

# Data-Driven Computational Framework for Personalized Training Plans in Cycling

Tom Van Deuren <sup>1,2,\*</sup>, Louis Van Gaelen <sup>1,2</sup>, Thomas Servotte <sup>1</sup> and Tim Verdonck <sup>1,2</sup>

Received: 1 March 2026

Accepted: 15 March 2026

Published: 30 October 2026

<sup>1</sup> Department of Mathematics, University of Antwerp, Middelheimlaan, Antwerp, 2020, Belgium.

<sup>2</sup> IDLab, University of Antwerp - imec, Sint-Pietersvliet 7, Antwerp, 2000, Belgium.

### Correspondence

Tom Van Deuren

Department of Mathematics, University of Antwerp, Middelheimlaan, Antwerp, 2020, Belgium.

[tom.vandeuren@uantwerpen.be](mailto:tom.vandeuren@uantwerpen.be)

### Keywords

prescriptive sports analytics; performance optimization; wearable sensor data; machine learning; differential evolution

## 1 Context

The widespread adoption of mobile power meters and heart rate monitors has transformed professional and amateur cycling into a highly data-driven discipline. Despite this massive collection of longitudinal sensor data, day-to-day training prescription still relies heavily on manual expert judgment rather than formalized mathematical optimization. While the predictive modeling of athletic performance has made significant advances, the transition from descriptive monitoring to automated, data-driven training prescription remains an unrealized opportunity in endurance sports.

## 2 Material and Methods

To address this gap, this study introduces a modular computational framework designed to automate the generation of personalized cycling training plans. First, we extract a continuous performance metric,  $PO@HR_{target}$ , utilizing longitudinal submaximal power output and heart rate data. Specifically, this metric filters historical training data for efforts within a personalized target heart rate range and employs linear interpolation to normalize the power output to this target heart rate, bypassing the need for disruptive and infrequent maximal laboratory testing. Second, we extract cycling-specific features from daily training data to capture the dose-response



This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



relationship, quantifying training load and intensity distributions through metrics like Training Stress Score, Acute and Chronic Training Loads, Chronic Intensity Load, and the Polarization Index. Third, we model the relationship between these training inputs and performance outputs using four diverse regression architectures: Lasso, Random Forest, XGBoost, and a novel tree-based model called Random Forest Featuring Linear Extensions (RaFFLE). These models are rigorously evaluated using 10-fold cross-validation, incorporating a seven-day buffer to combat data leakage. Finally, we apply a differential evolution algorithm to invert these predictive models, generating optimized training prescriptions, defining weekly durations, training load increases, and time-in-zone distributions, designed to maximize future athletic performance.

### 3 Results

Validation against laboratory testing confirmed that the submaximal  $\text{PO@HR}_{\text{target}}$  metric effectively tracks true physiological trends in the first lactate threshold. During the predictive modeling phase, the models achieved an average out-of-sample Mean Absolute Error ranging between 5 and 10 W, which aligns closely with the intrinsic  $\pm 1.5\%$  measurement uncertainty of commercial power meters. Furthermore, feature importance analysis revealed that chronic training load (e.g.,  $\text{CTL}_{56}$ ) consistently emerged as the dominant predictor of performance

across all model architectures. Our analysis also revealed significant inter-subject variability across cyclists; while linear models like Lasso accurately predicted performance and provided the lowest test error on average ( $7.17 \pm 2.70$  W), other athletes required complex non-linear tree-based ensembles to accurately capture the dose-response relationship. During the prescriptive phase, the optimization algorithm demonstrated a potential average performance increase of 5.5% when generating individualized training schedules targeting the  $\text{PO@HR}_{\text{target}}$  metric.

### 4 Conclusions

This framework successfully bridges the gap between predictive analytics and prescriptive decision-making in endurance cycling. By combining robust feature extraction from daily training data with advanced optimization techniques, it supports the automated generation of training plans that adapt to an athlete's evolving physiology. Ultimately, this methodology offers a scalable, sport-agnostic pathway for coaches and athletes to transition from retrospective data analysis to individualized performance optimization based on historical training data. Furthermore, future iterations of this framework should examine the relationship between this performance metric and daily recovery indicators, such as heart rate variability (HRV), sleep quality, and fatigue, to better account for daily fluctuations in an athlete's capacity to execute a prescribed plan.