



Trinity College Dublin

Coláiste na Tríonóide, Baile Átha Cliath

The University of Dublin

School of Statistics and Computer Science

Machine Learning to go *nyoom*

Using Machine Learning to evaluate rowing training and predict
training outcomes or performances

Liam Junkermann

Supervisor: Dr. Lucy Hederman

February 19, 2024

A Final Year Project submitted in partial fulfilment
of the requirements for the degree of
BA(Mod) in Science in Computer Science

Declaration

I hereby declare that this Final Year Project is entirely my own work and that it has not been submitted as an exercise for a degree at this or any other university.

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at <http://www.tcd.ie/calendar>.

I have completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at <http://tcd-ie.libguides.com/plagiarism/ready-steady-write>.

I consent / do not consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Signed: _____

Date: _____

Abstract

A short summary of the problem investigated, the approach taken and the key findings. This should not be more than around 400 words.

This must be on a separate page.

Lay Abstract

Similar to the actual abstract in terms of the information, but written for a non-specialist. So no jargon, no acronyms. Explain to a member of the general public what this project entailed. Should be no longer than the actual abstract.

This must be on a separate page.

Acknowledgements

Thanks Everyone!

You should acknowledge any help that you have received (for example from technical staff), or input provided by, for example, a company.

Contents

1	Introduction	1
1.1	Contents of the Introduction	1
1.2	Contents of the background chapter	1
1.3	The Conclusions chapter	1
2	Background	2
2.1	A brief introduction to rowing	2
2.1.1	Training Principles	2
2.1.2	Energy Systems	4
2.1.3	Performance	5
2.1.4	Summary	6
2.2	A Review of Performance Modelling	6
2.2.1	Quantifying Training Load (Fatigue)	6
2.2.2	Impulse-Response Models	8
2.2.3	Alternative Models	8
3	Data Collection and Management	9
4	Data Analysis and Visualization	10
5	Machine Learning Applications	11
5.1	Model Considerations	11
5.2	The Implementation	11
6	Discussion	12
7	Conclusion	13
A1	Appendix	16
A1.1	Appendix numbering	16

List of Figures

2.1	Borg 20	7
2.2	Borg CR10	7

List of Tables

Vocabulary

Heart Rate Metrics

HRV	Heart Rate Variability
HR	Heart Rate
HR _{max}	Maximum Heart Rate

Training Zones

UT2	Basic Oxygen Utilisation Training
UT1	Oxygen Utilization Training
AT	Anaerobic Threshold Training
TR	Oxygen Transport Training
AN	Anaerobic Capacity Training
LT	Lactate Threshold

General Training Terms

Stroke Rate	The number of strokes completed per minute (like cadence in cycling or running)
Split	The standard split in rowing is time to complete 500m

Chapter 1

Introduction

This chapter will introduce the general approach and motivation for this project, the expected outcomes, and a structure for the remainder of the report.

1.1 Contents of the Introduction

The introduction presents the nature of the problem under consideration, the context of the problem to the wider field and the scope of the project. The objectives of the project should be clearly stated.

1.2 Contents of the background chapter

The second chapter is typically a literature review, or survey of the state of the art, or a detailed assessment of the context and background for the project. The exact nature of this chapter depends on the topic and/or methods of the project. It is essential that the work of other people is properly cited. This will be discussed in detail in Chapter 2 below. Note that you should use references wherever is appropriate through the report, not just in the literature review chapter.

1.3 The Conclusions chapter

The final chapter should give a short summary of the key methods, results and findings in your project. You should also briefly identify what, if any, future work might be executed to resolve unanswered questions or to advance the study beyond the scope that you identified in Chapter 1.

Chapter 2

Background

This chapter will cover the basic background of rowing, the sports science which guides rowing training, and how an athlete's body responds to training stimulus. Next, a review of performance modelling will explore the development of human performance modelling since the introduction of the basic Bannister model in 1975 [1]. The section will outline how training load and performance can be quantified, and explain the way these approximations are used in various performance models to date.

2.1 A brief introduction to rowing

Rowing is an Olympic sport, raced across a 2,000 metre course, typically lasting six to seven minutes. It is classed as a power-endurance sport, this means training is focused on building aerobic, anaerobic, and power while also developing rowing technique [2]. Most time is spent building endurance, next most time is spent building anaerobic capacity, finally building strength and power through Strength and Condition sessions [3]. The importance of power is more significant in rowing than in cycling given the relatively short duration of exertions, with longer distance racing typically only covering five to seven kilometers, or fifteen to twenty-five minutes. Road cycling, for example, tends to last for a longer period of time, where the shortest races might last two hours. There are many different approaches to how training is conducted and which energy systems are targeted. This section will discuss the basic training principles which guide training, the way athletes respond to different kinds of training loads, and how performance is evaluated in rowing.

2.1.1 Training Principles

Generally when a coach builds a training plan they have a few factors they can work with: volume, the amount of mileage or total time spent training, sessions intensity, how hard the given session is meant to be, and finally the frequency, or the time spent in different

intensity zones. There are various ways to measure intensity, including heart rate, blood lactate concentration, velocity at maximal oxygen uptake ($\text{VO}_2 \text{ max}$), and rate of perceived exertion (RPE) [4]. Rowers tend to use heart rate zones or blood lactate concentration depending on access to the equipment to test blood lactate. Typically when a rower uses calculated aerobic zones each zone will be a percentage of HR_{max} . Following this approach, zones are defined as follows:

Z1 "Very Light" intensity, 50% - 60% of HR_{max}

Z2 "Light" intensity, 60%-70% of HR_{max}

Z3 "Moderate" intensity, 70%-80% of HR_{max}

Z4 "Hard" intensity, 80%-90% of HR_{max}

Z5 "Maximum" intensity, 90%-100% of HR_{max}

The exact definition of these zones varies in the literature, as does the method to calculate HR_{max} without a specific test. However, most high level athletes will have completed some kind of stress test to determine their HR_{max} in order to train more effectively on their prescribed zones. A rower who uses lactate based training zones might use the following zones:

T1 basic oxygen utilization training (UT2) [lactate = 0-2 mmol/L]

T2 oxygen utilization training (UT1) [lactate = 2-3.5 mmol/L]

T3 anaerobic threshold training (AT) [lactate = 3.5-4.5 mmol/L]

T4 oxygen transport training (TR) [lactate = 4.5-6 mmol/L]

T5 anaerobic capacity training (AN) [lactate \geq 6 mmol/L] [5]

Depending on how rigorous the testing protocol was, Heart Rate zones may be calculated for each zone, these may vary from the aerobic zones calculated from HR_{max} .

The most basic zone approximation approach uses three zones based around certain physiological thresholds, like, lactate thresholds (LT_1 and LT_2) and ventilatory thresholds. Cyclists may use critical power to determine these three basic zones, although this practice has not become popular in rowing training. The zones become simply, low-intensity, moderate-intensity, and high-intensity.

There are a few different approaches for distributing intensity for endurance training. The three main methods are: polarised training, sweet spot or threshold training, and pyramidal training. This guides the final factor a coach considers when building a general training plan, frequency. For the purposes of comparing polarized training (POL), threshold training

(THR), and pyramidal training (PYR), the more basic three zones of intensity will be used. The breakdown per zone for each training method is as follows:

Polarised Training Far more time spent in the low-intensity zone [6].

Low-Intensity 75%-85% of total training volume

Medium-Intensity 5%-10% of total training volume

High-Intensity 5%-10% of total training volume

Threshold Training More time spent in the medium-intensity zone [6].

Low-Intensity 45%-55% of total training volume

Medium-Intensity 35%-55% of total training volume

High-Intensity 15%-20% of total training volume

Pyramidal Training Most time spent in low-intensity zone with progressively less time spent in higher zones [7].

Low-Intensity 75%-85% of total training volume

Medium-Intensity 15%-20% of total training volume

High-Intensity 5%-10% of total training volume

This report will not compare the effectiveness of different training distributions. Different distributions tend to be used by different sports, or depending on which energy system is being targeted. The use of polarized training is most common in rowing [4], although other training approaches have been used, especially around competition time.

2.1.2 Energy Systems

In the human body there are two major types of skeletal muscle fibres, fast twitch, and slow twitch. Slow twitch muscles are used for longer, slower contractions where relative strength is low. Conversely, fast twitch muscles are shorter, faster contractions where strength is relatively higher. These types of fibres store energy differently and respond to training differently. Slow twitch muscles can be adapted to longer, lower effort, sessions and become resistant to fatigue, while fast twitch muscles fatigue easily, due to their lower glycogen capacity. Slow twitch muscle fibres have a low anaerobic capacity, while fast twitch muscle fibres have high anaerobic capacities. These two kinds of fibres draw on two different energy systems: anaerobic and aerobic systems. These systems provide muscles with Adenosine Triphosphate (ATP) which is used by the mitochondria in the muscle fibre cells to produce energy, allowing the muscles to contract [8].

The anaerobic system is used to provide energy to muscles to produce power without the use of oxygen. The anaerobic system typically provides energy for shorter periods of time. An immediate energy system can provide energy for 1-2 seconds of maximal work, this is typically used for resistance based strength training. More commonly in rowers, the short term energy system is used to provide energy to muscle fibres under significant strain. However, each time the energy system produces (ATP), lactate is also produced. This system is limited by an athletes ability to flush lactate from muscles. If the muscles are unable to flush the lactate quickly enough, the muscles will fatigue until exercise needs to stop.

The aerobic, or long term energy, system is the slowest at providing energy to muscles. This system uses sugars, fats and oxygen to produce ATP. This system only works when large amounts of oxygen are available, which for rowers is typically during lower intensity sessions. Longer "steady state" sessions rely on the aerobic to provide energy to muscles, with these sessions also being used to build the aerobic system. By spending time in the appropriate training zones the aerobic system builds efficiency by creating new capillaries to support the slow twitch fibres.

Rowers develop both their anaerobic and aerobic systems. Much of the winter season (typically September-March) is spent building the aerobic system, this is due to the longer length of any races done during this period, and to build a larger base on which to build a strong anaerobic system. As the sprint racing season begins, or shortly before, more anaerobic sessions will be introduced to build the system. Rowers will build their lactate tolerance, and spend more time at just-below-maximal efforts to prepare for the shorter race format.

2.1.3 Performance

There are two primary ways to measure a performance in rowing. First, on the rowing machine, a 2,000 meter, 6,000 meter, or 30 minute at stroke rate 20 (30 r20) test is normally used, depending on when during the season the test is done. This can be used to determine a rowers fitness and can be used to track progress for an athlete. These sessions are typically maximum effort sessions with some degree of taper (typically 4-7 days maximum) to illicit a peak performance. On the water, times can also be used to judge an entire crew, with external factors like wind direction and intensity, current flow, and temperature considered. Some boats and squads may use on the water telemetry to quantify the impact and power output of each rower. On the water performances may not always be peak efforts depending on a squads seasonal focus or an event's progression (eg. a maximal effort will rarely make an appearance in the heat stage of a heat-semi-final progression). Ergometer scores tend to be the pureset way of quantifying rowers performance, however a combination of approaches can be used to quantify on the water performance.

2.1.4 Summary

Rowing training is quite time consuming for athletes. Many rowers will learn much of what has been discussed in this section to make better decisions themselves, or understand why training is done in a given way. The focus for a majority of the training cycle is to build a strong aerobic base in order to allow for higher "peak", ideally coinciding with the peak event for a season of training. For international athletes this will be a yearly cycle targetting World Championships, which is part of a four-yearly cycle targetting the Olympics. For national level athletes, national championships are typically the target, with Henley Royal Regatta featuring as a seasons target. Coaches will normally take into account season targets and rower proficiency and base fitness when generating training plans. The target of this project is to make it easier for coaches and more accessible for more independent athletes.

2.2 A Review of Performance Modelling

The first wide spread approach to modelling performance was developed by Banister *et al.* [1] with the Banister Fitness - Fatigue model, and further refined by Morton *et al.* [9] when defining training impulse to determine fitness and fatigue. The use of machine learning, specifically artificial neural networks has since been introduced in approaches to model performance. In preparation for the 2000 Olympics in Sydney Edelmann-nusser *et al.* [10] successfully predicted olympic performances within an error of 0.05 seconds across a total time of 2:12.64 (min:sec) for the 200m backstroke event. Edelmann-nusser *et al.* [10] specifically consider the limitation of using a linear model, such as the model proposed by Banister *et al.* [1], on training adaptation and performance; adaptation to training is inherently a complex non-linear process.

This chapter will review basic approaches to systems modelling, exploring metrics which are commonly used to guide models and some non-linear systems models which have been explored.

2.2.1 Quantifying Training Load (Fatigue)

Rate of Perceived Exertion (RPE)

Rate of perceived exertion (RPE) is a scale, originally introduced by Gunnar Borg, to measure an athletes effort. The original scale ranged from 6-20, where 6 would be no exertion at all, and 20 is maximal effort. This scale ranges from 6-20 to correspond more easily with heart rate, as the scale is used beyond just athletic settings. Therefore another scale, ranging from 0-10, was developed by Borg as well for use with "extreme intensity of activity", this is the scale normally used in athletic studies. The second scale is called a

Category-Ratio (CR) scale where the reference anchor is RPE CR10 of 10, meaning maximal effort/pain. Session RPE (sRPE), using CR10, can be used, in conjunction with duration to determine session load [11].

Borg RPE	
Score	Level of exertion
6	No exertion at all
7	
7.5	Extremely light
8	
9	Very light
10	
11	Light
12	
13	Somewhat hard
14	
15	Hard (heavy)
16	
17	Very hard
18	
19	Extremely hard
20	Maximal exertion

Figure 2.1: The original Borg 6-20 RPE Scale [11]

Borg CR10 scale	
Score	Level of exertion
0	No exertion at all
0.5	Very, very slight (just noticeable)
1	Very slight
2	Slight
3	Moderate
4	Somewhat severe
5	Severe
6	
7	Very severe
8	
9	Very, very severe (almost maximal)
10	Maximal

Figure 2.2: The Borg Category-Ratio 10 RPE Scale [11]

The two Borg RPE scales

Training Impulse (TRIMP)

Training Impulse (TRIMP) is a method to calculate training load and is defined as $\text{TRIMP} = \text{Training Volume} \times \text{Training Intensity}$ [12]. There are many methods to calculate TRIMP, using different metrics to calculate training volume and training intensity. For the purposes of this project volume will simply be minutes, and intensity will be average heart rate (bpm), as the simplest TRIMP method outlined. Other modifications of may apply a weighting against a heart rate metric to normalise longer sessions completed at a lower heart rate – considering the noted difference in heart rate responses to training in male and female athletes [9], alternatively load can be calculated by using the product of RPE and duration (ie. $\text{RPE} \times \text{Duration (minutes)}$) .

Acute Chronic Workload Ratio (ACWR)

Acute Chronic Workload Ratio (ACWR) can be used to monitor load in an athlete. It compares training load accumulated, in arbitrary units (AU), over the last seven days (acute workload) to the training load accumulated over the last twenty-eight days (chronic workload). The exact calculations used to determine acute and chronic workload depends on what type of ACWR model is selected. Additionally, the exact time periods considered for acute and chronic load can change depending on the application [13]. Typically ACWR is

used for injury management in team sports, but some articles have found it may not be effective for preventing injury as it is inaccurate way of approximating load [14]. Regardless of it's effectiveness in reducing injury, ACWR charts are easy to produce and provide easy feedback to rowers to see how their training load can change week-to-week.

2.2.2 Impulse-Response Models

Banister Fitness-Fatigue Model

Limitations to the Impulse-Response Model

2.2.3 Alternative Models

Performance Potential (PerPot)

Artificial Neural Network (ANN) approaches

Chapter 3

Data Collection and Management

Discussing approaches to data collection, key decisions in generating the model, eg. building a model which can be run with minimal effort by the users.

Chapter 4

Data Analysis and Visualization

Basic data analysis done, visualizations generated, comment on feedback given to users.

Chapter 5

Machine Learning Applications

What was the target, what did I actually make, maybe some results

5.1 Model Considerations

What led to the final model, difficulties, and data Considerations

5.2 The Implementation

Details about the model, including some examples of outputs (on my own data?)

Chapter 6

Discussion

Evaluation of methods and model, and a discussion of potential further step.

Chapter 7

Conclusion

Did we do what we set out to do.

Bibliography

- [1] E. W. Banister, T. W. Calvert, M. V. Savage, and T. Bach, "A systems model of the effects of training on physical performance," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-6, no. 2, pp. 94–102, Feb. 1976. DOI: 10.1109/tsmc.1976.5409179.
- [2] J. Mäestu, J. Jürimäe, and T. Jürimäe, "Monitoring of performance and training in rowing," *Sports Medicine*, vol. 35, no. 7, pp. 597–617, Jul. 2005. DOI: 10.2165/00007256-200535070-00005.
- [3] K. S. Seiler and G. Ø. Kjerland, "Quantifying training intensity distribution in elite endurance athletes: is there evidence for an "optimal" distribution?" *Scandinavian Journal of Medicine & Science in Sports*, vol. 16, no. 1, pp. 49–56, 2006, ISSN: 0905-7188. DOI: 10.1111/j.1600-0838.2004.00418.x.
- [4] M. A. Rosenblat, A. S. Perrotta, and B. Vicenzino, "Polarized vs. threshold training intensity distribution on endurance sport performance: A systematic review and meta-analysis of randomized controlled trials," *Journal of Strength and Conditioning Research*, vol. 33, no. 12, pp. 3491–3500, Dec. 2019. DOI: 10.1519/jsc.0000000000002618.
- [5] A. Das, U. S. Kaniganti, S. J. Shenoy, P. Majumdar, and A. K. Syamal, "Monitoring training load, muscle damage, and body composition changes of elite indian rowers during a periodized training program," *Journal of Science in Sport and Exercise*, vol. 5, no. 4, pp. 348–359, Nov. 2022. DOI: 10.1007/s42978-022-00197-7.
- [6] S. Seiler and G. Ø. Kjerland, "Quantifying training intensity distribution in elite endurance athletes: Is there evidence for an "optimal" distribution?" *Scandinavian Journal of Medicine and Science in Sports*, vol. 16, no. 1, pp. 49–56, Oct. 2004. DOI: 10.1111/j.1600-0838.2004.00418.x.
- [7] S. Selles-Perez, J. Fernández-Sáez, and R. Cejuela, "Polarized and pyramidal training intensity distribution: Relationship with a half-ironman distance triathlon competition," *Journal of Sports Medicine and Science*, vol. 18, no. 4, pp. 708–715, Nov. 2019.
- [8] A. S. Göktepe, "Energy systems in sport," in *Amputee Sports for Victims of Terrorism*. IOS Press, 2007, pp. 24–31.

- [9] R. H. Morton, J. R. Fitz-Clarke, and E. W. Banister, "Modeling human performance in running," *Journal of Applied Physiology*, vol. 69, no. 3, pp. 1171–1177, Sep. 1990. DOI: 10.1152/jappl.1990.69.3.1171.
- [10] J. Edelmann-nusser, A. Hohmann, and B. Henneberg, "Modeling and prediction of competitive performance in swimming upon neural networks," *European Journal of Sport Science*, vol. 2, no. 2, pp. 1–10, Apr. 2002. DOI: 10.1080/17461390200072201.
- [11] N. Williams, "The borg rating of perceived exertion (rpe) scale," *Occupational Medicine*, vol. 67, no. 5, pp. 404–405, Jul. 2017. DOI: 10.1093/occmed/kqx063.
- [12] M. Kent, *Trimp method*, 2007. DOI: 10.1093/acref/9780198568506.013.7418. [Online]. Available: <https://www.oxfordreference.com/view/10.1093/acref/9780198568506.001.0001/acref-9780198568506-e-7418>.
- [13] R. White, *Acute:chronic workload ratio*, Oct. 2023. [Online]. Available: <https://www.scienceforsport.com/acutechronic-workload-ratio/>.
- [14] F. M. Impellizzeri, M. S. Tenan, T. Kempton, A. Novak, and A. J. Coutts, "Acute:chronic workload ratio: Conceptual issues and fundamental pitfalls," *International Journal of Sports Physiology and Performance*, vol. 15, no. 6, pp. 907–913, Jul. 2020. DOI: 10.1123/ijsp.2019-0864.

Appendix A1

Appendix

You may use appendices to include relevant background information, such as calibration certificates, derivations of key equations or presentation of a particular data reduction method. You should not use the appendices to dump large amounts of additional results or data which are not properly discussed. If these results are really relevant, then they should appear in the main body of the report.

A1.1 Appendix numbering

Appendices are numbered sequentially, A1, A2, A3. . . The sections, figures and tables within appendices are numbered in the same way as in the main text. For example, the first figure in Appendix A1 would be Figure A1.1. Equations continue the numbering from the main text.