Estimating Running Performance Combining Non-invasive Physiological Measurements and Training Patterns in Free-Living

Marco Altini and Oliver Amft[1]

***Abstract*— In this work, we use data acquired longitudinally, in free-living, to provide accurate estimates of running performance. In particular, we used the HRV4Training app and integrated APIs (e.g. Strava and TrainingPeaks) to acquire different sets of parameters, either via user input, morning measurements of resting physiology, or running workouts to estimate running 10 km running time. Our unique dataset comprises data on 2113 individuals, from world class triathletes to individuals just getting started with running, and it spans over 2 years. Analyzed predictors of running performance include anthropometrics, resting heart rate (HR) and heart rate variability (HRV), training physiology (heart rate during exercise), training volume, training patterns (training intensity distribution over multiple workouts, or training polarization) and previous performance. We build multiple linear regression models and highlight the relative impact of different predictors as well as trade-offs between the amount of data required for features extraction and the models accuracy in estimating running performance (10 km time). Cross-validated root mean square error (RMSE) for 10 km running time estimation was 2.6 minutes (4% mean average error, MAE, 0.87 R^2), an improvement of 58% with respect to estimation models using anthropometrics data only as predictors. Finally, we provide insights on the relationship between training and performance, including further evidence of the importance of training volume and a polarized training approach to improve performance.**

I. INTRODUCTION AND RELATED WORK

Medium and long distance running is becoming popular, with millions of people worldwide participating in running events from 5 km to the marathon distances. While human performance in running has been analyzed in elite athletes as well as recreational ones for decades, the scientific community is still investigating different aspects behind the limits of human performance, as recently shown by the Nike breaking2 project [1]. Being able to accurately estimate running performance can be helpful at several levels. First, we could provide individuals with better race pacing strategies, which are often guessed based on limited data. Secondly, we could also tailor training plans to individual abilities, therefore reducing injury risk.

Different anthropometric, physiological, and training characteristics influence human performance in running. Low body fat has been associated with better times [2], similarly to low resting heart rate (HR) and higher heart rate variability (HRV) [3]. Other physiological parameters measured in the lab, for example lactate threshold and VO_2max, have also been linked to better running times. In our recent work we have shown how VO_2max estimated from running workouts highly correlates with running performance in events between the 10 km and the marathon [4]. Training related variables, such as training volume (distance per week), as well as average training speed have been associated to improved running performance too. Recently, interest has shifted to training patterns analyzed over weeks or months. For example, most elite athletes train in a so-called polarized regime, in which most workouts are carried out at low intensities, and a few at very high intensity, as opposed to moderate intensity training, more typical of recreational runners [5], [6]. Even in recreational runners, a shift to a polarized training regime resulted in performance improvements [7].

Most literature published on estimating running performance however is constrained by small sample size and a rather homogeneous sample (for example only men, or a narrow age range or performance range). Variables included in the model are acquired under laboratory conditions or supervised settings that are not practical (VO_2max, lactate threshold, biomechanics [2]) Finally, parameters are analyzed in isolation (for example the impact of polarized training on performance [5]) and running time estimates accuracy is suboptimal or has not been cross-validated.

In the past few years, we have witnessed fast technological developments and integrations between different platforms and services (e.g. public APIs), resulting in increased availability of multivariate data streams acquired from mobile applications and wearable sensors (e.g. GPS, accelerometer, physiological data). Such developments are providing scientists with data at a scale that is typically not manageable in regular laboratory studies, and therefore with the opportunity of providing additional insights on the relation between physiology, training and performance. While several mobile applications and wearable sensors have been released on the market, providing users with metrics reflecting behavior (e.g. steps taken, distance ran, etc.) limited work has been carried out to provide insights on the individual's performance ability outside of laboratory settings.

In this work we propose the first longitudinal, large scale analysis of running performance with respect to a wide set of variables either self-acquired or acquired automatically and non-invasively in free-living, without laboratory tests or supervision. Analyzed variables include anthropometrics, resting physiology, training physiology, training volume, training patterns and previous performance.

In particular, we used the HRV4Training app [8], [9] to collect data from 2113 individuals of different fitness levels, over a time period of 2 years, and developed multiple linear regression models to highlight the relative impact of

[1]M. Altini and O. Amft are with Friedrich-Alexander Universitat Erlangen-Nurnberg (DE) email: altini.marco@gmail.com

different predictors on running performance estimation. We show that running performance (10 km time) can be estimated accurately from data acquired in free-living, without supervision or specific laboratory tests. Besides providing practical estimation models that could be employed by recreational runners to tailor training programs, we also provide additional insights on the relationship between training and performance, including confirmative evidence of the importance of polarized training.

II. DATA ACQUISITION

A. Measurement Protocol

Users downloaded the HRV4Training app from the Apple Store or Google Play Store and agreed to provide collected measurements and annotations for research purposes via a consent form embedded in the application. The HRV4Training app allows a user to measure resting physiological data (HR and HRV) using the phone camera or external sensors, and was recently validated with respect to electrocardiography (ECG) [10]. The application instructed users to perform the measurement right after waking up while still lying, to limit the effect of other stressors. In addition, users could link the HRV4Training app to other services, such as Strava or TrainingPeaks, so that not only morning measurements, but also workouts performed during the day, and associated GPS and heart rate data, would be collected automatically in the app.

B. Dataset

Data was collected using the HRV4Training app during 2016 and 2017. A total of 2113 users (1891 male, 222 female) met the inclusion criteria: training with a heart rate monitor, linking the app via third party APIs to collect workouts data, including at least one 10 km, and taking at least a month of morning physiological measurements. A user's best 10 km time was automatically identified as the fastest 10 km workout over the 2 years, and used as reference for this analysis. For each identified best 10 km time, the previous 3 months of data were used to extract features related to a user training volume and patterns. Totally, 464809 morning physiological measurements and 296739 running workouts were acquired during the 2 years longitudinal study, for an average of 220 morning measurements and 140 workouts with heart rate data per person. Anthropometric characteristics and reference 10 km times of the users are listed in Table I.

TABLE I

USERS ANTHROPOMETRIC DATA AND REFERENCE 10 KM TIME

	mean	sd	min	max
age (years)	39.80	8.70	18.00	69.08
bmi (kg/m^2)	23.24	2.51	16.65	36.59
weight (kg)	73.34	10.53	42.00	122.00
height (cm)	177.41	7.69	149.00	205.00
10km time (minutes)	49.8	7.2	34.2	75.0

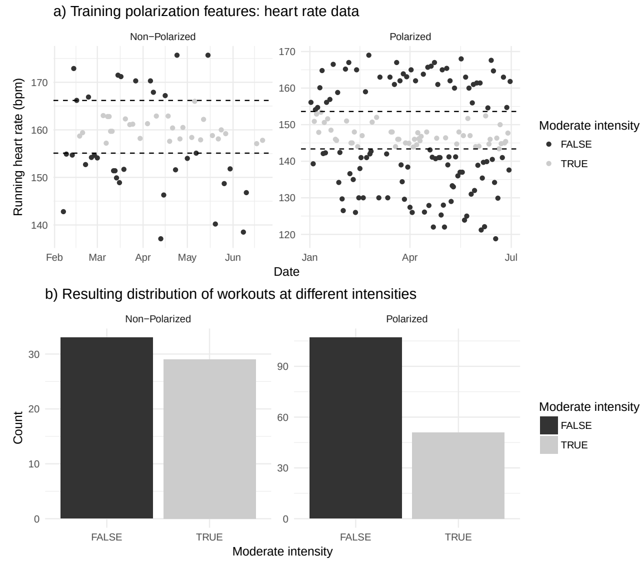

Fig. 1. a) Heart rate data for an individual training under a non-polarized approach (left) and an individual training under a polarized approach (right). Moderate intensity was defined as HR within 5% of a user's average. b) Resulting training intensities distributions highlighting how a more polarized approach involves less time spent at moderate intensities. Here we set an arbitrary threshold of 30% of workouts carried out at moderate intensity, to assign a user to the polarized or non-polarized category, only for visualization purposes.

III. DATA ANALYSIS

A. Features

We computed features representative of different aspects that may contribute to running performance. Then, we created the following sets to analyze their impact on estimation accuracy:

- *Ant*: anthropometrics data. A user's body mass index (BMI), age and gender.
- *Rest*: resting physiological data (HR and HRV). HR was computed as the mean HR during the daily morning measurement, while as HRV feature we used the square root of the mean squared RR intervals difference (rMSSD), a marker of parasympathetic activity [3], [9].
- *Vol*: training volume and speed. Average workout distance and speed.
- *TrPhy*: physiological data during training. We computed the speed to HR ratio, a feature that relies on the fact that a more fit (faster) runner would maintain a lower HR while running at a certain speed, with respect to less fit (slower) runners. This parameter is the main predictor behind VO_2max estimation models relying on sub-maximal tests or workouts data [4], [11].
- *Pol*: training polarization. Training polarization refers to training at different intensities, typically avoiding moderate intensity training. We derived features from workouts summaries to analyze the impact of training polarization on estimated performance. Features were: percentage of workouts performed at speeds 5% above or below a user's average workout speed and the percentage of workouts where HR was within 5% of a

user's average HR rate, a feature used to represent lack of polarization (see Fig. 1).

- *Performance*: past running performance. Finally, the last feature set included all previous features plus the best 10 km time found in the 3 months preceding the fastest time, as previous running performance is highly predictive of future running performance.

We chose to compute training polarization features as deviations from a user's average. By avoiding range, maximum and minimum values we maintain the algorithm's robustness to measurement inaccuracies and outliers which often occur in free-living. All features were computed over the 3 months preceding a user best 10 km time, depending on available data. An example of features collected is shown in Fig. 2.

B. *Running performance estimation*

Running performance (10 km running time) was estimated using multiple linear regression (MLR) models and the different sets of predictors listed in Sec. III-A. MLR was used so that coefficients could be easily analyzed and interpreted, in order to provide further evidence on the importance of different parameters in the context of running performance.

C. *Statistics and performance measures*

MLR models were validated using 10-fold cross-validation. At each iteration, 10% of the data was randomly selected for validation, while from the remaining 90%, all users with at least 20 workouts were used as training set. By selecting only users with a minimum amount of data, we ensured training patterns could be meaningfully estimated. However, we validated our models on all users, even the ones including little data, so that the relation between available data and accuracy could be investigated. We report Root Mean Square Error (RMSE) in minutes, Mean Percentage Error (MPE), in percentage, Pearson's correlation and explained variance (R^2) for all cross-validated MLR models.

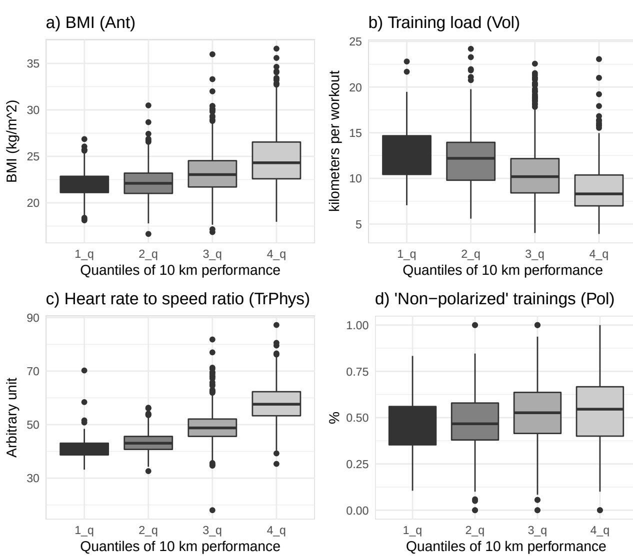

Fig. 2. Examples of features used to estimate 10 km running performance. Data were split into quartiles for clarity.

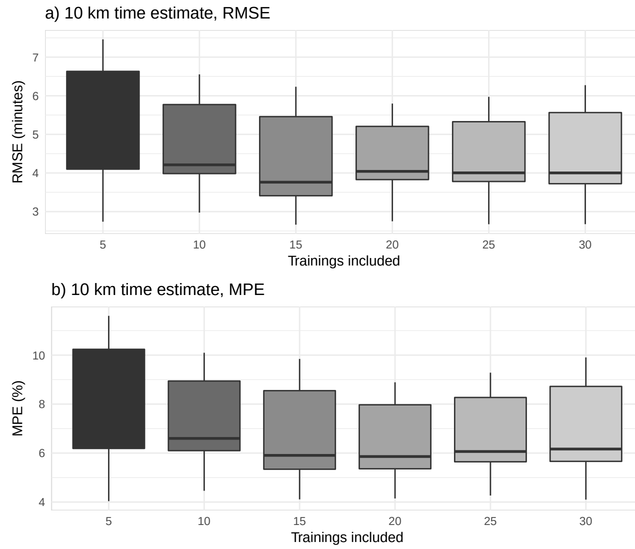

Fig. 3. RMSE and MPE when different amounts of workouts are included in the analysis. 15 workouts seem sufficient to extract meaningful features and minimize estimation error.

IV. RESULTS AND DISCUSSION

A. *Workouts number*

Results of running time estimation for different amounts of included workouts are shown in Fig. 3. From this analysis, 15 workouts seem sufficient to extract meaningful features and minimize estimation error.

B. *Feature sets*

Feature sets introduced in Sec. III-A were used as predictors in MLR models. Results were best for feature set *Performance*, with a RMSE of 2.68 minutes. Estimation was least accurate when using only anthropometrics data (set *Ant*, RMSE = 6.27 minutes) and improved progressively when adding resting physiological data (HR and HRV, set *Rest*, RMSE = 6.07 minutes), training volume and speed (average kilometers per workout and average speed per workout set *Vol*, RMSE = 4.04 minutes), physiological data during training (speed to heart rate ratio, set *TrPhys*, RMSE = 3.96 minutes) and training patterns (percentage of workouts at low and high intensities, set *Pol*, RMSE = 3.64 minutes). All results are reported in Table II while Fig. 4 shows Bland-Altman plots and R^2 for four MLR models.

TABLE II

RMSE, MPE AND PERASON'S CORRELATION FOR THE DIFFERENT MODELS DEVELOPED IN THIS STUDY.

Feature set	RMSE (minutes)	MPE (%)	r
Ant	6.27	9.91	0.53
Rest	6.07	9.56	0.57
Vol	4.04	6.25	0.84
TrPhys	3.96	6.08	0.85
Pol	3.64	5.52	0.87
Performance	2.68	4.10	0.93

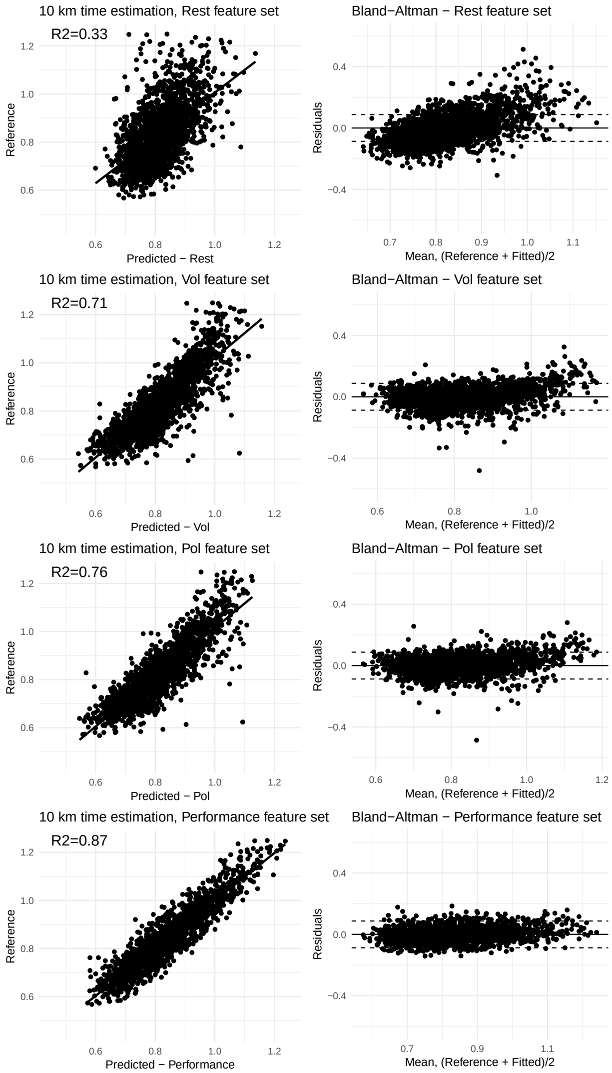

Fig. 4. Bland-Altman plots for four models: *Ant*, *Vol*, *Pol* and *Performance*.

C. Coefficients

MLR models coefficients could not be reported for all models due to space restrictions. We report here the sign of the model's coefficients as they provide insights on the relation between predictors and estimated performance. In particular, it is of interest to determine the impact of features representative of training patterns as derived from workouts. In our analysis, age, BMI, resting HR, speed to HR ratio and time spent at moderate HR intensity entered the model with a positive sign, meaning that a lower value for these predictors is associated with a faster 10 km. On the other hand, HRV (rMSSD), average distance and speed, percentage of workouts performed 5% faster or 5% slower than the average training entered the model with a negative coefficient. Thus, according to our dataset and analysis, a more polarized training regime, with a higher percentage of workouts preformed either faster or slower than the average workout, as well as a lower percentage of workouts performed at moderate HR intensity, is associated with improved performance.

V. CONCLUSIONS

In this work, we used data acquired longitudinally, in free-living, to provide accurate estimates of running performance on a dataset of 2113 runners of all levels. We investigated the relation between anthropometrics data, resting physiology, training patterns and performance, showing that running performance can be estimated accurately. The estimation models developed in this work do not require laboratory tests, and could be practically employed by the growing community of recreational runners to estimate performance and tailor training plans. Results for different feature sets are consistent with previous results from smaller studies [3], [7], [5], [4], showing a positive correlation between higher estimates of VO_2max, higher HRV, lower HR, higher training volume, higher training speed, a more polarized training regime and running performance.

Our analysis focused on readily available parameters that can be easily acquired and processed in free-living. However, more variables could be integrated as metrics linked to biomechanics and running power are also becoming available in the consumers market. Additionally, results could be backed up by an additional laboratory validation. Future work will aim at both including more parameters as well as looking at performance changes over time to determine the estimation model's ability to track such changes.

REFERENCES

[1] "Nike breaking2," https://news.nike.com/nike-breaking2-two-hour-marathon, accessed: 2018-01-20.

[2] P. Bale, D. Bradbury, and E. Colley, "Anthropometric and training variables related to 10km running performance." *British journal of sports medicine*, vol. 20, no. 4, pp. 170–173, 1986.

[3] M. Buchheit, A. Chivot, J. Parouty, D. Mercier, H. Al Haddad, P. Laursen, and S. Ahmaidi, "Monitoring endurance running performance using cardiac parasympathetic function," *European journal of applied physiology*, vol. 108, no. 6, pp. 1153–1167, 2010.

[4] M. Altini, C. Van Hoof, and O. Amft, "Relation between estimated cardiorespiratory fitness and running performance in free-living: An analysis of hrv4training data," in *Biomedical & Health Informatics (BHI), 2017 IEEE EMBS International Conference on*. IEEE, 2017, pp. 249–252.

[5] J. Esteve-lanao, A. F. San Juan, C. P. Earnest, C. Foster, and A. Lucia, "How do endurance runners actually train? relationship with competition performance," *Medicine & Science in Sports & Exercise*, vol. 37, no. 3, pp. 496–504, 2005.

[6] P. B. Laursen, "Training for intense exercise performance: high-intensity or high-volume training?" *Scandinavian journal of medicine & science in sports*, vol. 20, no. s2, pp. 1–10, 2010.

[7] I. Muñoz, S. Seiler, J. Bautista, J. España, E. Larumbe, and J. Esteve-Lanao, "Does polarized training improve performance in recreational runners?" *International journal of sports physiology and performance*, vol. 9, no. 2, pp. 265–272, 2014.

[8] "Hrv4training," https://www.hrv4training.com/, accessed: 2018-01-20.

[9] M. Altini and O. Amft, "Hrv4training: Large-scale longitudinal training load analysis in unconstrained free-living settings using a smartphone application," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2016, pp. 2610–2613.

[10] D. J. Plews, B. Scott, M. Altini, M. Wood, A. E. Kilding, and P. B. Laursen, "Comparison of heart rate variability recording with smart phone photoplethysmographic, polar h7 chest strap and electrocardiogram methods," *International Journal of Sports Physiology and Performance*, pp. 1–17, 2017.

[11] M. Altini, P. Casale, J. Penders, and O. Amft, "Cardiorespiratory fitness estimation in free-living using wearable sensors," *Artificial intelligence in medicine*, vol. 68, pp. 37–46, 2016.