Estimating Power without Measuring it:
a Machine Learning Approach

Science & Cycling
4^{th} July 2018

Guillaume Lemaître & Cédric Lemaître

guillaume.lemaitre@inria.fr
contact@apeira-technologies.fr

Inria — Paris-Saclay Center for Data Science
Apeira Technologies

Paris-Saclay
Center for Data Science

Introduction
Context

Saris PowerTap PowerCal

- Low-cost power-meter based on heart-rate
- ✓ Low-cost device
- ✗ Not suitable to track small power changes

Strava

- Large amount of data
- Mathematical model based on mechanic

Problematic

Use machine learning to find more complex models using heterogeneous data sensor

Introduction
State-of-the-art

Mathematical model based on mechanic

$$P_{meca} = \left(0.5\rho S C_x V_a^2 + C_r mg \cos\alpha + mg \sin\alpha\right) V_d \ . \qquad (1)$$

Model parameters

Name	Symbol	Unit
Air density	ρ	$\mathrm{kg\,m^{-3}}$
Frontal surface	S	$\mathrm{m^2}$
Drag coefficient	C_x	NA
Air speed	V_a	$\mathrm{m\,s^{-1}}$
Rolling coefficient	C_r	NA
Mass of rider and bike	m	kg
Gravitational constant	g	$\mathrm{m\,s^{-2}}$
Slope	α	rad
Rider speed	V_d	$\mathrm{m\,s^{-1}}$

Power estimation using machine learning

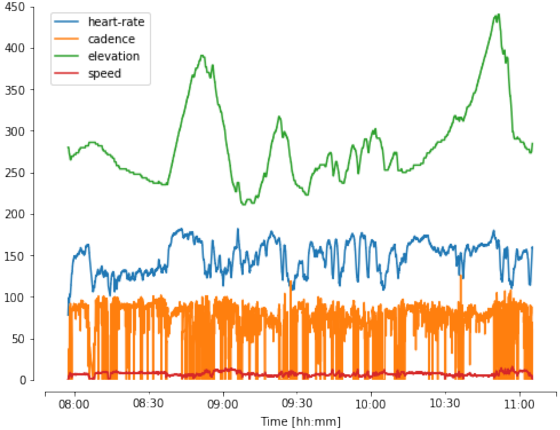

Figure: Original data

Power estimation using machine learning

Feature engineering

- Compute gradient for some data (acceleration, elevation, heart-rate)
- Compute derivative with different time periods (1 s — 5 s)
- Total of 48 features

Regressor

- Gradient boosting machine

Experiments
Setup

Data set

- 5 riders
- 4 power meters: Saris PowerTap, Rotor Power LT, Power2Max, and SRM
- 417 rides

Model validation

- Group k-fold cross-validation with $k = 3$

Model evaluation

- Coefficient of determination R^2
- Median absolute error (MAE)

Experiments
Model hyper-parameters

Mathematical model

Parameter	Value
Rider weight	Specific
Bike weight	6.8 kg
Rolling coefficient C_r	0.0045
Atmospheric pressure	1013 hPa
SC_x	$0.32\,\mathrm{m}^2$
Temperature	15 °C

Machine learning model

Parameter	Value
Number of decision tree	200
Depth of each decision tree	8

Results
Quantitative results

R^2 and MAE scores

Metric	Math	ML
R^2	-0.55	0.76
MAE	61.09	21.95

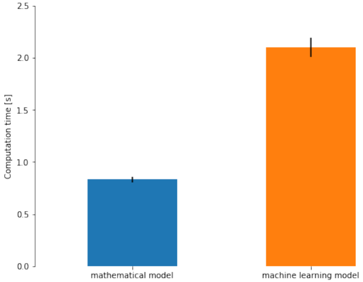

Figure: Computation time for estimation around 1 million samples

Results
Zoom-in

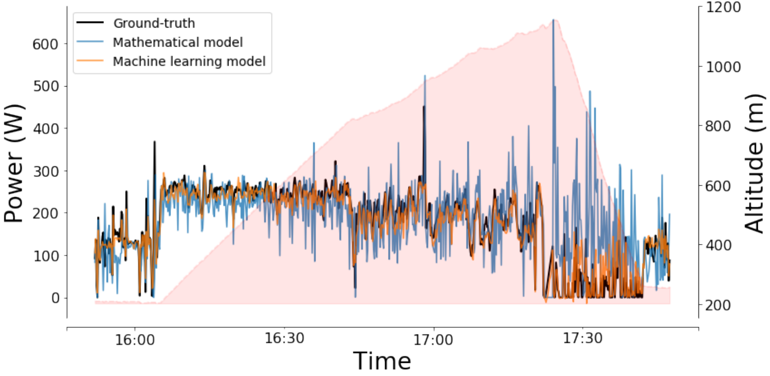

Figure: Power estimation for uphill and downhill

Results
Zoom-in

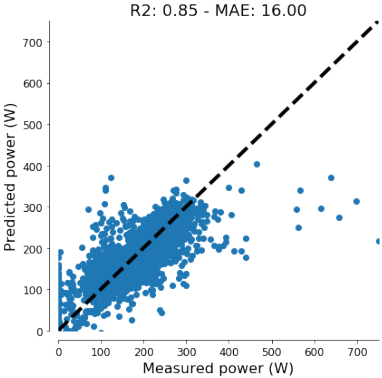

Figure: Left: Machine learning model — right: mathematical model

Results
Zoom-in

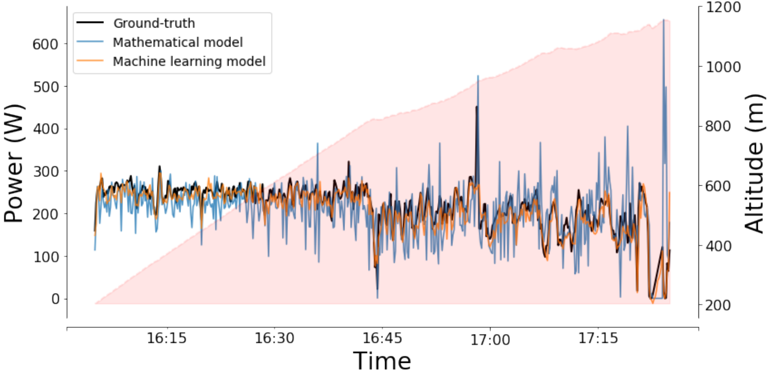

Figure: Power estimation for uphill

Results
Zoom-in

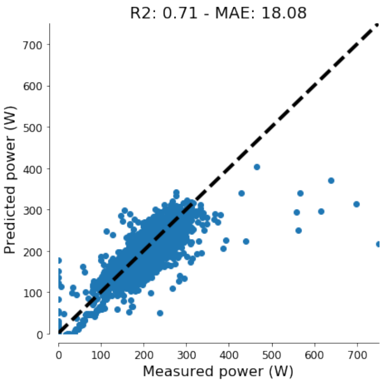

Figure: Left: Machine learning model — right: mathematical model

Results
Summary

Mathematical model

- ✓ Fast prediction
- ✗ Too much unknown parameters
- ✗ Too much variation in the estimation

Machine learning model

- ✓ Fast prediction
- ✓ Better prediction
- ✗ Difficulty to predict short power peak

Results
Analysis of the machine learning model

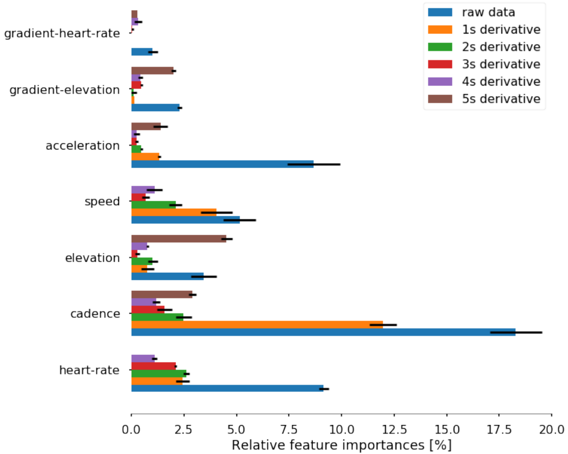

Figure: Feature importances of the different features used in the model

Future work

Extension of the current work

- ✓ Convolutional neural-network
- ✗ Larger data set

Open-source initiative via GoldenCheetah

- OpenData collection
- Development of `scikit-sports`

Purpose

- ✓ Reproducibility of analysis and methods
- ✓ Use data science tools to solve different problematic in cycling performance