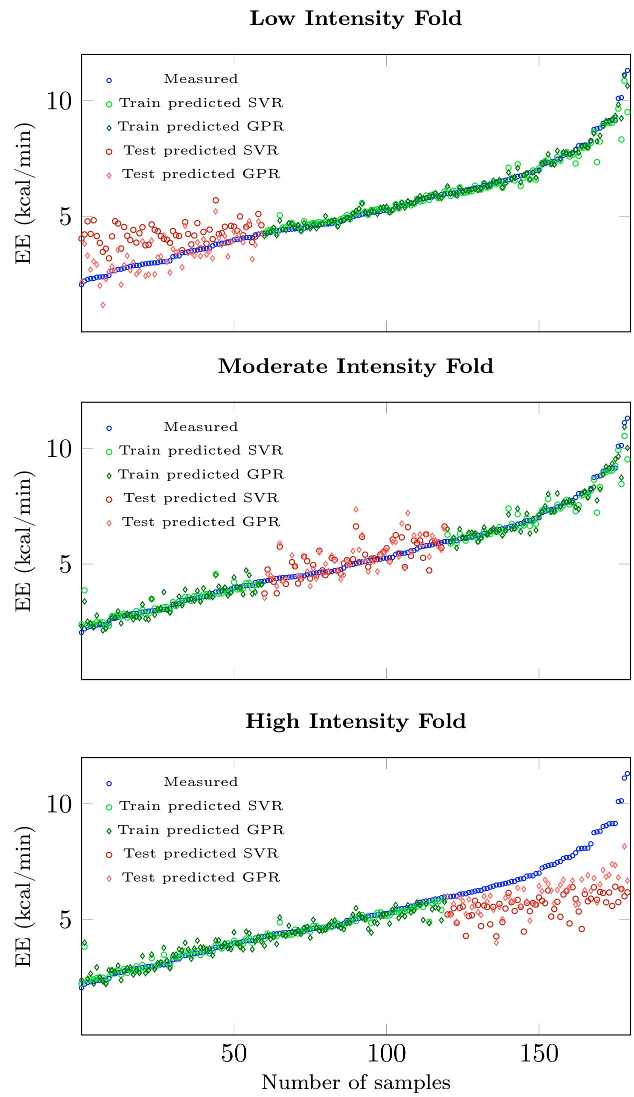

Figure 3. Regression results of intensity-wise 3-fold cross validation applied to estimate EE values sorted based on movement intensity.

5. CONCLUSION AND FUTURE WORK

This study analysed the prediction capabilities of Support vector and Gaussian process regression (SVR and GPR) approaches for estimating energy expenditure (EE) levels during wheelchair propulsion. More precisely, we considered estimators based on features extracted from Heart rate and Inertial measurement unit time series from four-minutes long experiments. Overall, the proposed regression approaches were precise in predicting EE values in various cross validation scenarios, proving that the proposed models are able to generalize on unseen data. The purpose of these analyses was also to explore future considerations for designing the experiments with wheelchair users (WCU), i.e., collecting the most informative data that includes both quantitative and qualitative indicators. This study shows that data collected in moderate intensity activities tends to be more informative. However, in order to increase the accuracy of EE predictions also for low and high

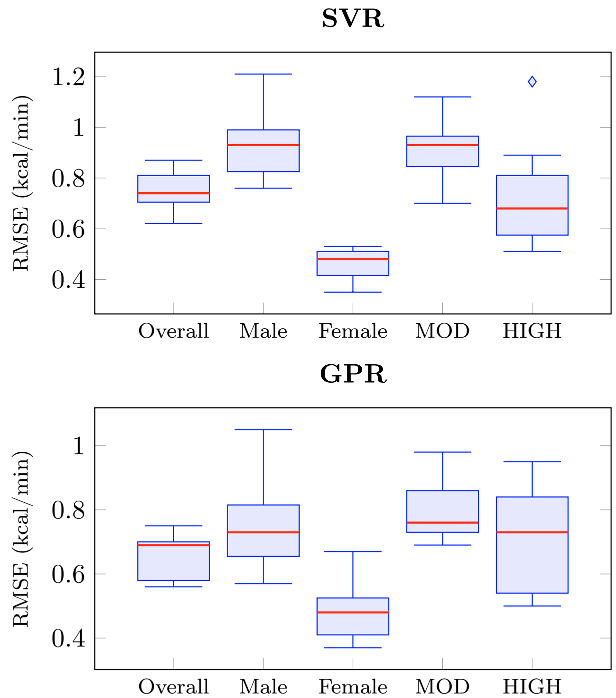

Figure 4. Comparison between SVR and GP results reporting participant-wise CV RMSE and standard deviations considering personalized models based on sex and IPAQ factor

intensity, the experiments need to be designed to cover as wide range of exercise intensities as possible.

Moreover, for personalizing the EE estimators, the dependency of the model performance on sex and IPAQ factors and likely other disability-related factors need to be considered when investigating WCU in the next experimental phase. More specifically, the type and level of the injury in manual wheelchair users, body mass and age are likely important factors to be considered in this strive for personalization. Thus, although the data collected for the current study is informative enough for developing algorithms for able-bodied participants, further experiment design modifications are necessary for developing EE prediction algorithms for WCU.

REFERENCES

Abdessalem, A.B., Dervilis, N., Wagg, D.J., and Worden, K. (2017). Automatic kernel selection for gaussian processes regression with approximate bayesian computation and sequential monte carlo. *Frontiers in Built Environment*, 3, 52.

Awad, M. and Khanna, R. (2015). Support vector regression. In *Efficient learning machines*, 67–80. Springer.

Barbé, K., Pintelon, R., and Schoukens, J. (2009). Welch method revisited: nonparametric power spectrum estimation via circular overlap. *IEEE Transactions on signal processing*, 58(2), 553–565.

Bauman, A., Ainsworth, B.E., Bull, F., Craig, C.L., Hagströmer, M., Sallis, J.F., Pratt, M., and Sjöström, M. (2009). Progress and pitfalls in the use of the international physical activity questionnaire (ipaq) for adult physical activity surveillance. *Journal of Physical Activity and Health*, 6(s1), S5–S8.

da Rocha, E.E.M., Alves, V.G.F., Silva, M.H.N., Chiesa, C.A., and da Fonseca, R.B.V. (2005). Can measured resting energy expenditure be estimated by formulae in daily clinical nutrition practice? *Current Opinion in Clinical Nutrition & Metabolic Care*, 8(3), 319–328.

Daffertshofer, A., Lamoth, C.J., Meijer, O.G., and Beek, P.J. (2004). Pca in studying coordination and variability: a tutorial. *Clinical biomechanics*, 19(4), 415–428.

Dannecker, K.L., Sazonova, N.A., Melanson, E.L., Sazonov, E.S., and Browning, R.C. (2013). A comparison of energy expenditure estimation of several physical activity monitors. *Medicine and science in sports and exercise*, 45(11), 2105.

Glasheen, E., Domingo, A., and Kressler, J. (2021). Accuracy of apple watch fitness tracker for wheelchair use varies according to movement frequency and task. *Annals of physical and rehabilitation medicine*, 64(1), 101382.

Klausen, B., Toubro, S., and Astrup, A. (1997). Age and sex effects on energy expenditure. *The American journal of clinical nutrition*, 65(4), 895–907.

Moreno, D., Glasheen, E., Domingo, A., Panaligan, V.B., Penaflor, T., Rioveros, A., and Kressler, J. (2020). Validity of caloric expenditure measured from a wheelchair user smartwatch. *International Journal of Sports Medicine*, 41(08), 505–511.

Nightingale, T.E., Rouse, P.C., Thompson, D., and Bilzon, J.L. (2017a). Measurement of physical activity and energy expenditure in wheelchair users: methods, considerations and future directions. *Sports medicine-open*, 3(1), 1–16.

Nightingale, T.E., Williams, S., Thompson, D., and Bilzon, J.L. (2017b). Energy balance components in persons with paraplegia: daily variation and appropriate measurement duration. *International Journal of Behavioral Nutrition and Physical Activity*, 14(1), 1–11.

O'Driscoll, R., Turicchi, J., Beaulieu, K., Scott, S., Matu, J., Deighton, K., Finlayson, G., and Stubbs, J. (2020). How well do activity monitors estimate energy expenditure? a systematic review and meta-analysis of the validity of current technologies. *British Journal of Sports Medicine*, 54(6), 332–340.

Schulz, E., Speekenbrink, M., and Krause, A. (2018). A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, 85, 1–16.

Smola, A.J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3), 199–222.

Van Der Maaten, L., Postma, E., Van den Herik, J., et al. (2009). Dimensionality reduction: a comparative. *J Mach Learn Res*, 10(66-71), 13.

Weil, E., Wachterman, M., McCarthy, E.P., Davis, R.B., O'Day, B., Iezzoni, L.I., and Wee, C.C. (2002). Obesity among adults with disabling conditions. *Jama*, 288(10), 1265–1268.

Weir, J.d.V. (1949). New methods for calculating metabolic rate with special reference to protein metabolism. *The Journal of physiology*, 109(1-2), 1.

Williams, C.K. and Rasmussen, C.E. (2006). *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA.