Estimate Power without Measuring it: a Machine Learning Application

Guillaume Lemaître[1], Cédric Lemaître[2]

[1] Paris-Saclay Center for Data Science, 1 Rue Honoré d'Estienne d'Orves, 91120 Palaiseau

[2] Apeira Technologies, 720 Avenue de L'Europe, 71200 Le Creusot

Keywords: Power estimation, Machine learning, Big Data

Context: Power is the core metric of performance in cycling. Trainers strive for maximizing it while engineers strive for minimizing its loss by improving equipement and aerodynamic drag. However, costly and specific equipments are required to optimize this process. Here, we show that power can be estimated from side data without the need to acquire power data itself. More specifically, we show a proof of concept that machine learning can serve to estimate metrics with limited equipment or missing data. In this work, we limit the scope at estimating power but aerodynamic drag could be estimated in the same spirit. The mechanical power produced by a cyclist can be mathematically modeled as:

$$P_{meca} = \left(0.5 \cdot \rho \cdot SC_x \cdot V_a^2 + C_r \cdot m\,g \cdot \cos(\alpha) + mg \cdot \sin(\alpha)\right) \cdot V_d \quad (1)$$

where ρ is the air density in kg.m^{-3}, S is frontal surface of the cyclist in m^2, C_x is the drag coefficient, V_a is the air speed in m.s^{-1}, C_r is the rolling coefficient, m is the mass of the rider and bicycle in kg, g in the gravitational constant which is equal to 9.81 m.s^{-2}, α is the slope in radian, and V_d is the rider speed in m.s^{-1}.

This model has been validated in [2]. Despite the reported results - a standard error of 2.7 W and R^2 score of 0.97 - the experiments fall short since they have been conducted in restricted conditions rather than in real field conditions.

Material and Methods: A data set of 417 activities have been collected from 5 riders. These riders have used different power-meters: (i) Saris PowerTap, (ii) Rotor Power LT, and (iii) Power2Max, (iv) SRM. All activities contain the following information: power, heart-rate, speed, cadence, distance, and elevation.

In addition to the data provided by the different sensors, the acceleration, slope, and derivative of the heart-rate are computed. To enable our model to anticipate some power variations, the data are augmented by computing the derivative of the original features for different time windows (i.e. from 1" to 5"). Therefore, each sample consists in 48 features and is associated to the power measured by the power-meter. Finally, the statistical model is built using a gradient boosting algorithm [1].

Experiments and results: To compare with our model, the power of cyclist is estimated using the mathematical model presented in Eq. (1). Apart of the cyclist weights, the parameters have been set to some default values as shown in Table 1.

Parameters	Values
Bike weight	6.8 kg
Rolling coefficient	0.0045
Atmospheric pressure	1013 hPa
SC$_x$	0.32
Temperature	15°C

Table 1: Default values of mathematical model.

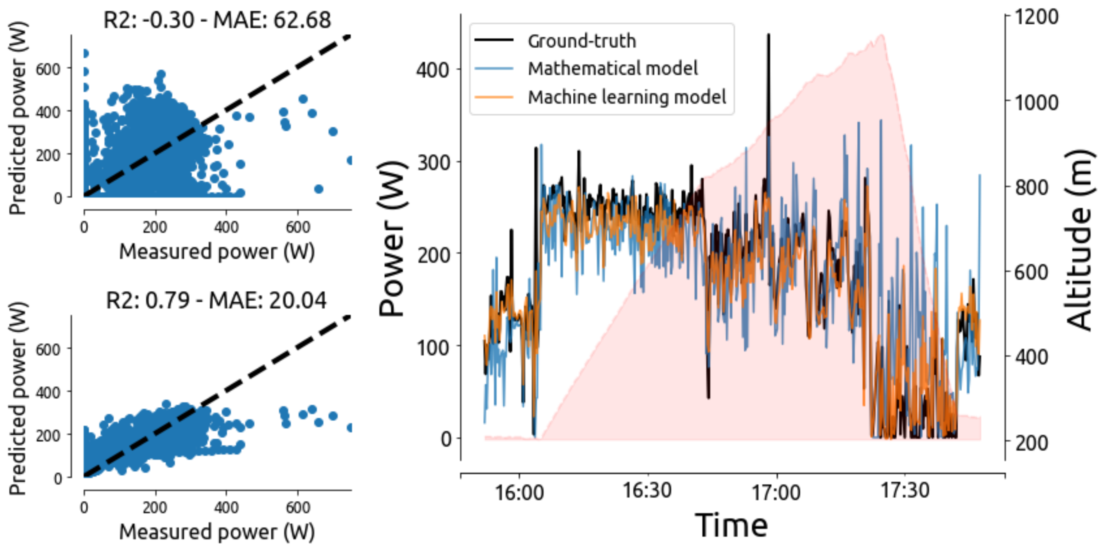

Figure 1: Cyclist power estimate for a single activity: *top-left and bottom-left* scatter plot of the measured and estimated power using the mathematical model and machine learning model, respectively ; *right* qualitative results obtained with the different models. The red shadow represents the elevation. Note that in this case, the machine learning model was trained on **all riders data**.

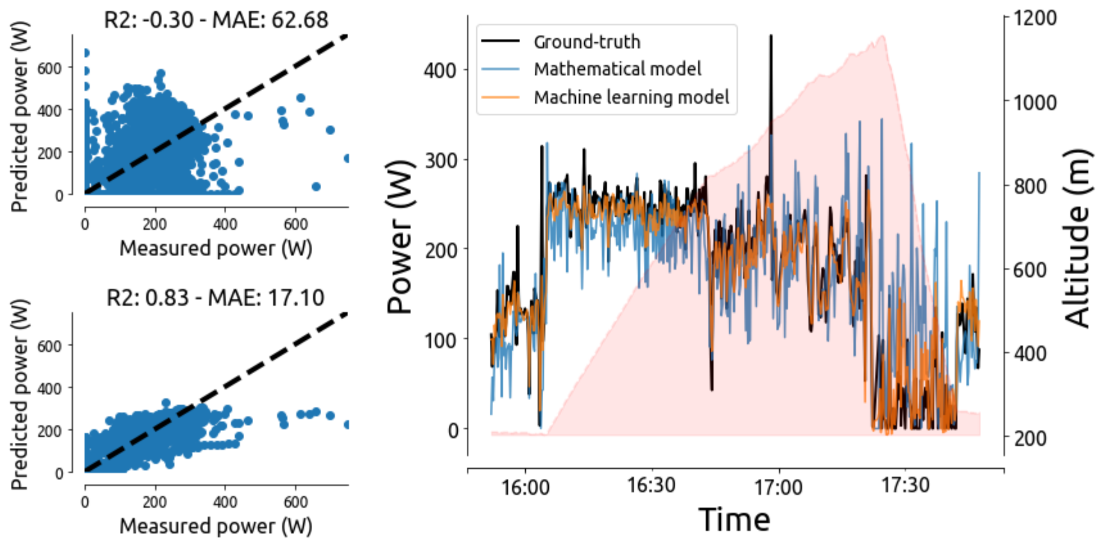

Figure 2: Cyclist power estimate for a single activity: *top-left and bottom-left* scatter plot of the measured and estimated power using the mathematical model and machine learning model, respectively ; *right* qualitative results obtained with the different models. The red shadow represents the elevation. The machine learning model has been trained specifically **on the rider data**.

These models are validated using a 3-fold cross-validation scheme. In these conditions, the machine learning model outperforms the mathematical models with a R^2 score of 0.71 and -0.26, respectively while the median absolute error (MAE) is 25.1 W and 55.2 W, respectively. An example is presented in Fig. 1. The mathematical modeling tends to be more unstable, especially in the descent and flat paths. However, the presented machine learning model does not allow to predict abrupt power changes (i.e. sprint).

We also train the machine learning model by stratifying the data by cyclist. We observed an improved R^2 and MAE scores for each cyclist. The same example as in Fig. 1 was used for which the machine learning algorithm has been trained specifically on the data of the given rider (see Fig. 2). We observe a decrease of the MAE from 20 W to 17 W and an increase of the R^2 from 0.79 to 0.83. It indicates that in practice, learning a model by cyclist will be more appropriate allowing to grasp particularities of each rider – e.g. specific cadence or heart-rate associated with slope. In addition, incremental learning algorithms could be used to grasp/forgive those particularities over time – e.g. weeks, months.

The mathematical model is implemented in scikit-cycling[1] while our experiments are available on GitHub[2].

Conclusion: This work shows the benefit of using machine learning to model mechanical power in cycling. It is expendable by using larger data sets which will be suitable to apply complex model (e.g. deep neural networks). Also, it opens new avenues for future research in which SC_x could be estimated in real field conditions using machine learning models. In this regard, machine learning model allows to estimate real-time power and SC_x without any important infrastructure like wind tunnel [3] and should lead to better performance than simple linear regression models [4].

1 Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics* 1189-1232.
2 Martin, J. C., Milliken, D. L., Cobb, J. E., McFadden, K. L., & Coggan, A. R. (1998). Validation of a mathematical model for road cycling power. *Journal of applied biomechanics*, *14*(3), 276-291.
3 Defraeye, T., Blocken, B., Koninckx, E., Hespel, P., & Carmeliet, J. (2010). Aerodynamic study of different cyclist positions: CFD analysis and full-scale wind-tunnel tests. *Journal of biomechanics*, *43*(7), 1262-1268.
4 Bouillod, A., Oggiano, L., Soto-Romero, G., Brunet, E., & Grappe, F. (2016, November). Preliminary study: A new method to assess the effective frontal area of cyclists. In *4th International Congress on Sport Sciences Research and Technology Support (IcSPORTS)*.

1 https://github.com/scikit-cycling/scikit-cycling

2 https://github.com/scikit-cycling/research/tree/master/power_regression