A mathematical model to determine optimum cadence for an individual cyclist using power output, heart rate and cadence data collected in the field

PhD Thesis

Robert REED

Salford Business School,
Universtity of Salford, Salford, UK

Submitted in Partial Fulfilment of the Requirements of the Degree of Doctor of Philosophy, September 2013

Contents

List of Tables

List of Figures

Abstract

We aim to develop a methodology to determine individual optimum cadences for competitive cyclists using field data. Cadence is the number of pedal crank revolutions per minute or pedalling rate. Currently athletes tend to select a cadence intuitively (choosing a gear that permits a cadence that feels comfortable), with some advice from coaches. Literature defines optimum cadence based on gross efficiency. However only power output, heart rate and cadence measurements from the field are available to us. Hence we determine an optimum cadence as the cadence that minimises heart rate for a given power output. In so doing we consider heart rate a reasonable proxy for gross efficiency. We fit statistical models of power output, heart rate amd cadence, with heart rate lagged behind changes in power output, at various lags (though we believe 30 seconds is appropriate). We consider the effect of fatigue on optimum cadence through calculation of training impulses or TRIMPs, but do not consider the effects of fitness, gradient, or whether athletes are standing or sitting. Optimum cadences are found for two athletes (83 and 70 revolutions per minute respectively); these cadences are similar to athletes' preferred cadences (82-92 and 65-75 rpm respectively). Optimum cadences do not vary by power output or heart rate in our study, and are relatively insensitive to TRIMP. Power output reduces by approximately 2% for cadences 10 rpm above or below optimum. The methodology we propose can be implemented by a wide range of competitive cyclists to calculate optimum cadence; cyclists need to collect power output, heart rate and cadence measurements from training sessions over an extended period (>6 months), and ride at a range of cadences within those sessions. Cyclists and their coaches can re-calculate optimum cadence, say every 6 months, to take account of possible changes in fitness.

Chapter 1: Introduction

1.1. Background

Cyclists are able to optimise their race performance by controlling the pace and hence the power output at which they ride at different stages within a race (Aisbett et al, 2009; Atkinson et al, 2003; Chaffin et al, 2008). For a target pace (a target power output), a cyclist can select a number of different combinations of gear and cadence. Cadence is the number of pedal-crank revolutions per minute, and is therefore proportional to the speed of rotation of the feet when pedalling. Having determined an optimum pace, the cyclist may then seek to further improve his or her performance by optimising the choice of gear and cadence combination. Cadence is linked to gear selection; the higher the gear, the more resistance encountered, and the harder it is to maintain the same cadence. An optimum gear and cadence combination may only bring a small level of improvement to performance, but at a professional level a small margin of improvement in performance (from optimising cadence) may be significant (Abbiss et al, 2009; Coast and Welch, 1985; Foss and Hallen, 2004). Currently athletes tend to select a gear and cadence combination intuitively (selecting a gear that allows them to ride at a cadence that feels most comfortable to them), with some advice from coaches. We have discussed the potential for performance improvement with sports scientists working with the same data from the same athletes. Optimising cadence has been identified as potential area where athletes can improve performance.

Current scientific studies into optimising cadence are typically conducted through laboratory tests on ergometers (indoor stationary cycling machines); athletes ride until exhaustion at a range of cadences, selecting the cadence at which they could maintain the highest power output before exhaustion (Foss and Hallen, 2004). However, the use of cycle ergometers may not necessarily accurately replicate conditions of road cycling (Jobson et al, 2007). While much research considers optimum cadence based on laboratory experiments with cycle ergometers and biomechanical and physiological theory, this research encompasses an empirical approach using power output, heart rate and cadence data obtained from a number of competitive cyclists' training sessions. The use of training data in calculating an optimum cadence potentially allows for more precision, as more data are collected, and athletes will have ridden at a greater range of cadences than in ergometer studies. In ergometer based studies, the purpose of the test is scientific enquiry rather than to improve an athlete's fitness, so the range of cadences tested is decided by an academic rather than a training coach – unless the test is carefully scheduled within an athlete's training schedule, researchers must assume that the athlete is not withholding effort or insufficiently motivated for a test that, unlike a real training session, is not for his or her benefit. To our knowledge, this is the first study that attempts to calculate an optimum cadence statistically from training data.

Training data from four competitive athletes are used. These data comprise measurements of power output, heart rate and cadence, recorded within-session and over a large number of sessions. Power output and cadence measurements are recorded using an SRM power meter on the bicycle (SRM is a power meter developed by the engineering company Schoberer Rad

Messtehnik.) Heart rate is calculated using a heart rate monitor. The SRM power meter calculates power output and cadence every half second and the heart rate monitor calculates heart rate every second, but due to limited storage capacity, athletes collect and store data at five second intervals (averages calculated over 5 second intervals, yelding one stored averaged measurement every 5 seconds). The data available to us are the calculated averages yielded and stored every 5 seconds. As such when we refer to 'raw' data throughout this thesis, we mean the data (calculated averages) presented to us at five second intervals. Hence we consider power output measured on the bicycle, rather than the physical power generated by the athletes themselves (the rate at which external mechancial work of the body is performed, Winter and Fowler, 2009).

We develop regression models of power output, heart rate and cadence, in order to calculate individual optimum cadences for the athletes in our analysis. Since heart rate can be considered a response to changes in power output (Churchill, 2009; Grazzi et al, 1999), we consider heart rate to be temporally a response to power output in our models. The statistical models we use imply the following definition of optimum cadence: the optimum cadence is that cadence which maximises power output for a given heart rate, or equivalently, it is that cadence that minimises heart rate for a given power output. We determine the optimum cadence under this definition. Ideally we would calculate gross efficiency, which is the amount of work done over the metabolic cost required to do that work (Winter and Fowler, 2009), and optimise cadence with respect to this measure. However laborotory testing is necessary to calculate gross efficiency. Although heart rate can be affected by additional factors such as dehydration and temperature, we consider heart rate to be a reasonable proxy for gross efficiency (Wingo et al, 2005). Furthermore, heart-rate data can be routinely collected in the field using a heart-rate monitor. Another commonly used indicator of effort exerted is rate of perceived exertion. While this quantity can be routinely measured at a session level, this will not be useful for optimising cadence when cadence itself varies during a session as is typical for field data.

There is a delay or time lag between changes in power output and the heart rate response, but the length of this delay is not clear from literature (Churchill et al, 2009). Therefore we investigate the size of this time lag. However a study by Stirling et al (2008) suggests a relatively low heart rate lag (less than 30 seconds) is appropriate for our data.

The only variables required to calculate an optimum cadence in our study are power output, heart rate and cadence. The gear in which an athlete is riding is not needed. Changing gear automatically enables the athlete to change cadence accordingly – the higher the gear, the more resistance provided by the bicycle; if a rider selects a lower gear he/she can increase cadence for the same level of exertion as he/she encounters less resistance.

Nevertheless we also consider additional variables that can be calculated from the power output, heart-rate, and cadence data, and explore how optimum cadence is related to these additional variables. For example, the training impulse developed in the session (up to a particular point) is one such explanatory variable, which we use to calculate training load. Training impulse refers to the average heart rate in a session multiplied by session duration, then multiplied by a weighting factor to give greater weighting to sessions where heart rate is particularly high.

We review sports science literature relevant to our study in chapter 2. We then describe the data we use in chapter 3. In chapter 4, we explore a range of regression models of power output, heart rate and cadence; we describe the methods for determining optimum cadence, and present the results and discussion of fitting such models. We then consider different data processing constants (such as heart rate lag) in chapter 5. In chapter 6, we include training impulse calculations (known as TRIMPs) in our regression models. With these calculations we aim to capture long term effects of fatigue, and observe how this affects optimum cadence. In chapter 7 we present a model of power output, heart rate excess and cadence – where heart rate excess is the current heart rate minus the resting heart rate. Finally we summarise the research undertaken in chapter 8 and conclude (in chapter 9) with key findings of the study and how the methodology can be implemented.

1.2. Aims and objectives

Our overall aim is to develop and apply a methodology to determine an optimum cadence for an individual cyclist using power output, heart rate and cadence data collected in the field, This leads to the folowing objectives:

- To develop a statistical model that relates heart-rate to power output and cadence and that allows an optimum cadence to be determined.
- To apply this statistical model and determine optimum cadences for a number of athletes for whom we have heart rate, power output and cadence data collected during competition and training.
- To explore what other factors may affect optimum cadence, where such factors may be calculated from field data, and to describe how such factors influence optimum cadence.
- To compare the optimum cadences yielded with the preferred cadences of the athletes in our study, and to compare optimum cadences with those typically found in experimental literature.
- To describe how the methodology should be used by a coach and athlete to determine optimum cadence.
- To describe what future work could be undertaken to further develop our methodology.

Chapter 2: Sports Science background

2.1. Introduction to sports science literature

In our study we aim to develop statistical models of power output, heart rate and cadence (collected from training data for competitive professional cyclists) that yield an individual optimum cadence – that is, a cadence that maximises power output for a given heart rate. To do this we use power output, heart rate and cadence measurements taken directly from power meters on the bicycles athletes in our studies have used. We expect optimising cadence in this way to bring a small but significant gain in performance as studies have suggested cadence is important in maximising power output (Abiss et al, 2009; Coast and Welch, 1985). In this section we present a review of existing literature into optimising cadence in road cycling, and related issues that affect cycling performance, including optimal pacing strategies, and, more broadly, investigations into the relationship between key variables such as power output and heart rate. We aim to develop our understanding of key physiological and sports science aspects that underpin professional road cycling, and to familiarise ourselves with relevant literature and theories that could aid the development of our statistical models. Firstly we present literature for optimising pace in 2.2. Next in 2.3. we present definitions used in the literature for optimum cadence and experimental conditions in such studies, and how these compare to our study. We explore relationships between other key variables in our analysis (such as power output and heart rate) in 2.4. In 2.5. we explore other factors that could potentially affect cadence. We summarise the key literature for our study in 2.6. Finally, in section 2.7., we describe the variables required to calculate a mathematical optimum cadence in our study.

Broadly, road cycling races can be classified as time trials, in which all athletes take turns to complete a course in the fastest time possible, or, alternatively, races in which all athletes start and compete together at the same time (de Koning et al, 1999). Professional cycling races can be run outdoors on roads or other terrain, or on tracks or circuits in an arena, in which athletes complete a number of laps of the circuit. The athletes in our study compete in road races and in individual time trials.

2.2. Optimising cycling pace

2.2.1. Introduction

Cyclists are able to optimise their race performance by choosing an appropriate pacing strategy, and regulating the power output they produce at different stages within a race (Aisbett et al, 2009; Atkinson et al, 2003; Chaffin et al, 2008). Indeed much research exists into optimising pacing strategy in cycling; studies have investigated a range of different pacing strategies, for a variety of different test durations. The terms ‘pacing strategy’ and ‘pacing pattern’ tend to be used in literature to broadly represent the distribution of variation in power output with a race. For example, pacing strategy was defined by Atkinson and Brunskill (2000) to be ‘the within-race distribution of work rate (power output)’. Hence throughout section 2.1, when we use the terms ‘pacing’ and ‘pacing strategy’, we are referring to the variation in power output throughout a race.

To our knowledge, studies that investigate optimising pacing strategy have tended to focus on time trials (in which athletes race individually against the clock), possibly because optimal performance in races in which athletes compete together at the same time is complicated by the importance of tactics, such as whether to lead or follow other athletes at different stages of the race (de Koning et al, 1999). Pacing strategy may be very important for professional cycling, with performances for different pacing strategies differing by 1%, which is representative of the typical differences between performances for gold silver and bronze winning performances in world championship cycling (de Koning et al, 1999).

2.2.2. Pacing strategies

Studies have investigated the use of an even pacing strategy (varying power output as little as possible within a race). Many of these studies have indeed concluded that even pacing may be optimal for performance (Chaffin et al, 2008; Foster et al, 1993; Nikolopoulos et al, 2001).

Foster et al (1993) tested nine subjects using a wind load simulator and attached bicycle, with a series of 2km time trials in which pacing strategy varied. Subjects controlled variation in power output such that they were able to complete the first kilometre in a time that was between 48% to 55% of the overall time for the trial. Foster et al (1993) concluded that their data suggest an even pace throughout what they consider middle distance time trials (2 to 4 km) is the optimal pace, and that deviations from an even pace have a significantly detrimental effect on the overall performance.

Atkinson and Brunskill (2000) suggested that even pacing is optimal, but that pacing strategy can be affected by wind conditions. They suggested that, for time trials in which the first half of the race is into a headwind and the second half of the races is into a tailwind, even pacing is appropriate (although they claim some athletes do not always adopt an even pacing strategy in such conditions). Athletes may have to vary the power output at which they ride if they encounter varying wind speed and directions, or for varying road gradients (Atkinson and Brunskill, 2000).

For an even pacing strategy, power output necessarily fluctuates only slightly throughout. Nevertheless there may be biological reasons for such fluctuations. Ulmer (1996) suggested that the slight fluctuations in power output may prevent cellular damage during intense exercise, and could also prevent internal organ damage.

Other studies have demonstrated that an early surge in power output before levelling to a steady power output for the rest of the race may form an optimal pacing strategy (Aisbett et al, 2009; de Koning et al, 1999; Nikolopoulis et al, 2001; Van Ingen Schenau et al, 1992).

Aisbett et al (2009) tested twenty-six well-trained male participants using time trials with varying starting stretegies. The study comprised six test sessions for each participant, with varying formats, and never on consecutive days. For three of the six tests, participants varied the start (first quarter of the test) with a fast start, slow start, or even paced race, in a randomized order amongst the participants. Participants also rode two 5 minute time trials in which they chose their own pace, to set as a benchmark finsihing time for the varying start trials. They found that participants starting with a high power output were able to sustain a higher mean power output for the duration of the test than were participants with a slow start or even paced start.

Moreover, Van Ingen Schenau et al (1992) suggested that, not only should athletes adopt a pacing strategy in which power output is very high at the beginning of the race, but that subsequent power output for the remainder of the race should be even (at least for 4000m pursuit races). For shorter races (1000m) they claimed that the high power output at the start of the race can be maintained, and there is no need to reduce power output to an even, and lowered, level for the remainder of the race.

De Koning et al (1999) also investigated the pacing strategy of an early surge in power output, in time trials of 1000m and 4000m. Whereas Aisbett et al (1999) used participants in laboratory experiments, de Koning et al used simulated races, using models to calculate a range of key variables, including power output production and rate of change in kinetic energy. De Koning et al (1999) considered the rate of change in kinetic energy to be equal to the power output production minus the power output dissipation to friction, which is the power output needed to overcome the friction encountered. The friction in the de Koning study had two components – the friction of the air against the clothes of the athlete, and the friction produced from air pressure differences between air in front of and behind the athlete. Kinetic energy was considered wasted energy as the athlete cannot use it to improve his or her own performance. They found an early surge in power output to be beneficial to performance. For the 1000m time trial, for the pacing strategy with the highest power output at the beginning, although the amount of energy lost to friction was highest (due to a high mean velocity over the race distance), the amount of leftover kinetic energy was lowest, and overall, athletes were able to complete the race most quickly. They also explored the benefits of an early surge in power output in the 4000m pursuit. They simulated a range of races in which the athletes started with a high power output, before reducing power output to a steady level for the remainder of the race. In their study the optimal pacing strategy for the 4000m pursuit was to begin with a high power output, before reducing power output after 12 seconds. Furthermore, they claimed that, not only is a high power output at the beginning of the 4000m pursuit race crucial for performance, but precisely how much time has elapsed before the athlete reduces to a steady, lower power output is also crucial.

Along with experimental data (laboratory sessions or simulated races), de Koning et al (1999) also observed that an early surge in power output corresponded to an optimal pacing strategy in 1998 world track championship cycling, amongst the top 8 riders, also in the 1000m time trial. There was a strong positive correlation (0.71) between time taken to complete the first 250m (the first lap) and overall race finishing time, whereas there was a negative correlation (albeit very low, -0.06) between time taken to compelte the last 250m (the last lap) and time taken to complete the whole race. They also explored the amount of time taken for successful athletes who rode with an early surge in power output to reduce to a lower, steady power output. In the 1998 world track championships 4000m pursuit race, athletes completed the first half of the race in 50.9% of the total time to complete the whole race – indeed this was very similar to the optimal pacing strategy in the simulated 4000m pursuit race in the de Koning et al (1999) study (50.7%).

A range of explanations have been suggested for why a high power output produced at the beginning of a race may lead to improved performance. Bishop et al (2002), in a study of kayaking, claimed that an early surge in power output leads to higher levels of oxygen consumption at the beginning of the race, and that that may lead to an increase in metabolic functioning, allowing for greater performance in the race. Alternatively de Koning et al (1999) suggested that the reduction in wasted kinetic energy may be why an early surge in power output is beneficial to overall race performance.

However Aisbett et al (2009) claimed that it is unclear how an early surge in power output may improve performance. They suggested the alleged improvement in performance associated with a fast start may to some extent be due to the reduction in the amount of time the athlete spends accelerating from a standing starting position. Aisbett et al (2009) acknowledged though that they were not able to monitor the amount of time spent accelerating in their study, and so were unable to test if this was indeed the reason for the improved performance associated with a high power output at the start.

However Wilberg and Pratt (1988) claimed that too high a power output at the beginning of a race may hinder performance, as it could increase the chance of fatigue occuring before the end of the race. They studied 222 male and female professional cyclists in 1000m time trials, finding that participants that adopted the extremely high early surge in power output were slower overall than those who did not. Thomson et al, (2003) also demonstrated (albeit in swimming not cycling) that fatigue occurs before the end of a race when maximal power output is adopted at the beginning of a race. Despite the possible risk of fatigue from a high power output fast start, Aisbett et al (2009) claimed that such fatigue is not likely to incur as long as the high power output at the beginning is not as high as the maximal power output that the athlete can achieve. They claimed that the extent to which athletes should produce an early power output surge may depend on aeorbic and anaerobic metabolism.

Chaffin et al (2008) claimed that existing studies in optimising pace in cycling tend to be of short duration (less than 30 minutes), and that there is a shortage of studies investigating pacing strategies for longer sessions or competitions. Indeed Aisbett et al (2009) also claimed that an optimal pacing strategy is largely dependent on duration of the race. Other authors however have considered race durations beyond such as short duration. Chaffin et al (2008) investigated optimal pacing strategy for a 30 minute maximal test. In their study they used twelve amateur (albeit well-trained) male and female cyclists between 19 and 45 years of age. They performed a 30 minute maximal cycling test, choosing their own pace. Power output, heart rate and oxygen uptake were recorded, along with the perceived exertion recorded by the athletes. The power output changed seldom over the course of the test (variations in power output tended to be less than 10W from minute to minute), although there was a great increase in power output in the final 30 seconds of the test.

Contrary to studies investigating performance over 1000m time trials or 4000m pursuits, in which participants complete races or tests in much less than 30 minutes, athletes in the Chaffin et al (2008) study did not produce an early surge in power output. Given that a very high power output at the beginning of a test or race can result in fatigue occurring before the end of the race (Aisbett et al, 2009), athletes therefore might not necessarily be able to adopt the pacing strategy of an early surge in power output for races of much greater length than the 4000m pursuit. Chaffin et al (2008) suggested that participants may have intentionally selected a slow, steady pace at first because they believed the work load associated with a high power output throughout the whole test would be too great. Instead Chaffin et al noted that the results of their study suggest a relatively even pacing strategy (albeit with a great surge in power output at the very end of the study) is optimal for long races. However they still suggested further work is needed for time trials of even longer duration and distance.

Indeed Thomas et al (2012) compared even and variable pacing strategies for long time trials (of 20km). In their study, ten well trained male cyclists rode three 20km time trials on a laboratory cycling ergometer. Subjects completed the time trials more quicly with an even pace than a varying pace, whilst even pacing also reduced the perception of effort required.

Wells et al (2013) have further suggested that, as race distance increases, so too does the importance of even pacing. They simulated a series of time trials for a hypothetical rider, for distances of 4, 16.1, 20 and 40km, for flat, windless conditions. For each distance, mean power output varied between 200 and 600W, with power output changed by between 5 and 15%. There were also separate trials depending on the frequency at which power output was changed within the simulated time trial – these were 2, 4, 8, 16 and 32 times. For the 4km time trial, variation in power output tended helped to improve performance. However for the longer time trials, as race distance increased, there were fewer combinations (of variation in mean power output and frequency of variation) that yielded improved performance. Wells et al (2013) concluded that for long race distances (over 16km), performance is optimised by adopting a constant power output and pace.

Most academic studies into cycling pacing strategy tend to be conducted such that the pacing strategies adopted by the participants are selected by the experimenters or academics (Nikolopoulis et al, 2001). There are also studies in which participants select their own pacing strategy. This enables experimenters to investigate how performance is affected by not only the actual distance of the race but also by perceived distance of the race (Nikolopoulis et al, 2001). Such studies have tended to focus on relatively long race distances (34km or longer) compared to other pacing strategy studies, which is understandale as it is implausible that athletes would perceive race distances incorrectly for events as short as the 1000m time trial or even the 4000m pursuit.

Palmer et al (1998) tested participants who rode three time trials of different distances (34km, 40km and 46km). Participants selected their own pacing strategy. In order to compare performance differences for different race distances for which athletes are unaware of differences, the experimenters informed participants that all three time trials were of equal distance – 40km, the middle distance of the three races. In fact participants performed similarly for all three races, as power output did not vary significantly for the three different race distances, suggesting that the perceived distance of the race has a significant effect on how athletes perform (Palmer et al, 1998).

Nikolopoulis et al (2001) further investigated the phenomenon of perceived and actual race distances. They tested 5 male cyclists, who were trained to a greater extent than were the participants in the Palmer et al (1998) study. They attempted to recreate the findings of Palmer et al (1998), subjecting each participant to 34km, 40km and 46km time trials whilst informing them that each race is 40km, to examine how participants perform when perceived distance does not match the actual distance. In addition to this, Nikolopoulis et al (2001) then repeated the shortest and longest time trial (this time informing participants of the race distances), in order to compare how participants performed with and without correct knowledge of the race distance. Indeed Nikolopoulis et al (2001) found similar results to Palmer et al (1998). When informed that all races were 40km, participants did not vary power output significantly between trials in which actual distance varied. Participants had access to heart rate measurements (which could act as an indication of intensity of exercise (Nikolopoulis et al, 2001)) and hence had an opportunity to change their pace if they so wished, but they did not change pace. Power output however changed accordingly for the second experiment in which athletes were informed of race distance. Power output changed only slightly though, and not statistically significantly, although possibly by enough to make a significant contribution to performance differences over the entire race durations (Nikolopoulis et al, 2001).

Nevertheless the less than expected power output differences for the trials in which athletes knew the race distance, and that such results were obtained even for well-trained athletes (the participants in the Nikolopoulis et al study), emphasises how important perceived race distance is for pacing strategy. Nikolopulis et al (2001) claimed that athletes select a pacing stretegy based not necessarily on the race distance but on their subjective perception of how much effort is required for the race. However they note that only a small sample was used, and that larger samples may yield more statistically significant differences between performances in different trials.

Indeed de Koning et al (2011) have suggested that the tendency of athletes to change pace in a race is related to their race understanding of race distance, or how much of the race remains. They have also argued that perceived exertion influences changes in pace in races.

Although a range of studies have suggested that different pacing strategies are optimal for different circumstances, a study by Aisbett et al (2003) has suggested that different pacing strategies do not necessarily have a significant impact on performance. Aisbett et al (2003) tested six male participants setting them ergometer cycling tests. Each participant performed two practice trials to gain familiarity with the cycling ergometer equipment, before performing three six-minute trials, over a period of two weeks. Three different pacing strategies were adopted by the participants – fast-start, even pacing and slow-fast. The strategies were based on the practice trials. For the fast-start trial, participants were instructed to maintain 104% of the mean power output obtained during the fastest practice trial for the first 120 seconds of the main trial. For the even pacing strategy they maintained 101% of the mean power output for the first 240 seconds, and for the slow-fast pacing strategy 98% for the first 240 seconds. Aisbett et al (2003) tested for statistical differences in power output, along with other variables related to oxygen consumption (such as peak VO_2). Although mean power output generated was higher for the fast-start condition compared with the other strategies, it was not significantly higher. Also, even pacing snd slow-fast strategies were almost identical in terms of mean power output. However as literature has suggested that pacing strategy may impact upon performance and power output that can be produced (albiet in different ways according to different authors), it may be therefore beneficial for athletes to try explore different pacing strategies. Indeed, Aisbett et al (2003) suggested there is limited research to suggest that even pacing is an optimum pacing strategy.

Air resistance can also have an impact on performance for athletes in road cycling in terms of pacing strategy; the perception of air resistance could enable athletes to select and modify their pacing strategy during a road race (Nikolopoulis et al, 2001). Nikolopoulis et al (2001) noted that, in their laboratory study in which athletes selected their own pacing strategy for different time trials (which participants perceived to be all the same distance though actual distance varied), some physical feedback clues were absent in the laboratory, such as levels of air resistance, or speed at which participants cover the ground (Nikolopoulis et al, 2001).

A number of different authors have found different relationships between pacing strategy and performance – this may be due to differences in the types of tests deployed or exercise duration (Aisbett et al, 2003; Aisbett et al, 2009; Chaffin et al 2008). Indeed the fast-start pacing strategy tends to vary greatly between different studies, in terms of the intensity and duration of the fast start (Aisbett et al, 2003). For the Aisbett et al (2003) study, in which no significant differences were obtained regarding power output produced from different strategies, the intense start was maintained for approximately 2 minutes. However, simulated cycling trials of 1000m and 4000m from de Koning et al (1999) and van Ingen Schenau (1992) suggest that, for a fast-start pacing strategy, the maximal start should last for a brief time, up to 12 seconds. Athletes in our study compete in time trial races of longer duration than 4000m. Even pacing strategies tend to be the best pacing strategies for long time trials, say 34 km and longer (Chaffin et al, 2008; Palmer et al, 1998; Thomas et al, 2012; Wells et al, 2013).

2.2.3. Existing mathematical models of performance prediction in cycling

A range of mathematical models have been used in academic studies of cycling, typically to predict pacing strategy (power output) an athlete needs to ride at in given circumstances (Di Prampero et al, 1979; Martin et al, 1998; Olds et al, 1995). Models of power output are also useful when estimates of power output are required but no direct measure of power output is available, and in comparing different cycling performances from the past in which there are no available data for power output (Bassett et al, 1999). We explore and summarise different models used. To our knowledge, mathematical modelling in cycling has tended not to be used for optimising pedalling rate or cadence. When outlining equations used by other authors, we use the symbols used by the authors in their articles.

In our study, we use SRM cranks in the bicycles to measure power output. Indeed SRM cranks are considered a valid measurement of power output (Bertucci et al, 2005; Martin et al, 1998; Paton and Hopkins, 2001). Nevertheless some authors have developed mathematical models to predict power output, along with time trial performance. In studies that aim to optimise time trial performance, a measure of power output is important (Martin et al 1998). Without a measure of power output, Martin et al (1998) claimed it is difficult to validate these models. A range of physical and physiological measures have been included in mathematical models of performance in cycling (Martin et al, 1998; Padilla et al, 2000). Martin et al (1998) claimed a model of power output can be produced by having a series of parameters to capture aerodynamic and physical aspects that affect a cyclist’s movement. Of these, air resistance is the variable that is most likely to affect performance, at least for speeds of greater than 50km/h (Kyle, 1988; Padilla et al, 2000). Indeed air resistance, or drag, is calculated using the same equation by Bassett et al (1999), Di Prampero et al (1979) and Martin et al (1998), and a very similar equation is calculated by Padilla et al (2000). Using the nomenclature of Di Prampero et al, this resistance (D, measured in N) is

$$D = \tfrac{1}{2} C_{\mathrm{D}} A_{\rho} \rho V^2$$

where v is the air velocity (in m/s), ρ is air density (in kg/m^3), A_p is the area of the frontal plane of the bicycle (in m^2), and C_D is the drag coefficient., This equation is also used by

Martin et al (1998), albeit with different symbols used to represent the elements of the equation, and Bassett et al (1999) use a very similar equation. However authors differ in how they develop and modify these air resistance calculations, and in how power output relates to other variables in mathematical models. Studies have also tended to include measures of friction between wheels, tyres and bearings. We now describe these studies and the mathematical models used.

Di Prampero et al (1979) developed a model of power output based on air temperature, barometric pressure, gradient and body size. To calculate mechanical power output, Di Prampero et al calculated the air resistance and rolling resistance (a force encountered by tyres rolling over the ground). Subjects on bicycles were towed by cars. They measured speed (and towing force) at two points 100m apart. The rolling resistance is independent of speed, and is proportional to the total mass of bicycle and rider (Di Prampero et al, 1979). As mentioned previously in this chapter, to calculate the air resistance (D, in N), Di Prampero squared the air velocity (v, in m/s), and multiplied this by air density (ρ, in kg/m^3), area of the frontal plane of the bicycle (A_p, in m^2), drag coefficient (C_D), and by a half.

In their study, Di Prampero et al (1979) used linear regression based on air velocity (v^2) to obtain the total resistance (R_T, the air resistance and rolling resistance combined), resulting in the following equation:

$$R_T = 3.2 + 0.19v^2 .$$

To calculate mechanical power output, the rolling resistance and the air resistance were each multiplied by ground speed (s). The mechanical power output in their study (W, in W or watts) was then:

$$W = 3.2s + 0.19v^2 s .$$

Di Prampero (1979) further developed models of power output based on air temperature, barometric pressure, gradient and body size. Di Prampero claimed mechanical power output is directly proportional to air density. They took account of air density by measuring barometric pressure (P_B in Torr) and temperature (T, in K), as they claimed this was enough to analyse the effects of air density. Thus they developed a model of power output for different values of temperature and air density:

$$W = 3.2s + 0.0725(P_B / T)v^2 s .$$

They further developed their model of power output to account for different body sizes and gradient. They simplify the effect of area of the front of the bicycle (A_p), assuming it to be proportional to the surface area of the rider's body (SA, or surface area). This surface area can be calculated using height and mass of the rider (Padilla et al, 2000). When cycling uphill, the rider must overcome an additional force related to the gradient, whilst equivalently there is less force to overcome when cycling downhill (Di Prampero 1979). Assuming the gradient is less than 16 degrees, this additional force (associated with gradient) introduced to the equation was represented by the total mass of bicycle and rider (P) multiplied by the incline or gradient (i) and road speed (s). They calculated mechanical power need for riding uphill (W) to be:

$$W = (4.5\times 10^{-2} Ps)\times((P_B / T)SAv^2 s + 9.8P\times i\times s)$$

Martin et al (1998) also developed a mathematical model to predict power output using physical and aerodynamic measures. In their model of power output, they include measures of aerodynamic drag, rolling resistance, friction of drive chain and wheel bearings, and rate of change in potential and kinetic energy. Friction lost to wheel bearings, however, is very small (Di Prampero et al, 1979; Whitt, 1971). Nevertheless they calculated the power output needed to overcome each of these physical components. Firstly, Martin et al calculated the power output needed to overcome aerodynamic resistance or drag (F_D in N). This was the same as the calculation of drag outlined by Di Prampero et al (1979). Martin et al calculated air resistance to be a multiplication of the coefficient of drag (C_D), the area of the front of the bicycle (A, in m^2), air density (ρ, in kg/m^3), air velocity (V_a^2, in m/s) and by a half, in the following equation:

$$F_D = \tfrac{1}{2} C_{\mathrm{D}} A \rho V_a^2 .$$

However Martin et al also include a term for the power output needed to overcome the aerodynamic effect of rotating the wheels. They multiplied the drag coefficient by area of the front of the bicycle, before adding the incremental drag area of the spokes on the wheels (F_w). This is then multiplied by air density, air velocity and ground velocity of the bicycle (V_G), and all divided by 2, in the following equation of total aerodynamic power (P_{AT}, in W):

$$P_{\mathrm{AT}} = \tfrac{1}{2} \rho (C_{\mathrm{D}} A + F_{\mathrm{W}}) V_a^2 V_G .$$

To calculate the power output needed to overcome rolling resistance, Martin et al first calculated the cosine of the inverse tangent of road gradient (G_R), expressed as $\cos(\tan^{-1}(G^R))$, and shortened to C_{RR}. They then multiplied this by ground velocity, the coefficient of rolling resistance, total mass (mass of bicycle and rider, m_T) and acceleration due to gravity, g, to obtain the power output needed to overcome rolling resistance (P_{RR}) in the following equation:

$$P_{RR} = V_{\mathrm{G}} C_{RR} m_T g \; .$$

The next component measured by Martin et al was friction from bicycle wheel bearings. Wheel bearing friction is related to the load and rotational speed of the wheel (Dahn et al 1991, Martin et al 1998). To calculate power lost due to wheel bearing friction, Martin et al multiplied the ground velocity by 8.7, and added 91. They then multiply this by ground velocity again and by 10^{-3} to get the power needed to overcome friction in the wheel bearings (P_{WB}), in the following equation: $P_{WB} = V_{\mathrm{G}}(91 + 8.7 V_{\mathrm{G}}) 10^{-3}$.

Power is needed to overcome potential energy when a cyclist is riding on a gradient (Martin et al, 1998). To calculate the power output needed to overcome the changes in potential energy, Martin et al calculated the sine of the inverse tangent of road gradient, $\sin^{-1}(\tan^{-1}(G_R)$. They then multiplied this by ground velocity, total mass of bicycle and rider, and acceleration due to gravity, to get power needed to overcome changes in potential energy (P_{PE}), in the following equation: $P_{PE} = V_{\mathrm{G}} m_T g \sin^{-1}(\tan^{-1}(G_R))$.

Finally, Martin et al calculated the power needed to overcome changes in kinetic energy. This is the rate of change in kinetic energy of the bicycle and rider (Martin et al, 1998). They calculated the moment of inertia of the two wheels (I), along with the radius of the tyre (r). They also measured the ground velocity at two time points (t_i, first time point, and t_l, final time point) - they squared the ground velocity at the first time point, and subtracted the square of the ground velocity at the second time point (V_{Gl}^2 – V_{Gi}^2). Power needed to overcome changes in kinetic energy (P_{KE}) was then the following:

$$P_{KE} = [\tfrac{1}{2}(mT + (I/r^2)) \times (V_{Gl}^2 - V_{Gi}^2)]/(t_l - t_1) .$$

Power output in the Martin et al model was then a summation of each of the power outputs needed to overcome the five physical components - aerodynamic drag, rolling resistance, friction of drive chain and wheel bearings, and rate of change in potential and kinetic energy. In the Martin et al study, six experienced male cyclists performed a series of road time trials, on a course measuring 471.8m. Martin et al calculated power output using their model, and compared this with power output measured by SRM cranks on the bicycles. A paired student's t test indicated no significant differences between calculated power output (with a mean of 172.8W) and power output measured using SRM cranks (with a mean of 172.0W). Also, linear regression indicated a high correlation between both power output measurements. Martin et al further claimed that this model can easily be used to predict power output in different circumstances (different physical conditions). This is because the effect of each physical measure (aerodynamic drag, rolling resistance, friction of drive chain and wheel bearings, and rate of change in potential and kinetic energy) on power output is linear or very nearly linear.

Bassett et al (1999) analysed performances of various cyclists that had broken the world record for furthest distance covered during 1 hour of unaccompanied track cycling. They developed a model of power output to predict power output needed for future world record attempts. In modelling power output, they included measures of velocity, altitude, cycling equipment, body position and body weight. Bassett et al measured the effect of each variable in their model on aerodynamic drag by measuring power output using SRM cranks, and by empirically determining the relationship between frontal area of the bicycle and the body surface area of the cyclist. Bassett et al were particularly interested in how power output varies at different altitudes, and consequently in calculating an ideal altitude for future world record attempts.

The model of power output proposed by Bassett et al included constants such as static rolling resistance (the frictional resistance of wheels and tyres) and dynamic rolling resistance (friction from rotating wheels and bearings). These were termed A_1 and A_2 respectively. The third constant, A_3, was an aerodynamic measure; this was a function of air density, coefficient of aerodynamic drag, and frontal area of bicycle and cyclist. The calculation of the aerodynamic measure, A_3, was the same as that used by D Prampero et al (1979) and Martin et al (1998) except it did not include a measure of velocity:

$$A_3 = \tfrac{1}{2} C_D A_f \rho$$

where ρ is the air density (in kg/m^3), C_d is the coefficient of drag, and A_f is the frontal area of the bicycle and cyclist (in m^2). Power output (P, in Watts) was calculated by Bassett et al to be a summation of products of each type of resistance, with velocity, in the following equation:

$$P = A_1 M_t V + A_2 V^2 + A_3 V^2 \, ,$$

where M_t is the total mass of bicycle and cyclist (in kg), and V is velocity (in m/s).

Bassett et al considered the effect of altitude through changes in air density at different altitudes. They modified their model of power output to include factors that take account of roughness of track (K) and a correction for aerodynamic effects including altitude (K_1). K_1 is a multiple of the ratio of air density at an elevated height compared to air density at sea level (K_d), the effect of position of the body on the bicycle (K_{po}), the aerodynamic effects of the shape of the helmet (K_h), the aerodynamic effects of the shape of the bicycle (K_b), and the aerodynamic effects of clothing (K_c): $K_1 = K_b K_c K_d K_h K_{po}$. Power output was then calculated to be:

$$P = K[A_1 M_t V + A_2 V^2 + K_1 A_3 V^2] \ .$$

They calculated the power output requirements of five previous world record rides, and compared them to power output measurements collected from the rides through SRM cranks. In one ride, the SRM power output was 25% higher than the predicted power output from the Bassett et al model – however Bassett claimed this could have been due to the roughness of the track on which this ride took place, or the tightness of the turns of the track. Nevertheless the model provided reasonable predictions of power output, with an average absolute prediction error of 3.9% (Bassett et al, 1999).

Padilla et al (2000) also sought to break the 1 hour world cycling record by developing a mathematical model of power output based on physical measures. Similarly to Di Prampero et al (1979) and Martin et al (1998), they included the size of the area of the front of the bicycle, air density, coefficient of drag, ground speed, rolling resistance, the total mass of the athlete and bicycle, and acceleration due to gravity.

Padilla et al tested one subject, a 30 year old male cyclist. The subject trained in a laboratory on a cycle ergometer, until exhaustion – that is, until he could no longer sustain a cadence of 75 rpm. Padilla et al used data from this laboratory session to calculate a maximal power output for this subject – this was the highest power output the subject could maintain for a period of 4 minutes, with an added 9% to compensate for friction in the cycle ergometer. The subject then performed 4 track tests of increasing distance. Padilla et al then set a target speed for the rider in the world record attempt - indeed the rider successfully broke the world record for the greatest distance covered in 1 hour on a cycling track (53.040km).

Padilla et al (2000) used the surface area of the rider to estimate the size of the frontal area of the bicycle. In their study Padilla et al calculated the frontal area of the bicycle to represent 18.1% of the surface area of the rider's body. The estimated frontal area (FA, in m^2) was used to calculate an estimated power output (W_T, in W), along with drag coefficient (C_d, in m^2), air density (ρ, in kg/m^3), speed (V, in m/s), the coefficient of rolling resistance (C_R), the mass of the bicycle and rider (M, in kg), and acceleration due to gravity (g, 9.81m/s^2). The following equation was yielded:

$$W_T = [FA \times C_D (\rho/2) V_3] + (C_R MgV) \, .$$

Whilst different mathematical models of power output have shared common elements (for example in calculating air resistance), there have been disagreements between different authors. Authors have differed in the extent to which they consider physical characteristics of athletes, such as body size, to affect the power output calculated using mathematical models (Padilla et al, 2000). Basset et al (1999) considered the coefficient of drag to be the same whatever the body size of an athlete. However, Padilla et al claimed that this is not necessarily the case, and that Basset et al consequently underestimated power output in large athletes using their models. Indeed Padilla et al and Bassett et al provided similar estimates of power output for relatively small cyclists, but for larger cyclists, power output estimates are greater using the model outlines by Padilla et al than the model outlines by Bassett et al (Padilla et al, 2000).

Also, authors have used the size of the surface area of a rider's body to estimate the size of the frontal are of the bicycle (Di Prampero et al, 1979; Padilla et al, 2000). The size of the surface are of a rider's body is calculated, and the frontal area is a fraction of this calculation. Different authors have agreed calculated different fractions. For example, Padilla et al (2000) calculated frontal area to be 18.1% of the surface area of the rider's body. This is less than calculated by other authors, including Capelli et al (1993) who calculated frontal area to be 20% of the surface area of the rider's body. However this fraction may depend on the physical characteristics of the rider (Padilla et al, 2000) – a similar fraction (17.8%) was calculated by Swain et al (1987), who used riders with similar physical characteristics as those in the Padilla et al study.

Atkinson et al (2007) and Swain (1997) used the mathematical models of Martin et al (1998) and Di Prampero et al (1979) respectively to predict time savings that can be made by varying power output depending on wind conditions and elevations. Whilst both authors calculated that varying power output on a race course with varying wind speeds can result in time savings compared to an even pacing strategy, they differed in the amount of time that can be saved. Atkinson et al (2007) and Swain et al (1997) calculated the time taken for a hypothetical cyclist to complete a number of different types of races, for example races with varying changes in elevation and wind. They compared the time taken to complete races using an even pacing strategy with a pacing strategy in which the cyclist varied power output for changes in elevation and wind speeds. Swain (1997) found that, on a flat 40km race course with no wind, a cyclist that varied his/her power output by 15% would take 9 seconds longer to complete the course than a cyclist who adopted a constant power output. Atkinson et al (2007), using the Martin et al model, found this to be a 20 second increase. For a course with varying levels of wind and elevation, both authors calculated that varying power output would help a cyclist to complete the course more quickly. The results of the Swain (1997) study indicated that, for a cyclist with an average power output of 289W, varying this power by 10% accordingly (depending on wind conditions), would help the cyclist complete a 10km course 90 faster and a 40km course 30 seconds faster, compared with a cyclist with a constant power output throughout. Atkinson et al (2007) however, found these time savings to be 126 seconds and 51 seconds respectively. Atkinson et al (2007) suggest two possible reasons for the greater time savings of the more recent model of Martin et al (1998) compared to the model of Di Prampero et al (1979). Firstly, the more recent model of Martin et al (1998), compared to the model of Di Prampero et al (1979) used a more modern, aerodynamically efficient bicycle to calculate various terms in their model. Secondly, the models differ in how they were validated (Atkinson et al, 2007). The Di Prampero et al (1979) model was validated by towing the bicycle and cyclist behind a car, measuring the resultant changes in air resistance, whilst the Martin et al (1998) model was validated against SRM cranks measuring power output.

Whilst a number of authors have modelled the impact of physical and aerodynamic factors on performance in cycling, Olds et al (1993) also modelled performance in cycling based on energy requirements and energy expenditure. They developed a model to predict time taken to complete 4000m time trials.

During cycling riders have aerobic and anaerobic sources of energy available to them (Olds et al, 1993). The amount of anaerobic energy available to the rider is determined by the maximal accumulated oxygen deficit, a method developed by Medbo et al (1988). This is the accumulated O_2 demand minus the accumulated uptake of oxygen by the rider (Medbo et al, 1988). The amount of aerobic energy available to a rider depends on how far into a period of exercise he or she is; at the beginning of exercise, oxygen consumption (VO_2) increases, before levelling – when it levels, the rider is said to reach a 'steady-state'. The rider is then consuming oxygen at the highest rate his or her body can produce – VO_{2max}.

For a race of total duration *t* minutes, and assuming and athlete rides at VO_{2max} for duration *f* minutes, the aerobic and anaerobic energy available to the rider (E_{int}, or internal energy available, in litres of O_2) was calculated by Olds et al to be the following:

$$E_{\text{int}} = O_2 def + tfVO_{2\max} - \tau(VO_{2\max} f - VO_{2\text{init}})$$

where VO_{2init} was the oxygen consumption at the beginning of exercise, and τ was a constant. The internal power output of the athlete, or the maximal rate at which this energy can be supplied to the rider (P_{int}), was equal to the internal energy divided by exercise duration *t*:

$$P_{\text{int}} = E_{\text{int}}/t = [O_2 def + tfVO_{2\max} - \tau(VO_{2\max} f - VO_{2\text{init}})]/t\,.$$

Olds et al (1993) calculated the time taken for an athlete to complete an event to be:

$$t = (d + d_{\text{acc}})60v_{\text{ss}}\,,$$

where v_{ss} was velocity (in m/s) during steady-state exercise (when the rider is consuming maximal oxygen (VO_{2max})), *d* was the total distance of the event (in metres) and d_{acc} was the distance the rider rides before reaching v_{ss} (in metres). Therefore P_{int} was expressed as the following:

$$P_{\text{int}} = [O_2 def + tfVO_{2\max} - \tau(VO_{2\max} f - VO_{2\text{init}})]/[(d + d_{acc})60v_{ss})]$$

However, the external work that must be done by the rider to complete an event is greater than the available internal energy, or the sum of the aerobic and anaerobic energy (Olds et al, 1993). Olds et al (1993) calculated the power output required to ride in an event, based on VO_2. Subjects completed a series of trials on a cycle ergometer, and Olds et al (1993) plotted VO_2 (litres/minute) against the power output required to ride an event (WR, in Watts). Power output (WR) was then inferred from the plots: WR = (VO2 – intercept)/gradient, where gradient was the slope of the plotted VO_2/WR regression line, and intercept was where the regression line intercepts the y axis. The maximal power output the rider can then produce was the following: WR = (P_{int} – intercept)/gradient, or, with *i* to represent intercept,

$$WR = \{[O_2 def + tfVO_{2\max} - \tau(VO_{2\max} f - VO_{2\text{init}})]/[(d + d_{acc})60v_{ss})] - i\}/gradient\,.$$

The energy available to complete this work, in *t* minutes, was calculated by Olds et al (1993) to be the power output (WR) multiplied by $(d+d_{acc})/v_{ss}$.

Olds et al (1993) compared predicted time to complete 4000m time trials with actual time, for 18 riders. Predicted and actual times were positively correlated (r=0.803), and there was a 4.6 seconds mean difference between actual times and predicted times. The errors, though small, could be due to fatigue, which Olds et al (1993) claimed is difficult to model.

A range of mathematical models of mechanical power output have been developed by various authors including Bassett et al (1999), Di Prampero et al (1979) and Martin et al (1998). Indeed Bassett et al (1999) and Martin et al (1998) have been able to verify their models to be reliable when measured against SRM cranks. Authors have indicated how to calculate power output for different gradients, sizes and shapes of an athlete's body, and for different levels of air resistance and rolling resistance. Authors of mathematical models of power output have tended to focus on physical relationships of power (to air resistance, for example), and to some extent physiological measures such as oxygen consumption.

However to our knowledge mathematical models of mechanical power output in cycling have tended not to feature a direct measure of cadence, though authors have acknowledged the influence of cadence on performance. For example, Di Prampero (1979) et al do acknowledge the importance of cadence, along with power output, on mechanical efficiency. They then use mechanical efficiency to calculate oxygen consumption. In the study by Padilla et al (2000) discussed earlier in this section, the subject chose his own preferred cadence. MacIntosh et al (2000) do consider the possibility of calculating power output mathematically from cadence – they investigated the relationship between cadence and power output, calculating power output by using principles of muscle force / velocity relationships, as power output is a multiplication of force (resistance in the wheel) and velocity (velocity of the crank, controlled by cadence). They considered an optimum cadence to be the cadence that achieves a target power output using the least amount of muscle activation (we discuss this paper in more detail in a separate section about muscle force / velocity relationships). However in this paper subjects ride on a cycle ergometer at only five cadences (50, 60, 80, 100 and 120 rpm). Whilst MacIntosh et al (2000) expected that, theoretically, power output is a parabolic function of cadence, they inferred the power output cadence relationship from their own data by plotting power output against crank velocity (in m/s), converting velocity to cadence (rpm) by multiplying by 10. We aim to explore this relationship by developing statistical models of power output and cadence, using empirical data from many months of training sessions. We will then calculate optimum cadence (that is, the cadence that maximises power output for a given heart rate) mathematically using our models. We also aim to include a measure of cumulative fatigue (by including training impulses (explained in more detail in 2.15)), and to calculate the effect this has on power output and optimum cadence.

2.3. Optimising cadence

2.3.1. Introduction to optimising cadence

In addition to determining an optimum pacing strategy, the cyclist may then seek to further improve his or her performance by optimising cadence. Cadence is defined as the number of pedal-crank revolutions per minute, and is therefore proportional to the speed of rotation of the feet when pedalling. If cadence is too high, athletes are no longer able to sustain the correct technique as extra stabilization of the body is required, and hence this is detrimental to performance as athletes may have to reduce their power output. (Hagberg et al, 1981).

An optimum cadence may only bring a small level of improvement to performance compared to an optimum pace, but at a professional level a small margin of improvement may be significant (Abiss et al, 2009; Foss and Hallen, 2004). Indeed Coast and Welch (1985) claimed that the choice of cadence may be an important determinant of performance for cyclists that pedal with a high power output.

In section 2.3., we explore how optimum cadences are defined and calculated in literature. We also describe how we define optimum cadence in our study, and how this relates to literature. We then describe the typical methodology used in literature to calculate optimum cadences in laboratory settings and how these conditions compare to actual road cycling. Finally we describe the range of optimum cadences typically found in literature.

2.3.2. Optimum cadence and gross efficiency

Optimum cadence is broadly defined to be the cadence with the highest gross efficiency. Gross efficiency is the amount of mechanical work achieved relative to the metabolic energy required to do this work (Whipp and Wasserman, 1969) – whilst mechanical work is broadly considered to be the distance through which the point of application of the force (of the athlete's muscles) moves (Winter and Fowler, 2009). Gross efficiency decreases when cadence is too high, as the level of energy exerted by the athlete increases but the athlete is not able to produce higher power outputs. Gross efficiency decreases when cadence is too low as a greater energy expenditure is required to recruit the extra muscle fibres needed (Seabury et al, 1977).

However, the literature provides contradictory results regarding this relationship between gross efficiency and cadence (Samozino et al., 2006). In the study by Samazino et al (2006), gross efficiency was affected by an interaction of both power output and cadence, such that the effects of power output on gross efficiency changed with cadence and the effects of cadence on gross efficiency changed with power output. Chavarren and Calbet (1999) and Samozino et al (2006) both observed that, when an athlete was not riding at an optimum cadence, the decrease in gross efficiency at low power outputs was greater than the decrease in gross efficiency at higher power outputs – as power output increases, the influence of cadence on gross efficiency decreased linearly.

2.3.3. Optimum cadence based on heart rate

Principally we aim to calculate a cadence that minimises some measure of rate of energy expenditure or effort exerted for a given power output. Only power output, heart rate and cadence measurements collected from the field are available to us – heart rate therefore is the closest measure we have to a measure of rate of energy expenditure. Nevertheless different measures of energy expenditure exist in literature; we now outline what measures are typically used and how they relate to heart rate.

Gross eficiency is commonly used to indicate effort required, but it can be difficult to measure gross effficiency (Winter and Fowler, 2009). Typical definitions of gross efficiency involve measuring mechanical work done by the athlete, and dividing this by a measure of energy input by the athlete (Whipp and Wasserman, 1969; Winter and Fowler, 2009). Winter and Fowler (2009) claimed that it is potentially difficult to measure the mechancial work done by muscles. Indeed for cycling, Winter and Fowler (2009) argued that academics should differentiate between work done to move limbs and work done to move the bicycle itself – internal and external work done respectively. External mechancial work done can be calculated, but such calculations do not take account of friction (from bearings in the wheel or between the chain and sprockets), whilst internal mechancial work done is particulary difficult to assess (Winter and Fowler, 2009). Assessment of internal mechancial work done involves measuring changes in potential and kinetic energy in the segments of the limbs of the athlete – though crucial for measuring the mechancial work done by the athlete, such calculations are not always included in studies (Winter and Fowler, 2009). In order to use gross efficiency, we would need to calculate it at regular 5 second intervals, to be consistent with power output and cadence data. Wingo et al (2005) claimed that, as heart rate progressively increases over time, there is equivalently an increase in relative metabolic intensity, as maximal oxygen uptake ($V0_2$max) decreases. Since gross efficiency is related to the amount of metabolic energy required to do work, heart rate to some extent would therefore appear to correlate with gross efficiency.

Another method commonly used in literature to represent some form of effort exerted is rate of perceived exertion (Borreson and Lambert, 2009). This is a subjective rating of the difficulty of a session, obtained 30 minutes after the end of the session (Borreson and Lambert, Foster et al, 1996). However, rate of perceived exertion is qualitative and relies on athletes being accurate in their assessments (Borreson and Lambert, 2009). Little and Williams (2007) suggested that rating of perceived exertion does not always correlate well with heart rate for short duration, high intensity exercise (albeit in soccer). However Foster et al (2001) found positive correlations between variations in rating of perceived exertion and variations in time spent in different heart rate zones (which are zones based on what percentage heart rate is of maximum heart rate) for cycle ergometry. Foster (1998) found correlations of between 0.75 and 0.9 between variation in rating of perceived exertion and variation in time spend in heart rate zones. We do not have data avialable regarding rate of perceived exertion. It would not be also not possible for athletes to calculate a rate of perceived exertion at regular 5 second intervals (to be consistent with power output and cadence meaurements). Nevertheless heart rate appears to corerelate to some extent with rating of perceived exertion in cycling (Foster et al, 2001).

The maximum rate of oxygen uptake ($V0_{2max}$, or the maximum amount of oxygen that the body can utilise in a period of one minute) can also be used in literature to measure exertion (Pichot et al, 2000; Swain and Leutholtz, 1997). However $V0_{2max}$ may be limited as a measure of rate of energy expenditure, as $V0_{2max}$ is similar between elite and well-trained amateur athletes, but gross efficiency is typically higher for elite athletes compared to amateurs (Lucia et al, 2002). Also, an elite athlete can compensate for having a relatively low $V0_{2max}$ by having a very high gross efficiency (Lucia et al, 2002). Heart rate may be related to oxygen uptake; the percentage of heart rate reserve (the difference between current and resting heart rate, divided by the difference between maximum and resting heart rate) is commonly thoguht to be linked to $V0_{2max}$ (Swain and Leutholtz, 1997). However Swain and Leitholtz (1997) suggested that the percentage of heart rate reserve is linked to the percentage of $V0_2$ reserve (which is the difference between $V0_{2max}$ and resting $V0_2$, or rate of oxygen consumption at rest). In their study, Swain and Leutholtz (1997) measured heart rate an $V0_2$ at rest, at maximum and at the end of a number of cycle ergometer tests for 63 participants. They found that percentage of heart rate reserve was more closely linked to percentage of $V0_2$ reserve than it was to $V0_{2max}$. Furthermore, as fitness ($V0_{2max}$) increased, so to did the discrepancy between percentage of heart rate reserve and $V0_{2max}$ (Swain and Leutholtz, 1997). Oxygen consumption could be measured in laboratory setting, but it would be very difficult to measure rate of oxygen consumption at constant intervals during road cycling, either in training or races.

Since only power output, heart rate and cadence measurements from the field are available to us, heart rate is the most appropriate measure avialable to us regarding a measure of effort exerted. In our study we therefore use heart rate as a proxy for gross efficiency, to represent some form of effort exerted. Changes in heart rate can reflect changes in relative metabolic intensity (Wingo et al, 2005), thus we consider heart rate a reasonable proxy to gross efficiency. Heart rate also appears to correlate reasonably with rate of perceived exertion (Fost et al, 2001) and it appears to be linked to oxygen consumption, as percentage of heart rate reserve is related to percentage of $V0_2$ reseve (Swain and Leutholtz, 1997), thus further emphasising the importance of minimising heart rate. It is therefore important for cyclists in our study to minimise heart rate, in order to conserve energy and maximise the effort they can put in at the end of a race or training session. Therefore optimum cadence in our study is defined as the cadence that minimises heart rate for a given power output (for a given pace). We therefore propose that an individual cyclist should select a gear that allows them to ride at their individual optimum cadence, to minimise heart rate for whatever pacing strategy they have chosen.

2.3.4. Limitations of the use of heart rate

During prolonged exercise, there tends to be a gradual increase in heart rate over time, and a reduction in stroke volume – this is a phenomenon known as 'cardiovascular drift' (Ericson et al; Hamilton et al, 1991; Wingo et al, 2005). There is also a reduction in arterial, pulmonary arterial and right-ventricular blood pressure (Hamilton et al, 1991; Wingo et al, 2005). Cardiovascular drift tends to occur after the first few minutes of exercise at a constant work rate; heart rate rises progressively after this early stage of exercise (Hamilton et al, 1991).

Cardiovascular drift can be caused by dehydration (Wingo et al, 2005). Indeed authors have studied the effects of fluid ingestion during intensity of exercise on cardiovascular drift (Hamilton et al, 1991; Wingo et al, 2005).

Wingo et al (2005) explored the effects of fluid ingestion on cardiovascular drift in nine male subjects. Subjects rode on indoor cycling ergometers for four separate trials. Subjects rode at 60% of their maximal oxygen uptake ($V0_{2max}$, which was determined in a test ride before the experimental trials), for 15 or 45 minutes. For two trials (one at 15 minutes and the other at 45 minutes), subjects received fluid in the form of tap water, ingested before the trials started, and at regular intervals during the trials (after 10 minutes for the 15 minutes trial and after 10, 25 and 35 minutes for the 45 minute trial). Equivalently subjects also rode two trials (one at 15 minutes and the other at 45 minutes) with no fluid ingestion. To assess the possible effects of cardiovascular drift in each 45 minute trial, cardiovascular measurements were taken between 8 and 15 minutes and between 38 and 45 minutes. Body mass of each subject was also measured before and after each trial, to observe the effects of dehydration (as dehydration reduces body mass after exercise; Wingo et al, 2005). For the trials in which fluid was not replaced, there was indeed a greater loss in body mass compared to the fluid replacement trials – 0.3% for the 15 and 45 minute trials with fluid ingestion, and 1.4% and 2.5% loss in the no fluid trials of 15 and 45 minutes respectively. However, this fluid ingestion did not prevent cardiovascular drift; there was no significant difference between the fluid ingestion and no fluid trials in change in heart rate or change in stroke volume (from the beginning to end of the 45 minute trials).

Hamilton et al (1991) however found that some types of fluid replacement can help to prevent cardiovascular drift better than others. They studied the effects of fluid replacement on stroke volume and heart rate during prolonged exercise (cycling on a stationary cycling ergometer) of 2 hours. Their study was split into two separate parts. In the first experiment, ten subjects performed two trials. One trial involved fluid replacement – in which subjects drank water at twenty minute intervals during exercise – and in the other trial there was no fluid replacement. In a second experiment, eight subjects performed two trials again. This time the trials were fluid replacement (the same procedure as in the first experiment) and a glucose infusion trial, in which subjects were given a glucose solution intravenously. For the first experiment, during the first hour of exercise, there was no significant difference between heart rate in the fluid replacement trial and the no fluid trial. However heart rate was significantly higher in the second hour of exercise in the no fluid trial compared to the fluid replacement trial. Heart rate increased in the no fluid trial by twice the margin it increased by in the fluid replacement trial. Nevertheless heart rate still increased over the two hours in the fluid replacement trial, from 140 to 145 beats per minute. Stroke volume equivalently was significantly lower in the no fluid trial compared to the fluid replacement trial. Indeed fluid replacement had a greater impact in reducing the effects of stroke volume reduction than it did in reducing the effect of heart rate increase over time. Results were similar in the second experiment. Whilst heart rate increased by 5-6% during the trial with water ingestion, heart rate increase was significantly less in the glucose infusion trial in comparison. Indeed in the glucose infusion trial, after twenty minutes of exercise, heart rate did not increase by more than 2 beats per minute (Hamilton et al, 1991).

The results from the Hamilton et al (1991) study suggested that fluid (water) replacement helps to prevent reduction in stroke volume over prolonged exercise, but it is not enough to prevent cardiovascular drift (Hamilton et al, 1991). That fluid (water) replacement may help to prevent the reduction in stroke volume may be due to body temperature. Hamilton et al (1991) suggested that the contribution of fluid replacement to prevention of reduction in stroke volume may be because the fluid replacement helps to maintain core and skin temperature. Indeed reduction in skin and core body temperature has been found to help prevent reduction in stroke volume (Rowell, 1971).

The magnitude of cardiovascular drift may be affected by ambient temperature (Lafrenz et al, 2008). Lafrenz et al (2008) tested ten athletes (comprising cyclists and runners) using cycle ergometers, at 22 and 35 degrees Celsius ambient temperature. Heart rate was measured after 115 and 45 minutes of continuous moderate intensity exercise. Heart rate increased from 15 to 45 minutes by approximately 10 beats per minute in the hot conditions and 3 beats per minute in the cool conditions. We do not consider the effect of temperature in our study, as no such data are available to us. However, since our data comprise training sessions for road races in Britain, including winter months, temperatures are unlikely to be as high as those in the Lafrenz et al (2008) study. Cardiovascular drift for our athletes may therefore not be as great as in the Lafrenz et al (2008) study.

We do not consider whether an athlete is standing or sitting, or the angle at which they sit, as no such information is available. Indeed seat position may influence heart rate (Price and Donne, 1997). In a cycle ergometer experiment, Price and Donne (1997) measured the impact of three different seat tube angles (68, 74 and 80) at three different seat heights (96%, 100% and 104% trochanteric height) on a number of factors, including heart rate. Heart rate tended to be significantly lower for lower seat angles compared to higher seat angles, for the same power output and cadence. A limitation of our study is therefore that we cannot detect any influence of sitting position or seat height on heart rate. We also have limited information in athletes' training schedules regarding whether a data are from a race or training session, or types of races undertaken. We would expect heart rate to be higher in races compared to, say, low intensity training sessions (though heart rate would still be high in a high intensity training session). However we take account of this by comparing models fitted to all data available (albeit sampled at regular intervals), with the same models fitted only to a subset of data where heart rate is particularly high. We describe this in more detail in chapter 4.

In this study, we use empirical data from training sessions, in which many sessions are over 1 hour in duration. It is therefore conceivable that cardiovascular drift may occur in the athletes in our data. Sessions vary in intensity, hence in less intense sessions cardiovascular drift is less likely to occur, since research suggests cardiovascular drift occurs in exercise that is not only prolonged but also moderate to high in intensity (Wingo et al, 2005). Nevertheless heart rate is the most appropriate measure of rate of energy expenditure available to us, and we consider it a reasonable proxy for gross efficiency.

2.3.5. Optimum cadences from ergometer studies

Existing research into optimum cadence tends to involve the use of cycle ergometers in laboratory based experiments, in which lower limb muscles are worked in a simulation of a bicycle, whilst upper body movements are restricted. In fact cycle ergometers are commonly used in general investigations into sports science, as cyclists tend to sustain injuries less frequently than do runners (Atkinson et al, 2003). The accuracy of the ergometer in its simulation of a bicycle is ensured by being adjustable to match the athlete's own bicycle, along with the inclusion of pedals and handle bars similar to that found in a real bicycle (Cost and welch, 1985). The amount of mechanical work achieved and the metabolic energy that has been used to achieve this are both required to calculate gross efficiency. The mechanical work achieved through lower limb muscles is calculated through linear displacement of the flywheel on the ergocycle, whereas the energy expenditure is based on the measurement of oxygen uptake (Samozino et al, 2006).

Participants are tested either to exhaustion, with a specific definition of exhaustion in each study, or for a fixed time period. Foss and Hallen (2004), for example, defined exhaustion to be when an athlete rode for 10 seconds consecutively at a cadence that was more than 3 rpm below the predetermined cadence. If tested to exhaustion (known as a maximal test), performance measures tend to include time taken to reach exhaustion, the power output or heart rate at the time of exhaustion and the oxygen uptake at the time of exhaustion. These maximal tests are often given at progressively increasing power outputs for a given cadence, therefore giving researchers a range of power outputs and cadences that can be used to explore the power output / cadence relationship. If tested for a fixed time period, performance measures include mechanical energy achieved, energy expenditure and gross efficiency. The performance measure is plotted for different cadences, and the optimum cadence is inferred from the fitted curve. Participants are tested at a range of cadence intervals and power outputs. For example, Samazino et al (2006) tested athletes at cadances of 40, 60, 80 and 100, and Foss and Hallen (2004) tested athletes at a similar though slightly higher range of cadences – 60, 80, 100 and 120. Coast and Welch (1985) tested a slightly wider range of cadences (40, 60, 80, 100 and 120), at power outputs of 100W, 150W, 200W, 250W and 300W.

2.3.6. Relationship between laboratory settings and road cycling

Some authors have reported differences in performance between road cycling and laboratory cycling exercises (Jobson et al, 2007). Participants may not necessarily ride at the same speeds in road cycling and ergometer cycling for the same power output achieved (Jobson et al, 2007; Smith et al, 2001). Smith et al (2001) compared performances of athletes in 40km time trials on a cycle ergometer and on the road. They tested eight trained cyclists, setting them three laboratory time trials and three time trials on a road course. Although there were no significant differences in mean power output between the road and laboratory conditions (mean power output was 312W on the road and 303W in the laboratory), athletes were 2.4km/h slower on the road compared to the laboratory. Although Smith et al (2001) claimed that their study suggests that power output measurements are reproducable for both laboratory and road conditions, Jobson et al (2007) claimed the results of the Smith et al study suggest that performance in road cycling is not necessarily replicated accurately in laboratory conditions.

Jobson et al (2007) investigated differences between performance in laboratory and road cycling and how any such differences may be affected by body mass of the participants, claiming that body mass is known to affect performance in road cycling. They recruited twenty three male trained cyclists, setting them 25 mile time trials in a laboratory and on a road course. They also tested participants to exhaustion on a cycle ergometer, as part of a measurement of the physical characteristics of participants. Similarly to Smith et al (2001), Jobson et al (2007) found that participants were slower on the road time trial than in the laboratory, although in their study by a smaller margin, of 1.7km/h. Furthermore Jobson et al (2007) developed a regression equation to represent the difference between road and laboratory speeds – they found road speed to be equal to 10.1 + (0.708*laboratory speed). Although greater body mass of the participants significantly reduced speed in the road time trial, it had no effect on speed in the laboratory time trial. Jobson et al (2007) were also able to develop a regression equation to predict road cycling speed (km/h) from laboratory cycling speed and body mass (kg) – road speed was equal to 13.8 + (0.821*laboratory speed) – (0.106*body mass). Indeed the inclusion of body mass as a variable increased explanatory power of their regression analysis from 69.3% to 78.3%.

Laboratory based ergometer studies tend to control gradient and air resistance to be constant, whereas they are likely to vary in race conditions (Atkinson et al, 2003; Hickey at al, 1992). Air resistance may have a significant impact on performance in actual race conditions, as 90% of total power output produced by the athlete is used to overcome the air resistiance (Grazzi et al, 1999; Olds et al, 1993). Air resistance can also have an impact on performance for athletes in road cycling in terms of pacing strategy – we have described this in more detail in the pacing strategy section earlier in the chapter.

In ergometer studies, the range of cadences tested is often decided by an academic rather than a training coach, and for the convenience of that academic (Coast and Welch 1985) – unless the test is carefully scheduled within an athlete's training schedule, researchers must assume that the athlete is not withholding effort or insufficiently motivated for a test that, unlike a real training session, is not necessarily devised for his or her benefit. Although optimum cadence could plausibly be within the ranges often tested, the athletes are tested at far fewer different cadences that they would use in a series of actual training sessions. In a study by Samazino et al (2006), athletes were tested for periods of 15 minutes with rests between sessions. The duration of such tests are far less than those found in the training schedules of the athletes in this research, perhaps limiting the scope for ergometer studies to test an athlete's maximum capabilities in race conditions. Indeed Lepers et al (2001) claimed that the effect of exercise duration on cadence has not been studied to much extent.

Differences in performance between road cycling and laboratory cycling are not fully understood (Jobson et al, 2007). For cycle ergometer studies, further research may be needed to investigate the relationship between performance in road races and results obtained from cycling ergometer laboratory experiments (Faria et al, 2005; Jobson et al, 2007). For our studies, the use of empirical training data eliminates any issues related to laboratory cycling and the possibility of performance in such studies not being representative of actual performances in road cycling.

2.3.7. Optimum cadence from field data

Only a small range of cadences are used (typically 4 or 5 different cadences at intervals of 20 rpm) in ergometer studies, compared to the number of different cadences at which athletes in our data have ridden. The inferred optimum cadence may occur somewhere between two tested cadences, at a cadence at which the athlete has not actually ridden in the study. In our study, the use of field data from training sessions provides us with a great range of cadences, potentially allowing us to be more precise in calculating an optimum cadence, as it is less likely that a theoretical optimum cadence is a long way above or below a cadence at which a participating athlete has ridden.

To our knowledge, there are very few studies investigating an optimum cadence using field data from training sessions. The only research we know of to use data from real cycling training sessions is a study by Sassi et al (2009). In this research Sassi et al investigated the relationship between freely chosen cadence and the gradient of the road, using data collected from training sessions of 10 athletes. However Sassi et al did not calculate a cadence that is optimal for a given set of circumstances, rather they analysed how the cadence chosen by athletes varies by gradient. The cadences chosen by the athletes may or may not have been the respective cadences that maximised their performances. We seek to develop a statistical model of power output, heart rate and cadence, which gives us a mathematically optimum cadence. To our knowledge no studies have attempted to calculate a mathematically optimum cadence in this way using field data.

2.3.8. Optimum cadences in literature

A wide range of optimum cadences have been reported throughout the literature (Coast and Welch, 1985), ranging from 30–60 rpm (Eckermann and Millahn, 1967) up to 80–90 (Hagberg et al, 1981), with other studies finding optimum cadence between these ranges, such as Michielli and Stricevic (1977), who found an optimum cadence range of between 50 and

60. Abbiss et al (2009) suggested that an optimum cadence for sprint cycling is 100-120 rpm, but an optimum cadence for long time trials (over 40km) is 90-100 rpm.

2.4. Relationships between power output, heart rate, cadence, speed and velocity

2.4.1. Introduction

Since we do not have speed or velocity data available, we consider how closely power output is related to speed. Nevertheless it is very important to be able to ride at a high power output in road cycling (de Koning et al, 1999). We also review literature that investigates realtionships between variables such as power output, heart rate and cadence.

2.4.2. Relationship between power output and speed/velocity

Increasing power output tends to increase the velocity at which the athlete rides (de Koning et al, 1999). However a number of factors can potentially affect the relationship between power output and velocity (Atkinson et al, 2003; Jeukendrupp and van Diemen, 1998). Environmental conditions, such as the angle of gradient or strength of winds, influence the relationship between power output produced and the overall velocity of a race (Atkinson et al, 2003, Jeukendrupp and van Diemen, 1998). The extent to which winds affect the power output velocity relationship are affected by the position of the rider and the pacing strategy adopted (Atkinson et al, 2003).

If a cyclist is riding at a high power output and begins to ride uphill, power output (and also heart rate) remain high, but velocity decreases, whereas for a downhill gradient, power output decreases but the athlete is able to maintain a high velocity (Jeukendrupp and van Diemenm 1998). Air resistance and drag also affect the power output velocity relationship (Atkinson et al, 2003; Bassett et al, 1999). Atkinson et al (2003) claimed that power output is typically a cubic function of speed. However, this may be influenced by air resistance (Atkinson et al, 2003; Basset et al, 1999). In an indoor cycling experiement (in which there was no air resistance), Basset et al (1999) found the exponent of power output and speed to be 2.6.

Similarly the amount of air drag also affects the power output velocity relationship if one cyclist rides directly behind another - known as slipstreaming (Atkinson et al, 2003; Broker et al, 1999; Jeukendrupp and van Diemen, 1998; McCole et al, 1990). Even it the two athletes ride at the same velocity, the following cyclist will not have to produce as high a power output as the leading cyclist, due to drag. The decrease in power output for the following cyclist may be as high as 30% (Atkinson et al, 2003; Broker et al, 1999). In a study by Broker et al (1999), seven male U.SA. national team members were recruited and split into two pursuit teams, for 2000m pursuit tests. Participants tended to benefit from slipstreaming, and indeed Broker et al (1999) also found that a third athlete, following the second, may also benefit from the slipstreaming with a 6% decrease in power output compared to the second athlete. The exact reduction in power output however depended on individual athletes and circumstances. Broker et al (1999) suggested that a shorter rider following a taller rider may experience greater benefit than the reverse circumstances.

Nevertheless in this study we seek to optimise power output in our models of power output, heart rate and cadence, by finding the cadence that optimises power output for a given heart rate. We do not have speed or velocity measurements – however, Jeukendrupp and van Diemen (1998) claim speed to be a poor indicator of exercise intensity. However by selecting a cadence that maximises power output, athletes can ride more quickly (de Koning et al, 1999).

2.4.3. Relationship between power output and heart rate

Conconi et al (1982) investigated the relationship between speed and heart rate in running, finding that the speed / heart rate relationship is linear at low speeds and curvilinear at higher speeds. This relationship has been replicated in a number of equivalent studies in cycling, in which the power output / heart rate relationship has been investigated. At low intensity exercise, the power output / heart rate relationship is said to be linear as increases in heart rate are proportional to increases in power output. As intensity of exercise increases, eventually increases in heart rate are not proportional to increases in power output – hence the curvilinear power output heart rate relationship (Grazzi et al 1999). The heart rate at which this relationship ceases to be linear and becomes curvilinear is known as the deflection point.

Researchers have investigated the existence of the deflection point in cyclists and whether or not it could be determined mathematically. Athletes ride at a range of different power outputs, with the heart rate recorded. The heart rate power output relationship is plotted from data collected.

Jeukendrupp and van Diemen (1998) claimed the heart rate deflection point may be an artefact of academic studies, and that the power output / heart rate relationship may therefore be linear. They argued that, for short durations of exercise (less than 1 minute), heart rate lags behind as the cirulatory system is unable to adjust or adapt to that intensity of exercise quickly enough, and that a deflection point is always found in such circumstances. When exercise comprises periods of intensity that last for longer durations, however, Jeukendrupp and van Diemen (1998) claimed that it is much more difficult and potentially not even possible to detect a deflection point, as the circulatory system in the body is able to adapt somewhat to changes in exercise intensity. Coast and Welch (1985) also found heart rate to increase with power output linearly.

However Grazzi et al (1999) were able to detect the deflection point. They tested 15 athletes at incremental cadences using a wind-load simulator. Initially cadence was increased every 30s (beginning at 60 rpm and increasing by 1 rpm each time), before being increased at shorter time intervals when the athlete appeared to be exhibiting signs of fatigue. After heart rate and power output data were plotted, a linear relationship was inferred from data for which the line of best fit had a correlation coefficient of 0.98 or greater, whilst a cuvilinear relationship was inferred otherwise. In 484 of 500 tests analysed, deflection point occurred; the deflection point was inferred to exist if the point at which the power output heart rate relationship became curvilinear occurred before the point in the test when time intervals between increases in cadence increased. If the deflection point had occurred after the point at which time intervals between increases in cadence increased, it could have been claimed that the deflection point only existed as a result of, or was caused by, the shorter time intervals between cadence increases (and the consequent increases in effort the athlete would have exerted).

Studies in which cadence is kept constant have not always detected a deflection point in the power output heart rate relationship, whilst the deflection point is detected much more easily in studies with increasing cadences (Grazzi et al, 1999). In studies in which cadence is increased progressively, Grazzi et al (1999) claimed the activation of anaerobic lactacid mechanisms to be the physiological cause of the deflection. Grazzi et al (1999) also claimed that the muscular power output required to cycle at maximal effort is twice as high in studies with a fixed cadence compared to those with increasing cadence, and that this could be why deflection point is not always detected in such studies. This increased muscular power required for fixed cadence tests could in turn lead to an earlier activation of anaerobic lactacid mechanisms (Grazzi et al, 1999), hence the accumulation of blood lactate enables the athlete to continue to increase his heart rate in line with his power output porportionally in these fixed cadence tests.

The Grazzi et al (1999) study involved subjects riding a stationary ergometer indoors. Outside of the laboratory, athletes encounter changing wind conditions and biomechancial resistance; we may expect such elements to affect the power output / heart rate relationship. In our study we include measurements of power output, heart rate and cadence from numerous training sessions - we analyse (amongst other things) the relationship between power output and heart rate empirically using this data. In fitting a regression model of power against heart rate and cadence, we observe the value of the coefficient of heart rate to explore the power output / heart rate relationship in training sessions outside laboratory conditions.

2.4.4. Heart Rate lag

Heart rate acts as a response to power output; when an athlete increases the power output at which he/she rides, his/her heart rate must subsequently increase in order to support and sustain that power output (Churchill et al, 2009; Grazzi et al, 1999). There is a time lag between the change in power output and the heart rate response, although it is not clear from existing literature precisely how long this heart rate lag is (Churchill et al, 2009). Moreover Jeukendrup and van Diemen (1998) claimed that there is a heart rate lag in response to increased intensity of exercise for periods of exercise of short duration, as the circulatory system is not able to fully adapt to change in exercise intensity, although they do not indicate how much of a time lag there is. However, studies by Stirling et al (2008) and Cheng et al (2007) appear useful in understanding the lag between power output and heart rate response. They studied the time taken for heart rate to change in response to beginning exercise (albeit running and fast walking, rather than the mechanical power output of bicycles in cycling).

In the study conducted by Stirling et al (2008), a 33 year old male subject ran a number of laps around around a track running circuit, 400m in length. His exercise consisted of five sessions (each session comprising four laps around the track), with 10 minute rest periods between the sessions. A Polar S810i heart rate monitor was used to record heart rate measurements (Stirling et al, 2008). Stirling et al plotted a series of graphs of heart rate against time. The maximum heart rate was approximately 160 for a typical session. The athlete's heart rate increased from 80 to 160 in approximately 60 seconds, and after the session finished, the heart rate took around 60 seconds again to decrease to 80. The change in heart rate was consistently around 60 seconds for each exercise session – there was therefore no longer term effect of fatigue on the size of the heart rate lag (the length of time taken for heart rate to respond to power). Both the increase and decrease in heart rate were almost linear, albeit the change in heart rate became slightly curvilinear at the end of the increase or decrease. Nevertheless for small heart rate changes, say anything less than 60 beats per minute, the change in heart rate appears linear.

Cheng et al (2007) investigated the physiological heart rate respopnse to exercise by setting 5 healthy male subjects a series of treadmill exercises, which involved walking at speeds of 5km/h, 6km/h and 7km/h. Each session lasted 15 minutes, with 20 minute rest periods between sessions. Heart rate during three minutes of inactivity before each session, along with heart rate measurements during exercise sessions and the recovery periods after the sessions, were plotted against time. In fact smoothed heart rate measurements were used for the plots, using a moving average over a 5 second window. Initially heart rate increases from resting heart rate, and levels off after approximately 60 seconds. Similarly after an exercise session has finished, heart rate decreases to just above the resting heart rate, and takes approximately 60 seconds to level off. Inevitably there are slightly different plots for each athlete, but for four of the five athletes there appears to be a very clear, sudden increase or decrease in heart rate (that appears to be approximately linear) over 60 seconds at the beginning and end of each exercise session.

Although the exercise sessions in the Cheng et al study involved fast walking rather than more intense running, patterns in heart rate time plots were arguably similar to those found in the Stirling et al study. Changes in heart rate (of, say, 80bpm) in the Stirling et al study occurred over 60 seconds. In the Cheng et al study, changes in heart rate of less, around 40-50 beats per minute, occurred over the same period, but this is for much less intense exercise. Therefore we may expect that, if athletes in the Cheng study were to exercise at a higher intensity, their heart rate may indeed be able to increase by more than 40-50 bpm over a 60 second period.

For the data recorded from the athletes in our research, heart rate tends to change quite gradually. Heart rate tends not to increase by the amount seen in the Stirling et al study. This is not surprising as the exercise periods in the Stirling et al and Cheng et al studies are short in duration, just 1600m of running and 15 minutes of brisk walking respectively in each session, whereas training sessions from the athletes in our research tend to be quite long, each lasting between approximately 1 hour and 5 hours. Therefore an athlete in our study is more likely to pace himelf accordingly, with some periods of high intensity and other peiods of lower intensity. Also, since the athletes in our study cycle over mixed terrain outdoors, they are likely to encounter numerous changes of conditions and intensity, causing heart rate to increase or decrease a little even during periods of relatively consistent amounts of physical exertion. We therefore investigate the length of time lag between changes in power output and heart rate response.

2.4.5. Relationship between cadence and heart rate

Although some researchers have investigated the impact of power output and cadence on gross efficiency (through submaximal tests), some have also investigated the impact of power output and cadence on heart rate (through both maximal and submaximal tests), and in doing so explored the relatipnship between cadence and heart rate. Indeed Churchill et al (2009) and Grazzi et al (1999) suggested that heart rate can be considered a response to power output.

Coast and Welch (1985) tested a group of five male athletes to exhaustion at a range of five different cadences (40, 60, 80, 100 and 120), at power outputs of 100W to 300W in intervals of 50W. They inferred a pedalling rate to be optimal if it minimised heart rate, for a given power output, which is equivalent to our definition of an optimum cadence. Heart rate increased with power output (which is expected as an athlete must work harder if he/she increases his/her power output). The heart rate increased linearly with power output, from approximately 105 beats per minute up to aproximately 175, in intervals of approximately 20 beats per minute between each power output increase of 50W. Heart rate also varied as a function of cadence for a given power output – the minimum heart rate (and therefore the optimum cadence) occurred at increasing power outputs as cadence increased, with optimum cadences ranging from approximately 57 to approximately 67.

However Lepers et al (2001) suggested that heart rate is not significantly affected by cadence. They tested eight tri-athletes using submaximal tests, with each test 30 minutes in duration, recording heart rate continuously. Although performance measures (including heart rate) indicated a drop in performance towards the end of a test, there was no evidence of any impact of different cadences on heart rate; performance levels dropped equally for each cadence. They suggested that tri-athletes can use the selection of gear ratios to adapt to changes in cadences. However they used a smaller range of cadences compared to some other studies already outlined, as they only tested the athletes at a cadence chosen by the athlete, the cadence 20% lower than this and the cadence 20% higher. Also, athletes were tested at 80% of their inferred maximum power output (the maximum being calculated as the highest power output sustained for a period of 2 minutes in a 30 minute submaximal test), whereas Coast and Welch (1985), who tested athletes to exhaustion, did find a significant effect of cadence on heart rate – therefore it could be that cadence only influences heart rate at very high power outputs. Indeed Lepers et al (2001) used tri-athletes as participants rather than professional cyclists – the presumably slightly inferior cycling skills of tri-athletes (as opposed to athletes whose sole specialty is cycling) may have further reduced the power outputs being reached compared to the Coast and Welch study.

The relationship between cadence and heart rate is not necessarily clear from literature. We seek to explore this relationship in our study, specifically whether or not there exists a range of optimum cadences depending on the heart rate of the athlete.

2.5. Additional factors that could influence cadence

2.5.1. Muscle force / velocity relationships

Broadly, the greater the force needed to overcome, the slower the shortening of the muscle (Hill, 1922; Wilkie, 1950). This is known as the muscle force velocity relationship (Hill, 1922, Wilkie, 1950). In cycling, the muscle force / velocity relationship is equivalent to the relationship between the force needed to overcome the resistance of the flywheel, and the cadence (MacIntosh et al, 2000). Wilkie (1950) calculated the velocity of shortening of elbow flexors, for a range of different afterloads. McCartney et al (1983) however argue that, in cycling, as large muscle groups with multiple joints are involved in the exercise, such experiments cannot be conducted. They instead argued it is better to study the force-velocity relationships involved in cycling by controlling the velocity of movement of the crank and measuring the external force generated (the power output generated by the athlete). As such, authors have investigated the possibility of an optimum cadence based on studying muscle force velocity relationships (MacIntosh et al, 2000, McCartney et al, 1983). Indeed Sargeant (1994) argued that a specific optimum cadence is likely to exist for individual athletes.

Whilst we define optimum cadence to be the cadence that maximises power output for a given heart rate, MacIntosh et al (2000), through studying force velocity relationships of muscles involved in cycling ergometry, defined an optimum cadence to be the cadence that requires the least muscle activation to generate a target power output. Given that an optimum cadence in our study could also be expressed as the cadence that minimises heart rate for a given power output, this appears somewhat similar to the optimum cadence defined by MacIntosh et al (2000), with heart rate or level of muscle activation representing some measure of physical exertion in each case. As power is a multiplication of force by velocity, in cycling this means power output is the multiplication of resistance and cadence – this is force multiplied by velocity of the crank (MacIntosh et al, 2000). For a given power output therefore, as resistance increases, cadence must decrease. MacIntosh et al (2000) argued that a given power output can be achieved at various combinations of cadence and resistance, but that one combination of cadence and resistance will likely require less muscle activation than the others - therefore for various levels of muscle activation, there will be a cadence that maximises power output in each case. MacIntosh et al (2000) recruited 8 male subjects to investigate this power output cadence relationship. They rode a cycle ergometer at cadences at 50, 60, 80, 100 and 120 rpm, each at a range of power outputs (100, 200, 300 and 400W), for short trials of not more than 30 seconds in duration (excluding time taken to warm up). MacIntosh et al (2000) inferred the power output cadence relationship by plotting peak power output against crank velocity (in m/s), multiplying crank velocity by 10 to get cadence (in rpm), rather than differentiating an equation of power output and cadence and equating to zero to gain a mathematical optimum cadence (as we aim to do in our study). Nevertheless for each targeted power output, power output was a parabolic function of cadence – similarly in our study we expect that, as cadence increases, power output will increase to a maximum before decreasing.

A study by McCartney et al (1983) however has suggested that the cadence that produces the highest power peak power output in brief maximal effort trials may not necessarily be the same as the cadence that produces the highest power output over longer periods of exercise. In the McCartney study, rather than measuring cadence of the athlete, crank speed was controlled by a motor in the bicycle, with athletes exerting as much force as they could on the moving pedals (such that no matter how much effort is exerted, pedalling rate or cadence could not exceed the pre-set crank speed on the bicycle). McCartney et al (1983) investigated torque-velocity relationships in cycling, exploring the effects of crank speed on power output. In the study 13 male subjects rode on a cycle ergometer at a range of crank speeds, for approximately two minutes at each crank speed. Crank speeds ranged from 60 rpm to 160 rpm, in intervals of 20 rpm. Subjects also rode in separate trials for 30 seconds at maximal effort at crank speeds of 60, 100 and 140 rpm, in which the experimenters studied how quickly power output declines after maximal effort at those crank speeds. Peak torque produced decreased linearly with increasing crank speed, though from plotting the relationship between peak torque and crank speed, the angle of the slope differed between subjects. Meanwhile the peak power output (the highest power output recorded in each trial for a given crank speed) produced by all but one of the subjects was a parabolic function of crank speed. However for one subject, peak power output increased with crank speed such that the highest peak power output occurred at the highest crank speed of 160 rpm; for this athlete, peak power output was hypothesised to occur at approximately 170 rpm, although this is higher than crank speeds typically chosen by athletes (Foss and Hallen, 2004). By plotting peak power output against crank speed, peak power output produced was typically greatest for crank speeds of approximately 120 rpm. Also in the second set of trials in which athletes rode at maximal effort for 30 seconds at 60, 100 and 140 rpm, peak power output occurred at the highest crank speed of 140 rpm (1050W at 140 rpm, compared to 457W and 964W at 60 and 100 rpm respectively). In addition to crank speed, as volume of thigh muscle increased, so too did peak power output produced. However, as crank speed increased, power output declined more quickly – rate of decline in power output was 32.4W/s at 140 rpm, compared to 11.6W/s at 60 rpm and 24.5W/s at 100 rpm respectively. After 17 seconds, power output was greater for a crank speed of 60 rpm than for 100 rpm or 140 rpm. McCartney et al argue this greater decline in power output at high crank speeds is due to decreased muscular efficiency at higher crank speeds.

Studies investigating muscle force velocity relationships have supported the notion of a specific optimum cadence for an individual athlete (MacIntosh et al, 2000; Sargeant, 1994). In our study we define an optimum cadence to be the cadence that maximises the power output for a given heart rate. MacIntosh et al (2000) found that power output was a parabolic function of cadence. McCartney et al (1983), who plotted power output against automated crank speed, also found that, as crank speed increases, power output reaches a maximum, before decreasing again. However in both studies this was mostly for exercise of very short duration. For exercise of longer duration, some authors argue that cyclists prefer to adopt a cadence of 90 – 105 rpm (Foss and Hallen, 2004), whilst others argue that a lower optimum cadence (around 60 rpm) is most efficient for exercise of long duration (Jordan, 1979).

We aim to calculate an optimum cadence for individual athletes in our study using field data from training session. We have more data available than in the McCartney et al study – hence we aim to investigate an optimum cadence for exercise of longer duration than in the McCartney et al study; an optimum cadence that an athlete can adopt in training and in races. By using SRM crank measurements we have power output measurements already available in our data. Hence we require only power output, heart rate and cadence measurements to calculate a mathematical optimum cadence for athletes in our study.

2.5.2. Muscle fibre type

Muscle fibres used during physical activity can be broadly classified into two types – type I and type II (Brooke and Kaiser, 1970; Coyle et al, 1992). Type I fibres are known as slow twitch, and type II are known as fast twitch – this refers to the speed of contraction of the muscles. Type II muscle fibres contract faster than type I (Crow and Kushmerik, 1982; Ingjer, 1978). Furthermore Type II can be split into type IIA, type IIB and type IIC, with increasing speeds of twitching or contraction from A to C – however they are broadly similar to each other; the greatest difference in characteristics occurs between type 1 fibres and type II fibres (Ingjer, 1978).

Different muscle fibres types are used depending on rate of oxygen consumption (Barstow et al, 1996; Poole et al, 1994). During heavy exercise oxygen uptake tends to continually increase until it reaches a maximum – before this maximum is reached, this is known as the 'VO_2 slow component', whilst exercise after this maximum has been reached is known as the 'VO_2 fast component' (Billat et al, 1998; Poole et al, 1994). Typically, during the VO_2 slow component, athletes tend to use a higher proportion of type II muscle fibres (relative to type I) than during the VO_2 fast component (Barstow et al, 1996; Poole et al 1994).

Whilst type II muscle fibres contract faster than type I muscle fibres, they are also less efficient than type I fibres (Coyle et al, 1992; Crow and Kushmerik, 1982; Inger, 1978). Indeed muscular efficiency may be related to the proportion of type I muscle fibres an athlete has (Coyle et al, 1991; Coyle et al, 1992). This was investigated by Coyle et al (1992), who studied 19 experienced cyclists with similar levels of maximum oxygen consumption (VO_{2max}). In order to determine what proportion of muscle fibres were type I in each athlete, experimenters obtained biopsies of samples of thigh muscles from both legs of each athlete (Coyle et al, 1992). The proportion of type I muscle fibres was then the number of type I muscle fibres found divided by the total number of muscle fibres found. Subjects varied in the proportion of their muscle fibres that were type I - between 32% and 76%. Experimenters measured maximum oxygen consumption (VO_2max) of subjects during exercise on a cycle ergometer. They also measured lactate threshold. They measured blood lactate concentration at exercise where oxygen consumption was between 50% and 60% of VO_{2max} – this is known as baseline blood lactate concentration. The lactate threshold is then the intensity of exercise that leads to an increase in blood lactate concentration of 1mM above the baseline blood lactate concentration (Coyle et al, 1992). Cycling efficiency was then calculated for exercise below the blood lactate threshold. Subjects exercised on a cycle ergometer at varying work rates below blood lactate threshold. Coyle et al (1992) determined cycling efficiency from gross efficiency and delta efficiency. Gross efficiency was defined in this study as the ratio of work completed to the rate of energy expenditure (in kcal/minute), and was averaged over the different work rates. Delta efficiency is the ratio of the change in rate of work completed and the change in rate of energy expenditure, over the different work rates (Coyle et al, 1992). Gross efficiency and delta efficiency both increased with proportion of muscle fibres that were type I. Gross efficiency was correlated with proportion of type I muscle fibres (r=0.75). An even higher correlation (r=0.85) was found between delta efficiency and proportion of type I muscle fibres. Indeed delta efficiency is a reliable measurement of cycling efficiency as it involves an accurate measurement of work completed; this accuracy means the calculation is not skewed by influence of metabolic processes (Coyle et al, 1992; Gasser and Brooks, 1975). Although there is a high correlation between efficiency and type I muscle fibres, this may not necessarily be a direct influence of type I muscle fibres on efficiency. In the Coyle et al (1992) study, the number of years of experience in endurance cycling an athlete has was positively correlated with proportion of type I muscle fibres. Coyle et al (1992) therefore acknowledged that the increased muscular efficiency of athletes with a high proportion of type I muscle fibres may be to some extent an ability developed over time by the athletes. However, whilst the proportion of type I muscle fibres are significantly positively correlated with the number of years of endurance cycling experience, they are not significantly correlated with total number of years of cycling experience. Coyle et al (1992) therefore claimed it is likely that the type I muscle fibres do contribute to muscular efficiency in cycling.

Whilst Coyle et al (1992) found a positive correlation between the proportion of type I muscle fibres and muscular efficiency, a study by Hansen et al (2002) suggests this relationship may depend on pedalling rate or cadence. Subjects in the Coyle et al (1992) study pedalled at a cadence of 80 rpm, whilst in the Hansen et al (2002) study, subjects (20 male semi-elite cyclists) exercised in a number of sessions with at a range of cadences – some preset and some sessions at freely chosen cadence. Subjects had between 21% and 97% type I muscle fibres. Exercise sessions were performed on a bicycle mounted on a treadmill. Before completing the treadmill cycling sessions, subjects exercised on a cycling ergometer in order for the experimenters to determine VO_{2max} of the subjects (to determine at what power output VO_{2max} occurs for each athlete). The treadmill cycling comprised 10 sessions for each athlete (5 minutes exercise followed by 5 minutes rest between sessions). In the first two sessions subjects chose their own cadences and rode at 90% and 70% of the power output at which VO_{2max} occurs. In the remaining 8 sessions, subjects rode at 61 rpm, 88 rpm, 115 rpm and freely chosen cadence, at 40% and 70% of the power output at which VO_{2max} occurs (Hansen et al, 2002). Subjects chose cadences between 56 rpm and 88 rpm at 40% of power output at which VO_{2max} occurs, and between 61 rpm and 102 rpm at 70%. Gross efficiency was calculated by dividing power output by rate of energy expenditure (Hansen et al, 2002). Hansen et al (2002) found a significant positive correlation between gross efficiency and proportion of type I muscle fibres for all preset cadences. However, when subjects chose their own cadence, there was no such correlation. There were also differences in the relationship between peak crank power (highest power output achieved by the subjects) and cadence depending on whether subjects pedalled at preset cadences or freely chosen cadences. At preset cadences, there was a significant negative correlation between proportion of type I muscle fibres and the peak crank power (r=-0.47), and also a significant negative correlation between the proportion of type I muscle fibres and the cadence at which peak crank power occurred (r=-0.81). However there were no such correlations when subjects pedalled at their own freely chosen cadence. When cycling at freely chosen cadence, subjects with a higher proportion of type I muscle fibres tended to choose a higher cadence (Hansen et al, 2002).

In estimating the proportion of muscle fibres that are type I or type II, experimenters tend to take a biopsy of small samples of muscle tissue, typically from vastus lateralis muscles in the thigh (Elder et al, 1982; Coyle et al, 1992). By taking samples from this muscle group, this provides an estimate of the muscle fibre distribution of the whole body (Coyle et al, 1992). Whilst there is little variation in muscle fibre distribution within different regions of the same sample of muscle tissue (Elder et al, 1982), in a study by Elder et al (1982), experimenters attempted to estimate proportion of type I and type II muscle fibres by sampling muscle tissue more extensively than in previous studies, including muscle tissue from biceps and triceps muscles. In this study, samples of muscle tissue were taken at autopsy from males who had died between 20 and 27 years of age. Autopsies were performed between 5 and 22 hours after death – in taking muscle samples in such short time after death, it is still possible to distinguish between type I and type II muscle fibres (Elder et al, 1982). Also, in taking autopsies rather than biopsies, this reduces the possibility of neuromuscular disease being present in the muscle tissue samples (Polgar et al, 1973). The study confirmed that, within the vastus lateralis in the thigh, there is little variation in distribution of muscle fibre type. However, within other muscles, there is more variation in distribution of muscle fibres (Elder et al, 1982). Nevertheless, this indicates that, in taking samples of muscle tissue for determining distribution of muscle fibre type, fewer samples are needed if tissue is taken from the vastus lateralis muscle than with other muscles (Elder et al, 1982). Indeed, in the Elder et al (1982) study, in order to take muscle tissue samples where muscle fibre distribution varied with a 5% standard deviation, 4 samples were required for tricep muscles, 5 for bicep muscles, but only 3 samples were needed if taking tissue samples from the vastus lateralis muscle. The findings from Elder et al (1982) therefore suggested that the methodology typically used in muscle fibre studies – i.e. taking samples from the vastus lateralis muscle – is not unreasonable.

In our study, we seek to find an optimum cadence – the cadence that maximises power output for a given heart rate – using statistical models of power output, heart rate and cadence. The relationship between power output and cadence could be influenced by muscle fibre type (McCartney et al, 1983). McCartney et al studied the effect of different crank speeds on power output (such that the pedals moved automatically at a certain rate, with the athletes exerting as much force on the pedals – pedals could not be moved at a faster rate than the speed to which the crank is set). Athletes with a higher concentration of type II muscle fibres (which contract more quickly) generated their highest power output at a higher crank speed than athletes with a lower concentration of type II muscle fibres (McCartney et al, 1983). For example, in the McCartney et al study, one subject with 72% type II fibres generated his highest power output at a crank velocity of 162 rpm, whilst another subject with 53% type II fibres generated his highest power output at 119 rpm. This suggests an individual athlete's optimum cadence could therefore be influenced by the prevalence of type II muscle fibres in that individual athlete. However, we aim to calculate an optimum cadence for athletes in our study empirically by using data from training sessions, such that we do not require knowledge of the composition of the athletes' muscle fibre types. Nevertheless, future research into optimising cadence using our methodology, combined with an analysis of the composition of different muscle fibre types of the athletes involved, could be useful in understanding the relationship between composition of muscle fibre type and cadence.

2.5.3. Optimum cadence and athletic skill

Athletes are unlikely to choose to ride at very low cadences, such as 50 (Lepers et al, 2001), despite this cadence being optimal in some ergometer studies (Foss and Hallen, 2004). Elite professional cyclists tend to prefer higher cadences, around 90-105 (Foss and Hallen, 2004). Foss and Hallen (2004) claimed that these higher cadences are due to athletes not being sufficiently trained in such ergometer studies. Indeed Coast and Welch (1985) demonstrated that optimum cadence increases with power output; the insufficiently trained athletes would not be able to produce high enough power outputs for optimum cadence to increase to the level favoured subjectively by the athletes. Belli and Hintzy (2001) also suggested professional cyclists choose to ride at higher cadences (around 90 – 110), although they claimed this is equally true for most professional athletes regardless of the level of training they have received.

Experimentation has also tended to include participants with varying levels of skill (Coast and Welch, 1985). Indeed Coast and Welch (1985) suggested that the reason why results from their study differed from results by Hagberg et al (1981) was due to the higher level of skill amongst the athletes in the study by Hagberg et al (optimum cadence from the Hagberg et al study was 91 rpm, compared to 83 in the Coast and Welch study). Athletes may in fact have their own individual optimum cadences, depending on the level of skill of the athlete; the greater the level of skill, the higher the cadence that can be maintained by the athlete (Coast and Welch, 1985).

2.5.4. Training impulses (TRIMPs)

Quantification of training stimuli in aerobic sports is typically done using training impulses, known as TRIMPs (Jobson et al, 2009; Morton et al, 1990; Borreson and Lambert, 2009). They are a multiplication of external training load by training intensity (Jobson et al, 2009). Put simply, to calculate TRIMP, the average heart rate for a session is multiplied by the duration of a session, and then by a weighting factor, to give a greater weight to sessions where heart rate is particularly high (Jobson et al, 2009; Morton et al, 1990).

TRIMPs were first proposed by Banister et al (1975), to represent a measure of physical effort, providing one measurement for a training session. Calculation of the TRIMP outlined by Banister et al requires measures of training duration, average heart rate during a session, resting heart rate and maximal heart rate, which form the following equation, and a weighting factor to give greater weight to periods of exercise of high intensity. Without such a correction, the equation had a bias towards extremely long training sessions where heart rate was relatively low (Borreson and Lambert, 2009). The TRIMP equation was:

$$TRIMP = dfY$$

where d is duration of a session (in minutes), f is fraction of heart rate reserve $((H_{av} - H_{rest}) / (H_{max} - H_{rest}))$ in which H_{av} is the average heart rate for that training session, H_{max} is maximal heart rate and H_{rest} is resting heart rate, and Y is the weighting factor, which is $0.64e^{1.92f}$ for males and $0.86e^{1.67f}$ for females. The weighting factor was based on different responses amongst men and women to changes in blood lactate concentration (in mM) in response to increasing heart rate (Borreson and Lambert, 2009). Morton et al (1990) then modified this TRIMP equation to

$$TRIMP = dfe^{fb}$$

where b is 1.92 for males and 1.67 for females.

TRIMP calculations have been adapted by Edwards (1993) to account for interval training. Edwards (1993) split data into five zones for heart rate, based on percentage of maximal heart rate (50-60%, 60-70%, 70-80%, 80-90% and 90-100% of maximal heart rate form the 5 zones respectively). Duration of exercise in each zone was multiplied by the zone number (from 1 to 5), and then summated.

This has been further modified by Lucia (1999) such that heart rate zones were based on exercise above and below aerobic and anaerobic thresholds. The three zones were below aerobic threshold, between thresholds, and above anaerobic threshold (zones 1, 2 and 3 respectively). Duration spent in each zones was multiplied by the number of that zones (1, 2 or 3), and then summated. For example, 60 minutes in the zone below aerobic threshold (zone

1) was equivalent to 30 minutes at the zone between thresholds (zone 2) or 20 minutes above anaerobic threshold (zone 3). This method is known as 'Lucia's TRIMP' (Impellizzeri et al, 2004).

Increases in weighting from one heart rate zone to another are linear, in both the summated heart rate zone method and in Lucia's TRIMP (Borreson and Lambert, 2009). For exercise above the anaerobic threshold, this linear increase does not reflect physiological responses to exercise (Borreson and Lambert, 2009). Both of these methods also do not account for differences in duration of pauses in exercise after periods of exercise of different intensity; they do not weight these pause times (Cejeuela-Anta and Esteve-Lanao, 2011).

In our study we also attempt to modify the TRIMP equation outlined by Morton et al (1990) to place further weight on periods of exercise of high intensity (where heart rate is higher). We do this by calculating a TRIMP value for each measurement point – each time power output, heart rate and cadence are recorded in the data. We multiply the heart rate at that measurement point by duration that has elapsed at that point, fraction of heart rate reserve and by the weighting factor e^{fb}. Therefore the weighting factor e^{fb} is included for each intermediate TRIMP measurement within a session. Rather than splitting exercise into a small number of different zones for heart rate, and use a weighting factor giving only as many options as there are zones, we use a weighting that is based on the exact heart rate measurement.

By calculating an intermediate TRIMP for any recorded measurement point within a training session, we can also include the intermediate TRIMP measurement as a variable itself. This allows us to explore how training load varies within a session (whereas in current literature training load tends to represent a single value to summarise one session (Jobson et al, 2009)).

We investigate the impact of the classic Morton et al (1990) TRIMP for each session, our modified version of TRIMP with a greater weight on periods of exercise of high intensity for each session (based on each data measurement), and the intermediate TRIMP for measurement points within sessions, on the power output / cadence relationship. We investigate how these TRIMP measures affect the optimum cadence – the cadence that maximises power output for a given heart rate - in our statistical models of power output, heart rate and cadence. We include models with TRIMP measurements in chapter 5. Equations used in calculating TRIMP are outlined again in the methodology of chapter 5.

2.5.5. Riding mode and race conditions

During cycling athletes may spend some of the time sitting in the saddle, and some time standing out of the saddle – we term this riding mode (whether they are sitting or standing). We do not have such information available, so we cannot consider whether optimum cadence varies with riding mode. In this study we also do not consider how cadence could vary with body orientation (the angle the athlete is sitting in the seat). In races athletes tend to tilt forwards slightly although athletes may be able to produce a higher power output when sitting upright (Welbergen and Clijsen, 1990). Leirdal and Ettema (2011) investigated the cadence-gross efficiency relationship for different body orientations. Athletes rode until exhaustion at different positions, sitting at a naturally chosen position, tilting further forward than the chosen position, and sitting with the seat adjusted such that the seat is slightly further back, thus increasing the distance between the seat and the pedals. However no significant differences in the cadence-gross efficiency relationship were found between different body orientations.

During road races and time trials athletes are likely to spend some time not pedalling, such as when going downhill, or when approaching tight corners. When the athlete starts pedalling again, he/she accelerates back up to a chosen cadence. As power output, heart rate and cadence data are collected on the bicycle at regular 5 second intervals, our field data includes time spent accelerating (from a low cadence back up to the athlete's chosen cadence). By defining optimum cadence as the cadence that maximises power output for a given heart rate for such field data, the optimum cadence in our study takes account of time spent accelerating. Note we do not include time spent not pedalling in our analysis. We do not consider whether an optimum cadence is affected by what proportion of time is spent pedalling vs not pedalling.

Cadence could potentially be affected by the type of race undertaken. During time trials, an athlete can choose a pacing strategy for the duration of the event without being on the track at the same time as other athletes. However for road races, more tactics are involved as athletes attempt drafting (riding behind another athlete), as the reduced air resistance means the athlete can ride at a lower power output but retain the same speed (de Koning et al, 1999). In this study we do not consider whether an optimum cadence could be affected by drafting.

2.6. Summary of literature reviewed

In this chapter we have reviewed literature not only about optimum cadence but also about general performance in road cycling. Different authors have claimed different pacing strategies to be optimal, including a fast start (de Koning et al, 1999) and even pacing throughout (Chaffin et al, 2008; Wells et al, 2013). For athletes in our study, who compete in long time trials (of 34km or over), literature suggests an even pacing strategy is optimal (Chaffin et al, 2008; Palmer et al, 1998; Wells et al, 2013), although Chaffin et al (2008) have argued that more studies should be condicted regarding pacing strategy in long distance time trials.

Studies in optimum cadence have tended to focus on the cadence that maximises gross efficiency, which is the amount of work completed relative the amount of metabolic energy required to do that work (Winter and Fowler, 2009; Whipp and Wasserman, 1969). However, gross efficiency is difficult to calculate (Winter and Fowler, 2009). For our study we consider a cadence to be optimal if it maximises power output for a given heart rate. Professional cyclists tend to adopt cadences of 90-105 (Foss and Hallen, 2004), although lower cadences are typically found to be optimal in laboratory experiments (Samazino et al, 2006; Coast and Welch, 1985). Such laboratory experiments however do not necessarily reflect the environment of outdoor road cycling (Jobson et al, 2007). The relationship between power output and speed differs between laboratory conditions and road cycling (Jobson et al, 2007) although, broadly, as speed increases, so too does power output (de Koning et al, 1999).

Power output and heart rate tend to be linearly related for low power outputs, but curvilinear for high power outputs (Grazzi et al, 2009). Heart rate is lagged behind changes in power output (Churchill et al, 2009). A study by Stirling et al (2008) suggests that for athletes in our study, the length of this lag in heart rate is likely to be 30 seconds or less.

Existing statistical models in cycling performance have tended to focus on aerodynamic features of performance, such as the power output needed to overcome air resistance (Di Prampero et al, 1979; Martin et al, 1998). However we fit regression models of power output, heart rate and cadence (detailed in chapter 4), in which heart rate lags behind power output – we experiment with different heart rate lags in chapter 5.

Muscle fibre type can also affect cycling performance (Coyle et al, 1992; Hansen et al, 2002). Muscle fibres are broadly classified as type I, which are slow to contract but efficient, and type II, which contract quickly but are less efficient than type I (Coyle et al, 1992). The more type I muscle fibres and athlete has, the higher the cadence he or she can adopt (Hansen et al, 2002).

We quantify accumulation of fatigue by calculating training impulses or TRIMPs. TRIMPs are the multiplication of the average heart rate for a session and the duration of a session, further multiplied by a weighting factor to give greater weight to sessions where heart rate is particularly high (Morton et al, 1990). TRIMPs have been commonly used for quantifying training load or fatigue (Borreson and Lambert, 2009). Morton et al (1990) have claimed that ideally TRIMPs should place an even greater emphasis on sessions where heart rate is particularly high – we therefore modify TRIMP calculations accordingly. We include TRIMPs as a variable in our regression equations – hence we observe how optimum cadence varies by fatigue.

2.7. Variables required to calculate optimum cadence in this study

We only consider variables that can be calculated from the power output, heart-rate, and cadence measurements —no other information was available to us at the time of the study. Nonetheless these raw data measurements of power output, heart rate and cadence are sufficient for the analysis. The speed at which an athlete travels is to some extent represented by the power output, as increasing power output tends to increases speed (de Koning et al, 1999). Nevertheless power output is an important measure of performance, as sustaining high power outputs is crucial for performance in both road and track cycling (Winter and Fowler, 2009). The gear in which an athlete is riding is not needed, as this is controlled by cadence. Changing gear automatically enables the rider to change cadence accordingly – if the rider rides up a hill for example, he or she can select a lower gear and decrease resistance, maintaining the same cadence at no extra effort. We do not study the physical power generated by the athlete (the rate at which external mechancial work of the body is performed (Winter and Fowler, 2009)), rather we include measurements of power output collected from SRM cranks on the bicycle. We seek to calculate the cadence that maximises power output for a given heart rate, for each athlete in our analysis.

Other variables can be calculated form the raw power output, heart rate and cadence mesaurements, and those that might influence the heart rate response are also included in our regression modelling approach. We take account of training load by calculating the training impulse within a session up to a particular time point.

Chapter 3: Description of the data

3.1. Introduction

In this chapter we describe the research we intend to undertake, the data available to us and how we intend to utilise the data. Firstly we describe the participants who took part in this study, and their characteristics. We then outline the research design and length of the study period. Next we describe how data were collected and how the data became available to us. We then summarise the data presented to us and which athletes we choose to use in this analysis. Finally we describe how we process the data in order to fit mathematical models to the data.

3.2. Participants and data structure

Ten competitive male athletes, along with their coaches, gave written, informed consent to participate in our study, providing us with data (power output, heart rate and cadence measurements) from their training. They are typically category II and III cyclists competing at county championship level. They compete in road races and time trials, but mainly in road races. They therefore train for long periods, with sessions ranging between 1 and 6 hours. These athletes tend to train on their own or in small groups of two or three. Athletes' age, height and weight are presented below in table 1. The ten athletes have mean (plus standard deviation in parentheses) age of 36 (9) years, height of 1.79 (0.46) metres, and weight of 74.3 (6.8) kg. Athletes are numbered for confidentiality reasons.

Table 1: Age, height and weight of the ten athletes.

Athlete	Age (years)	Height (m)	Weight (kg)
1	21	171.4	60.9
2	40	177.5	75.5
3	52	175	74.5
4	45	183	74.3
5	34	182	77
6	42	178.5	78.2
7	27	183.7	71.8
8	35	181	71
9	34	185.5	88.2
10	29	174.5	71.5

3.3. Research Design

This study took place over a period of five years, from October 2008 to September 2013. We used data comprising power output, heart rate and cadence measurements from field data (training sessions and races) for ten cyclists, riding between October 2006 and January 2008.

These are measurements taken using hardware fitted to the bicycle. We attempt to fit mathematical models of power output, heart rate and cadence that yield optimum cadences – this is a cadence that maximises power output for a given heart rate – for individual athletes using data provided. This is done using SPSS and R statistical packages. Our aim is to determine whether such models can be fitted using field data in this way.

We assess and compare different mathematical models through the apparent validity of the optimum cadences (how plausible they are) and by model fit criteria – explanatory power, Akaike information criteria and standard errors of coefficients (explained in more detail in section 4.1.10). If plausible optimum cadences are yielded, we compare these optimum cadences yielded from the models with cadences typically preferred by the athletes in our analysis, and with optimum cadences found in literature.

3.4. Measurement methods

The athletes rode with an SRM power meter which calculates power output and cadence every half second (SRM, 2012). (SRM is a power meter developed by the engineering company Schoberer Rad Messtehnik.) The SRM calculates power output through a strain gauge, measuring the amount of twisting or torsion, sending this information to a microprocessor, where it is converted to wattage (SRM, 2012). It has been verified in literature to be very reliable (Martin et al, 1998). On the online SRM shop SRM power meters are priced between €1892 and €3808. Heart-rate data were measured using a heart-rate monitor.

An example of an SRM power meter crank is illustrated in figure 1 below.

Figure 1: An example SRM power meter crank. This is the SRM Canondale MTB 2x10 model, which weighs 521g and costs €1892 (SRM, 2012).

SRM also provides a monitor on the handlebars of the bicycle. Monitors typically display power output (Watts), heart rate (beats per minute, providing the heart rate monitor used is compatible with SRM software), cadence (revolutions per minute), speed (miles per hour or kilometres per hour), temperature (Fahrenheit or Celsius) and altitude (metres) at the current time. Athletes can therefore check they are riding at a desired power output and cadence whilst riding. The monitor also displays rolling average power output up to the current time point (this is the mean power output for the session up to that point, excluding any time where power output was zero, when the athlete was not pedalling).

An example of an SRM monitor is illustrated below in figure 2.

Figure 2: An example SRM monitor. This is the Power Control 7 model, with a battery life of 120 hours (SRM, 2012).

SRM calculates power output and cadence once every half second, and the heart rate monitor calculates heart rate every second. However due to limited storage capacity, athletes typically collect data at wider intervals than half a second. For each variable (power output and cadence, along with heart rate measured from the heart rate monitor), the calculations are averaged to yield one value at regular intervals, typically 5 seconds. The calculations are averaged to yield one measurement every 5 seconds for the four athletes we chose, whilst for other athletes, this measurement interval varies between 5 and 15 seconds. Hence the data available to us are these averaged calculations at regular intervals, which we term 'measurements'. Later in chapters 4, 5, 6 and 7 in this thesis whenever we use the terms 'measurements' or 'raw data', we are indeed referring to the data originally presented to us (the calculations that are averaged at regular intervals to produce 1 measurement at each interval).

After a training session or race, athletes use software provided with SRM to upload power output, heart rate and cadence data on to a spread sheet, along with the data of the session. This results in a spread sheet of power output and cadence measurements where each column represents the session undertaken for that day.

For each athlete, training schedules contain all training and competitions over a period of approximately 12 months. The study we carry out in this thesis determines optimum cadence using the within session measurements.

3.5. Description of training undertaken

Training schedules range from 5 to 13 months, from October 2006 to January 2008, with mean power output for a session typically between 120W and 250W. The length of the study period - that is, the length of the data collection period - varied across athletes – as shown in table 2 (month, year – month, year).

Table 2: Start date, end date and duration of training schedules for each of the 10 athletes.

Athlete	Start	End	Duration (months)	Duration (days)
1	01/11/2006	30/09/2007	11	333
2	24/10/2006	07/10/2007	11	348
3	04/03/2007	11/08/2007	5	160
4	28/10/2006	12/12/2007	13	410
5	06/12/2006	04/09/2007	9	272
6	19/04/2007	31/01/2008	9	287
7	02/11/2006	02/08/2007	9	273
8	27/10/2006	30/09/2007	11	338
9	21/11/2006	28/07/2007	8	249
10	10/11/2006	23/09/2007	10	317

There was also variation in the completeness of the data record among the 10 athletes. Figure 3 shows the number of days in the study period and the number of days on which training data were collected for each athlete in the sample. Across athletes, training data were available on approximately half of the number of days in the study period. Absence of data on a particular day might be due to either a lack of recording or if there was no ride that day. For athlete 4, for example, there were 24 days in which heart rate was recorded, but not power output and cadence - such days were not included in our analysis and were treated as no activity. However, even if full data had been provided and these 24 days were treated as sessions, athlete 4 would appear to have trained on only 274 days of the 480 within the study period. Clearly an athlete cannot train every day for such a sustained period of time, but there were long periods of successive days in which no activity took place. For example, the first athlete included in the analysis did not record any training for a period of 52 days (between day 219 and 271). This lack of activity might be due to injury, illness, or equipment failure.

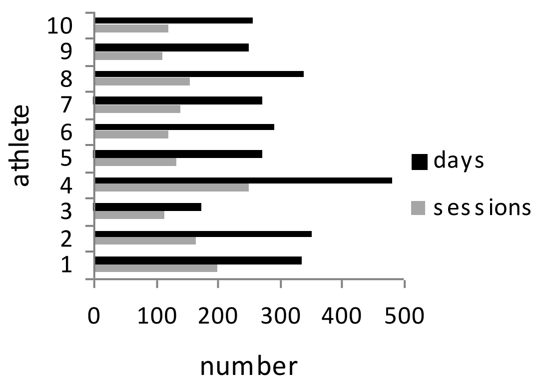


Figure 3: The number of days and sessions (with complete data for power output, heart rate and cadence) in each athlete's training schedule.

A further complication was that the sampling interval varied for some athletes - for athletes 5, 8 and 9, power output, heart rate and cadence were recorded at 7- or 15-second intervals for some sessions. Although we used within session measurements, it was still thought best to concentrate our analysis on those athletes with the most complete records. We chose four athletes for our analysis, who were labelled 1 to 4. In figure 4 we present examples of power output, heart rate cadence measurements yielded at 5 second intervals for athletes 1 and 3 from two training sessions. Their mean heart rate, power output, and cadence in each recorded session over the study period are shown in figures 5-7.

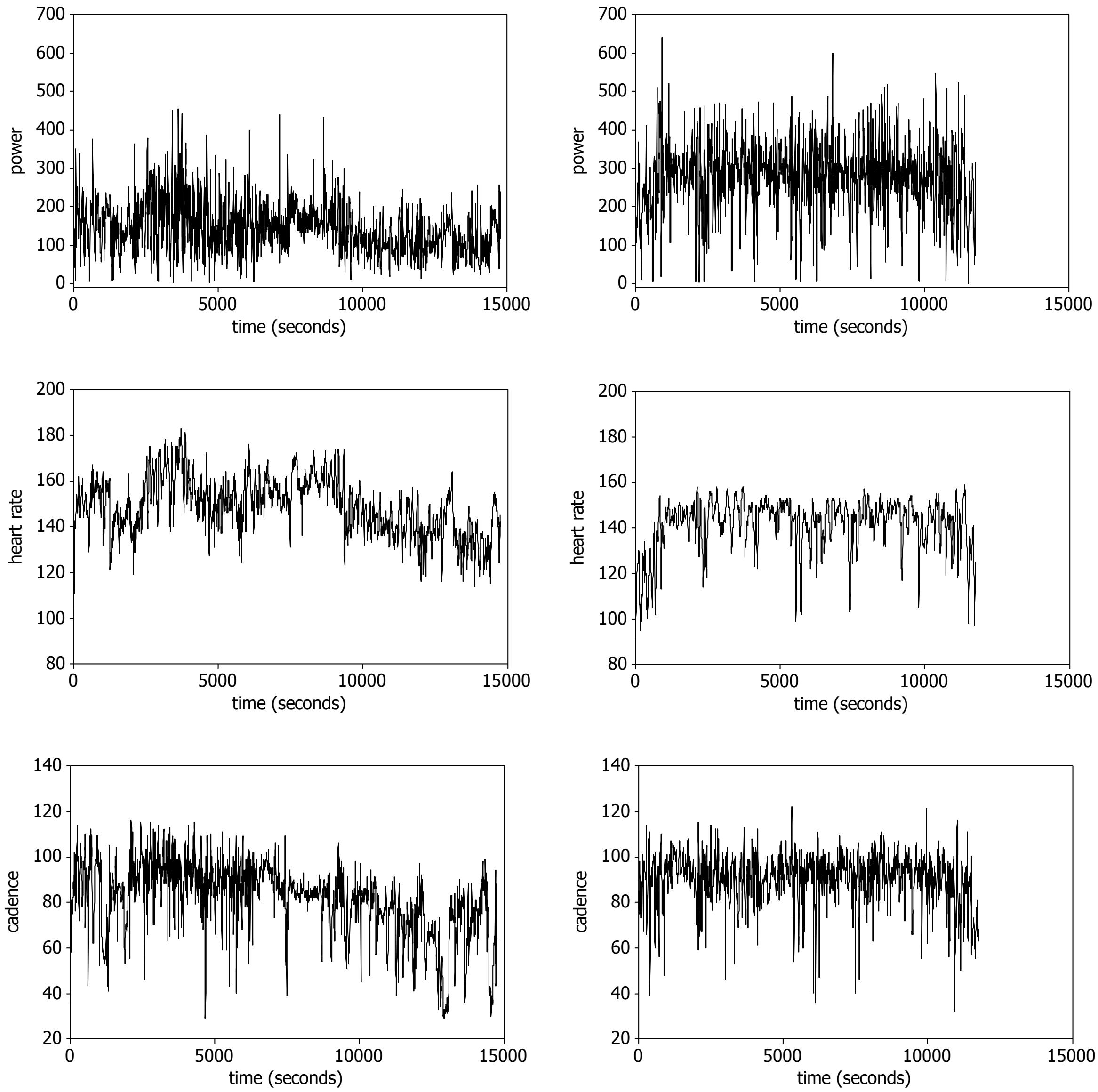


Figure 4: An example of power output (watts), heart rate (beats per minute) and cadence (revolutions per minute) traces from one session for athletes 1 (left) and 3 (right).

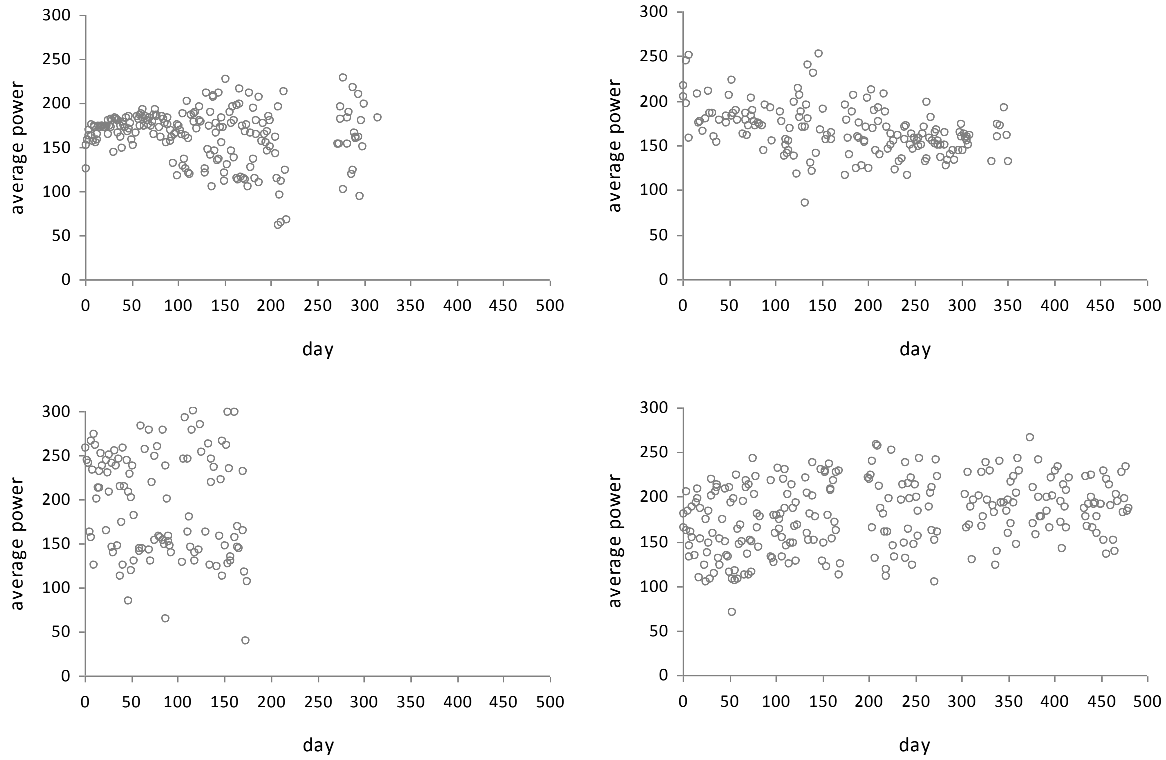


Figure 5: Mean output (watts) for each training session for athletes 1-4 (read lexicographically). Power output measurements (taken every 5 seconds) are averaged to form the 'mean output' for that session.

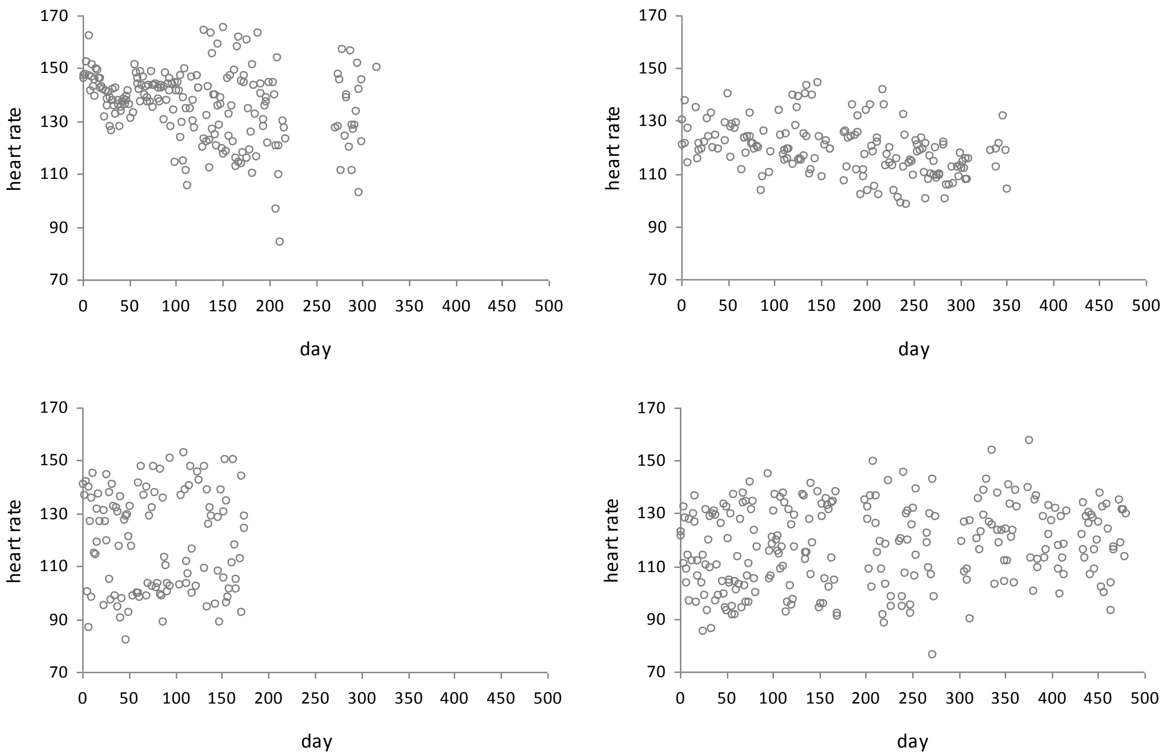


Figure 6: Mean heart rate for each training session for athletes 1-4.

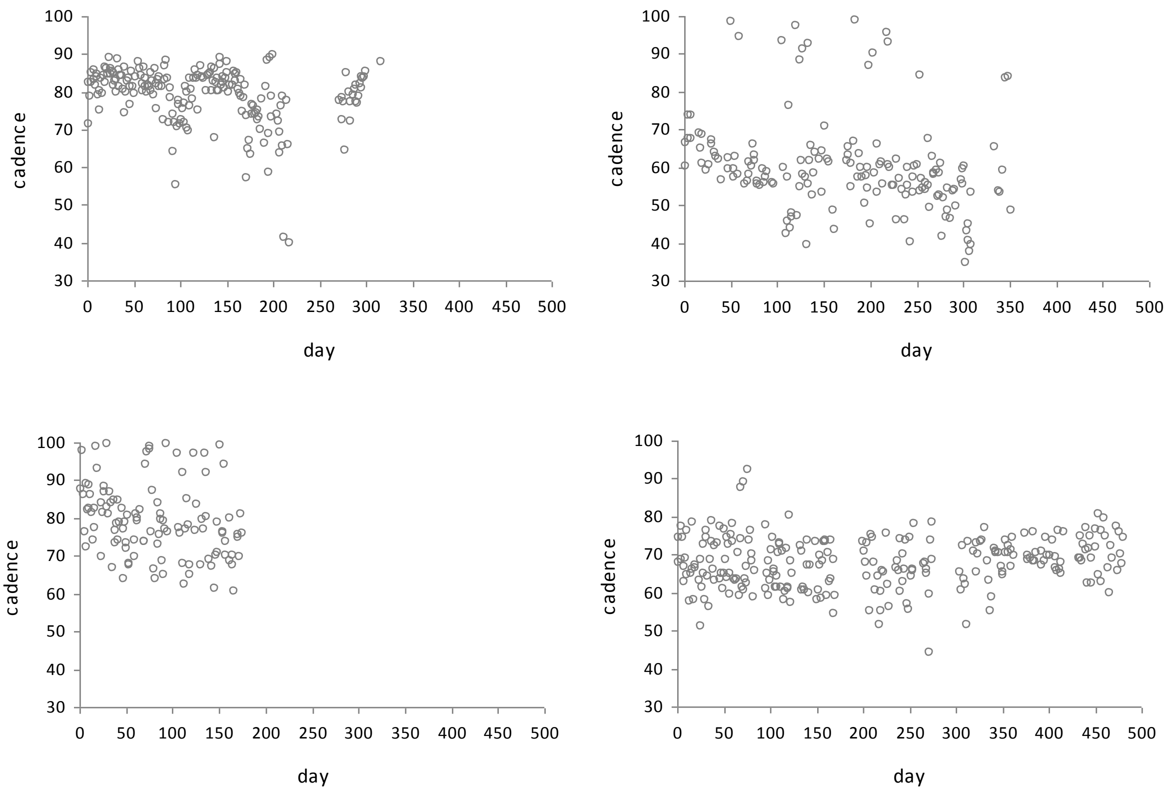


Figure 7: Mean cadence for each training session for athletes 1-4.

The data include training sessions and some races. Some athletes provide brief descriptions of races undertaken in the training diaries. Without such descriptions, we do not know if data come from a race or from training sessions. However there are likely to be many more training sessions than races. During winter months, athletes are likely to train for endurance, building up fitness slowly. After winter, there is an increase in sessions in which they work on speed skills (ability to ride at a high cadence), and force skills (ability to ride at a high gear). In spring and summer athletes take part in selected races. After races, athletes rest before building up training again. At the onset of the following winter, this cycle of training and races begins again. However descriptions of types of training session undertaken are not provided in the data.

3.6. Participants chosen for analysis

We selected the four athletes that appeared to have the most complete and consistent training diaries. Of the four athletes we chose, three had the longest schedules – these were athletes 1, 2 and 4, whose schedules were 333, 348 and 410 days respectively. The four athletes also had sessions on between 50% and 75% of the days of their schedules, which was slightly higher than the remaining 6 (athlete 3 had sessions on 75% of days, the highest of any athlete). For the six athletes not chosen, on some days one of power output, heart rate and cadence was missing for the entire day – such missing data rarely occurred for our four chosen athletes. Recording intervals for data were 5 seconds for athletes 1 to 4, but for some sessions for the other athletes, the recording interval changed to 7 or 15 seconds, which would complicate the process of sampling data to use. Athletes chosen had some description of races undertaken –

distance and position at the end of the race, and sometimes the type of race undertaken (although there was little description about the types of training session undertaken for all ten athletes).

We do not believe that the four athletes chosen are a biased sample of the ten available. Mean age, height and weight of the four chosen (33 years, 1.78m and 73.9kg) is similar to the mean age, height and weight of the six athletes not chosen (38 years, 1.79m and 74.6kg). Nevertheless, since our aim is to develop a methodology for calculating individual optimum cadences, we do not extrapolate optimum cadences found to apply to a more general population of cyclists, nor even for the group of ten cyclists whose data are given to us.

Amongst the four athletes with the most complete training diaries, we predict that it may be easier to calculate an optimum cadence for athletes 1 and 2 than for athletes 3 and 4. This is because the data suggest that athletes 1 and 2 ride at a variety of low, medium and high cadences within their training, whereas athletes 3 and 4 appear to seldom ride at high cadences.

3.7. Summary of data provided

We are presented with power output, heart rate and cadence measurements collected from training sessions and races for ten cyclists, over a period of approximately 12 months. The measurements are typically collected at 5 seconds intervals from SRM power metres fitted to the bicycles. For some athletes however data are collected at 7 or 15 second intervals for some sessions. There is limited information available regarding types of training session undertaken, or reasons for long periods where no training took place.

We select a sample of four of these ten athletes; we chose the four athletes with the most consistent training diaries (with the least amount of missing data and where data are recorded at regular 5 second intervals).

3.8. Processing the raw power output, heart rate and cadence measurements

We develop a range of regression models of power output, heart rate and cadence. Before the parameters in the regression models are estimated, we consider processing the raw measurements of power output, heart rate and cadence. By raw, we mean the data that were given to us at 5 second intervals. This processing addresses three issues: (i) smoothing the raw data; (ii) sampling the smoothed data in order to reduce serial dependence; (iii) determination of the most appropriate time-lag between power output and the heart-rate response.

We consider smoothing the data because there are sometimes extreme changes in power output (and to a lesser extent cadence and heart rate). Indeed Jeukendrupp and van Diemen (1998) suggest it is quite common for power output to vary more than heart rate. For heart rate, changes tend to be quite gradual, but there are still occasional moments in recording where there is a sudden increase or decrease in heart rate. For our power output, heart rate and cadence measurements, we use a moving average over k measurements (this is a window width of k time units). This results in less extreme changes in the variable from one measurement to the next. Denoting the raw measurements by $Y_1, Y_2, \ldots$, the smoothed measurements are given by $\bar{Y}_1, \bar{Y}_2, \ldots$ where, at time *t*,

$$\bar{Y}_t = (Y_t + Y_{t-k} + \ldots + Y_{t-k+1}) / k \,,\ (i \geq k)\,.$$

Since the data available to use are power output, heart rate and cadence measurements at regular 5 second intervals, 5 seconds is the unit of time. Smoothing will produce power output, cadence and heart rate measurements again every 5 seconds, albeit smoothed measurements.

Next we seek to reduce the serial dependence in each series by sampling every *m*th smoothed measurement. The choice of *m* is more arbitrary, although serial dependence in the smoothed series can be investigated and we should confine ourselves to $m \geq k$. The resulting processed measurements are then $\bar{Y}_l, \bar{Y}_{l+m}, \bar{Y}_{l+2m}, \ldots$ Finally, it is necessary to specify the time-lag between the power output and the heart rate response, which we denote by *l.* Thus for a given *k* we regard the processed heart rate measurement $\bar{H}_{k+tm+l}$ as a response the processed power output and cadence measurements $\bar{P}_{k+tm}$ and $\bar{C}_{k+tm}$, l time units earlier (for all *t*). For example, taking *k*=12, m=12 and *l*=6, implies that raw data would be smoothed using a moving average with a window of 12 time units (60 seconds); the resulting series would then be sampled once every 12 time units (60 seconds), but with heart rate sampling out of phase with the power output and cadence sampling so that the time-lag between heart rate response and power output would be 6 time units (30 seconds). Thus the smoothed heart rate at time 1 minute 30 seconds is assumed to be a response to the smoothed heart rate at 1 minute, the smoothed heart rate at time 2 minutes 30 seconds is assumed to be a response to the smoothed heart rate at 2 minutes, etc.

We present the models fitted, and optimum cadences yielded, in chapters 4, 5 and 6.

When fitting a range of models we first focus on the success of the model in representing the true relationships between heart rate, power output and cadence, and seek to find a range of optimum cadences from the models. We initially consider smoothing data with a window width of 12 time units (a period of 60 seconds), as we consider this to be a reasonable amount of time to enable extreme changes in power output, cadence and heart rate to be reduced somewhat. We choose a heart rate lag of 1 time unit (5 seconds, as this seems appropriate from literature), and sample once every 24 time units (120 seconds). Later, after fitting a range of models with some success, we will consider the impact of different data processing constants (*k* and *l*) on the models – in other words, the sensitivity of the models to the data processing constants.

Sessions are combined to provide one large stream of data for an individual athlete. Alternatively, the session effect might be modelled as a random effect in a mixed model, but we do not pursue this further here. We instead use other within and between session explanatory variables to represent this effect. This makes the analysis somewhat simpler and easier to interpret. The important thing to note is that our analysis can only consider variation in the optimum cadence of an athlete between sessions through the variables relating to the sessions that we measure. A description of such variables follows in chapters 4, 5 and 6; broadly, in chapter 4 we explore different models of power output, heart rate and cadence, whilst in chapters 5 and 6 we expand and modify the models.

Chapter 4: Exploration of linear and non-linear model fitting

4.1. Methodology

4.1.1. Introduction to methodology

In this chapter we explore a range of models of power output, heart rate and cadence data. We firstly present the methodology for fitting different models, along with the calculations needed to calculate optimum cadence from these models. We also describe statistical measures used to assess such models.

We suppose that heart rate principally acts as a response to power output, but also in response to cadence, and other training related variables. We further assume that the heart rate response occurs after some delay. The length of this delay is not necessarily clear from literature (Churchill et al., 2009). However, as described in the sports science background chapter, a low heart rate lag (below 30 seconds (Stirling et al, 2008)) appears most appropriate.

We fit a range of linear and non-linear models. In this chapter we include methodology, results and discussion for linear and non-linear models fitted (including models with extra covariates other than power output, cadence and heart rate).

4.1.2. Methodology for linear models of power output, heart rate and cadence

Whilst modelling heart rate as a response to power output and cadence, we also explore the relationship between power output and cadence. Initially we consider a linear relationship between power output and cadence within a regression model featuring heart rate as a response. In this first regression model the power output developed, *P*, the cadence, *C*, and other training related variables, summarised through the (possibly vector valued) explanatory variable *X*, are related to the expected heart rate at some time-lag, E(*H*), through the equation

$$E(H) = h_0 + \alpha P + \beta C + \gamma C^2 + \delta X \ . \tag{1}$$

Here h_0 is the baseline heart rate.

To determine the optimum cadence from this model, consider the question: for a given expected heart rate response, what value of cadence maximises the power output? To answer this question we regard *P* and *C* in equation (1) as variables, other quantities as fixed, and then express the power output *P* in terms of the cadence thus:

$$P = \{E(H) - h_0 - \delta X - \beta C - \gamma C^2\} / \alpha \tag{2}$$

We define the optimum cadence to be the cadence that maximises power output for a given heart rate, and therefore we imply that a cyclist will seek to ride in a manner that maximises power output for a given heart rate. Differentiating equation (2) with respect to *C* we get

$$\frac{\mathrm{d}P}{\mathrm{d}C} = \frac{1}{\alpha}(-\beta - 2\gamma C) \ .$$

Setting this to zero and solving for *C* gives

$$C^* = -\beta / 2\gamma \ . \tag{3}$$

This value of C maximises P if

$$\frac{\mathrm{d}^2P}{\mathrm{d}C^2} = -2\gamma / \alpha < 0\,. \tag{4}$$

The coefficients could be positive or negative in each case, but an optimum cadence from the regression model would only be found if a number of conditions regarding polarity of these coefficients are satisfied. Assuming that $\alpha > 0$ (which is this case if the heart rate increases as power output increases), condition (4) requires that $\gamma>0$. This is most simply explained by noting that, in equation (2), power output is a quadratic function of cadence C. Thus, there will exist a maximum power output only if the coefficient of C^2 is negative. Furthermore, for the cadence at which the power output is maximised to be positive, the model requires $\beta<0$. These conditions will need to be satisfied when statistically estimating the parameters in equation (2).

Alternatively, for a given power output we can seek that cadence that minimises the expected heart rate. Now assuming P in equation (1) is fixed, and that E(H) is a function of C, differentiating with respect to C we get $\mathrm{d}H / \mathrm{d}C = \beta + 2\gamma C$. Equating this to zero and rearranging gives $C^* = -\beta / 2\gamma$ which is the same as (3). This value of C minimises H if $\mathrm{d}^2 H / \mathrm{d}C^2 = 2\gamma < 0$. This requires that $\gamma>0$, and further that $\beta<0$ for $C^*>0$ as before.

So, the cadence that maximises power output for a given heart rate and the cadence that minimises heart rate for a given power output are equivalent. Thus we suppose that a rider can control cadence in order to either (i) vary heart rate while keeping power output fixed; or (ii) vary power output while keeping heart rate fixed. While (ii) may be a more plausible control mechanism, the implication for optimum cadence is the same in both cases.

We may instead posit a cubic model for the power output / cadence relationship in place of the quadratic model in equation (1):

$$E(H) = h_0 + \alpha P + \beta C + \gamma C^2 + \delta C^3 + \varepsilon X$$

so that

$$P = \{E(H) - h_0 - \varepsilon X - \beta C - \gamma C^2 - \delta C^3\} / \alpha$$

Then

$$\frac{\mathrm{d}P}{\mathrm{d}C} = \frac{1}{\alpha}(-\beta - 2\gamma C - 3\delta C^2)\,.$$

So setting this derivative to zero and solving for C gives

$$C^* = \frac{-\gamma + \sqrt{\gamma^2 - 3\beta\delta}}{3\delta} = \frac{1}{3}\left(-\frac{\gamma}{\delta} + \sqrt{\left(\frac{\gamma}{\delta}\right)^2 - \frac{3\beta}{\delta}}\right).$$

Conditions for the existence of an optimum, positive cadence in this cubic case then follows from an analysis of the second derivative similar to in the quadratic case.

We also consider other explanatory variables or covariates that are themselves determined from the raw measurements. Note, no data other than the power output, heart rate and cadence measurements, the date of the session, and the athlete were available. Additional covariates are designed to measure medium and long term fatigue. Initially for the linear models we consider the following additional variables, for each athlete:

X_{1t} = the normalized power output developed in the current session (up to the current sampling point, *t*).

X_{2t} = the proportion of time that has elapsed in the current session, (up to the current sampling point, *t*).

X_3 = the mean overall power output of a session for sessions within the previous seven days.

X_4 = the mean duration of all sessions within the previous seven days.

X_5 = whether or not there was a session on the previous day (1 if there was a session, 0 otherwise).

These covariates aim to capture the possible effect of previous sessions on the current heart rate response and also the effect of the cumulative load of the current session. Normalized power output is calculated as Jobson et al. (2009). Individual power output measurements are raised to the fourth power, the mean calculated (up to the current sampling point, *t*), and then the fourth root of this mean is the normalized power output up to the current sampling point, X_{1t}. It is a measure of the cumulative training load in the session so far.

4.1.3. Initial non-linear models of power output and cadence

Alternatively if the linear power output cadence relationships do not reflect the true power output cadence relationship, or do not provde a close enough fit between power output and cadence to enough of an extent that would allow us to accept any resulting optimum cadences to be valid, a more sophisticated model could be required. A model that provides a non-linear relationship between power output and cadence could provide a more valid power output cadence relationship.

There are a number of characteristics of the relationship between power output and cadence that a more complex model should include. Firstly, when power output is equal to zero, cadence should also equal zero as the athlete is not riding at this point. For a given heart rate the power output-cadence curve should then increase, possibly at a linear rate - Grazzi et al, (1999) suggest it is initially a linear rate of increase before levelling to a maximum (which indicates that the athlete is riding at a cadence that maximises power output for his given heart rate - the optimum cadence). After the maximum, power output decreases, as at this point the athlete is riding at too great a cadence, at the detriment to his technique and ability to control the bicycle. We would expect the power output to decrease at a gradient less steep than the initial increase in the curve. The steeper the gradient of the decline in power output, the more sensitive the power output that the athlete can produce is to the cadence at which he rides. If the decline in power output after the maximum is not very steep, this implies that the power output changes relatively little with a change in cadence. This may in turn suggest there is a range of cadences over which power output is maximised rather than one specific cadence, or alternatively it could merely emphasise the level of precision involved in such a professional level of sport; a change in cadence may only effect a small change in power output of a few watts, but a difference of only a few watts may still have a significant impact on performance. A relatively flat curve may occur if the athlete does not vary his cadence much during riding, as a lack of variation in cadences would therefore mean there would be little scope for any model to detect an effect of cadence on power output.

If in the model the power output does not decrease at the highest cadences (in other words the maximum is not reached), it is possible the athlete has not reached the highest theoretical cadence he can sustain within that session for the given set of conditions.

The shape of the curve in a gamma distribution would satisfy the aforementioned conditions of the power output-cadence relationship. This can be achieved by setting power output to equal cadence in the following equation for each athlete:

$$P = C_t^{\alpha} e^{-\beta C_t} \tag{5}$$

at sampling point t, where α and β are two constants.

The first constant, α, refers to the shape of the curve, whereas the second constant, β, refers to the scale of the curve. This means that changes to alpha fundamentally change the relationship between power output and cadence (the shape of the power output/cadence curve produced), whereas changes in β do not change the shape of the curve but may stretch the curve (such that overall power output output and the cadence at which power output is maximised can change). If the shape constant α is greater than 1, the curve increases at an increasing rate continuously. If α is less than 1, the curve initially increases at a constantly decreasing rate.

The power output an athlete can produce, the cadence at which he rides, and the relationship between the two may vary with heart rate. Therefore heart rate must also be included in the analysis. Initially for the non-linear models we smooth power output, heart rate and cadence using a moving average over 60 seconds, sampled every 120 seconds, with heart rate lagged at 5 seconds behind power output and cadence. The data are separated into different subsets depending on heart rate, at intervals of every 10 beats per minute. For example, data in which heart rate is between 120 and 130 are treated as one data subset. The gamma curve is fitted for each data subset for each athlete, thus allowing for the exploration of the relationship between power output and cadence at different heart rates or different levels of sustained effort. Therefore the model will identify an optimum cadence for each heart rate group for each athlete.

Clearly there exist more data for some heart rate data subsets than others. The athlete will seldom ride with an extremely low or high heart rate, as the low heart rate may imply minimal effort is being exerted, and an extremely high heart rate will be very difficult to sustain. Indeed there are markedly more data in heart rate subsets between 70 and 170 than there are for heart rate data subsets above or below these figures for all athletes. Data in which heart rate is below 40 are not included as this is conceivably below the athlete’s baseline heart rate (resting pulse), but there are only negligible amounts of data for such a low heart rate so very little is omitted. Equally the negligible amount of data in which heart rate is above 200 is also omitted.

Table 3 indicates the size of each data set for each athlete.

There is also some variation in data set size between athletes. The heart rate subset with the most data is 140 – 150 for athlete 1, 110 - 120 for athlete 2, 140 - 150 for athlete 3, and 130 - 140 for athlete 4. For athletes 1, 2 and 4, the size of the data subsets tends to increase up to the maximum heart rate before decreasing again, but the data subset sizes are more sporadic for athlete 3.

Table 3: Data are split into groups based on the range of recorded heart rates (each group contains a range of 10 beats per minute), for each athlete separately and also for all four athletes together – sample sizes (n) are displayed for each data set.

HR range	Athletes 1 to 4	Athlete 1	Athlete 2	Athlete 3	Athlete 4
40 - 50	136	224	20	27	65
50 - 60	192	43	27	35	82
60 - 70	340	45	81	46	168
70 - 80	915	64	364	110	377
80 - 90	2006	104	803	35	754
90 - 100	3330	119	1335	696	1180
100 - 110	5071	532	1844	1007	1688
110 - 120	6094	1272	1986	773	2063
120 - 130	7171	2056	1876	802	2437
130 - 140	8663	3229	1539	1095	2800
140 - 150	9151	4594	1049	1463	2045
150 - 160	5065	2734	823	674	834
160 - 170	2992	908	300	303	1481
170 - 180	587	476	43	21	47
180 - 190	184	150	16	3	15
190 - 200	28	14	5	2	7
ALL	51787	16570	12194	7569	15453

4.1.4. Methodology for non-linear models of power output, heart rate and cadence

Next we develop non-linear models of power and cadence to include heart rate, rather than splitting the data into subsets based on heart rate. Note we continue to sample data once every 24 time units. We seek to include heart rate at some lag in the model (such that if an athlete increases his power output, his heart rate must increase in order to meet the demands of the increased power output), along with a coefficient (γ) to model this effect of heart rate on power output. Heart rate is multiplied by the cadence effect, and it is raised to a power of γ. Since heart rate is raised to the power of γ, this implies that if an athlete increases his power output, this would require a polynomial increase in heart rate (assuming γ is positive). For example, if γ is two, it implies that an increase in power output would require heart rate to increase by a factor of only a fourth (two squared) of that increase in power output. However we expect γ to be close to 1. A separate coefficient to model athlete effect is also included (μ). Power output can be modelled against athlete effect multiplied by cadence effect multiplied by heart rate effect, in the following equation:

$$P_t = \mu \times C_t^{\alpha} e^{-\beta C_t} \times H_{t+l}^{\gamma} \times e_t \qquad (6)$$

at sampling point t, where α, β, γ and μ are constants, and heart rate is lagged at time l.

For the model (equation 6), α, which determines the shape of the power output-cadence curve, must be positive, so that initial increases in cadence produce increases in power output. We would also expect β to be positive. Since an increase in power output should require an increase in heart rate, we would expect γ to be positive.

When fitting linear models we initially settle for a smoothing window of 60 seconds (k=12), and a heart rate lag of 5 seconds (l=1). We continue with these data processing constants (as we will explore the effect of different data processing constants later in section 4.6).

When fitting the non-linear model of power output, heart rate and cadence (equation 6) we transform the model by taking the natural log of each element of the model. This makes the model easier to fit in statistical packages. We transform the model in the following steps:

$$\log P_t = \log \mu + \log(C_t^{\alpha} e^{-\beta C_t}) + \gamma \log H_{t+l} + \log e_t$$

The cadence term is then broken down to become:

$$\log P_t = \log \mu + \log C_t^{\alpha} - \log e^{-\beta C_t} + \gamma \log H_{t+l} + \log e_t$$

which then becomes:

$$\log P_t = \log \mu + \alpha \log C_t - \beta C_t + \gamma \log H_{t+l} + \log e_t$$

with $\log e_t = \varepsilon_t \sim N(0, \sigma^2)$ (independent), at sampling point t, with heart rate lagged at l time measurements.

Now, the coefficient of $\log C_t$ is α, the coefficient of C is $-\beta$ (hence the negative of the coefficient of C gives β), the coefficient of $\log H_{t+l}$ gives γ, and the exponential of $\log \mu_i$ gives μ.

Later we add other training related variables. Summarised through the explanatory variable X, the model is given by the equation

$$\log P_t = \log \mu + \alpha \log C_t - \beta C_t + \gamma \log H_{t+l} + \delta \log X + \log e_t$$

with $\log e_t = \varepsilon_t \sim N(0, \sigma^2)$ (independent). This equation implies that the expected value of power output is related to the other variables by the following equation:

$$\overline{P}_t = E(P_t) = [\mu \times C_t^{\alpha} e^{-\beta C_t} \times H_{t+l}^{\gamma}] + \exp(\sigma^2 / 2) .$$

Note that additional covariates for the non-linear models are considered in section 4.1.6.

4.1.5. Methodology for calculating optimum cadence in non-linear models of power output, heart rate and cadence

For a specific athlete, the non-linear model of power output, heart rate and cadence is given by

$$P_t = \mu \times C_t^{\alpha} e^{-\beta C_t} \times H_{t+l}^{\gamma}$$

which is log-transformed to become

$$\log P_t = \log \mu + \alpha \log C_t - \beta C_t + \gamma \log H_{t+l} \tag{7}$$

We define the optimum cadence to be the cadence that maximises the power output for a given heart rate and therefore we imply that a cyclist will seek to ride in a manner that maximises power output for a given heart rate. We can calculate an optimum cadence statistically by differentiating equation (7) with respect to cadence, and equating the result to zero.

Firstly we differentiate equation (7) with respect to cadence to obtain:

$$\frac{\mathrm{d} \log \overline{P}}{\mathrm{d}C} = \frac{\alpha}{C} - \beta$$

Setting this to zero and solving for *C* gives an optimum cadence (C*) of:

$$C^* = \alpha / \beta$$

The optimum cadence *C** is a maximum provided the second derivative is negative:

$$\frac{\mathrm{d}^2 \log \overline{P}}{\mathrm{d}C^2} = -\alpha / C^{*2} < 0 \qquad (8)$$

For the second derivative in equation 8 to be negative, we require α>0. Since, $C^* = \alpha / \beta$ we also therefore require β>0. Note that, in the context of equation (7), this in fact means a negative cadence term.

We also expect γ (the coefficient relating to heart rate) to be positive. This indicates that, as the athlete increases his power output, his heart rate also increases. Note the γ is exponential – for example, if γ is 2, this means that, as the athlete increases his power output, he only has to increase his heart rate by the square root of the margin of increase in power output. However we expect γ to be approximately 1 (Grazzi et al, 1999).

We also seek confidence intervals for any optimum cadences found. Since C* is a function of two parameters (equation 3), we require the delta method to calculate the variance of C*. The variance of C* is as follows:

$$\operatorname{var} C^* = \frac{2}{\beta^3}(\alpha \operatorname{var}\beta - \beta \operatorname{cov}(\alpha, \beta))$$

We then calculate 95% confidence intervals for optimum cadence, in the following way:

$$C^* \pm 1.96 \times \sqrt{\operatorname{var} C^*}$$

We can now calculate a theoretical optimum cadence along with 95% confidence intervals for a range of data processing constants (window widths, *k*, and heart rate lags, *l*).

The confidence intervals allow us to ascertain by how much an athlete can deviate from optimum cadence without detriment to performance. The greater the variance of the optimum cadence, the greater the confidence intervals, and therefore the wider the range of cadences are at which the athlete can ride optimally. We will compare optimum cadences along with 95% confidence intervals for a range of models featuring different covariates, and different data processing constants. In other words we will explore the sensitivity of the models to the different variables and data processing constants.

4.1.6. Exploring the effect of heart rate on optimum cadence in non-linear models

The initial non-linear model of power output and cadence yields a range of different optimum cadences for sixteen different subsets of data. These subsets of data are based on heart rate (for example, one group contains data in which heart rate is between 110 and 120 beats per minute). Although there is likely to be some degree of variation in optimum cadence with heart rate, such trends may vary by athlete. It is also not clear from literature what effect heart rate has on optimum cadence. Therefore we investigate further the possible effect of the heart rate on optimum cadence.

We then split the sampled data into two subsets based on heart rate. For each athlete separately, we calculate the mean of all sampled heart rate measurements – we call this the overall mean heart rate for each athlete. These subsets of data are then: data containing heart rate measurements below the overall mean heart rate for each athlete (termed the low heart rate data subset) and above the overall mean heart rate for each athlete (termed the high heart rate data subset). Though heart rate is not a direct measure of effort, if heart rate is particularly high, the athlete is likely to be exerting more effort than if heart rate is particularly low. Furthermore, since an athlete seeks to optimise his own power output in competitions in order to minimise the time taken to complete a race, we may argue that it is more important to find a valid model to fit the high heart rate data subset than for the low heart rate data subset or indeed for models fitted to all sampled data. We believe data in the high heart rate data subset are likely to be more representative of race conditions. The low heart rate data subset may in turn contain data from long low-intensity training sessions in which the athlete is not focussing on maximising the power output he can sustain for a given period of time, or training sessions in which the athlete is conserving some energy for later sessions or for competitions. It also allows us to monitor the explanatory power of the models for low and high heart rate, to see how they compare to each other and to models in which all data are included. Note that these subsets of data are not the same as for the initial non-linear model of just power output and cadence (equation 5 in section 4.1.2.); we now simply have low and high heart rate data subsets, rather than having sixteen subsets.

If there are significantly different results in terms of optimum cadence or shape of the fitted power output cadence curves we may conclude that heart rate has a significant impact on optimum cadence. We may therefore consider an interaction between heart rate and cadence in the model. This involves replacing the cadence term in the model with a cadence – heart rate interaction. Thus for an individual athlete the non-linear model becomes the following (equation 9):

$$\log P_t = \log \mu + \alpha \log(C_t H_{t+l}) - \beta(C_t H_{t+l}) + \gamma \log H_{t+l} \qquad (9)$$

In order to calculate optimum cadence for this model (equation 9), we require plots of power output against cadence. Such plots allow us to explore the power output cadence relationship yielded by the model. By setting heart rate to the mean heart rate found in the data for each athlete, we plot the fitted power output from the model we are fitting. The highest point on this power output cadence curve corresponds to the cadence that produces the highest power output for a given heart rate – therefore the optimum cadence. We also plot power output measurements for comparison; if the fitted power output cadence curve passes through the power output measurements accurately, we can be confident that the model has been fitted correctly and confident about the validity of the optimum cadence yielded.

4.1.7. Additional variables in non-linear models

For non-linear models of power output, heart rate and cadence, we also consider additional variables, to measure further medium and long term aspects of an athlete's riding. In chapter 4 we consider the following additional variables for the non-linear models:

X_{1t} = the normalized power output developed in the current session (up to the current sampling point, t).

X_{2t} = the proportion of time that has elapsed in the current session, (up to the current sampling point, t).

X_3 = the mean overall power output of a session for sessions within the previous seven days.

X_4 = the mean duration of all sessions within the previous seven days.

We also investigate interactions between covariates, such as a cadence and normalized power output interaction. We then seek various optimum cadences for different amounts of normalized power output. (Later, in chapter 5, we present models in which we consider more variables that represent an accumulation of fatigue.)

For the linear models (equation 1), we add extra covariates to the model in various combinations. The addition of extra covariates is more complex for the non-linear models though. The initial non-linear models are simply an exploration into the possible modelling of a non-linear relationship between power output and cadence, and thus only include power output and cadence as covariates. We expande these models to include heart rate and athlete effects. We now add X_{1t}, X_{2t}, X_3 and X_4 but we do not include X_5 - this covariate denotes whether or not there was a session the previous day, and therefore is binary, with a value of either 0 or 1. Since the non-linear models are log-transformed so they are more easily fitted (the natural log of each element is taken), X_5 cannot be included since the log of 0 returns an error. Nevertheless we include all other extra covariates. We take the natural log of each extra covariate, so the covariate is then multiplied by the coefficient and added to the non-linear model (equation 7). We add each of the additional covariates X_{1t}, X_{2t}, X_3 and X_4 in various combinations, observing the explanatory power and Akaike Information Criterion (AIC) at each stage. Adding all of the additional covariates to the non-linear model produces the following model (equation 10):

$$\log P_t = \log \mu + \alpha \log C_t - \beta C_t + \gamma \log H_{t+l} + \delta \log X_{1t} + \varepsilon \log X_{2t} + \eta \log X_3 + \theta \log X_4 \quad (10)$$

For this model optimum cadence is calculated statistically, in the same way we calculated mathematical optimum cadence for previous non-linear models of power output, heart rate and cadence.

We also consider a cadence / heart rate interaction, with the same additional covariates. Hence such a model is the following (equation 11):

$$\log P_t = \log \mu + \alpha \log(C_t H_{t+l}) - \beta (C_t H_{t+l}) + \gamma \log H_{t+l} + \delta \log X_{1t} + \varepsilon \log X_{2t} + \eta \log X_3 + \theta \log X_4 \quad (11)$$

We also consider interacting some other covariates with cadence, to explore the precise effect of these variables on optimum cadence. We considered interacting cadence with heart rate, normalized power output (X_{1t}) and the proportion of the session that has elapsed up to the current sampling point (X_{2t}). We then seek a range of optimum cadences for different normalized power outputs and different proportions of the session that have elapsed up to the current sampling point. These interactions (C*X_{1t} and C*X_{2t}) are added to the models thusly (represented by paramter ι) in equations 12 an 13, whilst retaining the cadence effect (C) as we do not expect cadence to fundamentally depend on X_{1t} or X_{2t} as much as it would be affected by heart rate:

$$\log P_t = \log \mu + \alpha \log C_t - \beta C_t + \gamma \log H_{t+l} + \delta \log X_{1t} + \varepsilon \log X_{2t} + \eta \log X_3 + \theta \log X_4 + \iota \log(C_t X_{1t}) \quad (12)$$

$$\log P_t = \log \mu + \alpha \log C_t - \beta C_t + \gamma \log H_{t+l} + \delta \log X_{1t} + \varepsilon \log X_{2t} + \eta \log X_3 + \theta \log X_4 + \iota \log(C_t X_{2t}) \quad (13)$$

Initially we fit these more complex models featuring extra covariates and interactions to all sampled data. However for the simple grand non-linear model with no additional covariates (equation 7) we also performed separate analyses of the model for low heart rate data (data in which heart rate was lower than the mean of all recorded heart rate data for an individual athlete) and high heart rate data (data in which heart rate was higher than the mean of all recorded heart rate data for an individual athlete).

We consider the high heart rate data to be a closer representation of real competitions, as the low heart rate data in some training sessions may indicate the athlete is deliberately withholding some physical effort. Indeed for high heart rate data, optimum cadences were yielded for more athletes, and the models for high heart rate data appeared to better capture the real power output cadence relationship than did the models for all sampled data. Therefore we also fit non-linear models with additional covariates (models in equations 10, 11, 12 and 13) for high heart rate data. This allows us to explore how other covariates, such as training load, specifically affect athlete performance (the power output the athlete produces) during very high levels of physical exertion.

For models in which cadence is interacted with heart rate, training load or proportion of the session that has elapsed up to the current sampling point, we plot power output against cadence to find the optimum cadence. The plotted optimum cadence would then vary with different values of heart rate, normalized power output or proportion of session that has elapsed up to the current sampling point. We started to explore fitting non-linear models with extra covariates before we conducted the sensitivity analysis. When we started exploring non-linear models with extra covariates we settled on a set of data processing constants for fitting these models. This is the same as the non-linear models (without additional covariates) that we fitted before the sensitivity analysis, that is a smoothing window of width twelve measurements (k=12), a heart rate lag of 5 seconds or 1 measurement (l=1), and a sampling of one measurement every 120 seconds or 24 measurements (m=24).

4.1.8. Specific athlete effects

The effect of an athlete can be handled in a number of ways. Given the small number of athletes in the study, it is sensible to consider athlete effects as fixed. A study with a larger number of athletes might investigate a random effects model (McCulloch and Searle, 2001), with athlete effects then arising from a probability distribution. For the fixed effects model, the full model would consider all covariates effects and the baseline heart rate as athlete specific - that is, a set of coefficients in the models that are unique to each athlete.

4.1.9. Importance of optimum cadence

For each optimum cadence we find in the models tested, we can explore what happens when the athlete does not ride at the optimum cadence. Fitted power output will decrease as the athlete rides at cadences above and below the optimum cadence. The further the cadence is from the optimum, the lower the fitted power output will be. We would expect an optimum cadence to be towards the end of the range of cadences at which each athlete has ridden, as we are using highly skilled athletes that, according to literature, should be able to maintain relatively high cadences compared to less professional and less skilled athletes. For statistical optium cadences, we also calculate 95% confidence intervals. Also, by examining by how much the fittted power output decreases for sub-optimal cadences for a range of athletes and models, we can further explore how important an optimum cadence is to athlete performance.

4.1.10. Model selection

Models that contravene coefficient polarity expectations cannot be considered valid models, as a plausible power output cadence relationship will not follow from the model. In the linear models, in order for an optimum cadence to exist, the coefficient for cadence (β) must be negative and the coefficient for the square of cadence (γ) must be positive. Also, the optimum cadence ought to take a sensible value. For non-linear models, coefficients relating to cadence (both α and β) must be positive. When non-linear models are developed further, the coefficient relating to lagged heart rate (γ) should be positve so that increasing power output requires an increase in heart rate, whilst the coefficient relating to athlete effect (μ) should be negative in the additive model and positive in the multiplicative model.

Furthermore, the fitted models can be compared statistically. A standard criterion for model choice is the Akaike Information Criterion (AIC), which is defined in the following way: $AIC = -2\ln(L) + 2p$, where L is the log-likelihood value for the model and p is the number of parameters in the model (Stuart et al., 1999, p.748). The minimum AIC model is then sought. Comparisons of AIC between different linear models can be made as each model uses the same data and therefore all linear models have the same sample size. An alternative is the Bayesian Information Criterion (BIC), $BIC = -2\ln(L) + p\log(n)$, which places a higher penalty on the number of parameters in the model, and leads to more parsimonious models.

As the AIC and BIC are largely affected by sample size, when developing the initial non-linear models it is difficult to compare different models as each data set (each heart rate subset for each athlete) has a different sample size. If the non-linear gamma models produce a lower AIC compared to the linear models, we can conclude that the non-linear models are a better fit of the data than the linear models. We would expect such a finding, but it is difficult to predict how great the difference will be.

We also calulate explanatory power R^2 – this ostensibly represents the amount of variation in the response variable that is explained by the predictor variable. For the linear models this means the amount of variation in heart rate that is explained by power output. For the grand non-linear models, the explanatory power refers to the amount of variation in power output that is accounted for by variation in cadence and heart rate. The larger the R^2, the more variation is explained, and we may consider a larger R^2 therefore to indicate better the model. We would expect the explanatory power to increase with the addition of each additional covariate. However, if some covariates offer only negligible increase in explanatory power, we may consider using a more parsimonious model that does not necessarily include all covariates.

As R^2 is not affected by sample size, comparisons can be made more easily between different linear and non-linear models in terms of R^2 rather than AIC. We fit models in statistical packages SPSS and R. For models fitted in SPSS the R^2 is given in the output –

linear models are fitted in SPSS so R^2 is imediately displayed for these models. We fit non-linear models of power output and cadence in R. In R, for non-linear models, R^2 is not given, as it is not a very reliable indicator of model fit for non-linear models. This is because, in non-linear models, the total sum of squares is not equal to the regression sum of squares plus the residual sum of squares. However it can be calculated nonetheless. The residual standard error is provided from the output in R for each model (each heart rate group and each athlete). In order to calculate the explanatory power (R^2) of the model the residual error from a null model is also required – the null model is simply a model with just one parameter applied to the same data set: $P = \mu$ for each athlete. The explanatory power (R^2) is then calculated in the following way:

$$R^2 = 1 - (\text{residual standard error} / \text{total standard error}),$$

where total standard error is the addition of the residual standard error of the non-linear model and that of the equivalent null model.

We can then check that R^2 (along with coefficient estimates and standard errors) calculated from the R program is the same as that found from the same model fitted in SPSS.

For non-linear models (where we take natural logs of each element of the model to transform it), we can fit models in both R and SPSS, hence R^2 is provided in SPSS output.

Although R^2 indicates the amount of variation in the response variable that is accounted for by variation in the predictor variables, we cannot solely use it (or indeed other statistical measures AIC and BIC) to determine whether or not we are satisfied with a given model. R^2 indicates how closely a fitted curve from a model fits the real data, or how small are the differences between points on the fitted curve and the real data points. For a regression model to provide an accurate, reliable indication of optimum cadence, the fitted curve from the regression model must accurately reflect the true relationship between the variables involved – for this the coefficient estimates must have certain polarities (as discussed at the beginning of this section 4.1.9.). Nonethless we can use statistical indicators such as R^2 to compare similar models (with similarly shaped fitted curves), for example in exploring the extent to which additional variables are worth including in regression models. Therefore overall we use a combination of statistical measures (R^2, AIC and BIC, and coefficient estimates) and check that fitted curves from regression models are a realistic shape when assessing different models.

The power output cadence relationship of a model can be assessed through plotting the power output calculated from a regression model, along with the power output measurements in our sampled data, against cadence. We observe the extent to which the fitted power output from the model overlaps with power output measurements in our sampled data, and the plot provdes a guide as to whether the model is a valid fit of the data. Should the estimates of the coefficients be of the expected polarity the fitted power output cadence curve should be a valid shape, but the plot allows us to observe the fit in more detail, and provides a visual demonstration as to the power output candence relationship that the model produces. For fitted power output / cadence curves for non-linear models of power output, heart rate and cadence, heart rate must be fixed to an appropriate value for the curve to be fitted (illustrating the cadence that maximises power output for a given heart rate). In some more complex non-linear models (such as those featuring an interaction between cadence and additional variables) it may not be immediately obvious from observing coefficient polairities whether the model represents the true relationship between power output and cadence. In such cases however the extent to which the model represents the true power output / cadence relationship found in the data becomes clear when the fitted curve has been plotted.

4.2. Results of linear and initial non-linear models

4.2.1. Linear models results

We first fit linear models of power, heart rate and cadence. These are models in which power output is related to cadence by a quadratic or cubic function. However such models do not yield reliable optimum cadences. We therefore do not include results of linear models in this report.

4.2.2. Results of initial non-linear models of power output and cadence

The power output cadence relationship appears too complex for linear models. We now fit a model with the non-linear power output-cadence relationship in equation 5 (page 45). Initially we omit heart rate from the model as we are exploring the power output cadence relationship of a model. We split data into different groups based on heart rates. We then fit the non-linear model for each data set (each heart rate range for each athlete).

As with the linear models, we sample data once every 120 seconds (m=24), and smooth data over a 60 second moving average window (k=12), for a 5 second heart rate lag (l=1).

We present the values for coefficients α and β (from equation 5), along with their standard errors, in tables 4, 5, 6, 7 and 8 – these are firstly for all four athletes together in the same regression analysis (table 4), and then for each athlete separately (tables 5 to 8).

An α of between 1 and 2 and a very small β suggest that, were power output plotted against cadence, power output would increase to a maximum before decreasing at a lesser rate than the earlier increase. The data sets featuring all sampled data have the lowest standard errors. For data sets grouped by heart rate range the α and β values with the lowest standard errors do not necessarily correspond to the largest sample size, but in general standard errors for α and β tend to decrease as sample size increases. There do not appear to be many detectable patterns in the variation of α and β, except for a slight increase in α for data sets including a heart rate of 150-160 or 160-170.

Mathematical optimum cadences are yielded for all athletes in most heart rate subsets. We present these in section 4.6.1. (Section 4.6. is where we summarise mathematical optimum cadences yielded for all models fitted). The residual standard error is obtained from output in R. This is used to calculate the AIC and, along with the residual standard error of the one parameter null model, the explanatory power R^2. We present the residual standard errors of the non-linear models of power output and cadence, along with the equivalent one parameter null models, in the appendix. AIC and BIC increase with sample size, with BIC slightly higher than AIC except for very small sample sizes. When all four athletes and all sampled data are included, AIC = 232067. This is considerably lower than the quadratic and cubic models (featuring the same data processing constants k=12, l=1, m=24), where AIC was approximately 850000 depending on how many covariates were included, suggesting that the non-linear models generally capture the relationship between power output and cadence much better than the quadratic and cubic models.

Table 4: Estimates for α and β and their standard errors obtained from fitting a non-linear model of power output and cadence (equation 5) in R, along with sample size of each data set (n), for all four athletes.

HR range	α	s.e.	B	s.e.	n
40 - 50	1.36	0.078	0.019	0.0050	136
50 - 60	1.28	0.065	0.011	0.0042	187
60 - 70	1.44	0.039	0.029	0.0030	340
70 - 80	1.28	0.023	0.016	0.0014	915
80 - 90	1.27	0.015	0.013	0.0009	2006
90 - 100	1.33	0.012	0.013	0.0007	3330
100 - 110	1.45	0.010	0.018	0.0006	5071
110 - 120	1.55	0.010	0.022	0.0006	6094
120 - 130	1.59	0.009	0.022	0.0005	7171
130 - 140	1.62	0.008	0.022	0.0005	8663
140 - 150	1.66	0.008	0.023	0.0004	9151
150 - 160	1.76	0.010	0.027	0.0005	5065
160 - 170	1.62	0.017	0.018	0.0009	2992
170 - 180	1.49	0.032	0.012	0.0016	587
180 - 190	1.40	0.052	0.008	0.0026	184
190 - 200	1.28	0.210	0.002	0.0108	28
ALL	1.45	0.004	0.013	0.0002	51787

Table 5: Estimates for α and β and their standard errors obtained from fitting a non-linear model of power output and cadence (equation 5) in R, along with sample size of each data set (n), for athlete 1.

HR range	α	s.e.	β	s.e.	n
40 - 50	1.22	0.169	0.007	0.0236	24
50 - 60	1.27	0.130	0.006	0.0138	43
60 - 70	1.53	0.086	0.031	0.0082	45
70 - 80	1.33	0.054	0.015	0.0043	64
80 - 90	1.47	0.056	0.030	0.0048	104
90 - 100	1.36	0.038	0.021	0.0026	119
100 - 110	1.34	0.019	0.016	0.0011	532
110 - 120	1.33	0.012	0.013	0.0007	1272
120 - 130	1.36	0.010	0.013	0.0005	2056
130 - 140	1.41	0.008	0.013	0.0004	3229
140 - 150	1.53	0.006	0.018	0.0003	4594
150 - 160	1.60	0.008	0.021	0.0004	2734
160 - 170	1.52	0.017	0.016	0.0009	908
170 - 180	1.57	0.025	0.017	0.0012	476
180 - 190	1.59	0.043	0.017	0.0022	150
190 - 200	1.20	0.380	0.002	0.0191	14
ALL	1.38	0.005	0.011	0.0003	16570

Table 6: Estimates for α and β and their standard errors obtained from fitting a non-linear model of power output and cadence (equation 5) in R, along with sample size of each data set (n), for athlete 2.

HR range	α	s.e.	β	s.e.	n
40 - 50	1.36	0.181	0.009	0.0108	20
50 - 60	1.39	0.106	0.013	0.0084	27
60 - 70	1.15	0.069	0.001	0.0052	81
70 - 80	1.05	0.043	0.006	0.0035	364
80 - 90	1.27	0.024	0.017	0.0018	803
90 - 100	1.39	0.014	0.019	0.0009	1335
100 - 110	1.51	0.011	0.022	0.0007	1844
110 - 120	1.61	0.010	0.025	0.0006	1986
120 - 130	1.66	0.010	0.025	0.0006	1876
130 - 140	1.72	0.010	0.026	0.0007	1539
140 - 150	1.76	0.011	0.027	0.0007	1049
150 - 160	1.84	0.013	0.030	0.0008	823
160 - 170	1.73	0.025	0.022	0.0015	300
170 - 180	1.52	0.167	0.016	0.0093	43
180 - 190	1.19	0.255	0.005	0.0141	16
190 - 200	1.99	0.887	0.044	0.0500	5
ALL	1.57	0.006	0.020	0.0004	12194

Table 7: Estimates for α and β and their standard errors obtained from fitting a non-linear model of power output and cadence (equation 5) in R, along with sample size of each data set (n), for athlete 3.

HR range	α	s.e.	β	s.e.	n
40 - 50	1.27	0.108	0.013	0.0087	27
50 - 60	1.23	0.229	0.009	0.0169	35
60 - 70	1.34	0.147	0.018	0.0114	46
70 - 80	1.36	0.085	0.019	0.0055	110
80 - 90	1.29	0.040	0.013	0.0023	345
90 - 100	1.34	0.027	0.013	0.0014	696
100 - 110	1.44	0.020	0.016	0.0010	1007
110 - 120	1.50	0.024	0.017	0.0013	773
120 - 130	1.55	0.021	0.017	0.0011	802
130 - 140	1.49	0.019	0.013	0.0010	1095
140 - 150	1.51	0.015	0.013	0.0008	1463
150 - 160	1.64	0.021	0.019	0.0010	674
160 - 170	1.60	0.027	0.015	0.0014	303
170 - 180	1.07	0.127	0.011	0.0061	21
180 - 190	1.51	0.784	0.015	0.0404	3
190 - 200	1.51	0.784	0.015	0.0404	2
ALL	1.33	0.010	0.006	0.0005	7570

Table 8: Estimates for α and β and their standard errors obtained from fitting a non-linear model of power output and cadence (equation 5) in R, along with sample size of each data set (n), for athlete 4.

HR range			α	s.e.	β	s.e.	n
40	-	50	1.39	0.154	0.026	0.010	65
50	-	60	1.17	0.122	0.005	0.007	82
60	-	70	1.42	0.059	0.029	0.004	168
70	-	80	1.35	0.033	0.021	0.002	377
80	-	90	1.41	0.025	0.020	0.002	754
90	-	100	1.41	0.025	0.017	0.002	1180
100	-	110	1.50	0.025	0.020	0.002	1688
110	-	120	1.53	0.027	0.020	0.002	2063
120	-	130	1.57	0.028	0.020	0.002	2437
130	-	140	1.61	0.028	0.020	0.002	2800
140	-	150	1.63	0.034	0.020	0.019	2045
150	-	160	1.57	0.053	0.016	0.003	834
160	-	170	1.68	0.045	0.020	0.002	1481
170	-	180	1.44	0.174	0.007	0.009	47
180	-	190	0.56	0.246	0.036	0.012	15
190	-	200	1.25	0.291	0.003	0.015	7
		ALL	1.33	0.010	0.006	0.001	15453

Having obtained optimal values of α and β from R, we further explore the power output / cadence relationship produced from a plot of power output against cadence. We use the non-linear model (equation 5) to obtain a fitted power output for each cadence value in a given data set (for a given athlete and a given heart rate range). We present such plots, with both the original power output values and the fitted power output against cadence, in the appendix. We include plots for each heart rate group, for each athlete.

For data sets with extremely low or high heart rates there is little scope for a model to provide a clear power output/cadence relationship. However for the majority of data sets (where heart rate is in the range of 60-70 up to 160-170) the gamma model provides a power output cadence relationship that appears realistic. Initially there is a sharp increase in power output with cadence. This rate of increase decreases and in most data sets a maximum power output is reached, after which power output decreases much more slowly than it initially increased.

The non-linear curves clearly tend to capture the true power output cadence relationship better than the equivalent linear quadratic and cubic curves. For athlete 1, the quadratic curves captured the slight decrease in power output when cadence increases from 60 to 80, but not the power output cadence relationship at lower or higher cadences than this. After splitting data into different data subsets based on heart rate, this decrease in power output at cadences from 60 to 80 occurs when heart rate is equal to 140-150 and 150-160. In these data subsets the gamma curves not only capture this but also provide plausible power output cadence relationships for all ranges of heart rate.

4.3. Preliminary results of non-linear models of power output, heart rate and cadence

In section 4.4.1, we present preliminary results of non-linear models of power output, heart rate and cadence. (Later, in section 4.4, and chapters 5 and 6, we present results of more sophisticated non-linear models in which we find an optimum cadence more successfully.) In chapter 6, this includes models which also contain measures of fatigue. In section 4.4.1 we present results of non-linear models with varying amounts of success in yielding optimum cadences.

4.3.1. Initial exploration of non-linear models of power output, heart rate and cadence

Previously, in chapter 4.3, we illustrated results of non-linear models of power output and cadence. For these models, we split data into multiple subsets based on heart rate. For that model (equation 5), we split sampled data into multiple subsets (sixteen in total), each containing sampled data in which heart rate varied by 10 beats per minute (for example, sampled data in which heart rate is between 110 and 120 beats per minute formed one subset). Hence, we plotted fitted power output against cadence for various heart rates (as heart rate was not included as a variable in the model in equation 5). The gamma curve produced from the non-linear model of power output and cadence (equation 5) provided a reasonable representation of the power output / cadence relationship in our data. We therefore build upon this model (equation 5) by adding heart rate as a variable, to form the model in equation 7 (see section 4.1.4. for equation 7). By including heart rate as a variable, we have a neater model; we no longer have to consider the somewhat laborious process of fitting the same model to sixteen different subsets of data.

Initially we choose a fixed set of data processing constants, as we are firstly concerned with ensuring a valid non-linear model of power output, cadence and heart rate can be fitted. We investigate different data processing constants at a later stage (chapter 4.5), though literature suggests a low heart rate lag is most appropriate for our data. We first chose a set of data processing constants (k=12, l=1 and m=24) – a smoothing window of 12 time units (60 seconds), a heart rate lag of 1 time unit (5 seconds), and sampling data once every 24 time units (120 seconds). We fit models for each athlete separately. In this sub-chapter (4.4.1.) we present results of non-linear models of power output, cadence and heart rate (model in equation 7), for the fixed set of data processing constants (l=1, k=12).

Results from the initial non-linear model of power output and cadence (equation 5), in which we split data into multiple subsets based on heart rate, suggested there may be a different optimum cadence at different heart rates. It may be that the different coefficient estimates for each subset of data resulted in a slightly different curve shape each time and therefore a different curve peak each time.

We therefore explore the effect of heart rate here on optimum cadence in the non-linear model of power output, heart rate and cadence (equation 7). We fit this model (equation 7) in three ways. We sample all raw data (power output, heart rate and cadence measurements given to us at 5 second intervals) once every 24 time units. Firstly we fit this model (equation 7) for all of these sampled data. We then split the sampled data into two subsets based on heart rate. For each athlete separately, we calculate the mean of all sampled heart rate measurements – we call this the overall mean heart rate for each athlete. These subsets of data are then: data containing heart rate measurements below the overall mean heart rate for each athlete (termed the low heart rate data subset) and above the overall mean heart rate for each athlete (termed the high heart rate data subset). Note that these subsets of data are not the same as for the initial non-linear model of just power output and cadence (equation 5); we now simply have low and high heart rate data subsets, rather than having sixteen subsets. Though heart rate is not a direct measure of effort, the high heart rate data subset may represent actual race conditions more closely than do sampled data from all training sessions, as the training schedules include some long, low-intensity rides.

We present coefficient estimates for the non-linear model of power output, heart rate and cadence for all sampled data, low heart rate data subset and high heart rate data subset in table 9. For all athletes an increase in power output requires slightly less than the equivalent increase in heart rate, as γ is greater than 1. Athlete 2 requires the least heart rate increase (compared to other athletes) for a given increase in power output, as he has the highest γ. In fact an increase in power output for athlete 2 only requires a heart rate increase of a magnitude that is the square root of the power output increase (as γ is almost 2).

Table 9: Estimates for coefficients α, β, γ and μ, from the simple non-linear model (equation 7) for each athlete, with heart rate lagged at 1 (5 seconds), for all sampled data, the low heart rate data subset and the high heart rate data subset.

	α	β	γ	μ	log μ
All sampled data					
A1	0.94 (0.02)	-0.005 (0.0004)	1.21 (0.02)	0.0102	-4.59 (0.08)
A2	1.41 (0.03)	0.016 (0.0007)	1.98 (0.03)	0.0001	-9.22 (0.14)
A3	0.65 (0.04)	-0.001 (0.0008)	1.34 (0.03)	0.0156	-4.16 (0.15)
A4	0.21 (0.02)	-0.006 (0.0005)	1.41 (0.02)	0.0093	-4.68 (0.10)
Low heart rate data subset					
A1	1.00 (0.03)	-0.004 (0.001)	0.52 (0.03)	0.187	-1.68 (0.15)
A2	1.19 (0.04)	-0.010 (0.001)	1.47 (0.07)	0.002	-6.43 (0.30)
A3	0.62 (0.06)	-0.003 (0.001)	0.63 (0.06)	0.162	-1.82 (0.24)
A4	0.16 (0.04)	-0.009 (0.001)	1.08 (0.03)	0.028	-3.58 (0.16)
High heart rate data subset					
A1	1.77 (0.04)	0.020 (0.001)	1.56 (0.03)	0.0002	-8.78 (0.27)
A2	1.84 (0.03)	0.024 (0.001)	1.57 (0.06)	0.001	-7.35 (0.31)
A3	0.68 (0.24)	-0.0007 (0.003)	1.66 (0.11)	0.005	-5.80 (0.96)
A4	0.14 (0.01)	-0.008 (0.001)	1.51 (0.05)	0.13	-2.07 (0.24)

Different coefficient estimates are produced when data are split into low and high heart rates. For athletes 1 and 2 α is much higher for high heart rates, whilst for all athletes γ is higher for the high heart rate data subset – the higher γ suggests that when the athlete is likely to be exerting more effort (heart rate is higher), an increase in power output requires relatively less of an increase in heart rate than it would if the athlete were exerting less effort. It may be that there is less scope for heart rate increase at high heart rate as heart rate is closer to the maximum an athlete can attain. The explanatory power varies amongst the models, as shown in tables 32 and 33. We present the explanatory power, AIC and BIC in table 10.

Explanatory power is higher for each athlete for the non-linear model of power output, heart rate and cadence (equation 7) than it is for the non-linear models of power output and cadence (equation 5). However AIC is also higher. For athletes 1, 2 and 3 explanatory power has reduced for both low and high heart rate data subsets compared to the same models fitted for all sampled data. For athlete 1 the explanatory power for low heart rate data is only marginally less than for the model including all sampled data (41.2 compared to 43.8), but there is a greater decrease in explanatory power for the other athletes. There is greater explanatory power for models fitted to the low heart rate data subset compared to the high heart rate data subset for all athletes.

When we fit the model (in equation 7) to all sampled data, a mathematical optimum cadence is yielded for athlete 2. When we split the data into low and high heart rate data subsets, mathematical optimum cadences are yielded for athletes 1 and 2 for the high heart rate data subset but not for the low heart rate data subset. We present these mathematical optimum cadences in table 11.

Table 10: AIC and explanatory power for the non-linear model (equation 7) fitted for each athlete separately, with a heart rate lag (*l*) of 1 (5 seconds), for all sampled data, the low heart rate data subset and the high heart rate data subset.

	n	AIC	BIC	R^2
All sampled data				
A1	16570	122335	122374	43.8
A2	12194	99804	99841	51.6
A3	7570	55515	55549	14.9
A4	15453	114790	114829	20.7
Low heart rate data subset				
A1	6752	47153	47187	41.2
A2	5330	41299	41332	43.6
A3	3204	22365	22395	12.2
A4	5269	37121	37154	20.9
High heart rate data subset				
A1	9641	58178	58214	37.2
A2	6181	45672	45705	46.5
A3	4164	29760	29791	10.9
A4	6504	61043	61079	16.7

Table 11: Optimum cadence for the simple non-linear model (equation 7), in which heart rate is lagged at 5 seconds (*l*=1).

Athlete 1	Athlete 2	Athlete 3	Athlete 4
Optimum Cadence	Optimum Cadence	Optimum Cadence	Optimum Cadence
All sampled data			
*	88.28 (87.46, 89.09)	*	*
Low heart rate data subset – no optimum cadences			
High heart rate data subset			
90.48 (90.06, 90.90)	75.85 (75.33, 75.36)	*	*

We fit plots of power output against cadence for the non-linear model (equation 7), for all sampled data, the low heart rate data subset and the high heart rate data subset. Black dots are the sampled data, and the red lines are the fitted power output based on the model (equation 7). For the plot for the model (equation 7) fitted for all sampled data, we set heart rate to the mean of all sampled heart rate measurements – the overall mean heart rate for each athlete. Additionally for the plot for all sampled data, we illustrate fitted power output (from the model) for heart rate measurements 40 beats per minute below and above the overall mean heart rate for each athlete.

For the low heart rate data subset, we have only included data in which heart rate is less than this overall mean heart rate for each athlete. For fitted power output cadence plots for the low heart rate data subset, we set heart rate to the mean heart rate measurement contained within the low heart rate data subset. Equivalently for the high heart rate data subset, we have only included data in which heart rate is greater than the overall mean heart rate for each athlete. For fitted power cadence plots for the high heart rate data subset, we set heart rate to the mean heart rate measurement contained within the high heart rate data subset. The mean heart rates for the low and high heart rate data subsets, along with the overall mean heart rate for each athlete are presented in table 12.

Overall mean heart rate is slightly closer to the high average, and the margin of difference is similar for each athlete. For each subset (low and high heart rate data subset) the mean heart rate of that subset is similar for athletes 2, 3 and 4 but higher for athlete 1.

Fitted plots for the model (equation 7) are presented for all sampled data, and low heart rate data and high heart rate data subsets, in figures 8, 9 and 10 respectively, following table 12.

The curves pass through the power output measurements, with a reasonable curve shape for athletes 1, 2 and 4. For the plot for all sampled data, an optimum cadence is yielded for athlete 2 (88.3 rpm). The gradient of the curve either side of the maximum (for athlete 2) is not particularly steep, which may suggest that that athlete can ride at a range of cadences either side of the statistically optimum cadence without much detriment to performance. However 95% confidence intervals for the optimum cadence are very narrow (from 87.46 to 89.09 for athlete 2 for all sampled data), suggesting this is not the case, emphasising the importance of riding at an optimum cadence. The relatively flat curve is misleading; although changes in power output are minimal when the athlete deviates from optimum cadence, over the course of an entire race or session the changes in power output accumulate and could have a more significant impact on performance than the curves suggest.

For the plot for all sampled data, with increasing heart rate of 40 bpm each time, there is a steady increase in fitted power output of approximately the same magnitude for athletes 1, 3 and 4. For athlete 2 though the magnitude of increase in power output for each increase in heart rate is slightly higher. This is likely to be due to the higher estimate for γ for athlete 2 compared to the other athletes; athlete 2 does not have to increase heart rate by as much as the other athletes for the same power output increase – therefore when equivalent heart rate intervals are plotted for each athlete, athlete 2 appears to produce a higher power output for each heart rate increase compared to the other athletes.

Table 12: Mean heart rate measurements from all sampled data, and from low and high heart rate data subsets.

	Heart rate averages (mean)		
	Low heart rate data subset	All sampled data (overall mean for each athlete)	High heart rate data subset
A1	121	138	151
A2	97	117	136
A3	97	122	142
A4	98	121	139

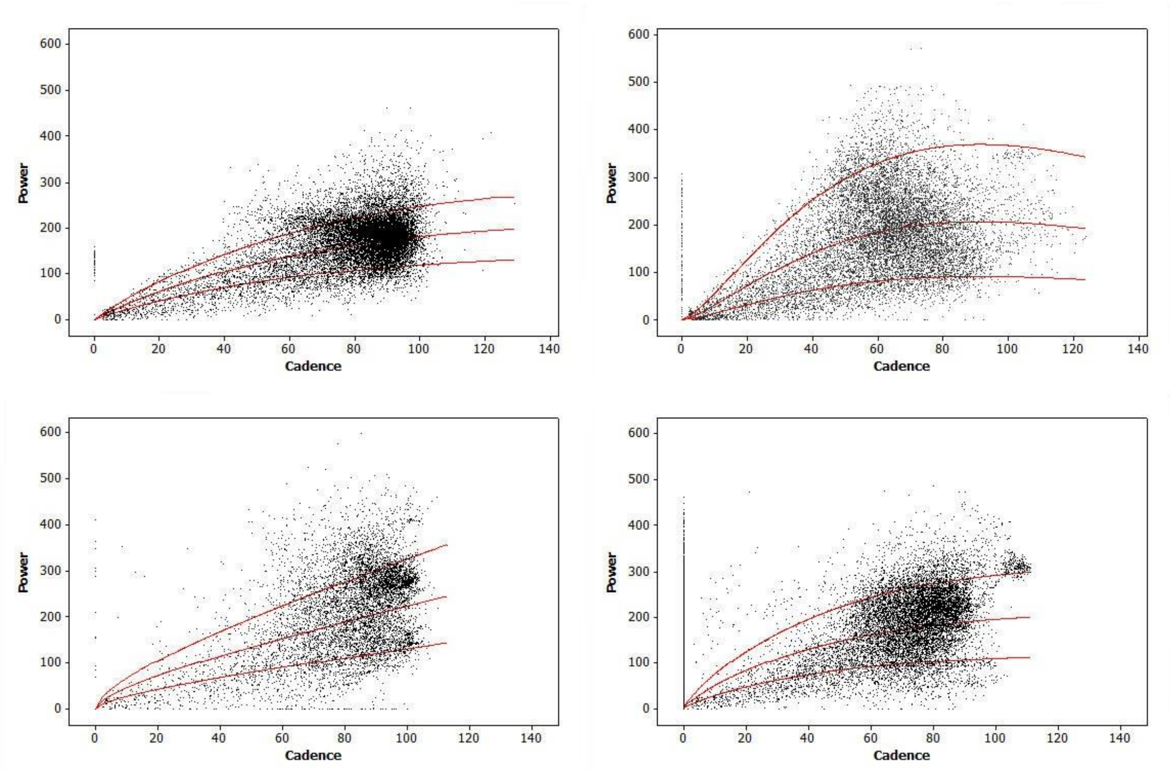


Figure 8: Fitted power output (red lines) from the simple non-linear model (equation 7) and power output measurements (black dots) against cadence, for athletes 1 to 4 (read lexicographically). Heart rate is lagged at 5 seconds, and the red fitted curves represent (bottom to top) heart rate 40bpm below mean, the mean, and 40 above mean.

For athlete 2 (the only athlete to have an optimum cadence when all sampled data are included in the regression model) cadence changes when data are split into the two heart rate subsets. For the non-linear model (equation 7), optimum cadence for all sampled data is 88.3, but it is 75.9 for the high heart rate data subset (there is no optimum cadence yielded for the low heart rate data subset for this athlete or indeed any athlete). As the higher heart rate may to some extent represent greater effort or intensity of training, the lower optimum cadence for the high heart rate data subset may suggest that the athlete cannot sustain such a high cadence when exerting great physical effort.

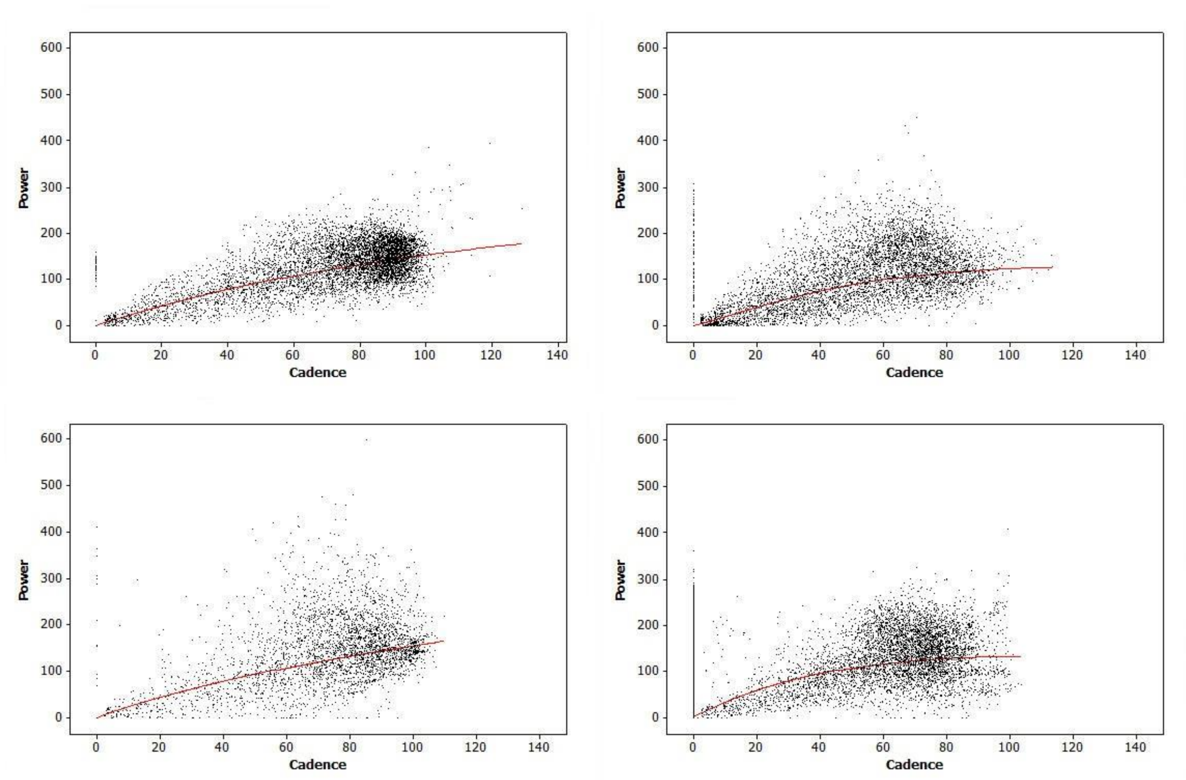


Figure 9: Fitted power output (red lines) from the non-linear model (equation 7) and power output measurements (black dots) against cadence, with heart rate lagged at 5 seconds, for athletes 1 to 4 (read lexicographically) – for the low heart rate data subset.

Confidence intervals for this model (equation 7) are also lower for the high heart rate data subset; for all sampled data confidence intervals are from 87.5 to 89.1, but for the high heart rate data subset this is lower, from 75.3 to 75.4. This suggests that the calculation of an optimum cadence may also be particularly important when the athlete is exerting great effort (which the high heart rate may represent). Precise calculation of optimum cadences could therefore be particularly important for intense training sessions, along with races and competitions, where the athlete is exerting more effort than in a less intense training session – in other words, for the most important competitions, cadence could therefore have a significant impact on performance, perhaps even more so than in training sessions.

Athlete 1 has an optimum cadence for the high heart rate data subset– a cadence of 90.5, with slightly wider confidence intervals than athlete 2 (from 90.1 to 90.9).

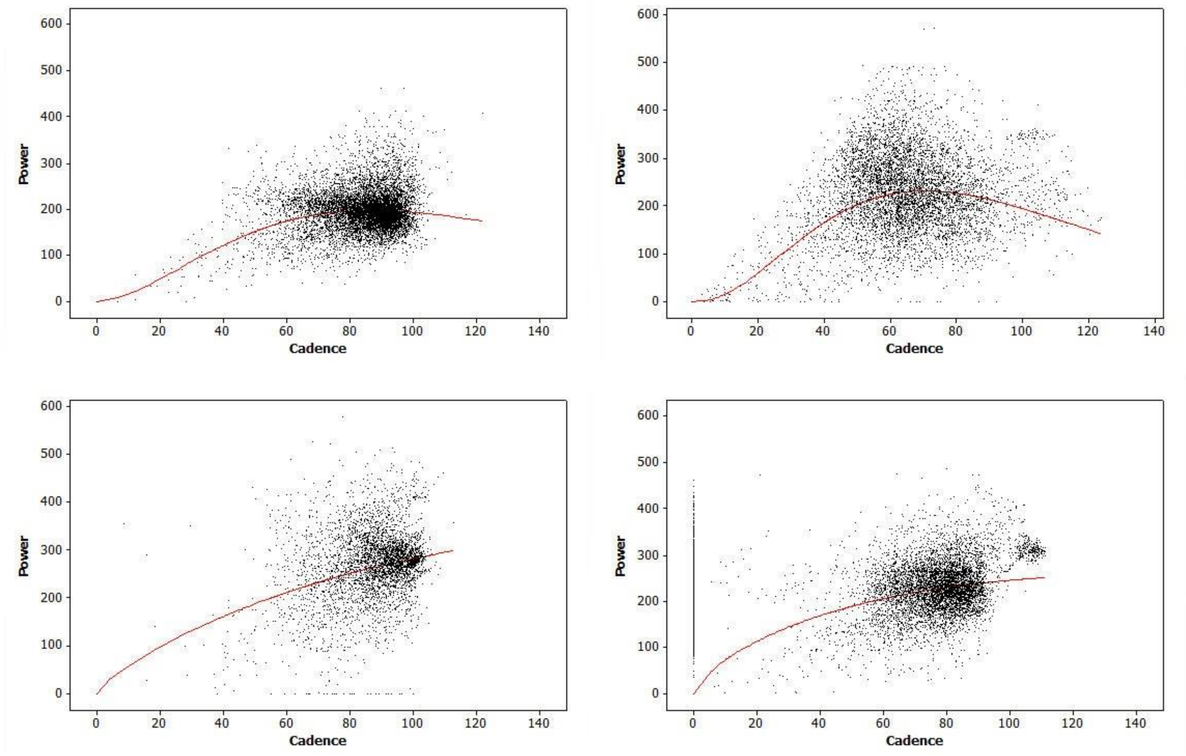


Figure 10: Fitted power output (red lines) from the non-linear model (equation 7) and power output measurements (black dots) against cadence, with heart rate lagged at 5 seconds, for athletes 1 to 4 (read lexicographically) – for the high heart rate data subset.

4.3.2. Nonlinear models of power output, heart rate and cadence with cadence / heart rate interaction

In chapter 4.4.1 we presented results of the non-linear model of power output, heart rate and cadence, fitted for data sampled once every 24 time units (m=24). We fitted this model for all sampled data, and also for two separate subsets of data base on heart rate. We calculated the mean of all sampled heart rate measurements, which we termed the overall mean heart rate for each athlete. The subsets of data are then: data in which heart rate is less than the overall mean heart rate for each athlete, and data in which heart rate is greater than the overall mean heart rate for each athlete (low and high heart rate data subsets respectively). Optimum cadence appears to be sensitive to heart rate to some extent; when we fitted the non-linear model (equation 7) to all sampled data an optimum cadence was yielded only for athlete 2, but for the high heart rate data subsets, optimum cadences were yielded for athletes 1 and 2. Next we consider interacting cadence with heart rate in the non-linear model of power output, heart rate and cadence (equation 9).

We fit this model with cadence / heart rate interaction (equation 9) using all sampled data, and also for low heart rate data and high heart rate data subsets separately. We present estimates for this model (equation 9) in table 13.

For model 24 fitted to all sampled data, for all athletes α has reduced compared to the simple non-linear model (equation 7). When we include all sampled data or only low heart rate data, athlete 2 has the greatest value for γ, meaning that when he increases his power output he does not have to increase his heart rate by as much of a margin as do the other athletes. Compared to the simple non-linear model (equation 7), γ is lower for all athletes. When we include all sampled data or only low heart rate data it is less than 1 for athletes 1, 3 and 4, meaning that when those athletes increase power output, this requires a disproportionately high increase in heart rate. In other words an increase in power output actually requires an increase in heart rate of a greater magnitude than the increase in power output. However this is possibly because, according to the model, heart rate has already increased when the athlete increases his cadence, since in this model (equation 9) cadence is interacted with heart rate.

Different coefficient estimates are yielded when data are split into low and high heart rate subsets. For athletes 1 and 2 α is much higher for high heart rate data subset, whilst for all athletes γ is higher when the model (equation 9) is fitted to the high heart rate data subset. The higher γ for the high heart rate data subset may suggest that when the athlete is exerting more effort or training intensity is greater, an increase in power output requires relatively less of an increase in heart rate than it would if the athlete were exerting less effort (as the higher heart rates may to some extent represent a greater effort or intensity of training than when low heart is particularly low). It may be that there is less scope for heart rate increase at high heart rate as heart rate is closer to the maximum an athlete can attain.

Table 13: Estimates for coefficients α, β, γ and μ, from the non-linear model with cadence / heart rate interaction (equation 9) fitted for each athlete separately, with heart rate lagged at 1 (5 seconds), for all sampled data, and low and high heart rate data subsets.

	α	β	γ	μ	log μ
All sampled data					
A1	0.62 (0.01)	-0.00002 (0.000002)	0.48 (0.02)	0.039	-3.23 (0.11)
A2	1.04 (0.02)	0.00006 (0.000004)	1.30 (0.04)	0.00004	-10.13 (0.20)
A3	0.47 (0.02)	-0.00005 (0.000004)	0.44 (0.04)	0.173	-1.76 (0.20)
A4	0.13 (0.02)	-0.00001 (0.000003)	0.78 (0.03)	0.031	-3.48 (0.14)
Low heart rate data subset					
A1	0.92 (0.03)	-0.004 (0.001)	0.40 (0.03)	0.295	-1.22 (0.17)
A2	1.09 (0.04)	-0.010 (0.001)	1.22 (0.07)	0.0009	-6.98 (0.33)
A3	0.52 (0.06)	-0.003 (0.001)	0.56 (0.06)	0.244	-1.41 (0.26)
A4	0.06 (0.04)	-0.009 (0.001)	0.89 (0.03)	0.103	-2.27 (0.19)
High heart rate data subset					
A1	1.77 (0.04)	0.020 (0.001)	1.26 (0.03)	0.0014	-6.56 (0.21)
A2	1.84 (0.03)	0.024 (0.001)	1.17 (0.06)	0.0005	-7.61 (0.34)
A3	0.58 (0.13)	-0.003 (0.002)	1.51 (0.06)	0.005	-5.28 (0.53)
A4	0.12 (0.05)	-0.005 (0.001)	1.45 (0.04)	0.011	-4.50 (0.29)

We present explanatory power, AIC and BIC for model 24 in table 14.

Explanatory power is lower than in the simple non-linear model not featuring the heart rate cadence interaction (equation 7). The explanatory power is lower for both low and high heart rate data subsets compared to models fitted to all sampled data, but to a much greater extent for the high heart rate data subset than for the low heart rate data subset.

We calculate optimum cadences for the model with cadence / heart rate interaction (equation 9) by plotting power output against cadence and finding the cadence that maximises power output. The inclusion of the cadence / heart rate interaction implies there should now be a different optimum cadence depending on what value heart rate is fixed to in the curve fitting. Therefore heart rate is set to a range of values in the plots. For the plot for all sampled data, we set heart rate to the overall mean heart rate for each athlete (the mean of all heart rate measurements in the sampled data for each athlete, and 40 beats per minute below and above this overall mean heart rate. The low heart rate data subset contains sampled data in which heart rate is less than the overall mean heart rate for each athlete. For plots for the low heart rate data subset we then set heart rate to the mean heart rate in this low heart rate data subset, and 20 and 40 beats per minute lower than the mean of this subset. The high heart rate data subset contains sampled data in which heart rate is greater than the overall mean heart rate for each athlete. For plots for the high heart rate data subset we then set heart rate to the mean heart rate in this high heart rate data subset, and 20 and 40 beats per minute greater than the mean of this subset. To summarise, the mean heart rates for each subset for each athlete are as follows, in table 15.

Table 14: AIC, BIC and explanatory power for the non-linear model (equation 9) fitted for each athlete separately, with a heart rate lag (*l*) of 1 (5 seconds).

	n	AIC	BIC	R^2
All sampled data				
A1	16570	122243	122275	41.4
A2	12194	99418	99452	48.7
A3	7570	55392	55436	13.5
A4	15453	113875	114920	20.4
Low heart rate data subset				
A1	6752	47150	47184	41.1
A2	5330	41372	41404	42.7
A3	3204	22336	22366	11.7
A4	5269	37202	37235	17.1
High heart rate data subset				
A1	9641	57638	57674	37.8
A2	6181	43489	43523	46.8
A3	4164	28564	28595	10.3
A4	6504	60382	60416	15.5

Table 15: Mean heart rate measurements from all sampled data, and from low and high heart rate data subsets.

	Heart rate averages (mean)		
	Low heart rate data subset	All sampled data (overall mean for each athlete)	High heart rate data subset
A1	121	138	151
A2	97	117	136
A3	97	122	142
A4	98	121	139

We now present the power output cadence plots for the model with cadence / heart rate interaction model (equation 9) in figures 11, 12 and 13.

When we fit model 24 to all sampled data, an optimum cadence is only yielded for athlete 2. For model 24, this is not a statistically calculated optimum cadence; rather it is inferred from plotting power output against cadence and finding the cadence that maximises power output. The optimum cadences for increasing heart rates (from 40 beats per minute below mean heart rate up to 40 above mean) are 226 rpm, 149, and 111 respectively. Clearly these cadences are unrealistic; only at extremely high heart rates (40 beats per minute above mean, or 191 beats per minute) is an optimum cadence yielded that is plausible (111 rpm). The model appears a poor fit for athlete 1 and 4 and particularly athlete 3, as the rate of increase in power output appears to accelerate markedly for all cadences.

Optimum cadences differ for different heart rate values in the cadence heart rate interaction model (equation 9). Optimum cadence is also lower for the simple non-linear model (equation 7) when this model is fitted to the high heart rate data subset (70.5 compared to 88.3 when all sampled data are included). This appears to suggest that increased heart rate significantly decreases the optimum cadence.

Similarly to the simple non-linear model (equation 7), optimum cadences are yielded for athletes 1 and 2 for the high heart rate data subset for the model with cadence / heart rate interaction (equation 9). These optimum cadences decrease with increasing heart rate. The margin of each decrease in optimum cadence decreases with heart rate – as heart rate increases, the optimum pedalling rate decreases, but by a lesser magnitude at each increase. Optimum cadences are lower than those found for all sampled data, again suggesting that optimum cadence decreases with increasing heart rate.

As expected, if the athlete is riding at a very high cadence, his power output reduces as he is unable to sustain such a high power output. However, at very high cadences, the fitted power output for the highest heart rate begins to decrease to a value less than that of fitted power outputs for lower heart rates. This is not plausible as the decreased heart rate for an increased power output implies here that the increased power output requires less physical exertion. However at such a high heart rate (40 beats per minute above the mean of 151 for athlete 1 is a very high 191 beats per minute), the athlete would be unlikely to ride at such a high cadence, and therefore the model need not be considered for such a high cadence when heart rate is very high. The model provides a reasonable fit of the data for most circumstances, and it indicates a clear decrease in optimum cadence as heart rate increases. (In section 4.4 and chapter 6, we include additional covariates to capture the effects of the accumulation of fatigue, and calculate how optimum cadence varies with respect to these additional variables.)

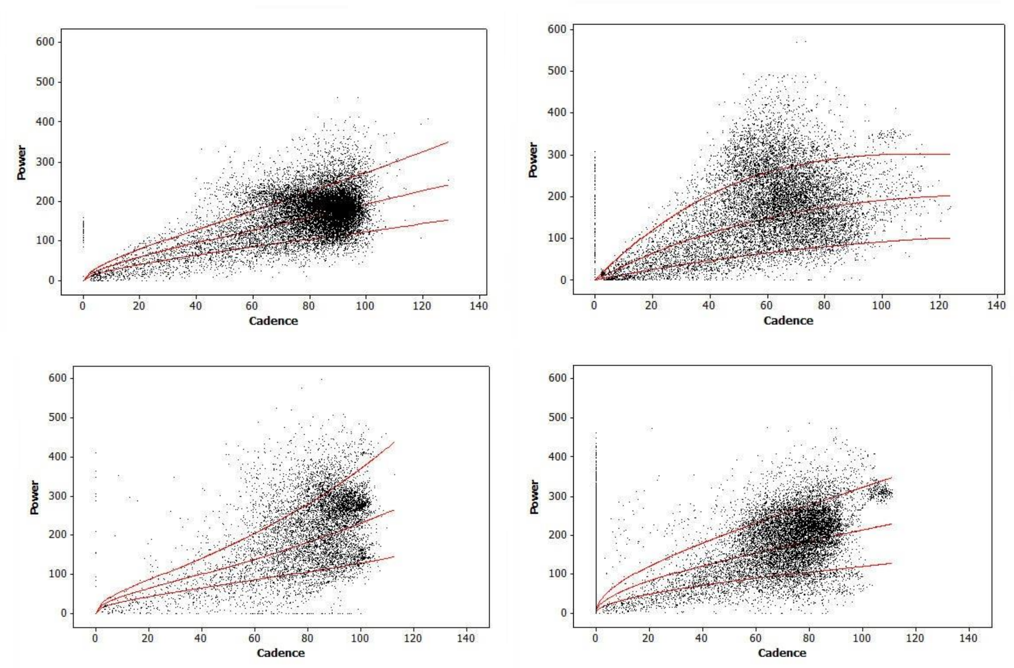


Figure 11: Fitted power output (red lines) from the non-linear model with cadence / heart rate interaction (equation 9) and power output measurements (black dots) against cadence, for athletes 1 to 4 (read lexicographically). Heart rate is lagged at 5 seconds, and the red fitted curves represent (bottom to top) heart rate 40bpm below mean, the mean, and 40bpm above mean.

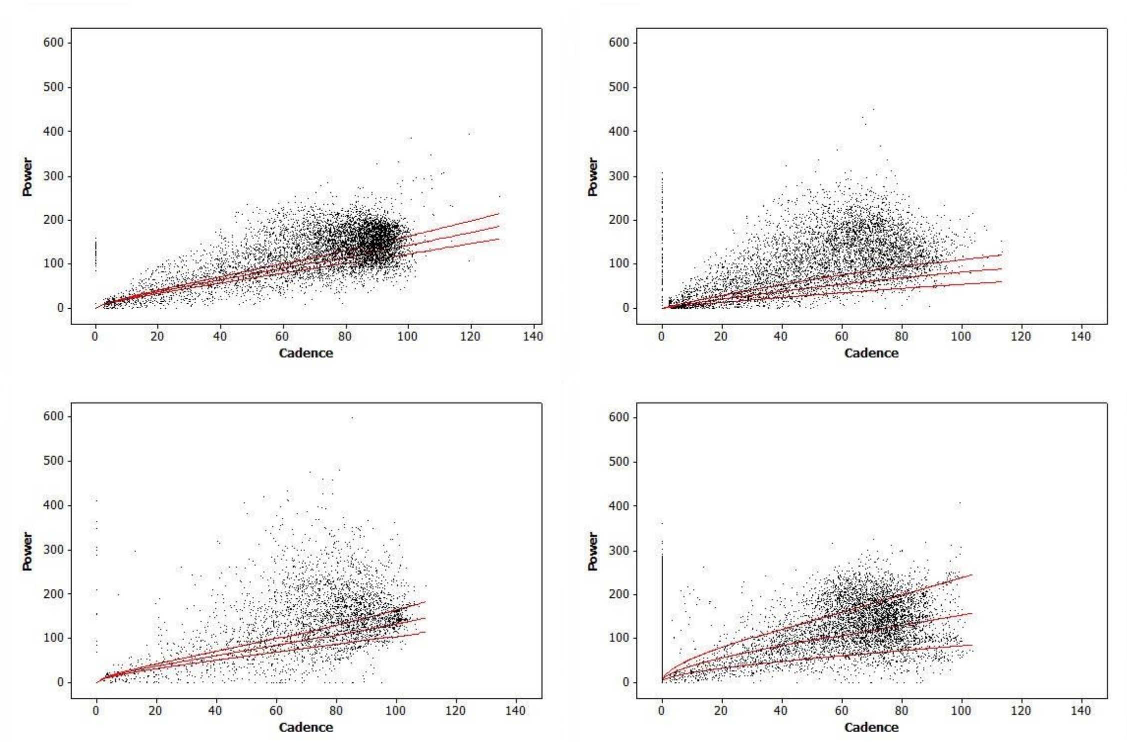


Figure 12: Fitted power output (red lines – from top to bottom, for mean heart rate, 20, then 40bpm below mean) from the non-linear model with cadence / heart rate interaction (equation 9) and power output measurements (black dots) against cadence, for athletes 1 to 4 (read lexicographically), for the low heart rate data subset. Heart rate is lagged at 5 seconds.

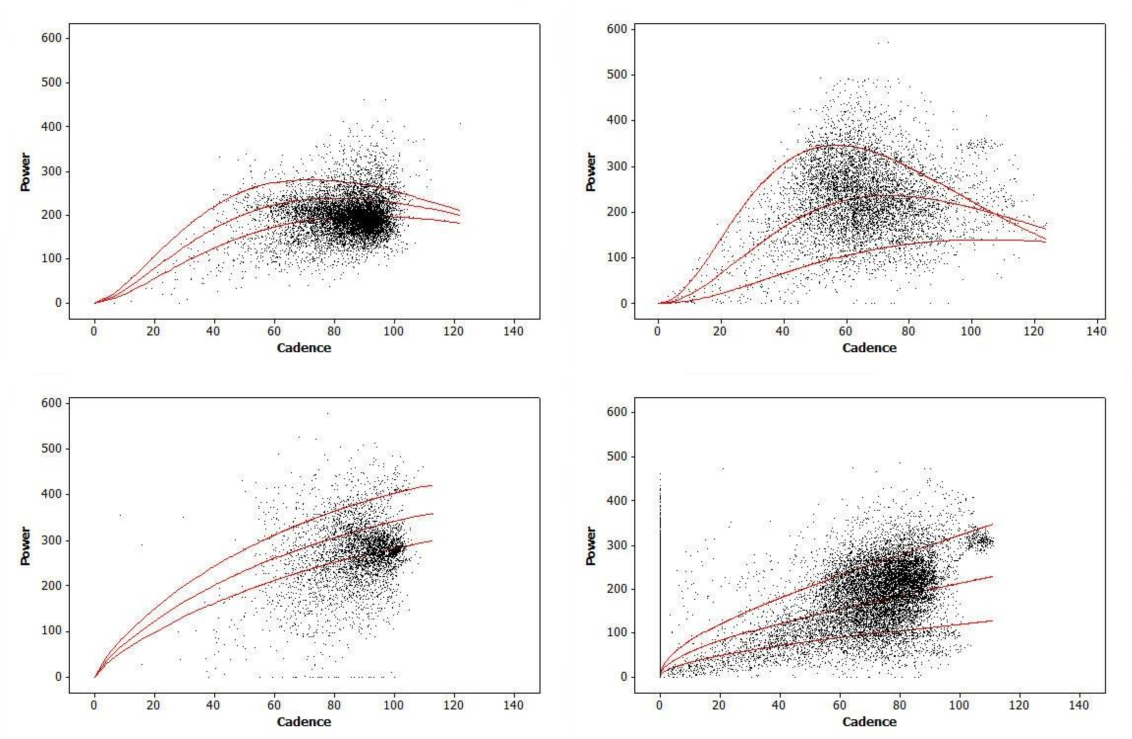


Figure 13: Fitted power output (red lines – from bottom to top, mean heart rate, 20, then 40bpm above mean)) from the non-linear model with cadence / heart rate interaction (equation 9) and power output measurements (black dots) against cadence, for athletes 1 to 4 (read lexicographically), for the high heart rate data subset. Heart rate is lagged at 5 seconds.

4.4. Results of non-linear models of power output, heart rate and cadence with additional covariates

4.4.1. Results of non-linear models with additional covariates but no interactions between variables

Initially when exploring non-linear models we include all sampled data (power output, heart rate and cadence measurements yielded at 5 second intervals which we sampled once every 24 time units). In this section we present such results first. Later we trim the data to focus on conditions in training sessions that may be more representative of race conditions. For each athlete we calculate the mean of all sampled heart rate measurements, which we term the overall mean heart rate for each athlete. Then we fit the same models but only including data in which the heart rate is higher than the overall mean heart rate for each athlete (which we term the high heart rate data subset) Though heart rate is not a direct measure of effort, we consider the high heart rate data subset to resemble race conditions more closely than do sampled data from all training sessions (as the training schedules include some very long low-intensity sessions).

We further develop the simple non-linear model of power output, heart rate and cadence. We add additional covariates that capture further aspects of an athlete's riding – X_{1t} (the normalized power output developed in the current session up to the current sampling point), X_{2t} (the proportion of time that has elapsed in the current session up to the current sampling point), X_3 (the mean of the normalized power outputs for all sessions in the previous seven days) and X_4 (the mean duration of all sessions in the previous seven days).

We add a range of combinations of the extra covariates to the non-linear model of power output, heart rate and cadence, producing the model in equation 10 when all extra covariates are included. In section 4.4. we present results of models in equations 10, 11, 12 and 13 - see section 4.1.7. for these equations.

We consider a heart rate lag of 1 time unit (5 seconds). Literature suggests a low heart rate lag is most appropriate for our data. Although it does not specifically suggest a 5 second lag, we settle for a lag of 5 seconds. Heart rate appears to change much more gradually in our data than in that of Stirling et al (2008), the study that implies a 60 second time lag for changes from resting heart rate to maximum heart rate in very brief exercise sessions. We settle on smoothing data using amoving average window of 12 time units (60 seconds). We continue to sample data one measurement every 24 time units (120 seconds). Hence models fitted in chapter 4.5 are fitted with data processing constants k=12, l=1 and m=24. We present the AIC and explanatory power of models with various combinations of extra covariates, in tables 16 to 19 for each athlete separately.

The normalized power output (X_{1t}) and the proportion of the session that has elapsed to the current sampling point (X_{2t}) broadly provide more explanatory power than do the remaining additional covariates. The proportion of the session that has elapsed up to the current sampling point (X_{2t}) increases the explanatory power the most, even more than the normalized power output (X_{1t}). Since X_{1t} and X_{2t} relate to the current session whilst X_3 and X_4 relate to previous sessions, it appears therefore that the power output the athlete can produce at the current sampling point is affected more by his riding in the session thus far than by his riding in previous sessions.

Adding the longer term covariates in addition to X_{1t} and X_{2t} further increases explanatory power a little.

Whether each covariate increases or decreases the power output the athlete can achieve, and the extent of this increase or decrease, are apparent from the coefficient estimates. We present coefficient estimates for each athlete in table 20.

Table 16: AIC, BIC and explanatory power for a range of models featuring additional covariates, fitted for athlete 1 (the sample size, n, is 16387).

Model	AIC	BIC	R2
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{l+1} + \delta\log X_{1t}$	120971	121035	44.1
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{2t}$	120940	120982	44.1
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_3$	120951	120993	43.9
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_4$	120981	120036	43.9
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t}$	120908	120947	44.0
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3$	120876	120914	44.0
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_4$	120907	120907	44.0
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3 + \eta\log X_4$	120878	120919	44.0

Table 17: AIC, BIC and explanatory power for a range of models featuring additional covariates, fitted for athlete 2 (the sample size, n, is 11480).

Model	AIC	BIC	R2
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{l+1} + \delta\log X_{1t}$	93956	93994	52.0
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{2t}$	93865	93903	52.2
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_3$	93955	93993	51.8
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_4$	93959	93997	51.8
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t}$	93865	93902	52.1
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3$	93863	93901	52.1
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_4$	93865	93902	52.1
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3 + \eta\log X_4$	93864	93904	52.1

Table 18: AIC, BIC and explanatory power for a range of models featuring additional covariates, fitted for athlete 3 (the sample size, n, is 7362).

Model	AIC	BIC	R2
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{l+1} + \delta\log X_{1t}$	53866	53904	15.3
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{2t}$	53868	53905	15.3
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_3$	53989	54027	15.0
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_4$	53990	54028	15.0
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t}$	53662	53697	16.8
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3$	53663	53696	16.8
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_4$	53663	53694	16.8
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3 + \eta\log X_4$	53664	53695	16.8

Table 19: AIC, BIC and explanatory power for a range of models featuring additional covariates, fitted for athlete 4 (the sample size, n, is 11770).

Model	AIC	BIC	R2
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t}$	112453	112492	21.2
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{2t}$	112116	112155	21.4
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_3$	112427	112465	20.9
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_4$	112385	112424	21.1
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t}$	111671	111710	22.6
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3$	111668	111708	22.6
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_4$	111630	111671	22.8
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3 + \eta\log X_4$	111626	111666	22.8

Athletes 1, 3 and 4 yield an α value of less than 1, whereas for athlete 2 as the α value is greater than 1 (1.42) meaning the power output increases with cadence at an increasing rate initially before steadying. All additional covariates have only a small impact on fitted power output (since estimates are very small for these additional covariates) but nevertheless there are differences between athletes in how the additional covariates affect power output. The coefficient for normalized power output (X_{1t}, the normalized power output developed to data) is positive for all athletes. Hence as normalized power output increases, rather than tiring, the athlete is able to increase his power output, possibly because the athlete is deliberately not exerting maximum effort for training sessions.

Table 20: Coefficient estimates for the non-linear model with extra covariates (equation 10), for athletes 1 to 4.

	Athlete 1 Estimate	Athlete 2 Estimate	Athlete 3 Estimate	Athlete 4 Estimate
α	0.945 (0.019)	1.425 (0.027)	0.641 (0.040)	0.285 (0.024)
β	-0.005 (0.0004)	0.016 (0.001)	-0.001 (0.001)	-0.005 (0.001)
γ	1.197 (0.020)	1.980 (0.030)	1.265 (0.032)	1.402 (0.019)
μ	0.01	0.00	0.01	0.01
$\log(\mu)$	-4.974 (0.091)	-9.613 (0.206)	-5.364 (0.164)	-4.683 (0.101)
δ	0.049 (0.009)	0.032 (0.026)	0.254 (0.018)	0.00008 (0.00001)
ε	-0.021 (0.003)	-0.056 (0.006)	-0.083 (0.006)	-0.156 (0.009)
η	0.014 (0.003)	0.009 (0.005)	0.004 (0.005)	0.007 (0.003)
θ	-0.005 (0.007)	0.007 (0.010)	0.005 (0.009)	-0.062 (0.007)

Estimates are negative for all athletes for the proportion of the session that has elapsed up to the current sampling point X_{2t}) – power output reduces as the athlete progresses towards the end of the session. Therefore the athletes may tire towards the end of the session; the further into the session the athlete is, the lower the power output that he can achieve. For athlete 4 this effect of proportion of time into the session is actually much greater than the normalized power output, as the estimate (along with standard error) is much higher (albeit negative) than that for normalized power output. The power output this athlete can achieve therefore is affected more by the session duration and how far into the session he is than by the calculated normalized power output. The power output and session duration averaged over the previous seven days do not present a particularly marked effect across athletes. For all athletes the power output he is able to produce increases with the mean output over the previous seven days. This again suggests the athletes are withholding some effort in some sessions, as the increased power output in previous days increases the power output he can achieve in the current session rather than inducing fatigue. For athletes 1 and 4 they have opposite polarity to each other.

The data appear to suggest that athletes fatigue towards the end of a given session (as estimates for X_{2t} are negative). However athletes are able to produce a higher power output when riding at a higher power output over the previous seven days. This could suggest the athletes are able to recover well after a tiring session.

Broadly the additional covariates yield similar estimates to the equivalent linear models of power output, heart rate and cadence (with the same set of additional covariates, also for a heart rate lag of 5 seconds). However it is difficult to compare the estimates in this respect as the linear models were based on heart rate (although both models regarding heart rate as a response to power output as both models consider heart rate lagged at some time after power output). This broad similarity suggests that the additional covariates affect the athlete's heart rate in a similar way to how they affect the power output he can currently achieve.

A mathematical optimum cadence is yielded for athlete 2, which we present in table 21.

We present fitted plots of power output against cadence for the non-linear model with extra covariates (equation 10) in the appendix, as the figures appear very similar to non-linear models of power output, heart rate, cadence and no additional covariates (equation 7).

Table 21: Optimum cadences for the non-linear model with additional covariates (equation 10), with heart rate lagged at 5 seconds (l=1).

Athlete 1	Athlete 2	Athlete 3	Athlete 4
Optimum Cadence	Optimum Cadence	Optimum Cadence	Optimum Cadence
*	87.70 (86.90, 88.49)	*	*

An optimum cadence is again yielded only for athlete 2. With extra covariates included (equation 10) optimum cadence for athlete 2 is 87.7, slightly less than it was for the non-linear model with no extra covariates (equation 7, with an optimum cadence of 88.3). Confidence intervals are still narrow (86.9 to 88.5), similar to the equivalent model without extra covariates (equation 7, for which confidence intervals were from 87.5 to 89.1). In other words, including other aspects of the riding, such as how far into a given session the athlete is, into the model, has had little impact on the optimum pedalling rate that the model yields.

After fitting non-linear models with and without the additional covariates, we decided to discard data which we believed to be unrepresentative of race conditions. We considered data in which recorded heart rate was quite low to be typical of training sessions in which the athlete is conserving energy, (something less likely to occur in actual races or competitions), or to represent long, low-intensity sessions, also not typical of race conditions. For each athlete we calculate the mean heart rate from the sampled heart rate measurements, which we term the overall mean heart rate for each athlete. We then discarded data in which recorded heart rate is lower than the overall mean heart rate for each individual athlete. The remaining data is then termed the high heart rate data subset. We therefore believe optimum cadences yielded from models featuring only high heart rate data may be more representative of the optimum pedalling rate an athlete may seek to ride at in actual race conditions. After splitting data into low and high heart rate data subsets, we found more plausible fitted power output cadence curves, and indeed optimum cadences for more athletes, for models fitted to the high heart rate data subset, compared to the low heart rate data subset or all sampled data. We consider low heart rate data is not significant in the investigation into the calculation of optimum cadence from field data, and focus instead on the high heart rate data subset.

We present results from the non-linear models fitted this time for the high heart rate data subset. These extra covariates are the following: X_{1t} (the normalized power output developed in the current session, up to the current sampling point, t), X_{2t} (the proportion of time that has elapsed in the current session, up to the current sampling point, t), X_3 (the mean of the normalized power outputs for all session in the previous seven days) and X_4 (the mean duration of all sessions in the previous seven days (hours)). We add extra covariates in various combinations to the simple non-linear model; adding all extra covariates to this model produces the model in equation 10.

We also fit models with cadence interaction terms. We consider interacting heart rate with cadence. Thus with all additional covariates included, producing the model in equation 11. We compare model fit of models in equations 10 and 11, in terms of AIC, explanatory power, covariate estimates and power output cadence plots.

We also explore the interaction of cadence with X_{1t}, the normalized power output developed in the current session, and X_{2t}, the proportion of time that has elapsed in the current session, up to the current sampling point, t. We consider such variables to have less of an impact on performance and indeed on optimum cadence than does heart rate, so we added the cadence interactions for these additional variables whilst retaining the solitary cadence effect in the model. Hence this produces the models in equations 12 and 13.

Explanatory power tends to increase with the addition of each explanatory variable. If the addition of a number of extra variables offers no improvement in model fit (through explanatory power or AIC) the principle of parsimony may be adopted. This would mean we choose not to include the extra covariates as we would choose to fit the smaller of two models with identical model fit. However it is unlikely that we will adopt the principle of parsimony as such as each covariate reduces AIC even if it does not increase explanatory power, and we still do not have a great number of extra variables in the models. Nevertheless we also present the explanatory power and AIC for several stages of adding covariates, to observe the extent to which each covariate improves model fit. We can compare the effects of extra covariates on model fit and athlete performance at high heart rate (when athletes are likely exerting great physical effort) and for all sampled data. We present the explanatory power and AIC for the addition of each extra covariates (added to the model in equation 7) in tables 22 to 25 for each athlete separately, for the high heart rate data subset.

Table 22: AIC, BIC and explanatory power for a range of models featuring additional covariates, fitted for athlete 1, for the high heart rate data subset. The sample size, n, is 9641.

Model	AIC	BIC	R2
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{l+1} + \delta\log X_{1t}$	57445	57484	37.4
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{2t}$	57549	57588	37.3
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_3$	57448	57487	37.4
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_4$	57418	57456	37.6
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t}$	57444	57483	37.7
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3$	57356	57394	37.7
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_4$	57334	57373	38.0
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3 + \eta\log X_4$	57308	57347	38.0

Table 23: Table 26: AIC, BIC and explanatory power for a range of models featuring additional covariates, fitted for athlete 2, for the high heart rate data subset. The sample size, n, is 6181.

Model	AIC	BIC	R2
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{l+1} + \delta\log X_{1t}$	43340	43377	47.1
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{2t}$	43340	43377	46.7
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_3$	43334	43371	46.6
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_4$	43330	43368	46.6
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t}$	43341	43376	47.2
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3$	43337	43374	47.2
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_4$	43333	43371	47.3
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3 + \eta\log X_4$	43333	43371	47.3

Next we present the explanatory power output and AIC for the model with cadence / heart rate interaction, with additional covariates added (in tables 26 to 29) for each athlete separately. When all extra covariates are added, this produces the model in equation 11.

Table 24: AIC, BIC and explanatory power for a range of models featuring additional covariates, fitted for athlete 3, for the high heart rate data subset. The sample size, n, is 4161.

Model	AIC	BIC	R2
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{l+1} + \delta\log X_{1t}$	24461	24492	11.3
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{2t}$	24557	24588	11.3
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_3$	24562	24594	11.1
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_4$	24564	24595	11.2
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t}$	24437	24469	11.9
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3$	24434	24465	12.0
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_4$	24427	24459	12.0
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3 + \eta\log X_4$	24427	24459	12.0

Table 25: AIC, BIC and explanatory power for a range of models featuring additional covariates, fitted for athlete 4, for the high heart rate data subset. The sample size, n, is 6504.

Model	AIC	BIC	R2
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{l+1} + \delta\log X_{1t}$	55517	55553	17.1
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{2t}$	55489	55524	17.5
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_3$	55562	55595	16.9
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_4$	55570	55604	17.0
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t}$	55210	55244	17.6
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3$	55206	55242	17.6
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_4$	55211	55245	17.6
$\log P_t = \log\mu_t + \alpha\log C_t - \beta C_t + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3 + \eta\log X_4$	55198	55234	17.7

With the exception of athlete 2, short term covariates (X_{1t} and X_{2t}) increase explanatory power more so than longer term covariates (X_3, X_4). For athlete 2 there is little difference in explanatory power between different extra covariates. Compared to the inclusion of all sampled data, when the models are fitted using only high heart rate data the explanatory power has reduced. However the explanatory power has increased with the addition of extra covariates compared to the models in equations 7 and 9, which were also fitted using only high heart rate data but did not include additional covariates.

Table 26: AIC, BIC and explanatory power for a range of models featuring additional covariates, fitted for athlete 1, for the high heart rate data subset. The sample size, n, is 9641.

Model	AIC	BIC	R2
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+l} + \delta\log X_{1t}$	57520	57556	37.9
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{2t}$	57638	57674	38.0
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_3$	57539	57576	37.9
$\log \mathrm{P}_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_4$	57509	57545	38.0
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t}$	57520	57556	38.1
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3$	57434	57470	38.1
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_4$	57413	57449	38.2
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{1t}\ \varepsilon\log X_{2t} + \zeta\log X_3 + \eta\log X_4$	57387	57423	38.2

Table 27: Table 26: AIC, BIC and explanatory power for a range of models featuring additional covariates, fitted for athlete 2, for the high heart rate data subset. The sample size, n, is 6181.

Model	AIC	BIC	R2
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+l} + \delta\log X_{1t}$	43488	43522	47.0
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{2t}$	43487	43521	47.0
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_3$	43482	43517	47.0
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_4$	43476	43513	47.1
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t}$	43489	43522	47.1
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3$	43484	43518	47.1
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_4$	43478	43515	47.2
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{1t}\ \varepsilon\log X_{2t} + \zeta\log X_3 + \eta\log X_4$	43478	43515	47.2

Table 28: AIC, BIC and explanatory power for a range of models featuring additional covariates, fitted for athlete 3, for the high heart rate data subset. The sample size, n, is 4164.

Model	AIC	BIC	R2
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+l} + \delta\log X_{1t}$	24461	24492	11.1
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{2t}$	24557	24588	11.1
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_3$	24562	24593	10.9
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_4$	24564	24595	10.9
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t}$	24437	24468	11.2
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3$	24434	24465	11.2
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_4$	24426	24457	11.2
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{1t}\ \varepsilon\log X_{2t} + \zeta\log X_3 + \eta\log X_4$	24427	24458	11.2

Table 29: AIC, BIC and explanatory power for a range of models featuring additional covariates, fitted for athlete 4, for the high heart rate data subset. The sample size, n, is 6504.

Model	AIC	BIC	R2
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+l} + \delta\log X_{1t}$	55504	55538	16.7
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{2t}$	55468	55502	17.0
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_3$	55547	55581	15.6
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_4$	55541	55573	15.8
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t}$	55288	55322	17.1
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_3$	55284	55318	17.1
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{1t} + \varepsilon\log X_{2t} + \zeta\log X_4$	55277	55311	17.1
$\log P_t = \log\mu_t + \alpha\log(C_t * H_{t+l}) - \beta(C_t * H_{t+l}) + \gamma\log H_{t+1} + \delta\log X_{1t}\ \varepsilon\log X_{2t} + \zeta\log X_3 + \eta\log X_4$	55258	55292	17.2

The extra covariates have a similar effect on explanatory power for high heart rate data as they did for all sampled data. However the mean session duration over the previous seven days (X_4) appears to yield a greater increase in explanatory power for high heart rate data than it did for all sampled data. For example, in the non-linear model with additional covariates (equation 10) for athlete 1, when all sampled data were included the covariates that yielded the highest explanatory power were jointly X_{2t} and X_3, whereas it is X_4 for the same model fitted using only high heart rate data. Therefore when heart rate is high and athletes are likely to be exerting great physical effort, the amount of time they have spent training in the last seven days has a relatively large effect on the power output he can achieve. This could represent a greater build up of fatigue at high heart rates, which is plausible. Despite the relatively high increase in explanatory power yielded from X_4, the proportion of the session that has elapsed up to the current sampling point (X_{2t}) does not yield such a high increase in explanatory power. For example, when the model with extra covariates (equation 10) is fitted using all sampled data to athlete 1, X_{2t} yields a similar explanatory power to the other extra covariates. However, when models in equations 10 and 11 are fitted using only high heart rate data, for athlete 1, X_{2t} yields the smallest increase in explanatory power of all the extra covariates.

There appears to be a long term effect of fatigue at high heart rates (resulting in the relatively high explanatory power of models that include X_4). However there does not appear to be a particularly great accumulation of fatigue within a given session (due to the relatively small increases in explanatory power when X_{2t} is added to the models).

We present the coefficient estimates for non-linear models with extra covariates (equations 10 and 11) for all athletes in tables 30 and 31.

Broadly for the model with additional covariates (equation 10) estimates for α, β, γ and μ are similar to the equivalent models not featuring extra covariates (that is, compared to the models in equations 7 and 9) that were also fitted for the high heart rate data subset. For athletes 1 and 2 α is greater than 1, implying an increase cadence (with power output) initially at an increasing rate. Thus far when α has been greater than 1, the power output cadence curve tends to reach a maximum (therefore producing an optimum cadence), but the curve has not reached a maximum when α has been less than 1. Therefore we expect there to be optimum cadences for athletes 1 and 2 again, and not for athletes 3 and 4. For athletes 1, 3 and 4 γ (the coefficient relating to heart rate) is lower when extra covariates are added. When extra covariates are included, the model (equation 10) suggests that when athletes ride at a higher power output, they now have to increase their heart rates by a greater amount compared to the model with no extra covariates. Athlete 2 however exhibits very similar estimates for α, β, γ and μ for models with or without extra covariates.

Table 30: Coefficient estimates for the non-linear model with extra covariates (equation 10), fitted for athletes 1 to 4 separately, for the high heart rate data subset, with a heart rate lag of 5 seconds.

	Athlete 1 Estimate	Athlete 2 Estimate	Athlete 3 Estimate	Athlete 4 Estimate
α	2.044 (0.067)	2.517 (0.053)	0.728 (0.130)	0.305 (0.049)
β	0.024 (0.001)	0.036 (0.001)	-0.001 (0.002)	-0.004 (0.001)
γ	1.417 (0.036)	1.672 (0.574)	1.184 (0.066)	1.310 (0.047)
μ	0.00005	0.00002	0.002	0.003
log (μ)	-9.891 (0.272)	-10.98 (0.385)	-6.416 (0.537)	-5.807 (0.295)
δ	0.125 (0.013)	-0.005 (0.038)	0.360 (0.031)	0.248 (0.021)
ε	-0.004 (0.003)	0.004 (0.006)	-0.038 (0.007)	-0.069 (0.005)
η	0.012 (0.002)	0.007 (0.005)	0.005 (0.004)	0.008 (0.002)
θ	0.040 (0.006)	0.027 (0.010)	0.022 (0.008)	-0.019 (0.006)

Table 31: Coefficient estimates for the non-linear cadence heart rate interaction model with extra covariates (equation 11), fitted for athletes 1 to 4 separately, for the high heart rate data subset, with a heart rate lag of 5 seconds.

	Athlete 1 Estimate	Athlete 2 Estimate	Athlete 3 Estimate	Athlete 4 Estimate
α	1.777 (0.061)	2.267 (0.049)	0.793 (0.119)	0.336 (0.045)
β	0.0001 (0.00001)	0.0002 (0.00001)	-0.00001 (0.00001)	-0.00002 (0.00001)
γ	1.364 (0.042)	1.492 (0.062)	0.602 (0.084)	0.896 (0.059)
μ	0.00000002	0.000000001	0.001	0.001
log (μ)	-17.72 (0.595)	-20.71 (0.560)	-7.477 (1.146)	-6.716 (0.516)
δ	0.134 (0.014)	0.023 (0.038)	0.360 (0.031)	0.252 (0.021)
ε	-0.004 (0.003)	0.007 (0.006)	-0.038 (0.007)	-0.069 (0.005)
η	0.012 (0.002)	0.008 (0.005)	0.004 (0.004)	0.008 (0.002)
θ	0.040 (0.006)	0.030 (0.041)	0.023 (0.008)	-0.018 (0.006)

Estimates for γ are lower for the cadence heart rate interaction model (equation 11) compared to the model with extra covariates but no interactions (equation 10), and is particularly marked for athletes 3 and 4. For the cadence heart rate interaction model therefore an increase in power output requires less of an increase in heart rate. This also occurred when extra covariates were not included in the models. This is possibly because in the cadence heart rate interaction model heart rate is assumed to increase when cadence increases, and therefore increases in cadence account for some of the increase in heart rate already. That this effect occurs for all four athletes further suggests that the interaction between cadence and heart rate would continue to yield lower estimates for γ for a range of different estimate combinations of coefficients.

Estimates for normalized power output (X_{1t}) and proportion of time that has elapsed in the current session up to the current sampling point (X_{2t}) vary between athletes, but again have opposite polarity to each other. For athletes 1, 3 and 4, the estimates for X_{1t} are positive and X_{2t} are negative. Therefore the power output the athletes can achieve increases with normalized power output and decreases towards the end of a session. This could indicate that the athletes are deliberately withholding some effort at first, and presumably tiring towards the end of the session.

However in the cadence heart rate interaction model (equation 11), the estimates for normalized power output (X_{1t}) are in fact positive for all athletes, whilst estimates for the proportion of the session that has elapsed up to the current sampling point (X_{2t}) remain negative for athletes 1, 3 and 4 and positive for athlete 2. Therefore for athlete 2 the power output he can achieve increases with both X_{1t} and X_{2t}. He does not appear to tire towards the end of a given session, and he may also be withholding effort for low normalized power outputs.

The interaction of cadence with heart rate, and indeed the absence of a measure of pure cadence, has changed the effect of normalized power output on fitted power output. Therefore there may be an effect of cadence on the power output produced at different normalized power outputs that is masked by the cadence heart rate interaction model (equation 11).

For the model with extra covariates but no interactions (equation 10), optimum cadences are yielded statistically for athletes 1 and 2, which we present in table 32.

Table 32: Optimum cadences for the non-linear model (equation 10), fitted for the high heart rate data subset. Heart rate is lagged at 5 seconds (l=1).

Athlete 1	Athlete 2	Athlete 3	Athlete 4
Optimum Cadence	Optimum Cadence	Optimum Cadence	Optimum Cadence
86.75 (86.29, 87.20)	70.60 (70.27, 70.94)	*	*

We present fitted power output cadence plots for non-linear models with extra covariates (equations 10 and 11) in the appendix, as these plots appear very similar to the equivalent plots for the non-linear model of power output, heart rate and cadence that does not include additional covariates (equation 7).

For all athletes, fitted power output overlaps well with the power output measurements. Optimum cadences are yielded for athletes 1 (86.8) and 2 (70.6). Confidence intervals for athlete 1 are from 86.3 to 87.2, and for athlete 2 they are from 70.3 to 70.9. Athlete 2 has a narrower range of cadences at which he can ride optimally compared to athlete 1. Riding at an optimal cadence appears slightly more crucial to performance for athlete 2 than for athlete 1, but confidence intervals are quite anrrow for both athletes, emphasising the importance of optimum cadence.

For athlete 2 however power output is very similar for both conditions (ie.when X_{1t} and X_{2t} are set ot zero or when they are set to mean values). It appears therefore that, when athletes are likely to be exerting a great amount of effort (that is, only high heart rate data are included), the extra covariates are less influential on performance for athlete 2 compared to the other athletes. There may be further evidence for this from the coefficient estimates and explanatory power. Indeed coefficient estimates are extremely similar for athlete 2 in particular between high heart rate models with or without extra covariates (i.e. between models in equations 7 and 10). Also, the relatively small increase in explanatory power for athlete 2 when extra covariates are added to the models further suggests that the power output athlete 2 can achieve is not greatly affected by normalized power output, proportion of time that has elapsed in the current session or covariates relating to performance over the previous seven days (mean output and session duration over the previous seven days). If athlete 2 therefore is able to produce similar power outputs for sessions of different durations, this may indicate he is withholding more effort in training sessions compared to other athletes.

We fit plots of power output against cadence for the cadence heart rate interaction model with extra covariates (equation 11), for the high heart rate data subset (data in which heart rate is greater than the mean heart rate for each athlete). We set heart rate to typical values for each athlete, based on the mean of the heart rate measurements included in the high heart rate data subset. We present these heart rates in table 33.

For the model with cadence heart rate interaction and extra covariates (equation 11) optimum cadences are calculated from the power output cadence plots (figure 14). Optimum cadence is the cadence that maximises power output – we present these optimum cadences in table 34.

Table 33: Range of heart rates considered for fitting power output cadence plots for the cadence / heart rate interaction model with extra covariate (equation 11), for the high heart rate data subset. The mean heart rate is the mean of all heart rate measurements included in the high heart rate data subset.

	Mean heart rate	Mean heart rate plus 20	Mean heart rate plus 40
A1	151	171	191
A2	136	156	177
A3	142	162	182
A4	139	159	179

Table 34: Optimum cadences for the non-linear cadence / heart rate interaction model (equation 11), for the high heart rate data subset.

	Athlete 1	Athlete 2	Athlete 3	Athlete 4
	Optimum Cadence	Optimum Cadence	Optimum Cadence	Optimum Cadence
H = mean	90.17	73.58	*	*
H = mean + 20	79.67	64.17	*	*
H = mean + 40	71.33	56.92	*	*

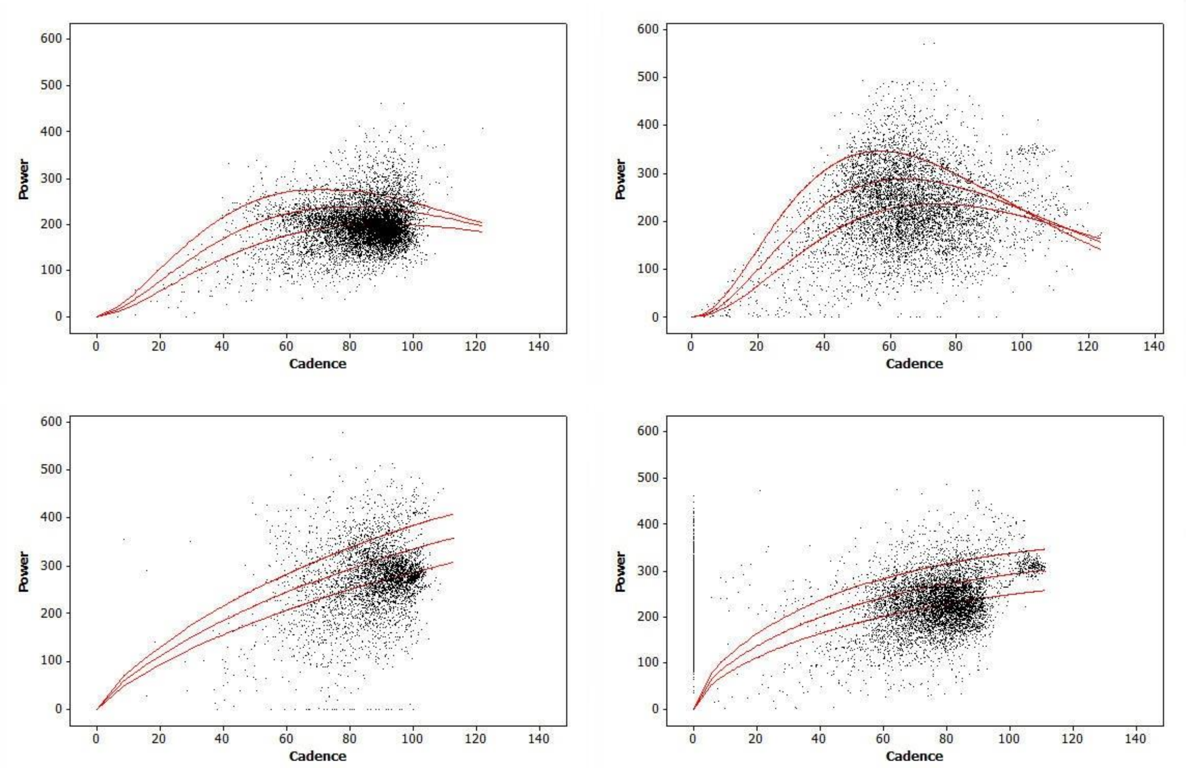


Figure 14: Fitted power output (red line) from the model with cadence / heart rate interaction and extra covariates (equation 11) and power output measurements (black dots) against cadence, for athletes 1 to 4 (read lexicographically), for the high heart rate data subset. Heart rate is lagged at 5 seconds. Heart rate and other covariates are set to the mean values found in the data for each.

4.4.2. Results of non-linear models with additional variables and interactions between variables

Next we further investigate the effects of normalized power output and proportion of the session that has elapsed up to the current sampling point (X_{1t} and X_{2t} respectively) by interacting them with cadence (models in equations 12 and 13).

We add the (C_t*X_{1t}) and (C_t*X_{2t}) interactions at the end of the model, to maintain an effect of cadence within the model. We fit these models (equations 12 and 13) for the high heart rate data subset only.

For the cadence heart rate interaction models, a plausible power output cadence relationship is yielded, though for cadences above 110, power output actually decreases as heart rate increases. Nevertheless even if this is not realistic, the model provides a realistic power output cadence relationship for most cadences; athletes seldom ride at over 110 rpm in the study.

We present the coefficient estimates for models in equations 12 and 13 in table 35.

There are some unexpected estimates in some circumstances. For example, α is not always positive. A positive α indicates that fitted power output (fitted against cadence) begins at zero and then increases, eventually at a decelerating rate. When α is negative, fitted power output is not zero at zero cadence – instead theoretically an infinity is produced. This is because the cadence of zero is raised to a negative number, hence α is divided by zero. Therefore the negative α does not provide as a good a fit of the data as does the positive α. In fact the same effect ocurs for a negative ι (the coefficient relating to (C_t*X_{1t}) or (C_t*X_{2t})). When ι is negative, and cadence or X_{1t} or X_{2t} are zero, they are divided by zero hence an infinity is produced. The negative α is yielded from the model in equation 12 for athlete 4 and the model in equation 13 for athlete 1.

A negative α implies a gamma curve that asymptotically approaches infinity as cadence reduces towards zero, with a theoretical value of infinity for zero cadence. However, whenever α is negative, ι is positive. Therefore when there is a negative effect of cadence on fitted power output, there is actually a positive effect for the cadence / normalized power interaction or cadence / proportion of the session that has elapsed up to the current sampling point interaction on power output. Therefore, whilst there is still a theoretical power output of infinity produced for a zero cadence, there is effectively both a positve and negative effect of cadence on fitted power output. The negative effect of one of the cadence variables is effectively cancelled out or nullified by the positive effect of the other cadence variable.

Indeed for the model with cadence / normalized power output interaction (equation 12), the coefficients for X_{1t} and (C_t*X_{1t}) tend to have opposite polarity to each other. The power output the athlete can produce is affected in different ways by the normalized power output and by the (C_t*X_{1t}) interaction – as they increase, one increases the power output he can produce whilst the other decreases the power output he can produce. Therefore normalized power output appears to have an effect on the power output the athlete can achieve that is independent of cadence. For the model with cadence interacted with proportion of the session that has elapsed up to the current sampling point (equation 13), there is also such opposite polarity (between estimates for X_{2t} and (C_t*X_{2t})). Estimates tend to be negative for X_{2t} and positive for (C_t*X_{2t}). As proportion of the session that has elapsed up to the current sampling point increases – that is, the further into the session an athlete is – the lower the power output he can achieve, possibly due to fatigue. However as cadence and proportion of the session that has elapsed up to the current sampling point both increase, the power output an athlete can achieve also increases. There appears therefore to be an effect of cadence that increases power output to such an extent that it negates the loss in power output that naturally occurs as the athlete progresses in the session.

Table 35: Coefficient estimates for the non-linear models with a cadence / normalized power output interaction (equation 12) and a cadence / proportion of the current session that has elapsed up to the current sampling point (equation 13), for the high heart rate data subset, with a heart rate lag of 5 seconds.

	Athlete 1 estimate	Athlete 2 Estimate	Athlete 3 Estimate	Athlete 4 Estimate
Cadence / normalized power output interaction (model in equation 12)				
α	2.492 (0.084)	0.349 (1.871)	1.266 (01.570)	-2.267 (1.157)
β	0.024 (0.001)	0.036 (0.001)	-0.0009 (0.002)	-0.004 (0.001)
γ	1.336 (0.037)	1.671 (0.056)	1.242 (0.065)	1.339 (0.046)
μ	0.0001	0.00002	0.002	0.003
$log\ (\mu)$	-9.984 (0.271)	-10.99 (0.385)	-6.402 (0.539)	-5.847 (0.295)
δ	0.608 (0.057)	-2.172 (1.870)	0.902 (1.575)	-2.725 (1.159)
ε	-0.004 (0.003)	0.004 (0.006)	-0.038 (0.007)	-0.069 (0.005)
η	0.011 (0.002)	0.007 (0.005)	0.005 (0.004)	0.008 (0.002)
θ	0.038 (0.006)	0.027 (0.010)	0.022 (0.008)	-0.019 (0.006)
ι	-0.423 (0.048)	2.169 (1.871)	-0.542 (1.575)	2.973 (1.159)
Cadence / proportion of the current session that has elapsed up to the current time point interaction (model in equation 13)				
α	-0.911 (0.699)	0.349 (1.871)	1.266 (1.570)	0.689 (0.053)
β	0.024 (0.001)	0.036 (0.001)	-0.0009 (0.002)	-0.004 (0.001)
γ	1.416 (0.036)	1.671 (0.056)	1.230 (0.066)	1.345 (0.048)
μ	0.0001	0.00002	0.002	0.003
$log\ (\mu)$	-9.983 (0.273)	-10.99 (0.385)	-6.402 (0.539)	-5.798 (0.295)
δ	0.125 (0.013)	-0.003 (0.038)	0.360 (0.031)	0.248 (0.021)
ε	-2.988 (0.702)	-2.165 (1.871)	0.504 (1.575)	-0.078 (0.013)
η	0.012 (0.002)	0.007 (0.005)	0.005 (0.004)	0.008 (0.002)
θ	0.040 (0.006)	0.027 (0.010)	0.022 (0.008)	-0.018 (0.006)
ι	2.984 (0.702)	2.169 (1.871)	-0.542 (1.575)	0.009 (0.013)

Athlete 2 exhibits very little difference in coefficient estimates between the two different interactions ((C_t*X_{1t}) and (C_t*X_{2t}), in models in equations 12 and 13 respectively). This appears to be further evidence that the power output this athlete is able to produce is only marginally affected by other covariates relating to his performance, compared to other athletes who are affected to a greater extent.

The coefficient estimates for the models including (C_t*X_{1t}) and (C_t*X_{2t}) interactions (equations 12 and 13 respectively) vary in the appparent validity of the model fit. We now present the explanatory power and AIC of these models in table 36. For those models with positive α (which is required for valid fit of the power output cadence relationship), we can observe trends in explanatory power. For athletes for whom α is negative (and therefore the model does not provide a valid representation of the power output cadence relationship), the results are italicised and asterisks added.

There is minimal difference between the two models (equations 12 and 13). There is also only a marginal increase in explanatory power for models 12 and 13 compared to the other non-linear models with extra covariates (equations 10 and 11).

When α is negative (italicised rows in tables 36) the fitted power output cadence relationship does not necessarily provide a valid representation of the true power output cadence relationship. For athlete 1, α is positive for the model with cadence interacted with normalized power output (equation 12) but negative for the model with cadence interacted with proportion of the session that has elapsed up to the current sampling point (equation 13). Given that these models (equations 12 and 13) yield approximately the same explanatory power, we might expect a lower explanatory power for the model in equation 13 with the negative coefficient estimate for α, but in fact AIC is slightly higher for the model in equation 13 than for the model in equation 12. This may mean that model fit is equally poor for both of these models, or that there is an extreme amount of noise in the data. Athlete 2 is not affected much by extra covariates. The explanatory power and even the AIC and BIC are very similar for both models (in equations 12 and 13).

Table 36: Explanatory power (R^2), AIC and BIC for the non-linear model with cadence / normalized power output interaction (equation 12) and cadence / proportion of the current session that has elapsed up to the current sampling point (equation 13) for the high heart rate data subset, with heart rate lagged at 5 seconds.

	N	AIC	BIC	R2
Cadence / normalized power output interaction (model in equation 12)				
A1	9641	57225	57261	38.3
A2	6181	43326	43359	47.4
A3	4164	24421	24453	11.3
**A4*	**6504*	**55587*	**55621*	**17.2*
Cadence / proportion of the current session that has elapsed up to the current time point interaction (model in equation 13)				
**A1*	**9641*	**57284*	**57319*	**38.2*
A2	6181	43326	43359	47.4
A3	4164	24421	24453	11.4
A4	6504	55596	55630	17.2

When cadence is interacted with normalized power output (X_{1t}) or porportion of the session that has elapsed up to the current sampling point (X_{2t}), the optimum cadence depends on the values of X_{1t} and X_{2t}. For these models (in equations 12 and 13), we calculate optimum cadence by plotting fitted power output against cadence for a range of cadences that the athletes have ridden at in their training sessions. The range of values found in the data for normalized power output varies by athlete. The mean values for normalized power output are 1113, 1370, 1568 and 1293 for athletes 1, 2, 3 and 4 respectively. We include only high heart rate data as we seek to find an optimum cadence for conditions that are reasonably close to actual competitions or races. Similarly when we consider normalized power output, we seek an optimum cadence for training conditions that are a reasonably close match to the normalized power output from a competition. Therefore we do not consider finding an optimum cadence for particularly low normalized power outputs. Instead, we consider a range of normalized power outputs that begin slightly below mean, in intervals up to the maximum recorded normalized power output for each athlete. We consider the following ranges of normalized power output, in table 37.

Table 37: Range of values normalized power output (X_{1t}, normalized power output up to the current sampling point) is set to for yielding optimum cadence in the model in equation 12, for each athlete.

Athlete	Mean	Maximum	Range			
1	1113	2041	1000	1300	1700	2000
2	1370	2218	1200	1500	1800	2100
3	1568	2646	1500	1850	2200	2550
4	1293	2814	1200	1700	2200	2700

The proportion of the session that has elapsed up to the current sampling point is simply a proportion ranging from 0 to 1. We set this to intervals of 0.3 – that is 0.1, 0.4, 0.7 and 1, to give a broad spread of different times or sections of a session.

We present optimum cadences in tables 38 and 39. Optimum cadences for models in equations 12 and 13 are not affected by the interactions – they are the same for each level of normalized power output (X_{1t}) or proportion of the session that has elapsed up to the current sampling point (X_{2t}).

Table 38: Optimum cadences for the non-linear model (equation 12), for the high heart rate data subset.

	Athlete 1		Athlete 2
Normalized Power output (X_{1t}) for Athlete 1	Optimum Cadence	Normalized Power output (X_{1t}) for Athlete 2	Optimum Cadence
1000	86.58	1200	70.58
1300	86.58	1500	70.58
1700	86.58	1800	70.58
2000	86.58	2100	70.58

Table 39: Optimum cadences for the non-linear model (equation 13), for the high heart rate data subset.

	Athlete 1		Athlete 2
Proportion of session that has elapsed (X_{2t}) for Athlete 1	Optimum Cadence	Proportion of session that has elapsed (X_{2t}) for Athlete 2	Optimum Cadence
0.1	86.50	0.1	70.58
0.4	86.50	0.4	70.58
0.7	86.50	0.7	70.58
0.9	86.50	0.9	70.58

We present plots of expected power output against cadence for models in equations 12 and 13 (in figures 15 and 16 respectively). Whenever there is a negative value for ι (the coefficient for the ($Ct*X_{1t}$) or ($Ct*X_{2t}$) interaction in models in equations 12 and 13 respectively), the fitted power output cannot be calculated for zero normalized power output or when no time has elapsed in the session, as we would have to raise zero to a negative power output. Therefore we begin the range of values for X_{1t} and X_{2t} for the curve fitting above zero. For athlete 2, fitted power output varies very little with either covariate interaction (models in equations 12 and 13). For athlete 2 fitted power output decreases with normalized power output but increases with the proportion of the session that has elapsed up to the current sampling point, whereas the opposite occurs for the remaining athletes. For athlete 2 therefore it appears that he does not fatigue within a typical session. We do not know how normalized power outputs are distributed within a session; high normalized power outputs could tend to be concentrated at the end of a session, with the athlete building up to an intense finish to the session – hence an effect of X_{2t} could potentially also represent an effect of normalized power output (X_{1t}). For the remaining athletes (1, 3 and 4) fitted power output appears to vary more with proportion of the session that has elapsed up to the current sampling point (X_{2t}) than it does with normalized power output (X_{1t}). This indicates that how far into a session the athlete is could be more important than is normalized power output.

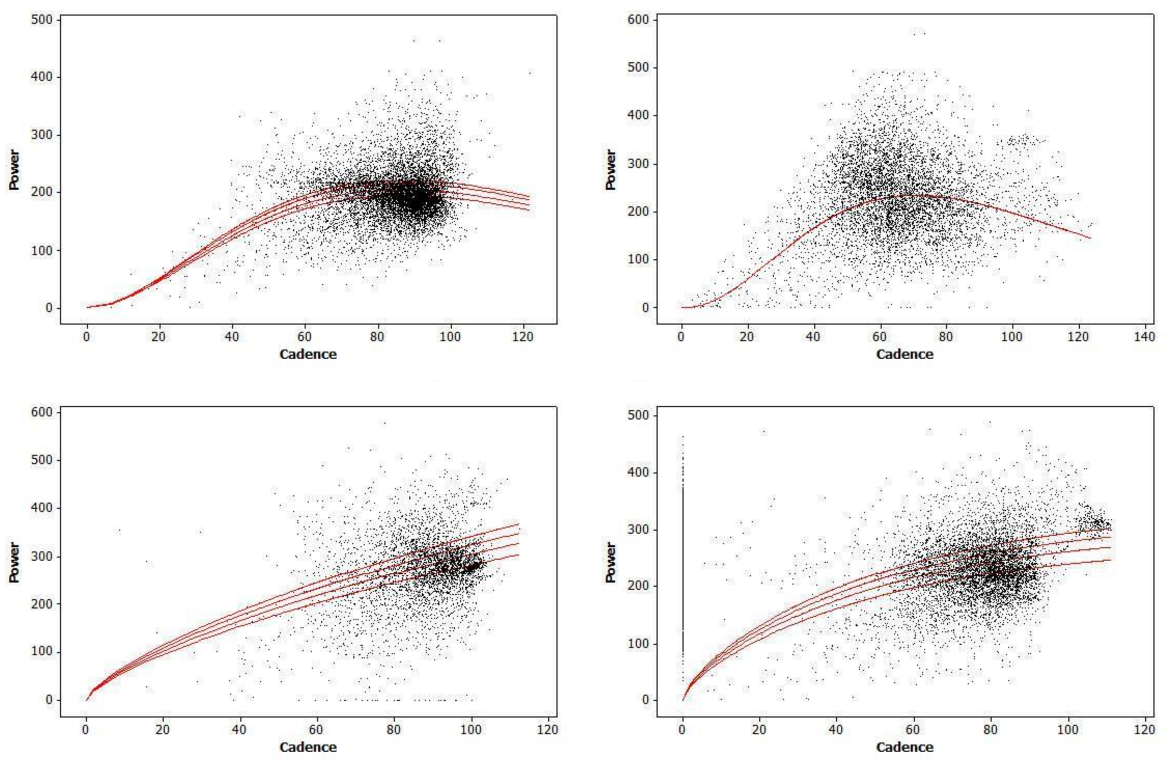


Figure 15: Fitted power output (red curves) from the non-linear model with cadence/normalized power output interaction (equation 12), and power output measurements (black dots) against cadence, for athletes 1 to 4 (read lexicographically), for the high heart rate data sub subset. The different red curves represent different values for normalized power output (X_{1t}). For athletes 1, 3 and 4, X_{1t} increases from the bottom red curve to the top (i.e. power output increases with X_{1t}), but for athlete 2 X_{1t} increases from top red curve to bottom. Heart rate is lagged at 5 seconds. Heart rate and other variables are set to typical values.

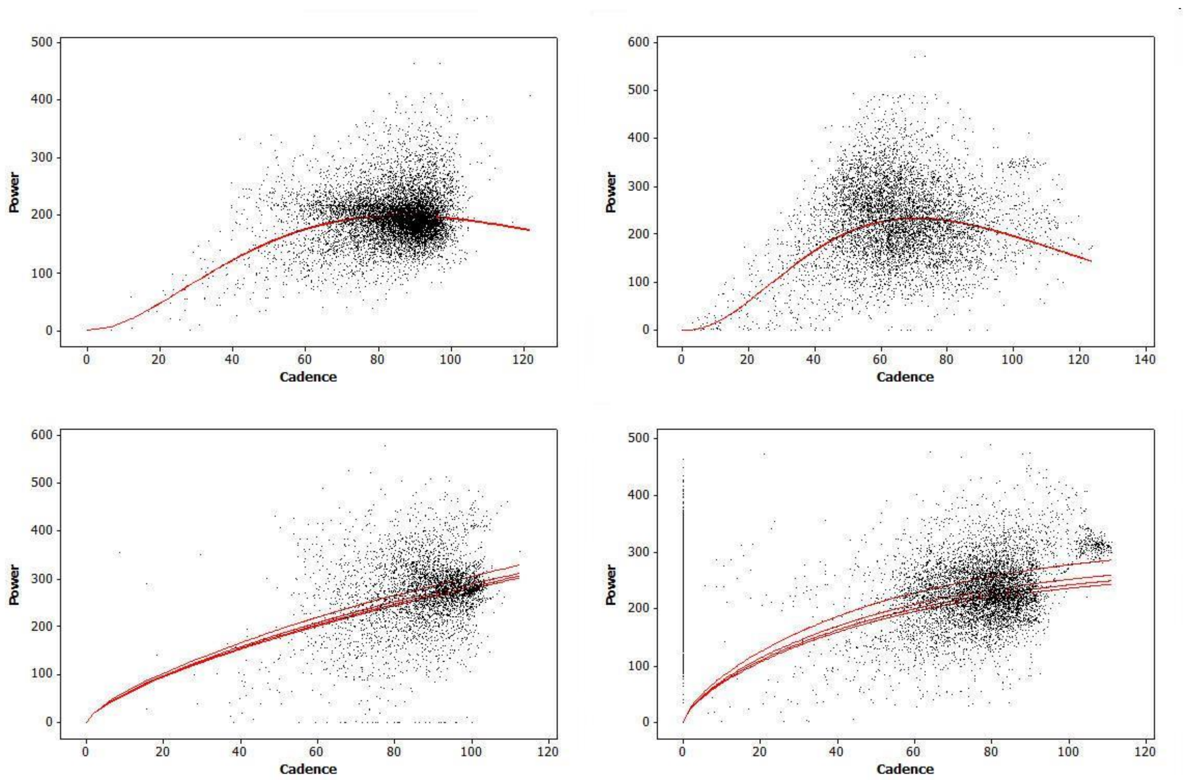


Figure 16: Fitted power output (red curve) from the non-linear model with cadence interacted with proportion of the session that has elapsed up to the current sampling point (equation 13), and power output measurements (black dots) against cadence, for athletes 1 to 4 (read lexicographically), for the high heart rate data subset. The different red curves represent values for the proportion of time that has elapsed in the current session up to the current sampling point (X_{2t} – 0.1, 0.4, 0.7 and 1). For athletes 1, 3 and 4, X_{2t} increases from the top red curve to the bottom (i.e. power output decreases as X_{2t} increases), but for athlete 2 X_{2t} increases from bottom red curve to top. Heart rate is lagged at 5 seconds. Heart rate and other variables are set to typical values.

4.5. Deviation from optimum cadence

The non-linear model of power output, heart rate and cadence (equation 7) yields an optimum cadence for athlete 2 when we include all sampled data. When we fit the same model using only high heart rate data (sampled data in which heart rate measurements are greater than the mean heart rate for each athlete) optimum cadences are yielded for athletes 1 and 2. Having found an optimum cadence (that cadence that maximises power output for a given heart rate), we may further investigate the importance of an optimum cadence.

The measure of importance of the optimum cadence may be represented by the change in performance that is affected by deviations from optimum cadence (that is, the level of decrease in power output that occurs when the athlete is no longer pedalling at the optimum pedalling rate). We can investigate this by comparing power output in the fitted power output cadence curve for various cadences above and below the optimum cadence for that particular athlete and that particular model. Cadences that are 5, 10 and 20 rpm below and above the optimum cadence are investigated. This is done for all models that yield an optimum cadence, which includes models fitted for all sampled data and also models fitted for just high heart rate data (where we only include data in which recorded heart rate is higher than the overall mean heart rate for each athlete). This also includes models in our sensitivity analysis, in which we considered a rage of data processing constants – window widths (*k*, or the size of the moving average smoother) and heart rate lag, *l*.

We present results for deviation from optimum cadence, starting with the simple non-linear model in equation 7. We present expected power output for a range of sub-optimal cadences for this model (equation 7) in table 40.

Table 40: Differences in fitted power output for athlete 2 using the simple non-linear model (equation 7) with a heart rate lags of 5 seconds, for various cadences above and below the optimum cadence.

change in C from C*	Cadence	Expected power output	change in Power output	% change in power output
l = 1				
-20	71.3	197.7	-8.3	-4.0
-10	81.3	204.0	-1.9	-0.9
-5	86.3	205.5	-0.5	-0.2
	91.3	206.0		
5	96.3	205.5	-0.4	-0.2
10	101.3	204.3	-1.7	-0.8
20	111.3	199.8	-6.2	-3.0

There is a greater reduction in power output for cadences below the optimum compared to cadences above the optimum. The higher the power output, the greater the reduction in power output for sub-optimal cadences, suggesting that it is more important for the athlete to select his cadence carefully if he is able to produce a high power output.

Next we fit the same model (equation 7) for conditions likely to represent race conditions. We calculate the mean of all sampled heart rate measurements (which we term the overall mean heart rate for each athlete). We retain data in which heart rate is greater than the overall mean heart rate for each athlete. We term this the high heart rate data subset. We present expected power output for sub-optimal cadences for this model (equation 7, using only high heart rate data) in table 41.

When we fit the model in equation 7 for the high heart rate data subset, peak fitted power output increases to 226.6, compared to 205 when all sampled data are included for this model. We would expect such an increase as power output tends to increase with increasing heart rate. Training schedules include some very long low-intensity rides, in which heart rate is likely to be lower than in a more intense training session or in a race. For the high heart rate data subset, in which fitted power output (from the model in equation 7) is higher, reduction in power output for sub-optimal cadences is greater. This emphasises the importance of optimal cadence selection in race conditions.

Table 41: Differences in fitted power output for athlete 1 (left) and athlete 2 (right) using the simple non-linear model (equation 7) for the high heart rate data subset, with a heart rate lag of 5 seconds, for various cadences above and below the optimum cadence.

change in C from C*	cadence	Expected power output	change in power output	% change in power output
-20	67.1	185.0	-12.1	-6.1
-10	77.1	194.3	-2.8	-1.4
-5	82.1	196.4	-0.7	-0.3
	87.1	197.1		
5	92.1	196.5	-0.6	-0.3
10	97.1	194.7	-2.4	-1.2
20	107.1	188.2	-8.9	-4.5

change in C from C*	cadence	Expected power output	change in power output	% change in power output
-20	50.5	203.8	-27.2	-11.8
-10	60.5	224.6	-6.3	-2.7
-5	65.5	229.5	-1.5	-0.7
	70.5	231.0		
5	75.5	229.5	-1.4	-0.6
10	80.5	225.6	-5.3	-2.3
20	90.5	212.0	-19.0	-8.2

For the high heart rate data subset we are also able to compare results for different athletes. Reduction in power output for sub-optimal cadences is greater for athlete 1 compared to athlete 2. In fact for 20 rpm below optimal cadence, the reduction in power output is approximately 50% greater for athlete 2 compared to athlete 1 (11.8% less compared to 6.1% less). This may be due to the higher overall fitted power output for athlete 2 compared to athlete 1 (at least for high heart rate data). Alternatively there may be some physiological differences between athletes that affects the respective ability of each athlete to produce power outputs at different cadences, or strategic differences in riding styles, which may not be captured by these statistical models. Such differences may affect an athlete's performance for different cadences. Conceivably the higher power output for athlete 2 and some other physiological differences may both combine to affect the penalty for deviation from optimum cadence differently for different athletes.

Next we interact cadence with heart rate in the model in equation 9. We present expected power outputs for a range of sub-optimal cadences for this model (equation 9) in table 42.

The greater the heart rate, the greater the reduction in power output for sub-optimal cadences. The greater heart rate could to some extent represent a greater effort or intensity; therefore this may emphasise the importance for optimal cadence selection in races. Athletes may vary in their knowledge of impact of cadence on performance, so it is conceivable for an athlete to ride 10 to 15 rpm below or above their individual mathematical optimum cadence and therefore fail to produce the maximum power output for the conditions.

For each heart rate, deviating from optimum cadence reduces power output more for athlete 2 than it does for athlete 1. There is a marked trend for more reduced power output in sub-optimal cadences for athlete 2 than for athlete 1. The fitted power output is consistently higher for athlete 2 than it is for athlete 1, for the high heart rate data subset.

Table 42: Differences in fitted power output for athlete 1 (left) and athlete 2 (right) using the non-linear model cadence heart rate interaction model (equation 9) for the high heart rate data subset, with a heart rate lag of 5 seconds, for various cadences above and below the optimum cadence, for a range of heart rates.

change in C from C*	cadence	Expected power output	change in power output	% change in power output
H = mean				
-20	70.9	187.7	-9.4	-4.8
-10	80.9	194.9	-2.2	-1.1
-5	85.9	196.6	-0.5	-0.3
	90.9	197.1		
5	95.9	196.6	-0.5	-0.3
10	100.9	195.2	-1.9	-1.0
20	110.9	190.0	-7.1	-3.6
H = mean + 20				
-20	60.3	222.9	-14.9	-6.3
-10	70.3	234.4	-3.5	-1.5
-5	75.3	237.0	-0.8	-0.3
	80.3	237.8		
5	85.3	237.1	-0.8	-0.3
10	90.3	234.9	-2.9	-1.2
20	100.3	227.1	-10.7	-4.5
H = mean + 40				
-20	51.8	258.7	-22.4	-8.0
-10	61.8	275.9	-5.2	-1.8
-5	66.8	279.8	-1.2	-0.4
	71.8	281.0		
5	76.8	279.9	-1.1	-0.4
10	81.8	276.8	-4.3	-1.5
20	91.8	265.6	-15.5	-5.5

change in C from C*	cadence	Expected power output	change in power output	% change in power output
H = mean				
-20	53.4	212.3	-23.1	-9.8
-10	63.4	230.0	-5.4	-2.3
-5	68.4	234.1	-1.3	-0.5
	73.4	235.4		
5	78.4	234.2	-1.2	-0.5
10	83.4	230.9	-4.5	-1.9
20	93.4	219.2	-16.2	-6.9
H = mean + 20				
-20	44.0	251.0	-38.0	-13.2
-10	54.0	280.2	-8.8	-3.1
-5	59.0	286.9	-2.1	-0.7
	64.0	289.0		
5	69.0	287.2	-1.9	-0.6
10	74.0	281.9	-7.1	-2.5
20	84.0	263.8	-25.3	-8.7
H = mean + 40				
-20	36.8	287.5	-58.8	-17.0
-10	46.8	332.7	-13.6	-3.9
-5	51.8	343.0	-3.2	-0.9
	56.8	346.3		
5	61.8	343.4	-2.8	-0.8
10	66.8	335.6	-10.7	-3.1
20	76.8	308.9	-37.3	-10.8

We also consider a range of additional covariates to capture further aspects of an athlete's riding. These are normalized power output in the current session up to the current sampling point, (X_{1t}), proportion of time that has elapsed in the currrent session up to the current sampling point (X_{2t}), mean of the normalized power outputs over the previous seven days (X_3) and mean session duration of the previous seven days (X_4). This produces the model in equation 10. We smooth data over a moving average of width 60 seconds (k=12), and settle for a heart rate lag of 5 seconds (l=1). First we tested this model for including all sampled data.We then split the data based on heart rate, into low or high heart rate categories. Low heart rate data refers to data in which heart rate was less than the mean of all recorded heart rate measurements for each athlete, whereas high heart rate data refers to data in which heart rate was higher than the mean of all recorded heart rate measurements. Optimum cadences are yielded for athletes 1 and 2 for high heart rate data, but no optimum cadences are yielded for low heart rate data. The following expected power outputs are yielded for the model with additional covariates (in equation 10) for all sampled data (which we present in table 43), then for high heart rate data (in table 44), for a range of sub-optimal cadences.

Compared to the simple non-linear model without extra covariates (equation 7), for the equivalent heart rate lag of 5 seconds, reduction in power output for sub-optimal cadences is slightly less with the inclusion of the extra covariates. The expected power output is lower compared to the model in equation 7, further suggesting that the importance of cadence selection increases with power output.

Table 43: Differences in fitted power output for athlete 2 using the non-linear model with extra covariates (equation 10), for all sampled data, with a heart rate lag of 5 seconds, for various cadences above and below the optimum cadence.

change in C from C*	cadence	Expected power output	change in Power output	% change in power output
-20	67.67	156.4	-7.0	-4.3
-10	77.67	161.8	-1.6	-1.0
-5	82.67	163.0	-0.4	-0.2
	87.67	163.4		
5	92.67	163.1	-0.4	-0.2
10	97.67	162.0	-1.4	-0.9
20	107.67	158.3	-5.2	-3.2

Table 44: Differences in fitted power output for athlete 1 (left) and athlete 2 (right) using the non-linear model with extra covariates (equation 10) , for the high heart rate data subset, with a heart rate lag of 5 seconds, for various cadences above and below the optimum cadence.

change in C from C*	Cadence	Expected Power output	change in power output	% change in power output
-20	67.1	185.0	-12.1	-6.1
-10	77.1	194.3	-2.8	-1.4
-5	82.1	196.4	-0.7	-0.3
	87.1	197.1		
5	92.1	196.5	-0.6	-0.3
10	97.1	194.7	-2.4	-1.2
20	107.1	188.2	-8.9	-4.5

change in C from C*	Cadence	Expected Power output	change in power output	% change in power output
-20	50.5	203.8	-27.2	-11.8
-10	60.5	224.6	-6.3	-2.7
-5	65.5	229.5	-1.5	-0.7
	70.5	231.0		
5	75.5	229.5	-1.4	-0.6
10	80.5	225.6	-5.3	-2.3
20	90.5	212.0	-19.0	-8.2

We expand the model with cadence / heart rate interaction to include extra covariates, producing the model in equation 11. We present expected power output for a range of sub-optimal cadences for the model in equation 11 in table 45.

We also interact cadence with normalized power output and the proportion of the session that has elapsed up to the current sampling point, producing the models in equations 12 and 13. However cadence is not affected by either additional variable, so we do not present tables of power output at sub-optimal cadences for these models (in equations 12 and 13)

Including extra covariates increases the explanatory power of the models but has little impact on optimum cadence. There is also little difference in the patterns of deviation from optimum cadence. The reduction in power output for sub-optimal cadences again increases with heart rate. It also appears to vary by athlete (reduction in power output is greater for athlete 2 than for athlete 1).

The greater reduction in power output for sub-optimal cadence for athlete 2 compared to athlete 1 may be partly due to the generally higher power output for athlete 2 compared to athlete 1. Nevertheless we would not expect all athletes to produce the same power output when riding. Therefore we may still conclude that the importance of an optimum cadence, (i.e. the amount by which power output reduces for sub-optimal cadences) may vary by athlete.

Table 45: Differences in fitted power output for athlete 1 (left) and athlete 2 (right) using the non-linear cadence heart rate interaction model (equation 11), for the high heart rate data subset, with a heart rate lag of 5 seconds, for various cadences above above and below the optimum cadence, for a range of heart rates.

change in C from C*	cadence	Expected power output	change in power output	% change in power output
H = mean				
-20	70.9	187.7	-9.4	-4.8
-10	80.9	194.9	-2.2	-1.1
-5	85.9	196.6	-0.5	-0.3
	90.9	197.1		
5	95.9	196.6	-0.5	-0.3
10	100.9	195.2	-1.9	-1.0
20	110.9	190.0	-7.1	-3.6
H = mean + 20				
-20	60.3	222.9	-14.9	-6.3
-10	70.3	234.4	-3.5	-1.5
-5	75.3	237.0	-0.8	-0.3
	80.3	237.8		
5	85.3	237.1	-0.8	-0.3
10	90.3	234.9	-2.9	-1.2
20	100.3	227.1	-10.7	-4.5
H = mean + 40				
-20	51.8	258.7	-22.4	-8.0
-10	61.8	275.9	-5.2	-1.8
-5	66.8	279.8	-1.2	-0.4
	71.8	281.0		
5	76.8	279.9	-1.1	-0.4
10	81.8	276.8	-4.3	-1.5
20	91.8	265.6	-15.5	-5.5

change in C from C*	cadence	Expected power output	change in power output	% change in power output
H = mean				
-20	53.4	212.3	-23.1	-9.8
-10	63.4	230.0	-5.4	-2.3
-5	68.4	234.1	-1.3	-0.5
	73.4	235.4		
5	78.4	234.2	-1.2	-0.5
10	83.4	230.9	-4.5	-1.9
20	93.4	219.2	-16.2	-6.9
H = mean + 20				
-20	44.0	251.0	-38.0	-13.2
-10	54.0	280.2	-8.8	-3.1
-5	59.0	286.9	-2.1	-0.7
	64.0	289.0		
5	69.0	287.2	-1.9	-0.6
10	74.0	281.9	-7.1	-2.5
20	84.0	263.8	-25.3	-8.7
H = mean + 40				
-20	36.8	287.5	-58.8	-17.0
-10	46.8	332.7	-13.6	-3.9
-5	51.8	343.0	-3.2	-0.9
	56.8	346.3		
5	61.8	343.4	-2.8	-0.8
10	66.8	335.6	-10.7	-3.1
20	76.8	308.9	-37.3	-10.8

4.6. Summary of optimum cadences found

We present a summary of optimum cadences for regression models explored in chapter 4. We begin with linear models (models in which the implied power output cadence relationship is linear), then present optimum cadences for non-linear models (models in which the implied power output cadence relationship is non-linear). The first non-linear models featured only power output and cadence in the regression equation. We then include heart rate, building non-linear models of power output, heart rate and cadence.

4.6.1. Optimum cadences from initial non-linear models of power output and cadence

Initially when exploring models of power output and cadence that yield a non-linear power output cadence relationship, we do not include heart rate in the model. Instead we focus on the power output cadence relationship and split the data into subsets of groups based on heart rate. We present the optimum cadence (the cadence that maximises fitted power output) for each data subset in tables 46 to 50, for all four athletes together in the same regression analysis and then each athlete (1 to 4) separately. In the tables C* refers to the optimum cadence, and P* refers to the corresponding power output obtained at that cadence.

Table 46: Optimum cadence (C*) for various data sets (different ranges of recorded heart rates) and the corresponding power output (P*) from the non-linear model of power output and cadence, for all four athletes.

HR range			C*	P*
40	-	50	69.92	83.62
50	-	60	*	116.10
60	-	70	50.42	67.54
70	-	80	81.00	76.49
80	-	90	99.58	97.01
90	-	100	100.08	122.37
100	-	110	80.75	139.03
110	-	120	72.17	161.18
120	-	130	72.83	185.50
130	-	140	73.42	209.75
140	-	150	71.17	229.24
150	-	160	64.67	260.69
160	-	170	90.75	290.03
170	-	180	*	287.37
180	-	190	*	296.35
190	-	200	*	290.73
		ALL	109.75	211.02

Table 47: Optimum cadence (C*) for various data sets (different ranges of recorded heart rates) and the corresponding power output (P*) from the non-linear model of power output and cadence, for athlete 1.

HR range			C*	P*
40	-	50	*	47.47
50	-	60	*	91.36
60	-	70	50.17	85.48
70	-	80	*	101.71
80	-	90	49.92	70.94
90	-	100	65.17	74.76
100	-	110	82.42	95.84
110	-	120	101.33	121.24
120	-	130	106.08	147.62
130	-	140	105.33	173.99
140	-	150	83.58	187.34
150	-	160	76.50	206.10
160	-	170	97.83	234.97
170	-	180	95.08	265.89
180	-	190	95.75	282.57
190	-	200	*	214.67
		ALL	*	197.89

Table 48: Optimum cadence (C*) for various data sets (different ranges of recorded heart rates) and the corresponding power output (P*) from the non-linear model of power output and cadence, for athlete 2.

HR range			C*	P*
40	-	50	*	194.87
50	-	60	*	143.21
60	-	70	*	113.21
70	-	80	*	66.63
80	-	90	73.75	66.50
90	-	100	72.67	96.05
100	-	110	67.58	129.99
110	-	120	65.25	168.51
120	-	130	66.50	202.46
130	-	140	65.92	238.84
140	-	150	66.08	274.46
150	-	160	62.08	317.26
160	-	170	77.83	337.19
170	-	180	96.42	228.00
180	-	190	*	135.20
190	-	200	71.83	203.80
		ALL	78.50	199.40

Table 49: Optimum cadence (C*) for various data sets (different ranges of recorded heart rates) and the corresponding power output (P*) from the non-linear model of power output and cadence, for athlete 3.

HR range			C*	P*
40	-	50	*	92.98
50	-	60	*	111.12
60	-	70	*	81.07
70	-	80	73.92	88.74
80	-	90	100.83	106.20
90	-	100	102.92	131.96
100	-	110	90.75	155.25
110	-	120	88.83	188.34
120	-	130	89.33	223.63
130	-	140	*	264.58
140	-	150	*	288.81
150	-	160	88.33	304.31
160	-	170	*	372.46
170	-	180	97.75	47.15
180	-	190	*	241.05
190	-	200	*	238.61
		ALL	*	277.94

Table 50: Optimum cadence (C*) for various data sets (different ranges of recorded heart rates) and the corresponding power output (P*) from the non-linear model of power output and cadence, for athlete 4.

HR range			C*	P*
40	-	50	52.42	63.38
50	-	60	96.33	126.85
60	-	70	50.33	61.18
70	-	80	66.25	75.10
80	-	90	71.00	98.30
90	-	100	84.92	126.82
100	-	110	76.25	147.09
110	-	120	78.42	173.238
120	-	130	78.75	199.53
130	-	140	78.92	224.39
140	-	150	80.25	249.34
150	-	160	101.00	293.46
160	-	170	85.75	324.68
170	-	180	*	391.36
180	-	190	20.58	2.55
190	-	200	*	217.99
		ALL	*	263.69

4.6.2. Optimum cadences from non-linear models of power output, heart rate and cadence.

Having found a range of plausible optimum cadences for non-linear models of just power output and cadence, we now seek to develop a non-linear model that also incorporates heart rate. Similarly to the linear models, we can calculate an optimum cadence mathematically (as outlined in the methodology section). We also present 95% confidence intervals for each mathematically calculated optimum cadence (the intervals are presented in parenthesis after the optimum cadences).

We fit the following non-linear model of power output, heart rate and cadence (equation 7). We initially smooth data using a moving average window of 60 seconds (k=12), and a heart rate lag of 5 seconds (l=1). We present optimum cadences in table 51 for this non-linear model (equation 7).

Table 51: Optimum cadences for the simple non-linear model (equation 7), in which heart rate is lagged at 5 seconds (l=1).

Athlete 1	Athlete 2	Athlete 3	Athlete 4
Optimum Cadence	Optimum Cadence	Optimum Cadence	Optimum Cadence
*	88.28 (87.46, 89.09)	*	*

We also explore the effects of heart rate on optimum cadence. When fitting power output cadence curves for the simple non-linear model (equation 7) heart rate is fixed to a range of values (40 bpm below mean, the mean itself, and 40 above mean). Mean indicates the mean heart rate from all recorded heart rate measurements, calculated for each athlete. Optimum cadence does not change with heart rate. However we also interact cadence with heart rate (equation 9). Optimum cadences are yielded for athlete 2 (displayed in table 52), but these are unrealistic. Only at very high heart rates (190 beats per minute – that is, mean heart rate plus 40 beats per minute) is a realistic optimum cadence produced.

Table 52: Optimum cadences for the cadence heart rate interaction non-linear model (equation 9), in which heart rate is lagged at 5 seconds (l=1).

	Athlete 1	Athlete 2	Athlete 3	Athlete 4
	Optimum Cadence	Optimum Cadence	Optimum Cadence	Optimum cadence
H = mean	*	226	*	*
H = mean + 20	*	149	*	*
H = mean + 40	*	111	*	*

Next we split data into to two groups based on heart rate. We calculate the mean of all sampled heart rate measurements (the overall mean heart rate for each athlete). One group contains only data in which heart rate is less than the overall mean heart rate for each athlete (the low heart rate data subset), and the other group contains only data in which heart rate is greater than the overall mean heart rate for each athlete (the high heart rate data subset). We fit models in equations 7 and 9. For the model in equation 7, optimum cadence is the same whatever value to which heart rate is fixed when plotting power output cadence curves. We present these optimum cadences (for low and high heart rate data separately) in table 53. Heart rate remains lagged at five seconds (l=1). For the model in equation 9, optimum cadence depends on heart rate – we therefore we consider a range of values for testing the sensitivity of optimum cadence to heart rate. Optimum cadences are found for athletes 1 and 2 for the high heart rate data subset, as presented in table 54.

Table 53: Optimum cadences for the simple non-linear model (equation 7), fitted separately for low and high heart rate data subsets.

Athlete 1	Athlete 2	Athlete 3	Athlete 4
Optimum Cadence	Optimum Cadence	Optimum Cadence	Optimum Cadence
Low heart rate data – no optimum cadences			
High heart rate data			
90.48 (90.06, 90.90)	75.85 (75.33, 75.36)	*	*

Table 54: Optimum cadences for the cadence heart rate interaction non-linear model (equation 9), for low and high heart rate data subsets.

	Athlete 1	Athlete 2	Athlete 3	Athlete 4
	Optimum Cadence	Optimum Cadence	Optimum Cadence	Optimum Cadence
Low heart rate data – no optimum cadences				
High heart rate data				
H = mean	90.92	73.42	*	*
H = mean + 20	80.25	64.00	*	*
H = mean + 40	71.83	56.75	*	*

The models yield a range of optimum cadences for two athletes when fitted to the high heart rate data subset, but do not yield any optimum cadences when fitted to the low heart rate data subset. We believe this is because high heart rate data is more representative of competitions and races, as the athlete is trying to maximise his power output in a race and this requires great physical exertion, which a high heart rate to some extent is likely to represent.

We also consider a range of additional covariates to capture further aspects of an athlete's riding. These are normalized power output in the current session up to the current sampling point, (X_{1t}), proportion of time that has elapsed in the current session up to the current sampling point (X_{2t}), Mean of the normalized power outputs over the previous seven days (X_3) and mean session duration of the previous seven days (X_4). This produces the model in equation (10). We smooth data over a moving average of width 60 seconds (k=12), and settle for a heart rate lag of 5 seconds (l=1). We present optimum cadences yielded from the model (in equation 10) in table 55, fitted for all sampled data.

Next we focus on data in which heart rate is greater than the overall mean heart rate for each athlete (high heart rate data) for the model with extra covariates (equation 10). We present optimum cadences yielded from the model in equation 10, in which only high heart rate data are retained, in table 56.

Table 55: Optimum cadences for the non-linear model with extra covariates (equation 10), with heart rate lagged at 5 seconds (*l*=1).

Athlete 1	Athlete 2	Athlete 3	Athlete 4
Optimum Cadence	Optimum Cadence	Optimum Cadence	Optimum Cadence
*	87.70 (86.90, 88.49)	*	*

Table 56: Optimum cadences for the non-linear model with extra covariates (equation 10), for the high heart rate data subset. Heart rate is lagged at 5 seconds (*l*=1).

Athlete 1	Athlete 2	Athlete 3	Athlete 4
Optimum Cadence	Optimum Cadence	Optimum Cadence	Optimum Cadence
86.75 (86.29, 87.20)	70.60 (70.27, 70.94)	*	*

We also interact cadence with heart rate, including a range of additional covariates, fitting this model (equation 11) to the high heart rate data subset. For this cadence heart rate interaction model (equation 11), we plot optimum cadence with heart rate set to a number of different values based on what is found in the data for each athlete. The optimum cadences for this model (equation 11) come from plotting fitted power output against cadence for a range of cadences the athlete has ridden at in his training sessions. We present optimum cadences yielded from this model (equation 11) in table 57.

Next we present the range of optimum cadences yielded from the models with cadence interacted with normalized power output (X_{1t}) and proportion of the current session that has elapsed up to the current sampling point (X_{2t}), equations (12) and (13) respectively. We fitted such models to the high heart rate data subset. The optimum cadences for models in equations 12 and 13 come from plotting fitted power output against cadence for a range of cadences the athlete has ridden at in his training sessions. We present optimum cadences yielded from models in equations 12 and 13 respectively in tables 58 and 59.

Table 57: Optimum cadences for the non-linear cadence heart rate interaction model (equation 11), for the high heart rate data subset.

	Athlete 1	Athlete 2	Athlete 3	Athlete 4
	Optimum Cadence	Optimum Cadence	Optimum cadence	Optimum Cadence
H = mean	90.17	73.58	*	*
H = mean + 20	79.67	64.17	*	*
H = mean + 40	71.33	56.92	*	*

Table 58: Optimum cadences for the non-linear model (equation 12), for the high heart rate data subset.

	Athlete 1		Athlete 2
Normalized Power output (X_{1t}) for Athlete 1	Optimum Cadence	Normalized Power output (X_{1t}) for Athlete 2	Optimum Cadence
1000	86.58	1200	70.58
1300	86.50	1500	70.58
1700	86.58	1800	70.58
2000	86.58	2100	70.58

Table 59: Optimum cadences for the non-linear model (equation 13), for the high heart rate data subset.

	Athlete 1		Athlete 2
Proportion of session that has elapsed (X_{2t}) for Athlete 1	Optimum Cadence	Proportion of session that has elapsed (X_{2t}) for Athlete 2	Optimum Cadence
0.1	86.50	0.1	70.58
0.4	86.50	0.4	70.58
0.7	86.50	0.7	70.58
0.9	86.50	0.9	70.58

4.7. Discussion of results

We discuss results of linear and non-linear models of power output, heart rate and cadence. We compare optimum cadences with literature, and finally summarise our findings, stating our preferred model from chapter 4.

4.7.1. Discussion of linear model results

For the linear models, the inclusion of the cubic term produces a slight increase in optimum cadence. Although optimum cadences are yielded for athlete 4, they are not plausible; they are much lower than we would expect (16 for the quadratic model and 32 for the cubic model). We may therefore conclude that these optimum cadences cannot be considered valid, and that they are producing a theoretical optimum cadence by chance. Optimum cadences have not been yielded at all for athletes 1, 2 and 3.

The lack of plausible optium cadences suggest that the quadratic and cubic models may not be valid methods for capturing the relationship between power output and cadence in the determinations of optimum cadence. The quadratic and cubic models provide a reasonable line of best fit through some of the more concentrated data points in scatterplots of power output measurements against real cadence values, but there are limitations to quadratic and cubic models. When power output is zero, cadence should also be zero, as the athlete is not riding at that point. This is not guaranteed from a quadratic or cubic model, and indeed this does not occur in the models in this analysis. A more sophisticated model may be required, that encompasses the intercept of zero (power output equals zero when cadence equals zero), and also provides a decreasing power output for very high cadences.

4.7.2. Discussion of initial non-linear models of power output and cadence

We sought a model of power output and cadence with a non-linear relationship. We fitted a gamma curve to represent the power output cadence relationship. So that we could focus on the power output cadence relationship of the model, we only model power output and cadence, and do not include heart rate in the model. We split the data into subsets based on different heart rates. A plausible optimum cadence is yielded for many heart rate subsets. However there tends not to be an optimum cadence at particularly low or high heart rates. At particularly low heart rates, though, the athlete may not necessarily be exerting much effort and therefore is not seeking to maximise power output when riding at such a low heart rate. For athletes 1, 3 and 4 there is no optimum cadence when we include all sampled (rather than splitting sampled data into subsets).

For athlete 1, optimum cadences are highest for heart rate data between 110 and 140, where they are over 100. For athlete 2, optimum cadences tend to be between 60 and 80 but increase to over 90 at high heart rates. Optimum cadences are slightly higher for athlete 3 compared to the others, as they are consistently over 88 and are over 100 for five of the heart rate data subsets. For athlete 4 the highest optimum cadences (101 and 109) occur at relatively high heart rates (150-160 and 170-180 respectively). For athletes 1, 3 and 4 the optimum cadence is higher when all sampled data are included than for any individual data subset.

In these gamma curves there is a general trend for optimum cadences to increase with heart rate (albeit more so for athletes 1, 2 and 4 than for athlete 3), possibly suggesting that when athletes exert more effort they are able to sustain higher cadences. This may explain why professional cyclists choose higher cadences in race conditions than those found to be optimal from ergometer studies – athletes may not necessarily be exerting as much effort in ergometer studies compared to races.

The corresponding fitted power output (produced from the model at the optimum cadence for the gamma curve) generally increases with heart rate (except for extremely high heart rates) for each athlete. This is expected as the athlete is likely to be exerting more effort when heart rates are higher. The highest corresponding power output is over 300 for athletes 2, 3 and 4 but less for athlete 1 (whose fitted power output is 283). For athlete 4, the highest optimum cadence for all data subsets (108, when heart rates are between 170 and 180) also corresponds to the highest power output (391). The gamma curves consistently produce plausible optimum cadences, much more so than the quadratic and cubic models.

The non-linear gamma curves clearly reflect the relationship between power output and cadence better than the quadratic and cubic models, yielding plausible optimum cadences, and producing power output in a realistic range for the given athlete. Quadratic and cubic models often failed to produce realistic power outputs and only produced optimum cadences by chance due to the shape of the curve.

4.7.3. Discussion of non-linear models of power output, heart rate and cadence

We incorporate heart rate into a non-linear model (equation 7). This model appears to provide a good representation of the power output cadence relationship, and provides the useful property that when power output is zero cadence is also zero. An optimum cadence is yielded only for athlete 2.

We split the data into low heart rate and high heart rate subsets (lower than or higher than the mean heart rate in the sampled data for each athlete). This has a marked impact on optimum cadence. Optimum cadence for all sampled data for the non-linear model (equation 7) is 88.3, but it is 75.6 for the high heart rate data subset (there is no optimum cadence yielded for the low heart rate data subset for any of the athletes in our study). Athlete 1 has an optimum cadence for the high heart rate data subset– a cadence of 90.5. The differences in optimum cadence between athletes 1 and 2 provide further evidence of the importance of having separate athlete coefficients and thus individual mathematical optimum cadences.

In our analysis we defined the high heart rate data subset to be simply data in which heart rate is higher than the overall mean heart rate for an individual athlete, effectively splitting data into two halves (a low half and a high half). This appears a reasonable method for splitting data as it allows for direct comparison between high and low heart rate data subsets, and still maintains a vast amount of data for the high heart rate subset. We could alternatively have sought an even more precise replication of race conditions by taking, say, the top 20% or 30% of sampled data (that is the sampled data that contain the highest 20% or 30% of heart rate measurements for each athlete). However during a race athletes pace themselves accordingly so, particularly for races with varying wind and terrain, and depending on strategy (whether an athlete is a team leader or whether they are effectively trying to disrupt the strategy of a rival team), and athlete’s heart rate may vary within a race. We could split heart rate data into a number of groups, say into five 20% groups (for example, the highest heart rate group would be data that contain the highest 20% of recorded heart rate measurements for each athlete). However this would presumably achieve a similar effect to interacting heart rate with cadence in the model; effectively each method yields models that produce a range of optimum cadences depending on the heart rate. For the cadence heart rate interaction models we sought optimum cadences for three different heart rates (the mean, 20 and 40 bpm above the mean heart rate for each athlete).

We also fitted the model with cadence / heart rate interaction (equation 9) to the high heart rate data subset. For this model optimum cadence depends on heart rate, and indeed decreases with heart rate. Also the margin of decrease in optimum cadence decreases with each increase in heart rate. As heart rate increases, this may to some extent represent an increase in effort; he is unable to sustain such a high cadence at great effort, so the highest cadence he can sustain without hinderence to his riding technique reduces. In chapter 5 we include measures of training impulses, which are a calculation of accumulation of fatigue. The impact of a measurement of fatigue (using training impulses) may be much more easy to interpret than the impact of heart rate on optimum cadence. The reduced optimum cadence for models fitted to the high heart rate data subset (compared to the same models fitted to all sampled data) could suggest that optimum cadence decreases as the athlete exerts more effort. Heart rate is not a direct measure of effort or training intensity; it can be influenced by other elements such as fitness. Nevertheless training schedules include some very long low-intensity sessions; heart rate is likely to be lower for such sessions compared to more intense sessions or races. Broadly, as heart racte increases, so too does power output. We believe the high heart rate data subset is likely to more closely represent race conditions than does the use of all sampled data.

We therefore consider it more important to find an optimum cadence for models fitted to the high heart rate data subset than for models fitted to all sampled data or for the low heart rate data subset. In fact for the high heart rate data subset optimum cadences are yielded for athletes 1 and 2, as opposed to only athlete 2 for all sampled data and none for the low heart rate data subset. That the presence of an optimum cadence is found more often for high heart rate data may suggests cadence selection in general is likely to be important in race conditions. Data within the low heart rate subset may represent long low-intensity training sessions. During such sessions, athletes are likely to be building fitness (rather than skill for maximising the power output they can achieve), or it may these sessions take place during cold winter months. That there are no optimum cadences yielded for the low heart rate data subset suggets that we may indeed have successfully excluded long low-intensity sessions when discarding data in the low heart rate subset.

Furthermore, in the exploration into deviation from optimum cadence, we noted that, broadly, as fitted power output increases, the reduction in power output for sub-optimal cadences increases. This suggests that increased power output also increases the importance of optimum cadence.

By calculating 95% confidence intervals for the mathematical optimum cadences, we can explore the importance of riding at optimum cadence. We do this for all models which do not feature an interaction between cadence and another variable.

When fitting models to the high heart rate data subset, confidence intervals become narrower, often twice as narrow (compared to the same models fitted using all sampled data). The precentage reduction in power output for sub-optimal cadences is also higher for models fitted to the high heart rate data subset than for models fitted for all sampled data. Also when interacting cadence with heart rate, whereby a range of different optimum cadences are yielded for different heart rates, precentage reduction in power output for sub-optimal cadences increases as heart rate increases. This may suggest that riding at a sub-optimal cadence has a greater effect on loss of performance (power output) when the athlete is exerting great physical effort than when he is exerting slightly less physical effort (assuming that an athlete is likely to be asserting greater effort if heart rate increases significantly). The importance of optimum cadence selection when heart rate is high may emphasise the importance of riding at an optimal cadence in race conditions (although heart rate is not a direct measure of effort, it is likely that heart rate is higher in a race than in long, low-intensity training sessions). Power output tends to reduce by approximately 2% for cadences 10 rpm above and below optimum. Even if power output reduces slightly, this could have a significant impact on performance obver a race distance.

Confidence intervals are consistently narrower for athlete 2 than for athlete 1. We do not expect that optimum cadences are necessarly the same for different athletes, due to the nature of the athletes (physiological and muscular differences). By the same principle the relative importance of optimum cadence (i.e. how precise the choice of optimum cadence needs to be) may not necessarily be the same. The narrower confidence intervals appear to suggest that choice of optimum cadence is slightly more important for athlete 2 than for athlete 1; in other words, athlete 1 is more likely to be able to sustain a high power output when deviating slightly from optimum cadence.

Variance for the mathematical optimum cadence tends to be quite low for all models. This in turn indicates narrow confidence intervals for the mathematical optimum cadence, meaning there is quite a narrow range of cadences at which the athlete can ride optimally. The athlete only needs to deviate slightly from an optimal cadence before he is not maximising the power output for the given heart rate. It may be the case that, due to the great amount of data for each athlete, even small deviations in optimum cadence are statistically significant. However, as this study is based on competitive professional athletes, very small changes in activity (in this case the cadence at which they ride) may have a great impact on performance level. The confidence intervals demonstrate the importance of riding at an optimum cadence.

We also added the following extra covariates to the simple non-linear model: X_{1t} (the normalized power output developed in the current session, up to the current sampling point, t), X_{2t} (the proportion of time that has elapsed in the current session, up to the current sampling point, t), X_3 (the mean of the normalized power outputs for all session in the previous seven days) and X_4 (the mean duration of all sessions in the previous seven days (hours)). We fitted the models in equations (10-13) – see section 4.1.6. for these equations in full. The impact of additional covariates on fitted power output depends on the athlete. For athletes 1, 3 and 4, power output increases with normalized power output (X_{1t}) and decreases the further into the session the athlete is (X_{2t} – the proportion of the session that has elapsed up to the current sampling point). However the opposite is true for athlete 2. Athlete 2 does appear to fatigue with the accummulation of normalized power output (as expected power output decreases as X_{1t} increases). He does not appear to fatigue as he approaches the end of a session (as expected power output increases as X_{2t} increases).

Normalized power output therefore does not necessarily appear to represent the accumulation of fatigue. We may have to include a measure of heart rate when quantifying training load in order to represent the accummulation of fatigue. Indeed we do this when fitting training impulse calculations to the model, which is explored in chapter 5.

We also fitted non-linear models with additional covariates for the high heart rate data subset. For the high heart rate data subset, the non-linear models with additional covariates yield optimum cadences for athlete 1 and 2 – this is the same set of athletes for whom optimum cadences are yielded in the simple non-linear model with no additional covariates (equation 7). For models in equations 12 and 13, the margin of decrease in power output for-sub optimal cadences is greater for athlete 2 than for athlete 1. This was also true for the same models not featuring the additional covariates. Such margins are only negligibly affected by the level of normalized power output or how far into the current session the athlete is.

The additional covariates do have a much greater effect on the power output an athlete can achieve. Optimum cadence appears to vary more with heart rate than it does with other additional covariates. The reduced optimum cadence at high heart rate may suggest that optimum cadence decreases as the athlete exerts more effort. Indeed when cadence is interacted with heart rate in the model in equation 11, for the high heart rate data subset, and heart rate is fixed to a range of different possible values for both athletes 1 and 2, optimum cadence decreases with heart rate. Also the margin of decrease in optimum cadence decreases with each increase in heart rate. As heart rate increases, he is unable to sustain such a high cadence, so the highest cadence he can sustain without hinderence to his riding technique reduces.

For athletes 3 and 4 in our analysis, no optimum cadences are yielded from non-linear models, even for the high heart rate data subset. We believe this is because athletes 3 and 4 may not have ridden at a wide enough range of cadences for a mathematical model to yield a cadence that optimises power output.

4.7.4. Discussion of optimum cadences in relation to literature, preferred cadence and riding mode.

The optimum cadences from most ergometer studies tend to be up to 90 rpm (Coast and Welch, 1985), though higher optimum cadences can be expected when using field data (Foss and Hallen, 2004), or when testing elite athletes The optimum cadences yielded from this study range from 68 to 120, although these are dependent on data processing constants. Foss and Hallen (2004) claim that cadences adopted by elite athletes are higher than those typically found in ergometer studies, around 90-105. Ergometer studies often include cyclists that are not necessarily elite athletes, or who may be triathletes (and therefore not necessarily as highly trained in cycling as are the athletes whose sole professional sport in cycling). In our study using field data we use competitive but not elite athletes – indeed optimum cadences for a non-linear model with sensible choice of data processing constants yields optimum cadences that are greater than some ergometer studies but less than those claimed by Foss and Hallen (2004), at around 83 and 70 rpm for athletes 1 and 2 respectively.

We also find some differences between optimum cadences for each athlete, supporting suggestions made by Coast and Welch (1985) that athletes may have their own individual optimum cadences. They claimed that this may be due to different individual skill levels. Coast and Welch (1985) suggested that increased athlete skill enabled athletes in their study to sustain higher cadences, of 91 rpm, than athletes in a study by Hagberg (1981), who achieved 83 rpm. Assuming no smoothing of raw power output, heart rate and cadence measurements (i.e. when k=1 in our study), cadences from our study are not as high as the 91 rpm found in the Coast and Welch study. Given the competitive but not elite levels of skill of the athletes involved in our study, it may be that the athletes in our study are not skilled enough to sustain higher cadences.

Differences between optimum cadences found in ergometer studies and those found in field data may be due to who devises the test or schedule. In ergometer studies the cadences at which athletes are tested tend to be chosen by academics (Coast and Welch, 1985). In the field data in our study, the training schedule is likely to be devised by coaches and tailored to the individual athlete, as the coaches are likely to have significant knowledge and understanding of the athlete's skill and objectives, and it is the athlete's success that forms the basis of the schedule, not necessarily knowledge that an academic can obtain. The high optimum cadences in our study compared to some ergometer studies may therefore suggest that field data may be useful in calculating an individual optimum cadence for competitive cyclists. The environmental differences between road cycling and ergometer tests (Jobson et al, 2007) may have been a factor in optimum cadences from this study being higher than those found in some ergometer studies.

Alternatively differences in optimum cadences may also represent higher levels of fitness; if an athlete's optimum cadence is higher than another athlete, the athlete with the higher fitness is able to maintain a higher cadence without tiring or his technique deteriorating – indeed each athlete trains according to his own individual training schedule, with each training schedule featuring different frequencies and durations of sessions. Athletes may be exerting great physical effort in training, but may not have developed a high enough level of fitness to sustain very high cadences; by the time they enter a major competition after the training schedule, their fitness may have improved to a high enough level to sustain such high cadences. However whether skill or fitness is the reason, they effectively have the same effect on cadence as they both appear to increase the cadence at which an athlete can ride.

Whilst optimum cadence does not vary with power output or heart rate in our study, athletes' typically preferred cadences do to some extent. We calculated the range, mean and mode cadence for athletes 1 and 2 for periods where power output was between 100 and 149W, 150 – 299W and so on, with the final group up to above 500W. We present these preferred cdences in table 60. Broadly, from 200W up to above 500W, as power output increases, preferred cadence (mean and mode) increase a little. Athletes therefore appear to favour slightly higher cadences at higher power output. Also, at the power output corresponding to the optimum cadence in the model (188W for athlete 1 and 190W for athlete 2), preferred cadences are similar for both athletes. Indeed for particlularly low power output (under 100-149W), preferred cadences are slightly higher than for power outputs of 200-250W. For athlete 2, mean chosen cadence at 100-149W is higher than mean cadence for any other range of power outputs.

Athlete 1 tends to ride just above his optimum cadence, with mean cadences for different ranges of power output between 82 and 92. Athlete 2 tends to ride slightly closer to his optimum cadence, with mean preferred cadences within 5 revolutions per minute either side of his optimum cadence of 70.

Nevertheless the margin of difference between optimum cadence and between preferred cadence are similar between athletes. Optimum cadence is higher for athlete 1 than for athlete 2 in our analysis (83 rpm comprared to 70 rpm), whilst athlete 1 also typically chooses to ride at a higher cadence (approximately 82-92 rpm) than does athlete 2 (approximately 65-75 rpm). (The ranges of 82-92 and 65-70 come from the mean chosen cadence calculated at different power outputs, between 150-199W, and 200-249W and so on). There may be physiological differences that result in different optimum cadences between athletes, such as muscle fibre type. Such physiological mechansisms may also result in different choices of optimum cadence.

We do not consider in this study whether athletes are standing or sitting on the seat (which we term riding mode), as no such information is available. An optimum cadence may vary depending on whether an athlete is standing or sitting, as different muscle groups may be being used for standing and sitting. For such information to be available in training sessions at 5 second intervals (to be consistent with power output, heart rate and cadence measurements), a pressure pad monitor would need to be developed for the saddle.

Table 60: Preferred cadences of athletes 1 and 2, for different power outputs

Power		Athlete 1 Cadence Range (Lowest to highest)		Mean	Mode	Athlete 2 Cadence Range (Lowest to highest)		Mean	Mode
100-	149	43	120	91	91	30	129	75	87
150-	199	30	127	88	92	29	125	73	75
200-	249	34	117	84	90	29	122	69	70
250-	299	42	178	82	96	29	118	65	56
300-	349	49	115	87	96	35	121	66	55
350-	399	50	124	88	96	39	124	68	57
400-	449	53	139	91	94	37	101	66	59
450-	499	65	128	92	86	43	92	66	58
500	+	69	134	89	94	54	139	73	57

4.7.5. Summary of discussion chapter

We have fitted a range of linear and non-linear models of power output, heart rate and cadence. Linear models do not appear useful in yielding a mathematical optimum cadence. However non-linear models do yield mathematical optimum cadences for two of the four athletes in our study, providing we retain only sampled data in which heart rate is higher than the mean overall heart rate for each athlete in our study. However additional variables (such as normalized power output) appear to have little impact on optimum cadence. As such our preferred model from chapter 4 is the simple non-linear model of power output, heart rate and cadence (equation 7). In chapter 5 we further develop this model (equation 7) by including training impulses, to represent the accumulation of fatigue.

Chapter 5: Sensitivity to data processing constants

5.1. Introduction to sensitivity analysis

We now consider what combination of data processing constants should be used for the data presented to us (the power output, heart rate and cadence measurements yielded at 5 second intervals). These are the width of a moving average smoothing window (k) and the time lag between power output and heart rate response (l). We have thus far used a moving average with window width 12 time units (60 seconds) to calculate smoothed power output, cadence and heart rate values, lagged heart rate at 1 time unit (5 seconds) behind power output and cadence, and then sampled the smoothed data every 24 time units (120 seconds). This processing is denoted by k=12, l=1 and m=24 respectively. We chose data processing constants that seemed plausible to us based on our knowledge and understanding of sports science literature at that time. We then also decided to use only data which we felt closely represented race conditions. We calculated the mean of all sampled heart rate measurements (which we term the overall mean heart rate for each athlete). We retained data in which heart rate is greater than the overall mean heart rate (which we termed the high heart rate data subset), to exclude data from very long, low-intensity training sessions. We now consider analysing the effects of different data processing constants in the simple non-linear model (equation 7).

We explore the sensitivity of the simple non-linear model (equation 7) to different data processing constants, for the high heart rate data subset.

We present results of coefficient estimate, AIC and R^2 for different smoothing window widths and heart rate lags, along with optimum cadences. We present power / cadence plots for selected data processing constants. We finally discuss the results, evaluating which set of data processing constants are most appropriate for our data.

5.2. Methodology for sensitivity analysis

We continue to sample measurements once every 24 time units – once every 5 seconds (denoted as m=24). (We consider smoothing data using a moving average with window widths of various time units; the window width is denoted by k - we consider k = 3, 6, and 12. We also considere the effects of not smoothing the data – this is represented by k=1 (in this case, we simply sample the power output, heart rate and cadence measurements once every 24 time units). We also consider different heart rate lags. Since literature suggests that heart rate responds to changes in power output, we lag heart rate at some time after power output. Stirling et al (2008) suggested that for large changes in power output, heart rate takes 60 seconds to respond. For our data changes in power output are much more gradual, so we chose a heart rate lag of 5 seconds. The heart rate lag, l, represents by how many time units heart rate lags behind power output. For this sensitivity analysis we initially considered l = 1, 2 and 3, but since explanatory power appears to increase with heart rate lag we also consider longer lags, 4, 6 and 12 (for window widths of 1 and 3 time units). We observe the effects of different data processing constants on explanatory power, optimum cadence, power output at optimum cadence, coefficient estimates and standard errors of the coefficient estimates. Note that explanatory power indicates the amount of variation in changes in power output that can be explained by variation in changes in heart rate and cadence. The usefulness of the fitted curve in predicting or providing an optimum cadence may not therefore necessarily coincide with high explanatory power.

5.3. Results of sensitivity analysis

5.3.1. Coefficient estimates and model fit

We present tables of coefficient estimates and their standard errors, in tables 61 to 64 (one table for each athlete). As k increases (that is the number of measurements included in a moving average increases), we expect covariates to fit less well. However with a greater window width there is less noise within the data, which could increase explanatory power.

For all combinations of data processing constants (window width k, and heart rate lag l), the model (equation 7) fits better for athletes 1 and 2 than for athletes 3 and 4. That is, the coefficient estimates for α and β are always positive for athletes 1 and 2, which means a mathematical optimum cadence is always yielded for these athletes. (Note this is for high heart rate data.) For athletes 3 and 4 however, the polarity of α and β are such that no statistical optimum cadences are yielded for these athletes for any combination of data processing constants.

For athletes 3 and 4, explanatory power is extremely low for all data processing constants. This could be because β is always negative, whilst for athlete 4 (for whom explanatory power is even lower than it is for athlete 3) α is often very low. The negative coefficients for β indicate a poor power output cadence fit for these athletes, hence only a small amount of the variation in fitted power output can be accounted for by variation in cadence or heart rate.

For a given window width (k), explanatory power tends to increase with heart rate lag (l). This suggests there is a better fit of covariates for higher heart rate lags. However we also observe that standard errors tend to increase with l for all covariates α, β, γ and μ, implying a greater fit of covariates for low heart rate lags. This may suggest a better model fit for lower heart rate lags. The increase in explanatory power with increasing heart rate lag could be an artefact of the model fitting, with possibly less noise for higher heart rate lags. The explanatory power and standard errors do not suggest there is one heart rate lag that yields the best fit of the covariates.

We expect there to be less noise in the data when smoothing power output, cadence and heart rate (when k = 3, 6 and 12). As k increases however, explanatory power decreases markedly. As k increases standard errors increase for α and β – the coefficients do not fit well for high window widths. Therefore the low explanatory power for high window widths could be due to a very poor fit of data at high window widths. However standard errors decrease with k for γ and $\log\mu$. Since explanatory power decreases as k increases, it therefore appears that standard errors for α and β have a stronger influence on the explanatory power of the model than do standard errors for γ and $\log\mu$.

For low window widths (k=1 or 3), the covariance beween α and β increases with heart rate lag (l). For the longest window width (k=12), it is less clear as covariance increases with l for athlete 4 but decreases with l for athletes 2 and 3.

Table 61: Coefficient estimates and standard errors for the non-linear model (equation 7) fitted for athlete 1, using a range of data processing constants k (moving average window width) and l (heart rate lag).

k	l	α	β	γ	μ	$\log \mu$	$cov(\alpha,\beta)$	R^2	AIC
1	1	1.58 (0.01)	0.020 (0.0005)	0.91 (0.06)	0.01	-4.59 (0.29)	0.000005	84.0	69664
1	2	1.56 (0.01)	0.020 (0.0005)	0.90 (0.06)	0.01	-4.51 (0.31)	0.000006	84.2	70919
1	3	1.56 (0.01)	0.020 (0.0005)	0.90 (0.06)	0.01	-4.53 (0.33)	0.000006	84.7	71547
1	4	1.55 (0.01)	0.020 (0.0005)	0.78 (0.06)	0.02	-3.91 (0.33)	0.000006	85.1	71846
1	6	1.53 (0.01)	0.018 (0.0006)	0.77 (0.07)	0.02	-3.88 (0.33)	0.000006	85.9	72600
1	12	1.51 (0.01)	0.017 (0.0006)	0.71 (0.07)	0.03	-3.55 (0.35)	0.000007	85.8	72727
3	1	1.49 (0.02)	0.014 (0.0004)	1.25 (0.05)	0.00	-6.44 (0.27)	0.000005	63.7	68233
3	2	1.47 (0.01)	0.013 (0.0004)	1.08 (0.06)	0.00	-5.60 (0.28)	0.000005	66.6	69354
3	3	1.45 (0.01)	0.012 (0.0004)	1.00 (0.06)	0.01	-5.23 (0.29)	0.000005	66.9	70194
3	4	1.42 (0.01)	0.011 (0.0004)	1.00 (0.06)	0.01	-5.13 (0.30)	0.000005	69.3	70846
3	6	1.39 (0.01)	0.010 (0.0004)	0.92 (0.06)	0.01	-4.74 (0.31)	0.000005	71.0	71436
3	12	1.36 (0.01)	0.009 (0.0004)	0.78 (0.06)	0.02	-4.01 (0.33)	0.000005	73.4	71169
6	1	1.68 (0.03)	0.018 (0.0005)	1.41 (0.04)	0.00	-7.69 (0.22)	0.000010	43.9	63180
6	2	1.69 (0.02)	0.018 (0.0005)	1.10 (0.04)	0.00	-6.19 (0.23)	0.000011	43.7	64333
6	3	1.58 (0.02)	0.015 (0.0004)	1.23 (0.05)	0.00	-6.62 (0.24)	0.000009	48.6	65045
12	1	1.77 (0.04)	0.020 (0.0006)	1.56 (0.03)	0.00	-8.78 (0.20)	0.000020	37.2	58178
12	2	1.78 (0.04)	0.020 (0.0006)	1.39 (0.03)	0.00	-7.95 (0.21)	0.000020	35.9	58641
12	3	1.79 (0.04)	0.019 (0.0006)	1.44 (0.04)	0.00	-8.30 (0.21)	0.000020	36.9	59179

Table 62: Coefficient estimates and standard errors for the non-linear model (equation 7) fitted for athlete 2, using a range of data processing constants k (moving average window width) and l (heart rate lag).

k	l	α	β	γ	μ	$\log \mu$	$cov(\alpha,\beta)$	R^2	AIC
1	1	1.66 (0.02)	0.024 (0.001)	0.75 (0.08)	0.024	-3.71 (0.42)	0.00001	74.7	50095.2
1	2	1.65 (0.02)	0.024 (0.001)	0.84 (0.09)	0.016	-4.15 (0.44)	0.00001	75.9	50659.5
1	3	1.64 (0.02)	0.024 (0.001)	0.83 (0.09)	0.017	-4.07 (0.45)	0.00001	75.9	50428
1	4	1.63 (0.02)	0.024 (0.001)	0.78 (0.10)	0.022	-3.83 (0.48)	0.00002	75.8	51223.9
1	6	1.61 (0.02)	0.023 (0.001)	0.61 (0.10)	0.050	-3.00 (0.49)	0.00002	76.5	51361.6
1	12	1.55 (0.02)	0.020 (0.001)	0.28 (0.11)	0.257	-1.36 (0.52)	0.00002	75.7	52217.6
3	1	1.57 (0.02)	0.016 (0.001)	1.30 (0.09)	0.001	-6.60 (0.44)	0.00001	59.0	50461.6
3	2	1.55 (0.02)	0.015 (0.001)	1.14 (0.09)	0.003	-5.80 (0.46)	0.00001	59.8	50841.5
3	3	1.50 (0.02)	0.014 (0.001)	1.09 (0.10)	0.004	-5.50 (0.48)	0.00001	62.8	51249.4
3	4	1.48 (0.02)	0.013 (0.001)	1.09 (0.10)	0.004	-5.48 (0.49)	0.00001	64.5	51755.9
3	6	1.45 (0.02)	0.011 (0.001)	0.80 (0.10)	0.018	-4.04 (0.51)	0.00001	65.9	51986.9
3	12	1.36 (0.02)	0.007 (0.001)	0.33 (0.11)	0.173	-1.76 (0.55)	0.00002	64.4	52796.8
6	1	1.67 (0.03)	0.019 (0.001)	1.42 (0.08)	0.001	-7.35 (0.39)	0.00002	49.9	48534.3
6	2	1.62 (0.03)	0.017 (0.001)	1.38 (0.08)	0.001	-7.13 (0.41)	0.00002	52.1	49073.2
6	3	1.59 (0.03)	0.016 (0.001)	1.27 (0.09)	0.001	-6.58 (0.43)	0.00002	53.6	49547.5
12	1	1.84 (0.03)	0.024 (0.001)	1.57 (0.06)	0.000	-8.42 (0.33)	0.00002	46.5	45672.3
12	2	1.79 (0.03)	0.023 (0.001)	1.46 (0.07)	0.000	-7.76 (0.34)	0.00002	46.2	46136.6
12	3	1.75 (0.03)	0.022 (0.001)	1.37 (0.07)	0.001	-7.25 (0.35)	0.00002	47.4	46684.8

Table 63: Coefficient estimates and standard errors for the non-linear model (equation 7) fitted for athlete 3, using a range of data processing constants *k* (moving average window width) and *l* (heart rate lag).

k	*l*	α	β	γ	μ	$\log \mu$	$cov(\alpha,\beta)$	R^2	AIC
1	1	1.17 (0.03)	-0.001 (0.001)	1.06 (0.14)	0.01	-5.19 (0.68)	0.00004	58.9	32595.4
1	2	1.13 (0.03)	-0.003 (0.001)	1.05 (0.14)	0.01	-5.14 (0.70)	0.00004	61.2	33023.7
1	3	1.10 (0.03)	-0.004 (0.001)	1.11 (0.14)	0.00	-5.38 (0.71)	0.00004	61.5	33175.1
1	4	1.05 (0.03)	-0.007 (0.001)	1.13 (0.15)	0.00	-5.48 (0.72)	0.00004	62.4	33366.4
1	6	0.99 (0.03)	-0.010 (0.001)	1.23 (0.15)	0.00	-5.95 (0.75)	0.00005	63.5	33228.3
1	12	0.95 (0.04)	-0.011 (0.002)	0.92 (0.15)	0.01	-4.45 (0.75)	0.00005	63.6	33155.5
3	1	0.87 (0.06)	-0.006 (0.001)	1.43 (0.14)	0.00	-6.08 (0.69)	0.00006	27.0	32504.3
3	2	0.90 (0.05)	-0.008 (0.001)	1.47 (0.14)	0.00	-6.54 (0.69)	0.00005	32.1	32661.9
3	3	0.84 (0.05)	-0.010 (0.001)	1.58 (0.14)	0.00	-7.05 (0.71)	0.00005	33.9	33017.2
3	4	0.86 (0.05)	-0.011 (0.001)	1.52 (0.14)	0.00	-6.88 (0.72)	0.00005	37.2	32845.4
3	6	0.88 (0.05)	-0.012 (0.001)	1.50 (0.14)	0.00	-7.04 (0.73)	0.00005	39.9	32992.8
3	12	0.82 (0.04)	-0.014 (0.001)	1.41 (0.15)	0.00	-6.48 (0.76)	0.00004	44.0	33248.3
6	1	0.14 (0.08)	-0.013 (0.001)	1.50 (0.12)	0.01	-4.40 (0.65)	0.00010	15.0	31382.2
6	2	0.21 (0.08)	-0.013 (0.001)	1.64 (0.12)	0.01	-4.77 (0.66)	0.00010	17.4	31554.2
6	3	0.19 (0.08)	-0.015 (0.001)	1.57 (0.13)	0.01	-4.44 (0.67)	0.00010	18.2	31437.8
12	1	0.68 (0.24)	-0.0007 (0.003)	1.66 (0.11)	0.00	-5.80 (0.96)	0.00077	10.9	29759.6
12	2	0.67 (0.24)	-0.002 (0.003)	1.62 (0.11)	0.00	-5.66 (0.95)	0.00075	11.6	29748.7
12	3	0.16 (0.24)	-0.010 (0.002)	1.58 (0.11)	0.02	-3.91 (0.68)	0.00025	12.9	30002.4

Table 64: Coefficient estimates and standard errors for the non-linear model (equation 7) fitted for athlete 4, using a range of data processing constants *k* (moving average window width) and *l* (heart rate lag).

k	*l*	α	β	γ	μ	$\log \mu$	$cov(\alpha,\beta)$	R^2	AIC
1	1	0.15 (0.03)	-0.0007 (0.002)	1.47 (0.14)	0.07	-2.59 (0.69)	0.000050	8.7	82794.1
1	2	0.09 (0.03)	-0.006 (0.002)	0.09 (0.16)	0.02	-4.08 (0.77)	0.000005	9.2	83475.8
1	3	0.17 (0.03)	-0.001 (0.002)	1.37 (0.15)	0.10	-2.26 (0.75)	0.000058	10.1	83641
1	4	0.16 (0.03)	-0.002 (0.002)	1.28 (0.15)	0.16	-1.83 (0.76)	0.000061	10.3	84039.3
1	6	0.18 (0.04)	-0.002 (0.002)	1.47 (0.16)	0.06	-2.84 (0.79)	0.000064	11.1	84747
1	12	0.20 (0.04)	-0.002 (0.002)	1.72 (0.17)	0.01	-4.25 (0.82)	0.000075	12.6	85994
3	1	0.20 (0.02)	-0.013 (0.001)	1.23 (0.08)	0.41	-0.88 (0.42)	0.000012	8.2	72127
3	2	0.21 (0.02)	-0.014 (0.001)	1.25 (0.09)	0.37	-1.00 (0.45)	0.000013	8.5	73441.3
3	3	0.21 (0.02)	-0.015 (0.001)	1.25 (0.10)	0.35	-1.05 (0.48)	0.000015	8.7	74486.2
3	4	0.22 (0.02)	-0.016 (0.001)	1.21 (0.10)	0.41	-0.90 (0.50)	0.000016	8.7	75677.9
3	6	0.21 (0.02)	-0.016 (0.001)	1.31 (0.11)	0.22	-1.50 (0.54)	0.000018	9.2	76486.9
3	12	0.24 (0.02)	-0.019 (0.001)	1.33 (0.12)	0.18	-1.72 (0.60)	0.000022	9.7	79057.6
6	1	0.21 (0.01)	-0.012 (0.001)	1.32 (0.06)	0.32	-1.14 (0.30)	0.000006	13.0	65271.8
6	2	0.22 (0.01)	-0.013 (0.001)	1.31 (0.06)	0.32	-1.14 (0.32)	0.000007	12.6	66640
6	3	0.25 (0.01)	-0.014 (0.001)	1.20 (0.07)	0.53	-0.64 (0.03)	0.000007	12.3	21766.1
12	1	0.14 (0.01)	-0.008 (0.001)	1.51 (0.05)	0.13	-2.07 (0.24)	0.000004	16.7	61043.4
12	2	0.16 (0.01)	-0.009 (0.001)	1.48 (0.05)	0.15	-1.91 (0.24)	0.000005	17.0	61189.3
12	3	0.18 (0.01)	-0.010 (0.001)	1.42 (0.05)	0.20	-1.61 (0.25)	0.000005	17.2	61573.7

Although literature suggests a low heart rate lag is appropriate, it is not necessarily clear exactly what the lag should be from literature. Also, model performance parameters (explanatory power and standard errors of coefficients) do not suggest that there is one heart rate lag that necessarily yields the best fit of covariates. Therefore we present optimum cadences for the model (equation 7) for all window width and heart rate lag combinations.

5.3.2. Optimum cadences

Tables 65 and 66 contain the optimum cadence (C*), 95% confidence intervals and corresponding power output for the optimum cadence, for various smoothing window widths (*k*) and various heart rate lags (*l*).

When we do not smooth data (*k*=1), plausible optimum cadences are yielded for all heart rate lags tested. For athletes 1 and 2, optimum cadence increases with increasing heart rate lag. That is, the longer the time lag between changing power output and heart rate response, the greater the pedalling rate an athlete can theoretically sustain, according to our model. For smoothed data (*k*=3, 6 and 12) however, optimum cadences are much higher. The optimum cadences become unrealistically high for very high heart rate lags.

Indeed for athletes 1 and 2 variance in optimum cadence increases with heart rate lag for all window widths. The increase in variance with each heart rate lag is quite low (around 0.01 to 0.05); there is nevertheless a clear trend for increasing variance of optimum cadence with increasing heart rate lag for athletes 1 and 2. The increase in variance of optimum cadence indicates wider confidence intervals; therefore confidence intervals for the mathematical optimum cadence increase with increasing heart rate lag. Wider confidence intervals suggest the athletes can ride at a wider range of cadences without detriment to performance. Nevertheless 95% confidence intervals are quite narrow for all combinations of data processing contants in the models. This suggests that the choice of cadence is very important for maximising power output (for a given heart rate).

Table 65: Explanatory power output (R^2), optimum cadence (C*) and corresponding power output at optimum cadence (P*) for the non-linear model (equation 7) fitted for athlete 1, using a range of data processing constants *k* (moving average window width) and *l* (heart rate lag).

k	*L*	C*	C* standard Deviation	C* variance	Confidence From:	Interval To:	P*
1	1	75.97	0.24	0.06	75.50	76.44	196.8
1	2	79.56	0.28	0.08	79.01	80.11	192.4
1	3	79.36	0.29	0.09	78.78	79.93	190.7
1	4	79.43	0.30	0.09	78.84	80.01	190.0
1	6	83.42	0.34	0.12	82.75	84.08	187.8
1	12	86.85	0.39	0.15	86.08	87.62	186.1
3	1	106.45	0.35	0.12	105.76	107.14	199.7
3	2	113.81	0.41	0.17	113.00	114.61	199.6
3	3	120.18	0.47	0.22	119.27	121.10	204.4
3	4	127.38	0.53	0.28	126.34	128.43	209.7
3	6	140.38	0.64	0.41	139.12	141.65	213.0
3	12	155.57	0.79	0.63	154.01	157.12	218.1
6	1	91.68	0.23	0.05	91.23	92.14	202.8
6	2	92.19	0.24	0.06	91.73	92.65	197.2
6	3	103.84	0.32	0.10	103.22	104.47	202.2
12	1	90.48	0.22	0.05	90.06	90.90	189.6
12	2	90.83	0.23	0.05	90.37	91.28	190.7
12	3	95.11	0.25	0.06	94.61	95.60	193.9

Table 66: Explanatory power output (R^2), optimum cadence (C*) and corresponding power output at optimum cadence (P*) for the non-linear model (equation 7) fitted for athlete 2, using a range of data processing constants *k* (moving average window width) and *l* (heart rate lag).

k	*L*	C*	C* standard Deviation	C* variance	Confidence From:	Interval To:	P*
1	1	68.35	0.35	0.12	67.67	69.02	214.0
1	2	68.10	0.36	0.13	67.39	68.81	205.3
1	3	68.29	0.37	0.14	67.55	69.02	200.8
1	4	68.57	0.39	0.15	67.81	69.34	195.8
1	6	70.41	0.44	0.19	69.55	71.26	189.6
1	12	76.09	0.55	0.30	75.01	77.16	179.2
3	1	96.67	0.58	0.34	95.53	97.80	237.3
3	2	100.98	0.65	0.43	99.70	102.26	237.1
3	3	109.67	0.80	0.64	108.10	111.23	222.0
3	4	116.83	0.91	0.84	115.04	118.62	222.1
3	6	130.11	1.14	1.29	127.88	132.34	243.1
3	12	190.35	2.35	5.54	185.73	194.96	299.5
6	1	89.12	0.44	0.19	88.26	89.97	226.8
6	2	96.04	0.53	0.28	95.01	97.08	228.5
6	3	102.15	0.62	0.38	100.94	103.36	221.8
12	1	75.85	0.26	0.07	75.33	76.36	231.8
12	2	77.95	0.29	0.09	77.37	78.52	222.1
12	3	81.21	0.33	0.11	80.56	81.85	214.6

We next present fitted power / cadence plots to provide a graphical demonstration of the power output cadence relationship implied by the model. We present power output cadence plots for *k*=1 (that is, power output, heart rate and cadence data are not smoothed), for a selection of heart rate lags (*l*=1, 2 and 6, or 5, 10 and 30 seconds respectively) in figure 17. For each figure, we present plots for athlete 1 on the left and athlete 2 on the right hand side. The first row is for the model (equation 7) fitted with a heart rate lag of 5 seconds (*l*=1), with increasing heart rate lag for each subsequent row in the figure (lags of 10 and 30 seconds).

Fitted power output cadence plots suggest a power output cadence relationship that provides a good representation of the true power output cadence relationship. As heart rate lag increases, so too does optimum cadence. For very long heart rate lags (*l*=12), optimum cadences are not plausible – that is, a mathematical optimum cadence yielded from the model (equation 7) is greater than any cadence at which the athlete has actually ridden during his training sessions.

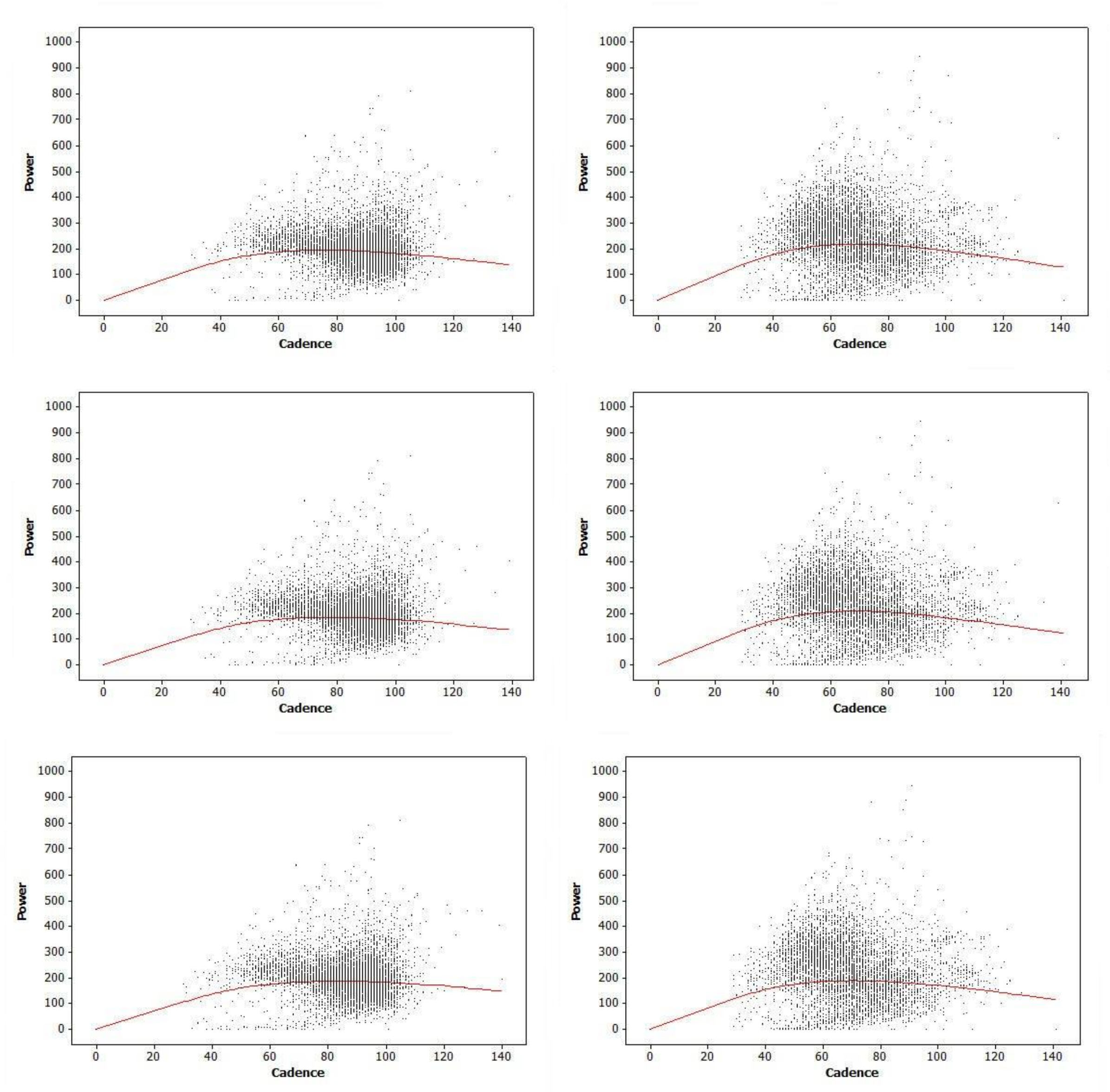


Figure 17: Plots of power output measurements (black dots) and fitted power output (red line) from the non-linear model (equation 7) fitted for athletes 1 (left) and 2 (right). Data are not smoothed (k=1). Heart rate (l) is lagged, from top to bottom, at 1, 2, and 6 time units (5, 10, then 30 seconds respectively).

5.3.3. Deviation from optimum cadence

We present a table for reduction in power output for sub-optimal cadences when data are not smoothed (k=1), for a selection of heart rate lags (l=1, 2, and 6), as low heart rate lags (l=6 or below) are most appropriate for our data. We fit the model (equation 7) for different data processing constants for the high heart rate data subset. We present expected power output for a range of sub-optimal cadences in table 67.

For all values of k (the size of the window width or moving average smoothing interval), the reduction in power output for sub-optimal cadences is greater for lower heart rate lags (l). When there is no smoothing (k=1), the margin of the reduction in power output for sub-optimal cadences is not particularly sensitive to heart rate lag, but nonetheless it decreases as *l* increases.

Table 67: Differences in fitted power output for athlete 1 (left) and athlete 2 (right) using the simple non-linear model (equation 7) for unsmoothed data (*k*=1) and various heart rate lags (*l*), for various cadences above and below the optimum cadence.

change in C from C*	cadence	Expected power output	change in power output	% change in power output
l=1				
-20	56.0	184.0	-12.7	-6.5
-10	66.0	193.8	-2.9	-1.5
-5	71.0	196.1	-0.7	-0.4
	76.0	196.8		
5	81.0	196.1	-0.6	-0.3
10	86.0	194.3	-2.5	-1.3
20	96.0	187.8	-9.0	-4.6
l=2				
-20	59.6	181.3	-11.1	-5.8
-10	69.6	189.8	-2.6	-1.3
-5	74.6	191.8	-0.6	-0.3
	79.6	192.4		
5	84.6	191.8	-0.6	-0.3
10	89.6	190.2	-2.2	-1.1
20	99.6	184.5	-8.0	-4.1
l=6				
-20	63.4	178.2	-9.6	-5.1
-10	73.4	185.5	-2.2	-1.2
-5	78.4	187.2	-0.5	-0.3
	83.4	187.8		
5	88.4	187.3	-0.5	-0.3
10	93.4	185.9	-1.9	-1.0
20	103.4	180.8	-7.0	-3.7

change in C from C*	cadence	Expected power output	change in power output	% change in power output
l=1				
-20	48.4	195.8	-18.2	-8.5
-10	58.4	209.9	-4.2	-2.0
-5	63.4	213.0	-1.0	-0.5
	68.4	214.0		
5	73.4	213.1	-0.9	-0.4
10	78.4	210.6	-3.4	-1.6
20	88.4	201.6	-12.4	-5.8
l=2				
-20	48.1	187.8	-17.5	-8.5
-10	58.1	201.2	-4.0	-2.0
-5	63.1	204.3	-1.0	-0.5
	68.1	205.2		
5	73.1	204.4	-0.9	-0.4
10	78.1	201.9	-3.3	-1.6
20	88.1	193.4	-11.9	-5.8
l=6				
-20	50.4	174.9	-14.7	-7.7
-10	60.4	186.2	-3.4	-1.8
-5	65.4	188.8	-0.8	-0.4
	70.4	189.6		
5	75.4	188.8	-0.7	-0.4
10	80.4	186.8	-2.8	-1.5
20	90.4	179.5	-10.1	-5.3

5.4. Discussion of sensitivity analysis

Having found a plausible range of optimum cadences for various non-linear models of power output and cadence, we re-visited the issue of data processing. We considered a range of window widths (the length of the moving average smoother for power output, cadence and heart rate; *k*) and heart rate lags (*l*) for processing the power output, heart rate and cadence measurements given to us.

Literature suggests that a low heart rate lag (only a brief delay between changes in power output and the heart rate response of the athlete) is appropriate for the data in our study (Stirling et al, 2008), say less than 30 seconds (*l*=6 or less). This means that, for small changes in power output, heart rate must increase accordingly very soon afterwards. Nevertheless we explored models with different heart rate lags. Model performance criteria (such as explanatory power and standard erors of coefficients) do not necessarily suggest there is one heart rate lag that necessarily yields the best fit of covairates. Whilst explanatory power tends to increase with *l*, implying a better fit for higher heart rate lags, standard errors for coefficients also increase with *l*, implying a better fit for lower heart rate lags. The increase in explanatory power with increasing heart rate lag could be an artefact of the model fitting, with possibly less noise for higher heart rate lags.

When we do not smooth data (denoted by k=1), optimum cadences range between 68 to 83, which is simlar to those found in literature (Foss and Hallen, 2004). However when k is greater than 1 (that is, when data are smoothed using a moving average), plausible mathematical optimum cadences are not always yielded. The mathematical optimum cadence decreases as k increases – the greater the window width, the lower the optimum cadence. The mathematical optimum cadence yielded by the model also increases with heart rate lag. Consequently mathematical optimum cadences for smoothed data (k = 3, 6 or 12) are sometimes unrealistically high for all but the lowest heart rate lags. In such circumstances the theoretical optimum cadence produced is extremely high (higher than any cadence at which the athlete has ridden in his training), and almost certainly greater than any athlete would be able to sustain. It may therefore be inappropriate to smooth the data in our study.

When we do not smooth data, any low heart rate lag (say less than 30 seconds, or l=6 or less) appears appropriate for our data – it supports literature that suggest a low heart rate lag, and yields plausible optimum cadences. For smoothed data, a heart rate lag of 5 seconds appears a reasonable choice - when heart rate lag is equal to 5 seconds (l=1), plausible optimum cadences are yielded whatever the smoothing window width (k), though are highly sensitive to smoothing window width. These optimum cadences for l=1 range from 68 to 110.

We considered smoothing data to reduce sudden changes and fluctuations in power output, heart rate and cadence. However sudden changes in heart rate do not occur to the same extent as they do in power output and cadence data. Explanatory power increases markedly as k decreases (as the smoothing window width decrease), and is highest when we do not smooth. For example, for a heart rate lag of 5 seconds (l=1), explanatory power for athlete 1 is 84, whilst it is only 63.7 for a window width of 15 seconds (k=3), 43.9 for k=6 and 37.2 for k=12. Plausible optimum cadences are yielded for a range of heart rate lags when we do not smooth data. Smoothing data may therefore not be necessary.

Patterns in optimum cadence variance are similar to patterns in optimum cadence. For both athletes 1 and 2 variance of optimum cadence increases with heart rate lag l. For k = 1, there is only marginal increase in optimum cadence variance with l. However, as k increases (i.e. when we smooth data using the moving average), optimum cadence variance increases by a much greater margin with increases in l (similarly to optimum cadence itself). Also, optimum cadence variance decreases with k (not including k=1). Confidence intervals for optimum cadence are therefore wider when we consider a longer time lag between changes in power output and heart rate response. Nonetheless all confidence intervals appear quite narrow, implying there is quite a narrow range of cadences at which the athlete can ride optimally.

We also calculated the percentage reduction in power output for statistically sub-optimal cadences. Reduction in fitted power output tends to coincide with confidence intervals; as one increase, so does the other. For each heart rate lag, fitted power output is lower for a given intervals below statistically optimum cadence than for intervals above statistically optimum cadence. This is not surprising, as the fitted power output cadence plots show a curve that increases, reaches a peak, and then reduces steadily.

Reduction in power output for sub-optimal cadences tends to be greater for athlete 2 than for athlete 1, typically by a margin of 30 – 50% (for example, for k=1, l=1, and a cadence 20 rpm below optimum, percentage reduction in power output is 6.5% for athlete 1 but 8.5% for athlete 2). The only exception to the athlete effect is for a moving average smoothing window of 30 seconds or 6 measurements (k=6), where reduction in power output tends to be of a similar magnitude for both athletes. However for this window width, unlike for other window widths, optimum cadences are very similar for both athletes. Indeed there is a clear trend for each athlete that as fitted power output and optimum cadence increase, the relative reduction in power output decreases. It is difficult to infer from this whether the fitted power output or the optimum cadence has the greatest impact on the margin of reduction in power output for sub-optimal cadences.

However, as heart rate lag (l) increases, relative reduction in power output for sub optimal cadences increases, along with a decrease in mathematical optimum cadence, but fitted power output does not always change. Across all combinations of data processing constants, for both athletes, the margin of the relative reduction in power output for sub-optimal cadences always reduces as the mathematical optimum cadence increases. However the margin of decrese in power output (for sub-optimal cadences) does not always reduce when fitted power output increases. We could infer that the margin of reduction in power output for sub-optimal cadences is merely a function of how high the optimum cadence is and not a function of the magnitude of the fitted power output. However, the trend for reduction in power output for sub-optimal cadences to not necessarily coincide with the magnitude of the fittted power output only occurs when we explore different heart rate lags beyond the 5 second lag we consider most appropriate from literature. We may alternatively infer therefore that this trend is merely an effect of the heart rate lag.

Nevertheless reduction in power output for sub-optimal cadences differs between the athletes. However this difference between the athletes tends to coincide with the nature of the differences between the athletes in terms of their optimum cadences and magnitude of fitted power output. Optimum cadence therefore appears to be important for each athlete according to our results. The results suggest that the margin of decrease in power output for sub-optimal cadences appears to depend on the optimum cadence and the magnitude of the power output, though to some extent such differences may be due to the physiological nature of the athletes.

A limitation of our study is that optimum cadence depends on heart rate lag. We consider a lag of 30 seconds or less to be appropriate for our data based on our reading of the literature.

Chapter 6: Including training impulse calculations in non-linear models

6.1. Introduction to training impulses (TRIMPs)

At this point in our analysis our favoured model is the non-linear model of power output, heart rate and cadence. We now consider adding a training impulse calculation, known as TRIMP, to measure training load and therefore fatigue. In this chapter we present the methodology for adding training impulse calculations to the model (section 6.2.). We then present model results in section 6.3., we assess the practical effects of optimum cadence with respect to training load (calculating power output for cadences below and above optimum cadence) and we then summarise optimum cadences yielded (for models with TRIMP calculations) in section 6.4. Finally in section 6.5. we discuss the extent to which cadence is affected by training load, and whether or not training impulses should be included in the model.

6.2. Methodology for including training impulse calculations (TRIMPs) in the non-linear model of power output, heart rate and cadence

TRIMP measures training load over a session, taking account of the intensity and duration of the session (Jobson et al, 2009; Morton et al, 1990). TRIMP is considered to be a standard measure of training load througout literature, although the exact calculation differs between authors (Jobson et al, 2009).

The calculation that is typically used for TRIMP, for a session of *dT*, where *T* is the number of sample points and *d* is the sampling interval (e.g. a session sampled every 5 seconds on 720 occasions would be of length 1 hour), is the following (Morton et al, 1990), in equation 14:

$$TRIMP \; = dT\,\overline{f}e^{b\,\overline{f}}\,. \tag{14}$$

where

$$\overline{f} = (H_{av} - H_{resting})/(H_{\max} - H_{resting})\,.$$

is the fractional average heart-rate reserve for a given session (H_{av} is the average (mean) heart rate for the given session), and b is a suitable constant, chosen by Morton et al to be 1.92. TRIMP is dimensionless, but the time units used in the calculation of d and $\overline{f}$ should be consistent. In this study we measure heart rates in beats per minute, so the classic TRIMP measure (heart rate * duration) is the number of heart beats in a session.

Morton et al (1990) suggest that, ideally, in the TRIMP calculation, more weight should be given to sessions where heart rate is high. Given that heart rate varies within a session it would appear to make sense to put more weight on periods of higher heart rate within a session. Therefore we propose to modify the classic TRIMP above as follows, in equation 15:

$$TRIMP' = d\sum_{t=1}^{T} f_t \; e^{b\,f_t}\,. \tag{15}$$

where t indexes the sample points and, in equation 16,

$$f_t = (H_t - H_{resting})/(H_{\max} - H_{resting}). \tag{16}$$

is the heart rate fraction reserve at sample point t within the session, and H_t is the heart rate at sample point t.

For the athletes in our study, there are long training schedules, up to twelve months. Due to the length of the training schedules, we seek to quantify the effect of training loads of previous sessions on the current session. Hence, we calculate a cumulative TRIMP, taking into account TRIMPs from previous sessions, to quantify a long term effect of fatigue. In our cumulative TRIMP calculation we include an exponential decay of all previous session TRIMPs. Firstly we calculate a TRIMP value for a whole session using either the Morton et al model (equation 14) or our modified model (equation 15). The cumulative TRIMP is then a weighted sum of previous session TRIMPs such that more or less weight is given to more recent TRIMPs – a decay constant ϕ determines how much weight is given to the most recent sessions. The lower the ϕ value, the more emphasis is placed on the most recent session. Let X_i be the overall session TRIMP for session i (i=1,…,N) calculated using either our modified TRIMP (equation 15) or the classic TRIMP (equation 14). Then, the cumulative TRIMP at session i is

$$Z_i = \sum_{k=1}^{i-1} \phi^{j_i - j_k} X_{i-k}.$$

where j_i is the day number of session i (taking the first day of the training schedule as day 1). Thus the TRIMP of a training session x days in the past is weighted by an amount ϕ^x. The models become the following, with cumulative TRIMP added in two ways, producing models in equations 17 and 18:

$$\log P_t = \log \mu + \alpha \log C_t - \beta C_t + \gamma \log H_{t+l} + \delta \log Z. \tag{17}$$

$$\log P_t = \log \mu + \alpha \log C_t - \beta C_t + \gamma \log H_{t+l} + \delta \log C_t \log Z. \tag{18}$$

Cadence is interacting with cumulative TRIMP in the model in equation 18, but not in the model in equation 17. For the interaction (the model in equation 18), the mathematical optimum cadence is calculated in the following way:

$$C^* = (\alpha + \delta \log Z)/\beta.$$

We choose the decay constant for each athlete based on which yields the lowest standard errors for the cumulative TRIMP coefficient estimate (Z).

Thus far TRIMP calculations have been for whole sessions; one TRIMP value is calculated to represent the training load for a given session. In addition to this, we calculate TRIMPs at intermediate points within sessions - thus we calculate a TRIMP for each 5 second measurement of time. Using our modified TRIMP (calculated using equation 15), as an alternative to calculating a TRIMP for all sampling points in a session (t to T), we do so for some of the session, say an intermediate point within a session, τ. The TRIMP at an intermediate point τ within a session is:

$$TRIMP_\tau = d \sum_{t=1}^{\tau} f_t \, e^{b f_t}.$$

Where f_t is the heart rate fraction reserve at sampling point t within a session (as in equation 16 presented earlier in section 5.1.), b is a constant equal to 1.92, and d is the sampling interval. Thus this within session TRIMP allows us to quantify training load at any sampling point within a session. We let X_τ be the TRIMP at intermediate point τ within a session. We add the within session TRIMP (X_τ) to the power output/heart rate model in two ways, producing models in equations 19 and 20:

$$\log P_t = \log\mu + \alpha\log C_t - \beta C_t + \gamma\log H_{t+l} + \delta\log X_\tau . \tag{19}$$

$$\log P_t = \log\mu + \alpha\log C_t - \beta C_t + \gamma\log H_{t+l} + \delta\log C_t \log X_\tau . \tag{20}$$

In the second model (equation 20), cadence is interacted with within session TRIMP, such that the mathematically optimum cadence depends on the TRIMP at any sampling point t. Thus the mathematically optimum cadence for this model (in equation 20) becomes:

$$C^* = (\alpha + \delta\log X_\tau)/\beta .$$

The mathematically optimum cadence therefore is dependent not only on α and β but also $\delta\log X_{\tau.}$ We can therefore explore how optimum cadence varies with respect to different TRIMP values, or different training loads. The time unit of the sampling interval (d) does not need to be defined. When selecting a range of values for X_τ (to explore the effect of X_τ on optimum cadence), we can select a range of values based on what is typical for a given athlete.

Before fitting regression models, we process the data available to us (the power output, heart rate and cadence measurements yielded at 5 second intervals). Literature suggests that heart rate lags behind changes in power output, and that a low heart rate lag is appropriate for our data (Stirling et al, 2008). We observe from fitting the non-linear model (equation 7) with different heart rate lags that any low heart rate lag (of 6 time units or less) yields a good fit of covariates. For all models with TRIMP calculations we settle for a heart rate lag of 1 time unit (5 seconds, denoted by l=1). We have previously considered smoothing the data available to us to reduce extreme fluctuations in the data (by taking a moving average of window width k time units). However we do not smooth data (denoted as k=1). Data are sampled once every 24 time units (once every 120 seconds, denoted as m=24).

We fit models with TRIMP (equations 17, 18, 19 and 20) to data we believe to closely resemble race conditions. We calculate the mean of all sampled heart rates, which we term the overall mean heart rate, for each athlete. We retain sampled data in which heart rate is greater than the overall mean heart rate for each athlete (which we term the high heart rate data subset). We believe the relatively high heart rate in the high heart rate data subset help to eliminate data from very long low-intensity training sessions. We fit models with TRIMP (equations 17, 18, 19 and 20) to the high heart rate data subset.

6.3. Results of non-linear models of power output, heart rate and cadence with training impulse calculations (TRIMPs)

We include training impulses, or TRIMPs, to our non-linear models. TRIMPs are a calculation to represent training load (Jobson et al, 2009; Morton et al, 1990). We fit models with TRIMP to the high heart rate data subset – that is, sampled data in which heart rate is greater than the mean heart rate for each athlete.

6.3.1. Models with cumulative TRIMP

We consider a cumulative TRIMP from session to session. By incorporating previous session TRIMPs into a cumulative TRIMP measure for the current session, we intend to capture a long term accumulation of fatigue over a training schedule.

We calculate a cumulative TRIMP for session *i* using our modified version of the Morton model (equation 15), which we denote as Z_{Xi}, and also using the Morton model (equation 14), which we denote Z_{Ti}.

Subsequently adding cumulative TRIMP to the non-linear model produces the models in equations 17 and 18 – see section 6.2. for these equations in full. We fitted models in equations 17 and 18, for a heart rate lag of 5 seconds (l=1), using both our modified TRIMP (calculated using equation 15) and the classic TRIMP (calculated using equation 14). The cumulative TRIMP contains a decay constant, φ. This determines how much weight is placed on the most recent session. The lower the constant, the more weight is placed on the most recent session. We consider the optimum φ value for each athlete; that is the φ value for each athlete that yields the lowest standard errors for the coefficient estimate Z. The optimum φ is 0.5 for athletes 1, 3 and 4, and 0.9 for athlete 2. This means that for athletes 1, 3 and 4, the optimum φ has a greater emphasis on the most recent session than it does for athlete 2. This means cumulative TRIMPs are typically greater for athlete 2 than for athletes 1, 3 and 4.

The following estimates, along coefficient standard errors, explanatory power and AIC, are yielded, in table 68. In this table, we use different model numbers to denote which type of TRIMP is used (our modified TRIMP or the classic TRIMP). The models featuring cumulative TRIMP (models in equations 17 and 18) yield optimum cadences for athletes 1 and 2 (since for these athletes both α and β are positive). Estimates and standard errors are broadly similar to the non-linear model with no additional covariates (equation 7). For athletes 1 and 2, δ is negative. This means that, as the cumulative TRIMP increases, expected power output decreases – this is what we expect, since the cumulative TRIMP should indicate a level of fatigue. However for athletes 3 and 4, δ is positive. For these athletes, as cumulative TRIMP increases, expected power output also increases. Nevertheless, since it is athletes 1 and 2 for whom mathematical optimum cadences are yielded, we are most concerned with results for athletes 1 and 2.

When calculating the optimum cadence in the model with cadence / cumulative TRIMP interaction (equation 18), we first choose a range of different values to use for the cumulative TRIMP. We base this range on values that are typically found in the data (the power output, heart rate and cadence measurements yielded at 5 second intervals). The following range of cumulative TRIMP values are calculated in the data, and we present these in table 69.

Table 68: Coefficient estimates (standard errors in parenthesis) and explanatory power for the cumulative TRIMP models, (equations 17 and 18 for modified and classic TRIMPs), for each athlete for a heart-rate lag of 5 seconds.

	Model	*α*	*β*	*γ*	*δ*	*log μ*	*R2*	*AIC*
Athlete 1	1	1.59 (0.01)	0.021 (0.0005)	0.87 (0.06)	-0.05 (0.01)	-4.06 (0.31)	84.0	69635
	2	1.67 (0.02)	0.021 (0.0005)	0.86 (0.06)	-0.014 (0.03)	-4.32 (0.29)	84.0	69629
	3	1.59 (0.01)	0.021 (0.0005)	0.87 (0.06)	-0.05 (0.01)	-4.05 (0.31)	84.0	69634
	4	1.67 (0.02)	0.021 (0.0005)	0.86 (0.06)	-0.014 (0.03)	-4.31 (0.29)	84.0	69627
Athlete 2	1	1.66 (0.02)	0.025 (0.0008)	0.75 (0.08)	-0.12 (0.03)	-2.79 (0.46)	74.8	50063
	2	1.88 (0.05)	0.025 (0.0008)	0.75 (0.08)	-0.029 (0.01)	-3.69 (0.42)	74.8	50063
	3	1.66 (0.02)	0.025 (0.0008)	0.75 (0.08)	-0.12 (0.03)	-2.83 (0.46)	74.8	50065
	4	1.87 (0.05)	0.025 (0.0008)	0.75 (0.08)	-0.029 (0.01)	-3.69 (0.42)	74.8	50064
Athlete 3	1	1.18 (0.03)	-0.0008 (0.001)	1.10 (0.14)	0.05 (0.02)	-5.64 (0.70)	59.0	32582
	2	1.13 (0.04)	-0.0008 (0.001)	1.10 (0.14)	0.011 (0.004)	-5.39 (0.68)	59.0	32581
	3	1.18 (0.03)	-0.0007 (0.001)	1.11 (0.14)	0.05 (0.02)	-5.868(0.70)	59.0	32581
	4	1.13 (0.05)	-0.0007 (0.001)	1.11 (0.14)	0.012 (0.004)	-5.42 (0.68)	59.0	32580
Athlete 4	1	0.14 (0.03)	-0.001 (0.002)	1.51 (0.14)	0.15 (0.03)	-3.54 (0.71)	9.1	82754
	2	0.12 (0.05)	-0.0008 (0.002)	1.48 (0.15)	0.005 (0.01)	-2.61 (0.69)	8.7	82784
	3	0.15 (0.03)	-0.001 (0.002)	1.50 (0.14)	0.17 (0.03)	-3.60 (0.71)	9.2	82744
	4	0.14 (0.05)	-0.0007 (0.002)	1.47 (0.14)	0.001 (0.01)	-2.59 (0.69)	8.7	82785

Model 1: $\log P_t = \log\mu + \alpha\log C_t - \beta C_t + \gamma\log H_{t+l} + \delta\log Z$ using our modified TRIMP
Model 2: $\log P_t = \log\mu + \alpha\log C_t - \beta C_t + \gamma\log H_{t+l} + \delta_t\log Z\log C$ using our modified TRIMP
Model 3: $\log P_t = \log\mu + \alpha\log C_t - \beta C_t + \gamma\log H_{t+l} + \delta\log Z$ using the classic TRIMP
Model 4: $\log P_t = \log\mu + \alpha\log C_t - \beta C_t + \gamma\log H_{t+l} + \delta_t\log Z\log C$ using the classic TRIMP

Table 69: Minimum, maximum and mean cumulative TRIMP calculations (for both our modified TRIMP and the classic TRIMP) for each athlete.

Athlete	Minimum	Maximum	Mean
Z (our modified TRIMP)			
1	111	568	281
2	114	1921	1448
3	57	781	170
4	47	362	145
Z (the classic TRIMP)			
1	105	557	270
2	105	1665	1303
3	55	769	162
4	38	303	134

Optimum cadences are yielded for athletes 1 and 2 from models in equations 17 and 18. Therefore for these athletes we choose a suitable range of cumulative TRIMP values for testing the sensitivity of optimum cadences to cumulative TRIMP in equation 18. We consider the cumulative TRIMP for hypothetically easy, moderate and hard months. We consider hypothetical cumulative TRIMPs for easy, moderate and hard months as follows:

- easy month—10 2 hour sessions with fraction of heart-rate reserve at 0.6;
- moderate month—as above plus 10 3 hour sessions with a fraction of heart-rate reserve of 0.7;
- hard month—as above plus 10 3 hour sessions with a fraction of heart-rate reserve at 0.8.

We present the range of cumulative TRIMPs, along with optimum cadences and expected power output, in tables 70 and 71, for the models in equations 17 and 18. For reference, a session TRIMP of 250 typically relates to a three hour session, whilst a cumulative TRIMP varies between athletes. For athlete 1, cumulative TRIMPs for easy, moderate and hard months, using our modified TRIMP calculation, are 223, 383 and 517 respectively. For athlete 2, these are 783, 1217 and 1733 respectively.

There is little difference in optimum cadences between the two TRIMP calculation methods (our modified TRIMP and the classic TRIMP), with marginally higher optimum cadences for the latter.

Table 70: Optimum cadences (C*), along with 95% confidence intervals, yielded for athletes 1 and 2 for models with cumulative TRIMP covariate (equation 17, using both our modified TRIMP and the classic TRIMP), with expected power output at optimum cadence (P*).

	Model	C*	C* standard Deviation	C* Variance	Confidence From:	Interval To:	P*
Athlete 1	1	75.45	0.24	0.06	74.99	75.92	192.42
	3	75.50	0.24	0.06	75.04	75.96	196.29
Athlete 2	1	67.71	0.34	0.11	67.04	68.37	212.32
	3	67.77	0.34	0.12	67.11	68.44	212.45

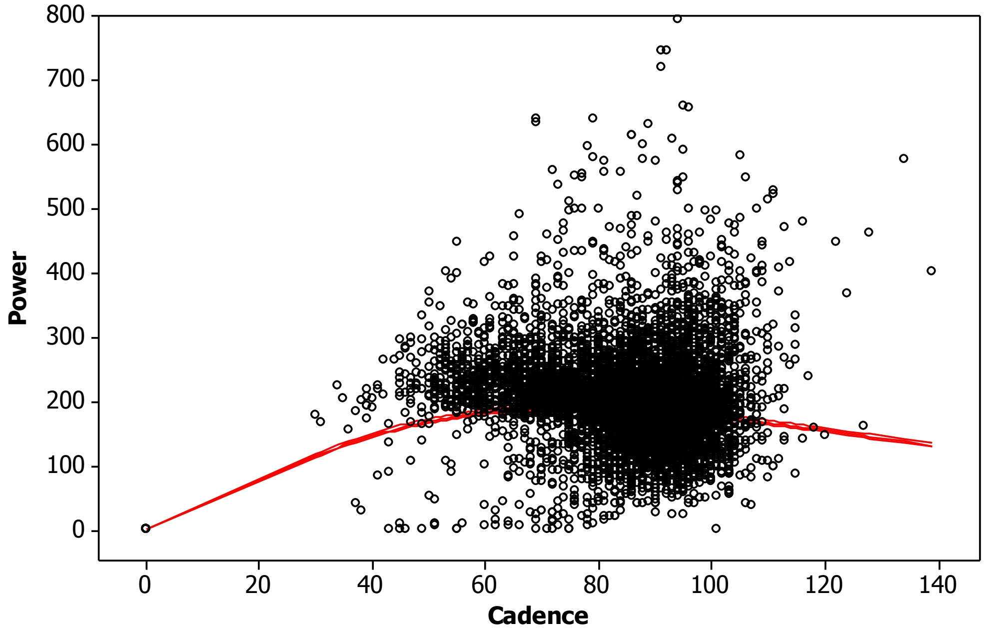
Model 1: $\log P_t = \log \mu + \alpha \log C_t - \beta C_t + \gamma \log H_{t+l} + \delta \log Z$ using our modified TRIMP
Model 3: $\log P_t = \log \mu + \alpha \log C_t - \beta C_t + \gamma \log H_{t+l} + \delta \log Z$ using the classic TRIMP

Table 71: Optimum cadences (C*), along with 95% confidence intervals, yielded for athletes 1 and 2 for models with cadence/cumulative TRIMP interaction (equation 18, using both our modified TRIMP and the classic TRIMP), with expected power output at optimum cadence (P*).

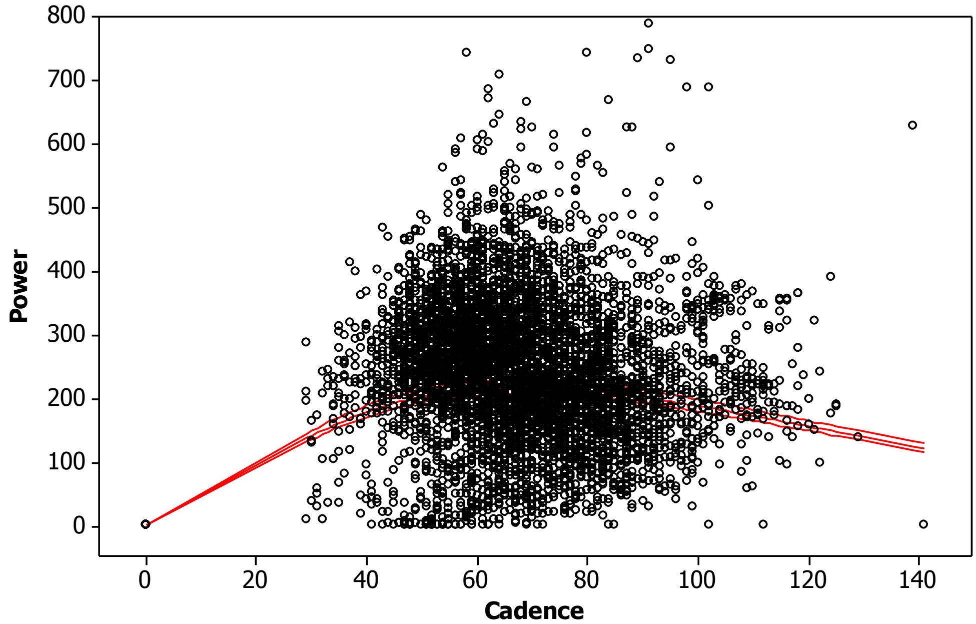
	Model	Difficulty	TRIMP	C*	P*
Athlete 1	2	Easy	223	75.44	198.46
		Moderate	383	75.12	192.72
		Hard	517	74.93	189.34
	4	Easy	233	75.47	197.99
		Moderate	367	75.18	192.70
		Hard	500	74.97	189.15
Athlete 2	2	Easy	783	68.39	229.13
		Moderate	1217	67.86	216.92
		Hard	1733	67.44	207.59
	4	Easy	767	68.34	226.55
		Moderate	1167	67.85	215.28
		Hard	1633	67.45	206.66

Model 2: $\log P_t = \log \mu + \alpha \log C_t - \beta C_t + \gamma \log H_{t+l} + \delta_t \log Z \log C$ using our modified TRIMP
Model 4: $\log P_t = \log \mu + \alpha \log C_t - \beta C_t + \gamma \log H_{t+l} + \delta_t \log Z \log C$ using the classic TRIMP

We present power output cadence plots, for athletes 1 and 2, for the models in equations 17 and 18 (in figures 18 to 21), for typically easy, moderate and hard months. When plotting the power output cadence graphs, we set heart rate to a constant (the mean heart rate of all heart rate measurements in the sampled data for each athlete). For the model with cadence / cumulative TRIMP interaction (equation 18), different TRIMPs also have an effect on the mathematical optimum cadences yielded.

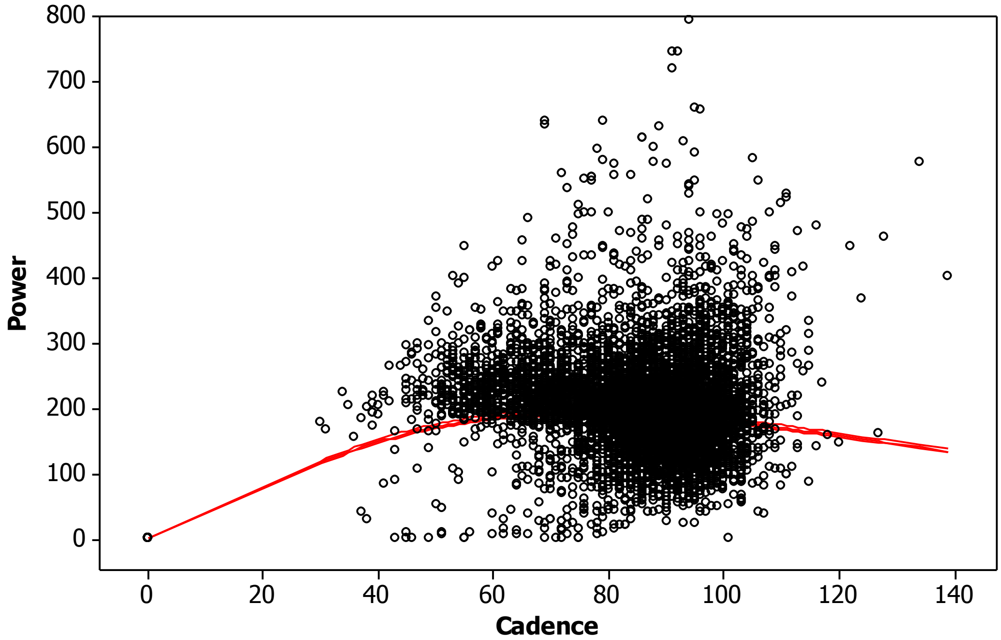


(a)

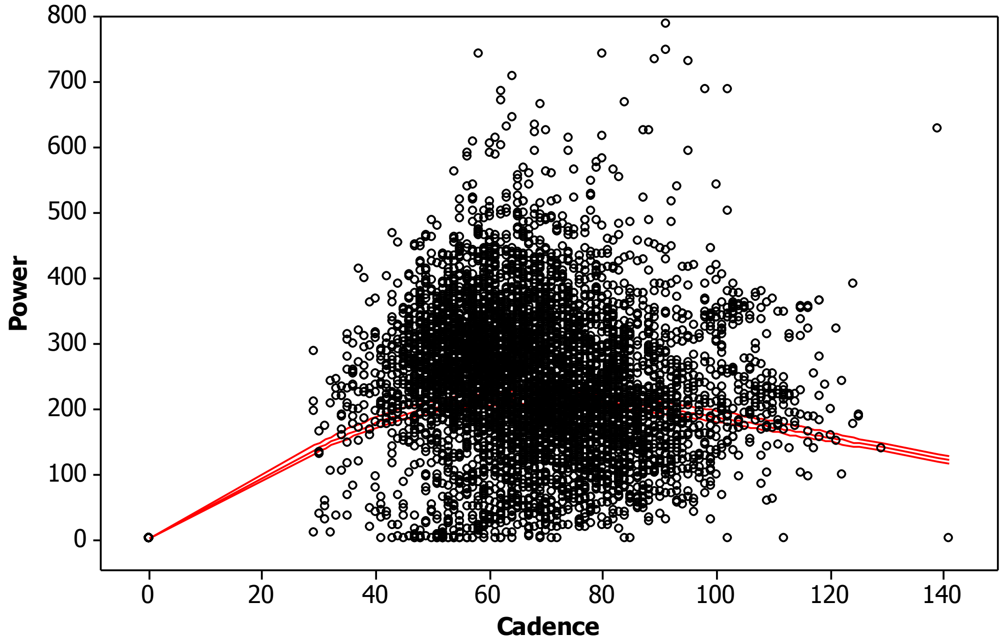


(b)

Figure 18: Fitted power output (red curve) from the non-linear model with cumulative TRIMP (equation 17, using our modified TRIMP) against power output measurements (black dots) for athletes1 (a) and 2 (b). The red lines represent cumulative TRIMPs from easy, moderate and hard months.

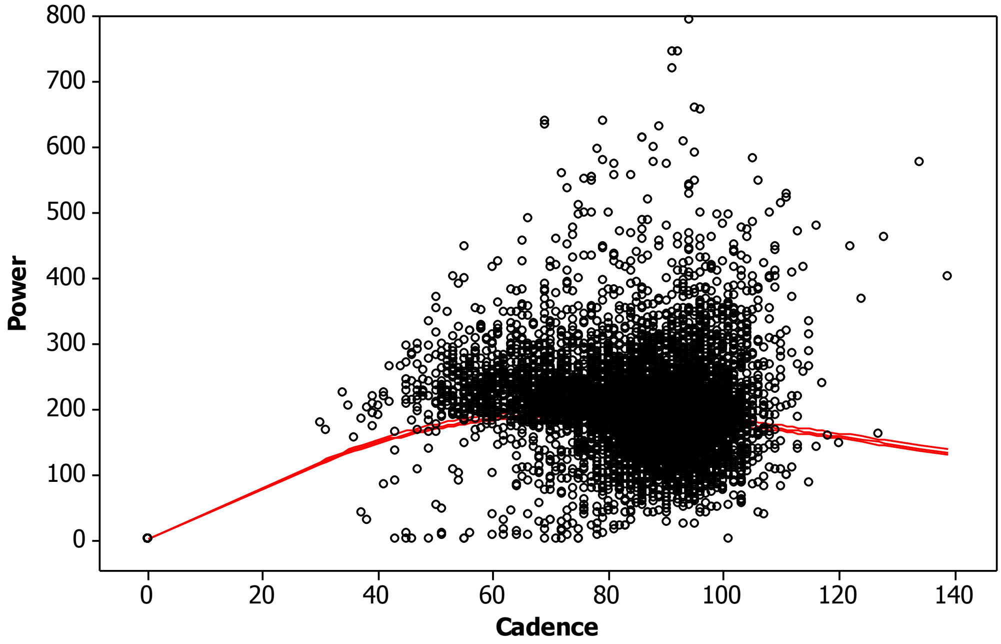


(a)

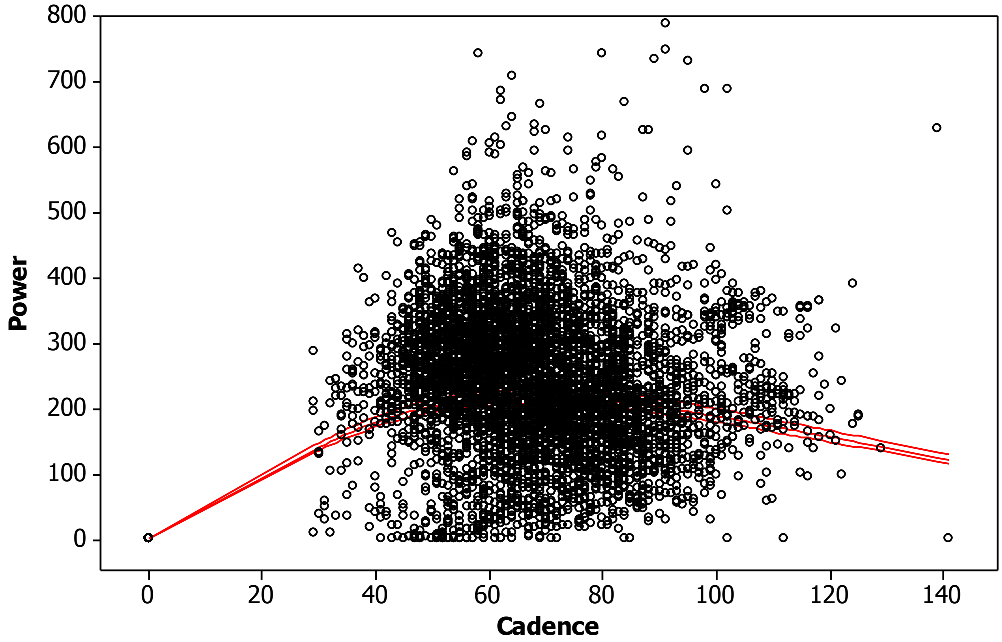


(b)

Figure 19: Fitted power output (red curve) from the non-linear model with cumulative TRIMP (equation 17, using the classic TRIMP) against power output measurements (black dots) for athletes1 (a) and 2 (b). The red lines represent cumulative TRIMPs from easy, moderate and hard months.

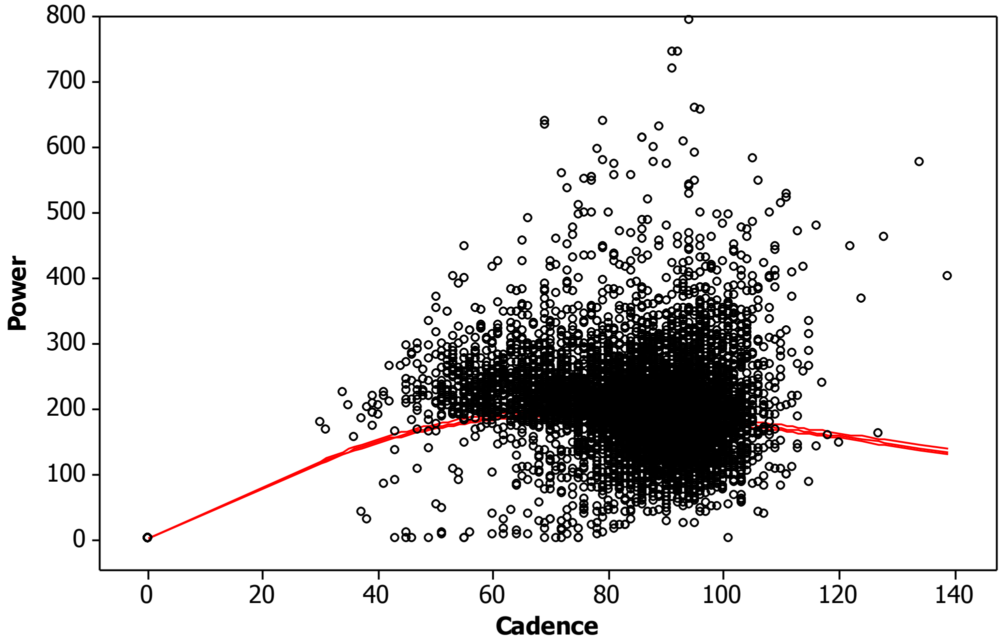


(a)

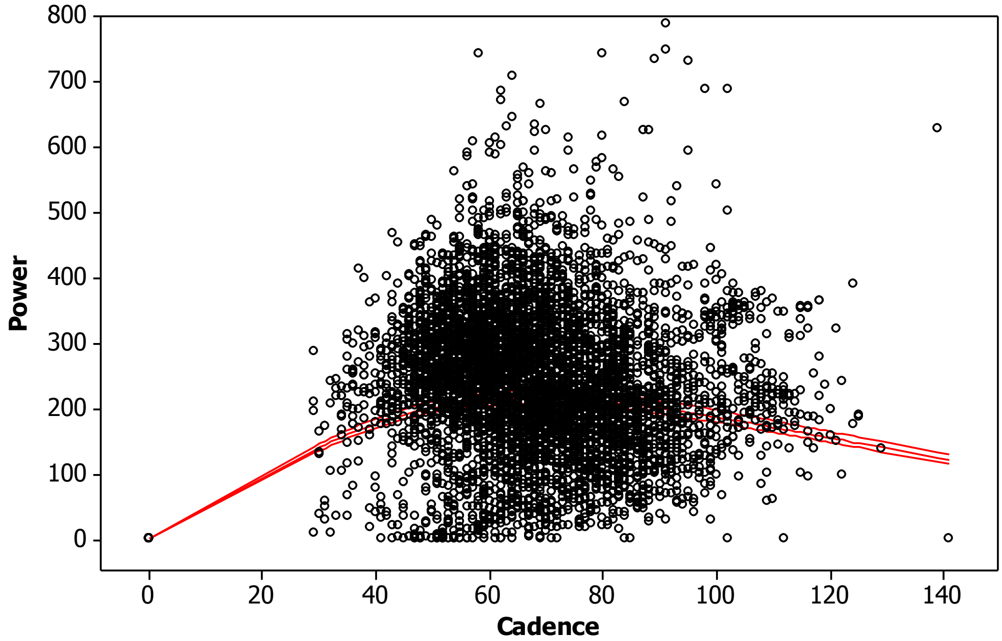


(b)

Figure 20: Fitted power output (red curve) from the non-linear model with cadence cumulative TRIMP interaction (equation 18, using our modified TRIMP) against power output measurements (black dots) for athletes1 (a) and 2 (b). The red lines represent cumulative TRIMPs from easy, moderate and hard months.



(a)



(b)

Figure 21: Fitted power output (red curve) from the non-linear model with cadence cumulative TRIMP interaction (equation 18, using the classic TRIMP) against power output measurements (black dots) for athletes1 (a) and 2 (b). The red lines represent cumulative TRIMPs from easy, moderate and hard months.

6.3.2. Models with within-session TRIMPs

We modify the TRIMP outlined in Morton et al (1990) so that it can be calculated for intermediate points within a session, in fact for each sampling point where power output, cadence and heart rate data have been collected. We also add greater emphasis to sampling points where heart rate is high. We fit models with intermediate TRIMP calculations (X_τ), producing models in equations 19 and 20 – see section 6.2. for these equations in full.

For the model with cadence / within-session TRIMP interaction (equation 20), optimum cadence depends on the level of the TRIMP X_τ, along with its coefficient δ. We settle for a heart rate lag of 5 seconds (l=1), as literature suggests a low heart rate lag is realistic for our data, and we do not smooth the data (k=1). We fit models with TRIMP to the high heart rate data subset, as we are concerned with optimising cadence for sessions which are likely to resemble race conditions

We present the coefficient estimates (with standard errors in parentheses), along with explanatory power and AIC, for models in equations 19 and 20 respectively, in table 72.

Table 72: Coefficient estimates (with standard errors in parentheses), R^2 and AIC for non-linear models with within-session TRIMP (equations 19 and 20), fitted for athletes 1 to 4 separately, with a heart rate lag of 5 seconds.

	Model	*α*	*B*	*γ*	*δ*	*log μ*	R^2	*AIC*
Athlete 1	1	1.58 (0.01)	0.021 (0.0004)	0.90 (0.06)	0.008 (0.004)	-4.57 (0.29)	84.0	69654
	2	1.58 (0.02)	0.021 (0.0005)	0.90 (0.06)	0.002 (0.001)	-4.53 (0.29)	84.0	69653
Athlete 2	1	1.66 (0.02)	0.025 (0.0008)	0.76 (0.08)	-0.03 (0.009)	-3.62 (0.42)	74.7	50085
	2	1.70 (0.02)	0.025 (0.0008)	0.77 (0.08)	-0.008 (0.002)	-3.77 (0.42)	74.7	50074
Athlete 3	1	1.17 (0.03)	-0.001 (0.001)	1.06 (0.14)	0.003 (0.01)	-5.19 (0.68)	58.9	32588
	2	1.17 (0.04)	-0.001 (0.001)	1.06 (0.14)	0.0006 (0.003)	-5.18 (0.68)	58.9	32588
Athlete 4	1	0.13 (0.03)	-0.001 (0.002)	1.60 (0.15)	-0.04 (0.01)	-3.03 (0.71)	8.8	82774
	2	0.18 (0.03)	-0.001 (0.002)	1.60 (0.15)	-0.04 (0.01)	-3.09 (0.73)	8.8	82776

Model 1: $\log P_t = \log \mu + \alpha \log C_t - \beta C_t + \gamma \log H_{t+l} + \delta \log X_\tau$

Model 2: $\log P_t = \log \mu + \alpha \log C_t - \beta C_t + \gamma \log H_{t+l} + \delta \log C_t \log X_\tau$

The coefficient of the TRIMP (δ), along with the polarity of the coefficient, differs depending on the athlete. The effect of within session TRIMPs on expected power output may vary depending on other aspects, such as how far into a session the athlete is. In addition to different polarities, the size of the effect of the within session TRIMP varies by athlete – for example, it tends to be 4 times greater for athlete 2 than for athlete 1. Estimates are similar for the two models – interacting cadence with the within session has had little effect on estimates. However the size of δ is greater when it is not interacted with cadence (and correspondingly the standard errors for δ are lower when it is not interacted with cadence). The within session TRIMP appears therefore to have a greater impact on expected power output when not interacted with cadence.

Although the polarity of δ varies by athlete, whether it is positive or negative may depend on typical session length. Sessions for athletes for whom δ is positive (athletes 1 and 3) are typically 25% longer than for those athletes (athletes 2 and 4) for whom δ is negative. Indeed these longer sessions may mean there is a longer warm up period, thus the effect of TRIMP on expected power output is harder to capture for athletes 1 and 3.

AIC and explanatory power (R^2), and standard errors for α and β are very similar to the non-linear model with no additional covariates (equation 7), with marginal reduction in AIC now that the within session TRIMP is added (for example, AIC has reduced from 69700 to 69654 for athlete 1 with the addition of X_τ to the model).

We calculate mathematical optimum cadences for models with within-session TRIMP calculations (models in equations 19 and 20. We assume typical values of all other variables, including heart rate and in this case also the within session TRIMP X_τ – we use mean values for each. However for the cadence within-session TRIMP interaction model (equation 20), optimum cadence depends on the value of the TRIMP and the coefficient of the TRIMP. We therefore present a range of optimum cadences for different levels of the within session TRIMP. We choose a range for each athlete based on what is typically found in the data.

The following range of within session TRIMP values are calculated in the data. We present this range in table 73.

Table 73: Minimum, maximum and mean within session TRIMP calculations (X_τ) for each athlete.

Athlete	Minimum	Maximum	Mean
1	0.0109	691	179
2	0.0041	399	101
3	0.0040	1090	141
4	0.0020	574	101

The mean X_τ tends to be much closer to the minimum value found than to the maximum value found. This is presumably because sessions vary in length, so in relatively short sessions, the TRIMP does not have enough time to accumulate to reach a very high value. Athletes 2 and 4 have similar typical values of X_τ but they are higher for athletes 1 and 3. We use a range of X_τ values to test the sensitivity of optimum cadence (to the within session TRIMP, X_τ) in the model with cadence / within session TRIMP interaction (equation 20). We also plot expected power output against cadence to demonstrate the shape of the curve. In model 40 this curve also depends on the range of values for X_τ, The range of values we consider for calculating an optimum cadence is based approximately on the table (table 73) - we present this range in table 74.

Table 74: Range of values the within session TRIMP calculations (X_τ) are set to, for testing sensitivity of optimum cadence to X_τ.

Athlete	Range of X_τ values					
1	83	167	250	333	500	667
2	50	100	150	200	267	333

We present optimum cadences yielded from models in equations 19 and 20 respectively, in tables 75 and 76.

Table 75: Optimum cadences yielded from the non-linear model with within-session TRIMP (equation 19) for athletes 1 and 2.

Athlete	C*	C* standard Deviation	C* Variance	Confidence From:	Interval To:	P*
1	75.99	0.24	0.06	75.53	76.46	196.88
2	67.55	0.34	0.12	66.89	68.22	211.33

Table 76: Optimum cadences yielded from the non-linear model with cadence / with-session TRIMP interaction (equation 20) for athletes 1 and 2, with expected power output at optimum cadence (P*).

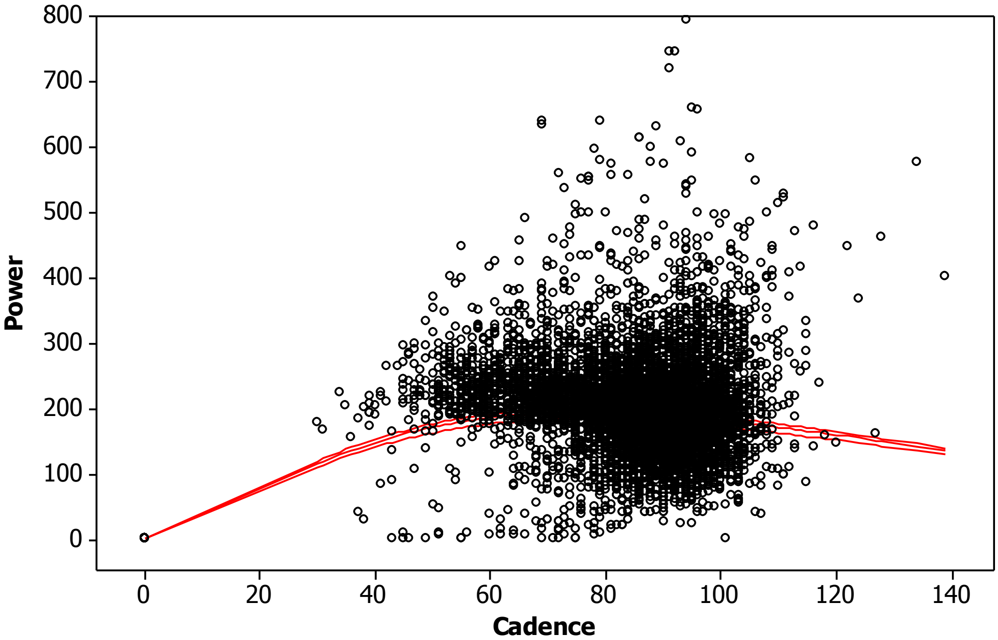
	X_t	C*	P*
Athlete 1	83	75.95	195.72
	167	76.01	196.78
	250	76.04	197.40
	333	76.07	197.84
	500	76.10	198.46
	667	76.13	198.91
Athlete 2	50	67.55	291.88
	100	67.32	285.01
	150	67.19	281.07
	200	67.09	278.31
	267	67.00	275.58
	333	66.92	273.48

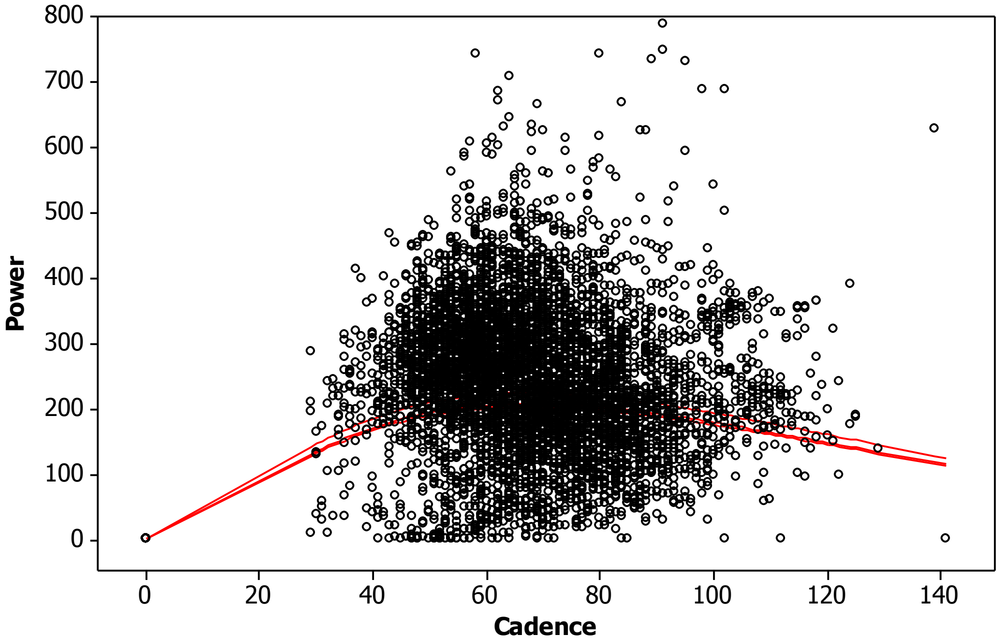
Optimum cadences are very similar to (though marginally higher than) those yielded from the cumulative TRIMP models (equations 17 and 18). For example, optimum cadence for athlete 1 is 76 in the within session TRIMP model (equation 19), but is 75.5 in the cumulative TRIMP models (equations 17 and 18).

We present the following power output cadence plots, in figures 23 and 24, for the models featuring within-session TRIMP. Optimum cadences have consistently been yielded for athletes 1 and 2 but not 3 and 4. We therefore decide to concentrate on power output cadence plots for athletes 1 and 2 only. We set TRIMP and heart rate to typical values. The range of values for within session TRIMPs for these plots is presented in table 76. In the model with cadence / within-session TRIMP interaction (equation 20), we observe the change in optimum cadence with changing X_τ. The fitted plots (figures 22 and 23) follow table 77. The plots illustrate that optimum cadence varies little with the within session TRIMP X_τ, although it does vary with X_τ more for athlete 2 than for athlete 1.

Table 77: Range of values of within-session TRIMPs (X_τ) considered for fitting power output cadence plots for the models with within-session TRIMP.

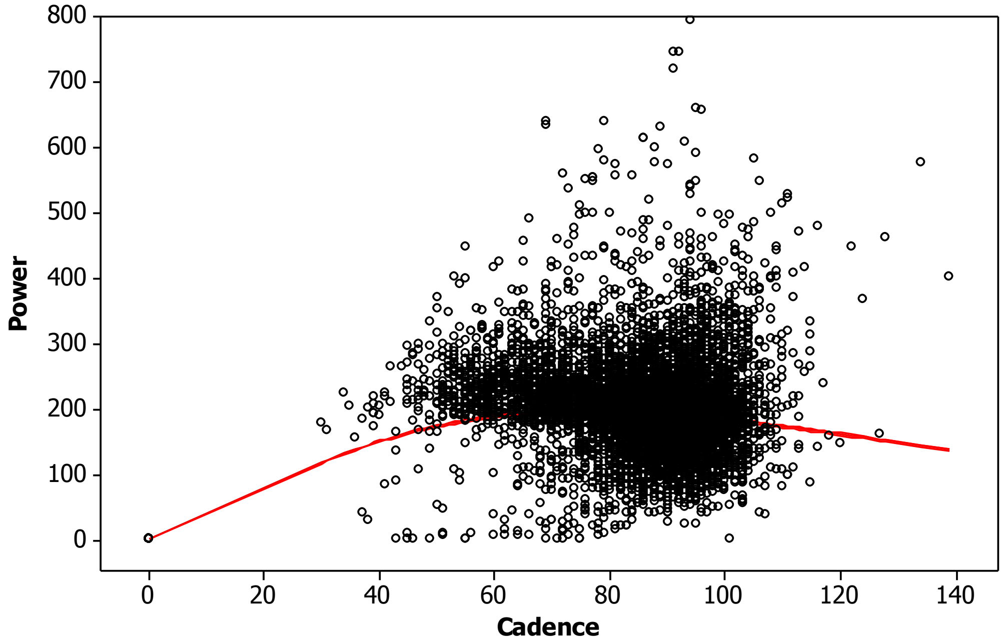
Athlete	Range of X_τ values		
1	17	200	667
2	17	200	400



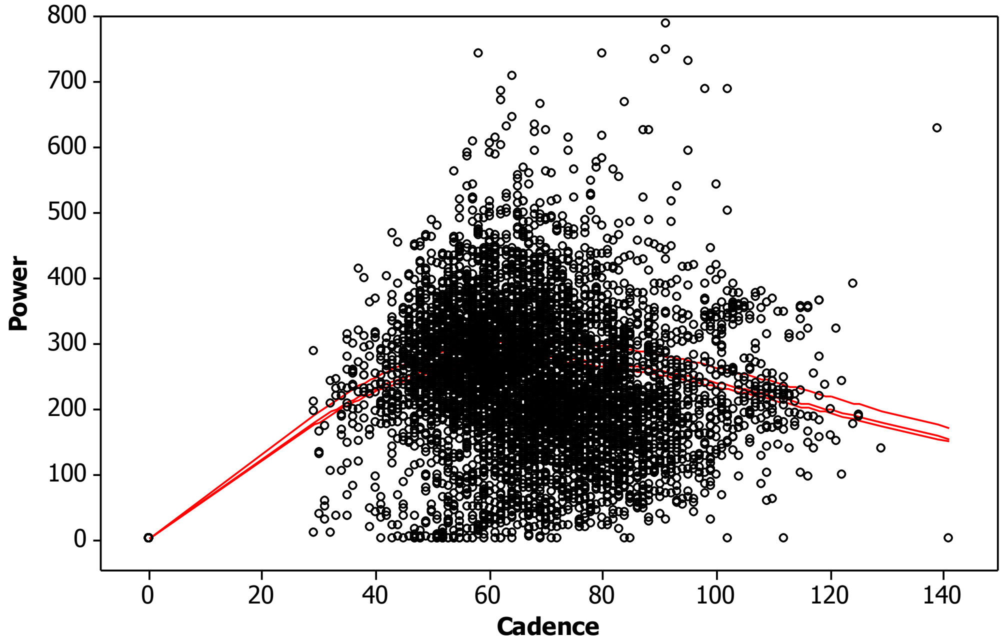


(b)

Figure 22: Fitted power output (red curve) from the non-linear model with cadence / within-session TRIMP (equation 19) against power output measurements (black dots) for athletes1 (a) and 2 (b). The red lines vary depending on the value of X_τ. For athlete 1, X_τ increases as power output increases, whilst the opposite occurs for athlete 2.



(a)



(b)

Figure 23: Fitted power output (red curve) from the non-linear model with cadence / within-session TRIMP interaction (equation 20) against power output measurements (black dots) for athletes1 (a) and 2 (b). The red lines vary depending on the value of X_τ. For athlete 1, X_τ increases as power output increases, whilst the opposite occurs for athlete 2.

6.3.3. Deviation from optimum cadence for models with TRIMP

For all models with TRIMP, we settle for a heart rate lag of 5 seconds (*l*=1), and we do not smooth data. We fit models with TRIMP to the high heart rate data subset. We fit models with training load impulse calculations (TRIMPs) to measure training load. TRIMPs are typically used in academic studies to quantify training load (Morton et al, 1990). We fit models with a cumulative TRIMP measure that takes into account previous session TRIMPs. This appears to represent some accumulated fatigue over time. We use two methods for calculating session TRIMPs – the classic TRIMP outlined by Morton (1990), and our modified TRIMP that gives greater weight to periods where heart rate is high. TRIMP is added to the non-linear model of power output, heart rate and cadence as in the models in equations 17 and 18 – see section 6.2. for these equations in full.

We present expected power outputs for a range of sub-optimal cadences for the model in equations 17 (using our modified TRIMP) in tables 78 and 79, as there are only marginal differences in expected power output between models with the modified and classic TRIMPs.

As level of fatigue (cumulative TRIMP) increases, the margin of decrease in expected power output for sub-optimal cadences increases slightly. Deviation from optimum cadence continues to result in a greater reduction in expected power output for athlete 2 than for athlete 1. However the margin of difference between the athletes is not as great as it is for the simple non-linear model with no additional covariates (equation 7).

Table 78: Differences in fitted power output for athlete 1 (left) and athlete 2 (right) using the cumulative TRIMP model (equation 17 using our modified TRIMP) with a heart rate lag of 5 seconds, for various cadences above and below the optimum cadence.

change in C from C*	cadence	Expected power output	change in power output	% change in power output
-20	55.5	180.0	-12.4	-6.5
-10	65.5	189.6	-2.8	-1.5
-5	70.5	191.8	-0.6	-0.3
	75.5	192.4		
5	80.5	191.7	-0.7	-0.4
10	85.5	189.9	-2.6	-1.3
20	95.5	183.3	-9.1	-4.8

change in C from C*	Cadence	Expected power output	change in power output	% change in power output
-20	47.7	193.8	-18.5	-8.7
-10	57.7	208.1	-4.2	-2.0
-5	62.7	211.3	-1.0	-0.5
	67.7	212.3		
5	72.7	211.4	-0.9	-0.4
10	77.7	208.8	-3.5	-1.6
20	87.7	199.8	-12.5	-5.9

Table 79: Differences in fitted power output for athlete 1 (left) and athlete 2 (right) using the model with cadence / cumulative TRIMP interaction (equation 18, modified TRIMP) with a heart rate lag of 5 seconds, for cadences above and below the optimum cadence, for various cumulative TRIMPs (Z).

change in C from C*	cadence	Expected power output	change in power output	% change in power output	change in C from C*	Cadence	Expected power output	change in power output	% change in power output
Z = 223					Z = 783				
-20	55.4	185.4	-13.1	-6.6	-20	48.4	209.4	-19.7	-8.6
-10	65.4	195.4	-3.0	-1.5	-10	58.4	224.6	-4.5	-2.0
-5	70.4	197.7	-0.7	-0.4	-5	63.4	228.1	-1.1	-0.5
	75.4	198.5				68.4	229.1		
5	80.4	197.8	-0.7	-0.3	5	73.4	228.2	-1.0	-0.4
10	85.4	195.9	-2.5	-1.3	10	78.4	225.4	-3.7	-1.6
20	95.4	189.2	-9.2	-4.7	20	88.4	215.7	-13.4	-5.9
Z = 383					Z = 1217				
-20	55.1	179.9	-12.8	-6.6	-20	47.9	198.1	-18.8	-8.7
-10	65.1	189.8	-2.9	-1.5	-10	57.9	212.6	-4.3	-2.0
-5	70.1	192.0	-0.7	-0.4	-5	62.9	215.9	-1.0	-0.5
	75.1	192.7				67.9	216.9		
5	80.1	192.1	-0.6	-0.3	5	72.9	216.0	-0.9	-0.4
10	85.1	190.2	-2.5	-1.3	10	77.9	213.4	-3.6	-1.6
20	95.1	183.7	-9.0	-4.7	20	87.9	204.1	-12.8	-5.9
Z = 517					Z= 1733				
-20	54.9	176.7	-12.6	-6.7	-20	47.4	189.4	-18.2	-8.7
-10	64.9	186.4	-2.9	-1.5	-10	57.4	203.4	-4.2	-2.0
-5	69.9	188.6	-0.7	-0.4	-5	62.4	206.6	-1.0	-0.5
	74.9	189.3				67.4	207.6		
5	79.9	188.7	-0.6	-0.3	5	72.4	206.7	-0.9	-0.4
10	84.9	186.9	-2.4	-1.3	10	77.4	204.2	-3.4	-1.6
20	94.9	180.5	-8.9	-4.7	20	87.4	195.3	-12.3	-5.9

We also fit models with a within session TRIMP caclulation (X_τ), as in models in equation 19 and 20. The effect of within session TRIMP on fitted power output is not as expected. We therefore believe the models with within session TRIMP (in equations 19 and 20) tells us little about the actual effect of training load on power output within a session. We do not present power output for sub-optimal cadences for these models.

6.4. Summary of optimum cadences yielded in non-linear models of power output, heart rate and cadence with TRIMPs

Firstly we add a cumulative TRIMP which quantifies the accumulation of fatigue (by using session TRIMPs from previous sessions). We use two methods for calculating session TRIMPs – the classic TRIMP outlined by Morton (1990), and our modified TRIMP that gives greater weight to period where heart rate is high. TRIMP is added to the non-linear model of power output, heart rate and cadence, producing models in equations 17 and 18.

We present optimum cadences for these models (equations 17 and 18 respectively) in tables 80 and 81. For the model with cadence / cumulative TRIMP interaction (equation 18), we consider hypothetical TRIMPs for easy, moderate and difficult months.

We also include a within session TRIMP calculation, X_{τ}, which is added to the non-linear model of power output, heart rate and cadence, producing models in equation 19 and 20. We present optimum cadences for these models (equations 19 and 20 respectively) in tables 82 and 83.

Table 80: Optimum cadences (and confidence intervals in parentheses) for the cumulative TRIMP model (equation 17, using firstly our modified TRIMP and then the classic TRIMP), for a heart rate lag of 5 seconds (l=1).

Athlete 1	Athlete 2	Athlete 3	Athlete 4
Optimum Cadence	Optimum Cadence	Optimum Cadence	Optimum Cadence
Modfied TRIMP:			
75.45 (74.99, 75.92)	67.71 (67.04, 68.37)	*	*
Classic TRIMP:			
75.50 (75.04, 75.96)	67.77 (67.11, 68.44)	*	*

Table 81: Optimum cadences for the model with cadence / cumulative TRIMP interaction (equation 18, using firstly our modified TRIMP and then the classic TRIMP), for a heart rate lag of 5 seconds.

Athlete 1			Athlete 2		
Modified TRIMP:			Modified TRIMP:		
Difficulty	Cumulative TRIMP (Z)	Optimum Cadence	Difficulty	Cumulative TRIMP (Z)	Optimum Cadence
Easy	223	75.44	Easy	783	68.39
Moderate	383	75.12	Moderate	1217	67.86
Hard	517	74.93	Hard	1733	67.44
Classic TRIMP:			Classic TRIMP:		
Difficulty	Cumulative TRIMP (Z)	Optimum Cadence	Difficulty	Cumulative TRIMP (Z)	Optimum Cadence
Easy	233	75.47	Easy	767	68.34
Moderate	367	75.18	Moderate	1167	67.85
Hard	500	74.97	Hard	1633	67.45

Table 82: Optimum cadences (and confidence intervals in parentheses) for the within-session TRIMP model (equation 19), for a heart rate lag of 5 seconds (l=1).

Athlete 1	Athlete 2	Athlete 3	Athlete 4
Optimum Cadence	Optimum Cadence	Optimum Cadence	Optimum Cadence
75.99 (75.53, 76.46)	67.55 (66.89, 68.22)	*	*

Table 83: Optimum cadences for the model with cadence / within-session TRIMP interaction (equation 20), for a heart rate lag of 5 seconds (l=1).

	Athlete 1		Athlete 2
Within session TRIMP calculation (X_τ)	Optimum Cadence	Within session TRIMP calculation (X_τ)	Optimum Cadence
5000	75.95	3000	67.55
10000	76.01	6000	67.32
15000	76.04	9000	67.19
20000	76.07	12000	67.09
30000	76.10	16000	67.00
40000	76.13	20000	66.92

6.5. Discussion of non-linear models of power output, heart rate and cadence with TRIMPs

TRIMPs are considered in literature to be a standard measure of quantifying training load, although the exact calculation varies by author (Jobson et al, 2009; Morton et al, 1990). We fit models with TRIMPs (models in equations 17, 18, 19 and 20), only to the high heart rate data subset (sampled data in which heart rate is higher than the mean heart rate for each athlete). We consider a cumulative TRIMP, taking into account session TRIMP values of all previous sessions. This is to capture possible longer term accumulation of fatigue (longer than the build up of fatigue in one session). We calculate a cumulative TRIMP using an exponential decay of previous session TRIMPs, so that data from the whole of an athlete's training schedule are included in the cumulative TRIMP calculation.

For athletes 1 and 2 (whose results we are most concerned with since the models yield statstical optimum cadences for them), increasing the cumulative TRIMP decreases expected power output. For these athletes therefore, the cumulative TRIMP appears to capture an effect of fatigue; when the cumulative TRIMP is high, meaning they have a high level of fatigue, the athletes (1 and 2) are not able to produce as high a power output. With such a large number of data, inevitably there is some noise along with signal for any measure or variable. Nevertheless it appears to provide a measure of accumulated fatigue for athletes 1 and 2 in our analysis. However, it appears difficult to pick up much signal for athletes 3 and 4, whether for calculating an optimum cadence or for any other indicator of performance such as fatigue.

For the cumulative TRIMP, we calculated individual session TRIMPs (which are included within the cumulative TRIMP caclulation) using two methods. We used the method outlined in Morton et al (1990) in equation 14, and also a modified version that placed a greater emphasis on high heart rates. For the cumulative TRIMP, estimates and standard errors are very similar for both methods. Over a long training schedule, the two methods for calculating an individual session TRIMP show little difference. Any such effect may have become negligible over the schedule, with the cumulative TRIMP possibly affected more by other measures such as the day of the most recent session or the value of the decay constant. Optimum cadences are approximately the same whether the classic TRIMP or modified

TRIMP are used – the margin of difference, less than 0.1 revolution per minute, is not practically significant as an athlete would not be able to modify cadence by such small margins. Optimum cadences differ by approximately 1 or 2 rpm between easy and hard months. Such a small change in cadence (1 or 2 rpm) is not of practical significance. Changing cadence requires an equivalent change in gear to keep a constant power output, but there would be no gear change small enough that would allow an athlete to change cadence by such a small margin and maintain the same power output.

We also fitted models with a within session TRIMP calculation, in order to quantify training load at each sampling point within a session. The effect of the TRIMP on the expected power output is not necessarily clear – the polarity of the coefficient for within session TRIMP (δ) varies by athlete. The coefficient for the TRIMP (δ) is negative for athletes 2 and 4. This means that for these athletes the expected power output for an athlete decreases as the TRIMP increases. For athletes 1 and 3, the expected power output actually increases with the TRIMP. The relationship between expected power output and within session TRIMP appears to depend on the athlete. Moreover, it appears quite a complex relationship. The specific time point within a session could affect the relationship. The intensity of a particular session may further exagerate this effect. Given that we are fitting models to real empirical training data, there may be a wide range of sessions with varying intensities. The range of intensity of sessions may vary by athlete. Some sessions may be of low intensity and have a fairly low training load.

Session length may complicate the effect of a within sesion TRIMP on expected power output. For example, at the beginning of a session, an athlete may be warming up, and may be producing a low power output – at this point the power output TRIMP relationship would be positive. However, later in a session, the athlete may have accumulated fatigue, and the TRIMP may decrease the expected power output. Sessions are typically longer for athletes 1 and 3 than for athletes 2 and 4. It is indeed for athletes 1 and 3 that increasing within session TRIMP inreases expected power output. This may suggest that the expected power output generally decreases as the within session TRIMP increases, but that it is complicated by the beginning of a session when power output is low. For athletes for which δ is positive, they may therefore possibly spend more time warming up in a session, before reaching a high power output. The longer sessions may suggest the athletes have more time to accumulate a higher range of TRIMP values. We would expect the presence of extremely high TRIMP values to yield a model where the TRIMP generally decreases expected power output, due to fatigue accumulated. However this is not the case for athletes with very long sessions. Possible longer warm up times for long sessions may explain this. Nonetheless the within session TRIMP calculation relationship with expected power output may depend on other factors.

When examining the range of within session TRIMP values calculated for each athlete, we observe that the mean X_τ is closer to the minimum X_τ than it is to the maximum X_τ. We would expect this to some extent as sessions vary in length so there is not always time for the TRIMP to reach a high value. The within session TRIMP is typically higher for athletes 1 and 3 than it is for athletes 2 and 4. This is most likely because sessions are typically longer for athletes 1 and 3 than for athletes 2 and 4, so they have time for a within session TRIMP calculation to accumulate. However the TRIMP is slightly higher for athlete 1 than for any other athlete, so typical session length is possibly not the only explanation for the range of

TRIMP caclulations yielded. Our results suggest that it is difficult to investigate the effect of within session TRIMP calculations alone on expected power output for competitive cyclists. Nevertheless the results appear to suggest that, broadly, after the athlete has completed a warm up period in a session, expected power output decreases a little as within session TRIMP increases.

Compared to the simple non-linear model with no additional covariates (equation 7), optimum cadences for the cumulative TRIMP models (equations 17 and 18) are slighlty lower. For example, in the model in equation 7 optimum cadences are 76 for athlete 1 and 68.4 for athlete 2, and in equation 35, optimum cadences are 75.5 for athlete 1 and 66.7 for athlete 2. Optimum cadences are higher for athlete 1 than for athlete 2. We interact cadence with cumulative TRIMP in the model in equation 18. For a given change in cumulative TRIMP, optimum cadence changes by a greater margin for athlete 2 than for athlete 1. The optimum cadence for athlete 2 therefore appears more sensitive to fatigue than does the optimum cadence for athlete 1. Over the same interval expected power output also changes by a greater margin for athlete 2. Optimum cadence decreases by increasing magnitude as the TRIMP increases. For example, for athlete 1 in the model with cumulative TRIMP using our modified TRIMP method (equation 18), we consider a range of TRIMPs at approximately level intervals, for easy, moderate and hard months. The optimum cadence decreases by 0.3 as TRIMP increases from 223 to 383 (easy to moderate), then by 0.2 as TRIMP increases from 383 to 517 (moderate to hard). For athlete 2, the same trend occurs; the optimum cadence decreases by 0.5 as TRIMP increases from 783 to 1217, then by 0.4 as TRIMP increases from 1217 to 1733. Optimum cadence is typically 1 revolution per minute lower for a session after a hard month compared to an easy month. Considering different sessions over the previous two or three month period (rather than the last month) may yield a greater difference in optimum cadences between sessions following easy or hard periods.

For the within session TRIMP models, optimum cadences are slightly higher than in the cumulative TRIMP models. When interacting cadence with X_τ, optimum cadences vary between different TRIMP calculations for athlete 2 more than for athlete 1. Athlete 2's optimum cadence appears more sensitive therefore to the within session TRIMP. However this may be due to the typical size of the TRIMP X_τ. It is typically approximately twice as large for athlete 1 than for athlete 2. Indeed the coefficient estimate of X_τ (and hence the margin of change in optimum cadence) is approximately twice as high for athlete 2 than for athlete 1.

The optimum cadences yielded by the models with and without TRIMP calculations are typically greater for athlete 1 than for athlete 2 by approximately 8 rpm. This margin cannot be explained purely by the typical values of additional variables. For example 8 rpm is a much greater margin of difference than the difference between an optimum cadence for a low TRIMP and an optimum cadence for a high TRIMP for a partiular athlete. Although the effect of TRIMPs on expected power output can be difficult to model, a margin of approximately 8 rpm occurs whether we include a cumulative TRIMP or a within session TRIMP. This may emphasise the importance of individual optimum cadences. Moreover, the results may emphasise that an optimum cadence is most sensitive to an individual's characteristics (physical and physiological) than to levels of fatigue or duration of exercise.

We did not consider a range of heart rate lags when including the training impulse measures to the model, and we did not smooth the data. The sensitivity analysis conducted previously however, along with the literature, suggest that a low heart rate lag is most appropriate for our data. The optimum cadences yielded in the sensitivity analysis are quite sensitive to the moving average window width, but by not smoothing we are able to compare our optimum cadences with literature.

Our study has demonstrated that a mathematical optimum cadence can be calculated for an individual athlete empirically using data from training sessions, providing he/she has ridden at a great enough range of cadences and power outputs. By including cumulative TRIMP in the regression models, we have observed how optimum cadence varies with fatigue (although there is little impact in our study). We conclude that training impulses do not need to be included in models for calculating optimum cadence. A greater impact of fatigue on optimum cadence may potentially be found with more data collected over a longer period.

Chapter 7: Non-linear models of power output, heart rate excess and cadence

7.1. Introduction to models with heart rate excess

Thus far we have fitted non-linear models of measurements of power output, heart rate and cadence, sampled such that we take one measurement for every 2 minutes. We lag heart rate at some time after power output and cadence. Whilst this has successfully yielded mathematical optimum cadences for two of our athletes, we now consider an alternative model in which power output is related to excess heart rate $(H_{t+l} - h_0)$, where h_0 is the resting heart rate.

7.2. Method

Resting heart rates are provided in the data for athletes 1, 2 and 4. We assume the same resting heart rate for athlete 3 as is found for athletes 2 and 4. The resting heart rates are presented in the results section in 6.2. The model with excess heart rate becomes the following (equation 21):

$$\log P_t = \log \mu + \alpha \log C_t - \beta C_t + \gamma \log(H_{t+l} - h_0) + \varepsilon_t \qquad (21)$$

This model contains the property that, when the athlete is not riding, the power output in the model is zero, i.e. when $H_{t+l} = h_0$, $P_t = 0$. Data provided to us are power output, heart rate and cadence measurements yielded at 5 second intervals. We consider heart rate to be a response to changes in power output, at some delay, or lag. We fit this model for different heart rate lags, comparing covariate estimates and optimum cadences with the equivalent non-linear model that includes H_{t+l} rather than excess heart rate. We consider heart rate lags of 5, 10, 15, 20, 30 and 60 seconds (l= 1, 2, 3, 4 6 and 12 respectively). We previously considered smoothing data using a moving average of window width k time units. We do not smooth the data (denoted by k=1), and we continue to sample data once every 24 time units (120 seconds, denoted by m=24).

We fit the model (equation 21) to the high heart rate data subset (data in which heart rate is greater than the mean heart rate for each athlete individually).

7.3. Results

We present coefficient estimate along with explanatory power and AIC for the model in equation 21 in table 84. Whilst explanatory power increases with increasing heart rate lag, so too do standard errors. It is therefore difficult to ascertain an optimum heart rate lag for models with excess heart rate. AIC is marginally lower for models with heart rate excess (equation 21) compared to equivalent models featuring lagged heart rate (equation 7).

We briefly explore the sensitivity of coefficient estimates to the resting heart rate. We take an example – athlete 2 at heart rate lag of 30 seconds – and consider alternative hypothetical resting heart rates. We present coefficient estimates in table 85. Estimates for α

and β for athlete 2 are not affected by choice of resting heart rate, although γ increases as resting heart rate increases.

Optimum cadences are yielded for athletes 1 and 2, presented in table 86. Optimum cadences are broadly the same as those yielded in the non-linear model of power output, heart rate and cadence (equation 7) that did not include heart rate excess.

Plots of power against cadence for models with heart rate excess appear identical to those for non-linear models of power output, heart rate and cadence, presented earlier in section 4.6. We therefore do not include power cadence plots for models with heart rate excess.

Table 84: Optimum cadences (C*), along with 95% confidence intervals, yielded for athletes 1 and 2 from the non-linear model of power output, heart rate excess and cadence (equation 21), for various heart rate lags (*l*).

	l	*A*	β	γ	μ	*log* μ	R^2	*AIC*
Athlete 1	1	1.58 (0.01)	0.021 (0.0005)	0.62 (0.04)	0.057	-2.86 (0.18)	84.0	69654
H_0 = 53,	2	1.56 (0.01)	0.020 (0.0005)	0.61 (0.04)	0.062	-2.78 (0.19)	84.2	70914
51, 56 *	3	1.56 (0.01)	0.020 (0.0005)	0.61 (0.04)	0.061	-2.79 (0.20)	84.7	71542
	4	1.55 (0.01)	0.020 (0.0005)	0.53 (0.04)	0.089	-2.42 (0.20)	85.1	71841
	6	1.53 (0.01)	0.018 (0.0006)	0.52 (0.04)	0.092	-2.39 (0.20)	85.9	72597
	12	1.51 (0.01)	0.017 (0.0006)	0.48 (0.05)	0.111	-2.20 (0.22)	85.8	72724
Athlete 2	1	1.66 (0.02)	0.024 (0.0008)	0.57 (0.06)	0.076	-2.58 (0.28)	74.7	50085
H_0 = 42	2	1.65 (0.02)	0.024 (0.0009)	0.62 (0.06)	0.060	-2.82 (0.29)	75.9	50653
	3	1.64 (0.02)	0.024 (0.0009)	0.61 (0.07)	0.063	-2.76 (0.30)	75.9	50423
	4	1.63 (0.02)	0.024 (0.0009)	0.57 (0.07)	0.076	-2.58 (0.31)	75.8	51220
	6	1.61 (0.02)	0.023 (0.0010)	0.45 (0.07)	0.131	-2.04 (0.32)	76.5	51359
	12	1.56 (0.02)	0.020 (0.0010)	0.21 (0.07)	0.388	-0.95 (0.34)	75.7	52217
Athlete 3	1	1.17 (0.03)	-0.001 (0.001)	0.77 (0.10)	0.031	-3.48 (0.45)	58.9	32593
H_0 = 42	2	1.13 (0.03)	-0.003 (0.001)	0.77 (0.10)	0.032	-3.45 (0.47)	61.2	33021
	3	1.10 (0.03)	-0.004 (0.001)	0.80 (0.10)	0.028	-3.57 (0.48)	61.5	33173
	4	1.05 (0.03)	-0.007 (0.001)	0.81 (0.10)	0.026	-3.63 (0.72)	62.4	33365
	6	0.99 (0.03)	-0.010 (0.001)	0.88 (0.11)	0.020	-3.92 (0.50)	63.5	33227
	12	0.95 (0.04)	-0.011 (0.002)	0.67 (0.11)	0.052	-2.95 (0.50)	63.6	33155
Athlete 4	1	0.15 (0.03)	-0.0007 (0.002)	1.05 (0.10)	0.872	-0.14 (0.45)	8.7	82791
H_0 = 42	2	0.15 (0.03)	-0.001 (0.002)	0.98 (0.10)	1.164	0.15 (0.47)	9.2	83474
	3	0.17 (0.03)	-0.001 (0.002)	0.98 (0.11)	0.945	-0.06 (0.49)	10.1	83639
	4	0.16 (0.03)	-0.002 (0.002)	0.91 (0.11)	1.348	0.30 (0.50)	10.3	84037
	6	0.18 (0.04)	-0.002 (0.002)	1.04 (0.11)	0.689	-0.37 (0.51)	11.1	84745
	12	0.20 (0.04)	-0.0012 (0.002)	1.23 (0.12)	0.257	-1.36 (0.53)	12.6	85992

* for athlete 1, resting heart rates is given as 53 for most of the schedule, but later changes to 51, then to 56.

Table 85: Coefficient estimates (standard errors in parenthesis) and explanatory power for the non-linear model of power output, heart rate excess and cadence (equation 21), for athlete 2 at *l*=6 (heart rate lag of 30 seconds), for different resting hear rates (H_0). The italicised row represents the resting heart rate provided in the data for athlete 2.

H_0	α	β	γ	μ	*log* μ	σ	*R2*	*AIC*
38	1.61 (0.02)	0.023 (0.0010)	0.46 (0.07)	0.119	-2.13 (0.34)	0.867	76.5	51359
42	*1.61 (0.02)*	*0.023 (0.0010)*	*0.45 (0.07)*	*0.131*	*-2.04 (0.32)*	*0.867*	*76.5*	*51359*
46	1.61 (0.02)	0.023 (0.0010)	0.43 (0.07)	0.143	-1.94 (0.31)	0.867	76.5	51359
50	1.61 (0.02)	0.023 (0.0010)	0.42 (0.06)	0.157	-1.85 (0.29)	0.867	76.5	51358

Table 86: Optimum cadences (C*), along with 95% confidence intervals, yielded for athletes 1 and 2 from the non-linear model of power output, heart rate excess and cadence (equation 21), for various heart rate lags (l).

	L	C*	C* standard deviation	C* variance	Confidence From:	Interval To:	P*
Athlete 1	1	75.99	0.24	0.06	75.50	76.44	196.81
	2	79.58	0.28	0.08	79.01	80.11	192.91
	3	79.37	0.29	0.09	78.78	79.93	191.41
	4	79.43	0.30	0.09	78.84	80.01	190.47
	6	83.41	0.34	0.12	82.75	84.08	187.95
	12	86.82	0.39	0.15	86.08	87.62	186.31
Athlete 2	1	68.42	0.35	0.12	67.67	69.02	214.24
	2	68.18	0.36	0.13	67.39	68.81	206.14
	3	68.35	0.37	0.14	67.55	69.02	201.24
	4	68.57	0.39	0.15	67.81	69.34	196.18
	6	70.41	0.44	0.19	69.55	71.26	189.90
	12	76.09	0.55	0.30	75.01	77.16	179.34

We explore how power output in the model (equation 21) changes for sub-optimal cadences. We consider cadences 5, 10 and 20 revolutions per minute above and below the mathematically optimum cadence for each athlete, in table 87, for a selection of heart rate lags.

Although fitted power output is slightly higher for the model with heart rate excess, riding at a sub-optimal cadence reduces power output by the same margin whether heart rate or heart rate excess is included in the model (equations 7 and 20 respectively). For example, for athlete 1 at a lag of 30 seconds (l=6), fitted power output is 187.8 at the optimum cadence for model 18, but it is 188.0 in the model with heart rate excess (equation 21) – however riding at a cadence 20 revolutions per minute below optimum results in a 5.1% reduction in power output in both models.

7.4. Discussion

By relating power output to heart rate excess, this ensures that, when heart rate is equal to an athlete's resting heart rate, power output is zero. However, including heart rate excess in the non-linear model of power output, heart rate and cadence has made little difference to coefficient estimates and optimum cadences. Explanatory power is the same whether heart rate or heart rate excess is used in the non-linear model, with AIC marginally lower for the model with heart rate excess. Coefficient estimates for α and β are also broadly the same whether heart rate or heart rate excess is used in the model – hence so too are statistical optimum cadences, which are yielded for athlete 1 and 2 in our study.

However, the greatest difference between non-linear models with heart rate or heart rate excess is that γ, the coefficient relating to heart rate or heart rate excess, is much lower in the latter. Also μ is much greater when we add heart rate excess to the model compared to when we use heart rate. The reduced γ indicates that, as the athlete increases his power output, heart rate excess does not have to increase by the same amount as would heart rate – this is because heart rate excess is necessarily greater than heart rate (for heart rates above resting heart rate).

For athlete 1 in our analysis, the resting heart rate provided in the data changes twice. The resting heart rate may therefore have been recorded at different stages within his training schedule. For other athletes (athletes 2 and 4 in our analysis), the resting heart rate does not change. We do not know if the resting heart rate was recorded at different stages and it happened not to change, or if resting heart rate was only recorded at the beginning of the training schedule.

Given the similarity in results between non-linear models of power output, heart rate and cadence whether heart rate or heart rate excess is included in the model (equations 7 and 20 respectively), in terms of optimum cadences, fitted power output of the models and AIC, we do not consider it necessary to expand equation 21 to include further covariates. Some coefficient estimates (though not those related to the calculation of optimum cadence) appear sensitive to the choice of resting heart rate. Nevertheless in future research into optimising cadence, either heart rate or heart rate excess could be included in non-linear models of power output, heart rate and cadence.

Table 87: Differences in fitted power output for athlete 1 (left) and athlete 2 (right) using the non-linear model of power output, heart rate excess and cadence (equation 21) for unsmoothed data (k=1) and various heart rate lags (l), for various cadences above and below the optimum cadence.

change in C from C*	Cadence	Expected power	change in power	% change in power	change in C from C*	Cadence	Expected power	change in power	% change in power
l=1					l=1				
-20	56.0	184.1	-12.7	-6.5	-20	48.4	196.1	-18.2	-8.5
-10	66.0	193.9	-2.9	-1.5	-10	58.4	210.1	-4.2	-1.9
-5	71.0	196.1	-0.7	-0.4	-5	63.4	213.3	-1.0	-0.5
	76.0	196.8				68.4	214.3		
5	81.0	196.2	-0.6	-0.3	5	73.4	213.3	-0.9	-0.4
10	86.0	194.3	-2.5	-1.3	10	78.4	210.8	-3.4	-1.6
20	96.0	187.8	-9.0	-4.6	20	88.4	201.9	-12.4	-5.8
l=2					l=2				
-20	59.6	181.8	-11.1	-5.8	-20	48.2	188.6	-17.5	-8.5
-10	69.6	190.3	-2.6	-1.3	-10	58.2	202.1	-4.0	-1.9
-5	74.6	192.3	-0.6	-0.3	-5	63.2	205.2	-1.0	-0.5
	79.6	192.9				68.2	206.1		
5	84.6	192.3	-0.6	-0.3	5	73.2	205.3	-0.9	-0.4
10	89.6	190.7	-2.2	-1.1	10	78.2	202.8	-3.3	-1.6
20	99.6	184.9	-8.0	-4.1	20	88.2	194.2	-11.9	-5.8
l=6					l=6				
-20	63.4	178.3	-9.6	-5.1	-20	50.5	175.2	-14.7	-7.7
-10	73.4	185.7	-2.2	-1.2	-10	60.5	186.5	-3.4	-1.8
-5	78.4	187.4	-0.5	-0.3	-5	65.5	189.1	-0.8	-0.4
	83.4	188.0				70.5	189.9		
5	88.4	187.5	-0.5	-0.3	5	75.5	189.2	-0.7	-0.4
10	93.4	186.1	-1.9	-1.0	10	80.5	187.1	-2.8	-1.5
20	103.4	181.0	-7.0	-3.7	20	90.5	179.8	-10.1	-5.3

Chapter 8: Summary of research undertaken

In this study we have developed a methodology for calculating a mathematical optimum cadence. This is the cadence that maximises power output for a given heart rate. By riding at an optimal cadence, an athlete can make small but significant gains in performance (Abbiss et al, 2009; Coast and Welch, 1985; Foss and Hallen, 2004).

We have developed a range of mathematical models of power output, heart rate and cadence to yield an optimum cadence for four competitive cyclists. These are typically category II and III cyclists competing at county championship level. Linear, quadratic and cubic models did not yield a mathematical optimum cadence. However non-linear models yielded optimum cadences for two of the four athletes. The optimum cadences yielded for the athletes in our study were approximately 83 and 70 revolutions per minute for the two athletes respectively. These optimum cadences are similar to preferred cadences, although athlete 1 tends to ride just above his optimum cadence (typically riding at 82-92 rpm) and athlete 2 tends to ride just above and below his optimum cadence (65-70 rpm).

Our analysis does not indicate that optimum cadence varies by power output. We fitted a model in which cadence interacted with heart rate, yielding different optimum cadences for different power outputs or different heart rates, but statistical meaures did not indicate that this model was superior to a model without such an interation.

Power output reduces by ~2% for cadences 10 revolutions per minute below or above optimum. These are in accordance with some literature from laboratory based ergometer studies, but less than the typical cadences chosen by elite athletes in competition (Foss and Hallen, 2004). The athletes in our data are below elite level, so it is possible that higher cadences would be yielded for elite athletes. Expected power outputs for our athletes are also lower than they would be for category I athletes; fitted power output is typically approximately 200W for the athletes in our analysis. However given the subjective nature of cadence selection amongst cyclists, it is not certain that cyclists always choose the most appropriate cadence. For the two athletes in our analysis for whom optimum cadences were not yielded, it is most likely that they did not ride a wide enough range of cadences for a model to pick up a peak cadence, above which power output decreases.

For non-linear models of power output, heart rate and cadence, we retain sampled data in which the heart rate is greater than the mean heart rate for each athlete. We believe this is reasonable as it is likely to more closely represent conditions in races, as it excludes data from long, low intensity training sessions in which athletes are not exerting a particularly great amount of effort, and athletes are likely to be exerting great effort in races. Though heart rate is not a direct measure of effort, we would expect heart rate to be higher in races than in low intensity training sessions. Optimum cadences yielded are indeed lower when we retain only high heart rate data, and they are yielded for two athletes for high heart rate data but for only one athlete when we include all sampled data.

There are two main limitations of our study. Firstly, optimum cadence depends on heart rate lag. A study by Stirling et al (2008) has suggested that heart rate lags behind changes in power output by approximately 30 seconds or less. We considered a range of heart rate lags, with explanatory power typically highest for a lag of 30 seconds (although standard errors were lowest for a lag of 5 seconds). In accordance with literature and patterns in explanatory power, we suggest 30 seconds is an appropriate time lag between changes in power output and heart rate response for training sessions in competitive cycling. Secondly, optimum cadence may depend on riding mode, which means whether an athlete is standing or sitting on the bicycle. There may even be two optimum cadences: one for when the athlete is standing and one for when the athlete is sitting. However no such information regarding riding model was available to us, and it is unlikely that such information could be collected and available at 5 second intervals without the introduction of new monitoring technology.

A further limitation of our study is that we do not measure or model gradient in this study. As such we do not measure any effect of gradient on optimum cadence. Future research could be conducted to relate gradient to optimum cadence.

We considered the effect of fatigue on optimum cadence by calculating training impulses (TRIMPs), which are a standard way of measuring training load. For the athletes for whom an optimum cadence is yielded in our study, fatigue reduces the optimum cadence, although only by a small margin; typically optimum cadence following an easy month of training is 1 rpm higher than optimum cadence following a hard month of training. It is possible that a bigger effect of TRIMP could be found if more data were used, over a longer period. The differences between the optimum cadences of the two athletes is typically 8 rpm, which is much greater than the influence of fatigue on optimum cadence in our study. Individual difference between the athletes, such as muscle fibre type composition, may therefore have a fundamental affect on optimum cadence, more than fatigue. We also considered the effect of TRIMP variation within a session, but some long, low intensity sessions made this effect difficult to pick up.

The athletes in our study are typically category II and III cyclists competing at County Championship level. Nevertheless this methodology, to calculate optimum cadence, can be adopted by athletes with a wide range of experience and skill. Athletes must ride with a power meter to collect power output, heart rate and cadence measurements from each training session undertaken, and upload such data after each session to a spreadsheet. Ideally this should be for a number of months to collect enough data, and they should ride a wide range of cadences within their training sessions. Hence any athlete who can commit to regular, long training sessions, who works with a professional coach and who can purchase a power meter can adopt this methodology, ranging from regional and national championship athletes, to Olympic and World Championship athletes.

Chapter 9: Conclusions and future work

We believe optimum cadences can indeed be calculated statistically using field data for individual athletes, provided athletes ride at a range of cadences either side of their preferred cadence, or either side of the cadence that they believe to be optimal. The optimum cadences yielded from our study were approximately 83 and 70 revolutions per minute for two athletes respectively, which is similar to literature (Abbiss et al, 2009; Coast and Welch, 1985). Higher cadences may have been found for elite, category I athletes (Foss and Hallen, 2004). Our analysis suggests that optimum cadence does not vary with power output or heart rate (or how hard the session is). Optimum cadences are close to the athletes' preferred cadences. Optimum cadence is higher for athlete 1 in our analysis than for athlete 2 (83 rpm compared to 70 rpm), whilst athlete 1 typically chooses to ride at a higher cadence (82-92 rpm) than does athlete 2 (65-75 rpm). This may suggest that physiological differences between athletes that underpin choice of cadence may also be the same physiological differences that lead to different optimum cadences between the athletes (such as muscle fibre type). Our analysis also suggests that optimum cadence does not vary with fatigue.

Our research has some limitations. Optimum cadence depends on the length of time delay between changes in power output and heart rate response, although we believe 30 seconds is an appropriate time lag. We also do not consider whether the athletes in our analysis are standing or sitting on the saddle.

In order to implement the methodology, athletes can give their collected data to their coach. The coach can then work with a performance analyst, to use statistical software to obtain the parameters that yield the value of the optimum cadence (parameters α and β in our non-linear models of power output, heart rate and cadence). Within a race, athletes can monitor their cadence on screen (on the bicycle), to ensure they are riding at their indiviaul optimum cadence. After calculating an optimum cadence, athletes can continue to collect power output, heart rate and cadence data from training sessions, and update their optimum cadence after, say, another six months, and so on. Although we do not model fitness in the statistical models used, if fitness changes, any influence of fitness on optimum cadence would yield itself as a change in the optimum cadence. Indeed a fitter athlete may be able to sustain a higher cadence than a less fit athlete.

Future work could be conducted to include a model with changes in fitness. An extra variable, readiness to perform, could be calculated. This would be an accumulation of fitness minus fatigue (Banister and Calvert, 1975). Currently we do not know what appropriate values should be used for the parameters in such a variable. If more research is conducted to find reliable estimates for these parameters, models could be fitted in which optimum cadence is dependent on readiness to perform. Future work could also fit our models to a larger group of athletes, with athlete effects considered random, rather than fixed.

Selection of cadence tends to be quite subjective, with athletes choosing what is comfortable, with some input from coaches and fellow athletes. This methodology allows for a more precise calculation of optimum cadence, and through repeated implementation, varying with changes in fitness. It is also calculated using data which can easily be gathered (from training sessions), without need for separate laboratory experiments.

References

Abbiss, C. R., Peiffer, J. J., and Laursen, P. B. (2009). Optimal cadence selection during cycling. *International SportMed Journal*, 10(1): 1-15.

Aisbett, B., Le Rossignol, P., McConell, G.K., Abbiss, C.R. and Snow R. (2009). Effects of starting strategy on 5-mon cycling time trial performance. *Journal of Sports Sciences,* 27(11): 1201-1209.

Aisbett, B., Le Rossignol, P., Sparrow, W. A. (2003). The Influence of pacing during 6-minute supra-maximal cycle ergometer performance. *Journal of Science and Medicine in Sport*, 6(2): 187-198.

Atkinson, G. and Brunskill, A. (2000). Pacing strategies during a cycling time trial with simulated headwinds and tailwinds. *Ergonomics*, 43 (10): 1449-1460.

Atkinson, G., Davison, R., Jeukendrupp, A. and Passfield., L. (2003). Science and cycling: current knowledge and future directions for research. *Journal of Sports Sciences,* 21: 767-787.

Atkinson, G., Peacock, O. and Passfield, L. (2007). Variable versus constant power strategies during cycling time trials: Prediction of time savings using an up-to-date mathematical model. *Journal of Sports Sciences*, 25(9), 1001-1009.

Banister, E. W., and Calvert, T. W. (1975). Planning for future performance: implications for long term training. *Canadian Journal of Applied Sports Science*, 5: 170-176.

Barstow, T. J., Jones, A. M., Nguyen, P. H., and Casaburi, R. (1996). Influence of muscle fibre type and pedal frequency on oxygen uptake kinetics of heavy exercise. *Journal of Applied Physiology*, 81: 1642-1650.

Bassett, J.R., Kyle, C.R., Passfield, L., Broker, J.P. and Burke, E.R. (1999). Comparing cycling world hour records, 1967-1 996: modelling with empirical data. *Medicine and Science in Sports and Exercise,* 31: 1665-1676.

Bertucci, W., Duc, S., Villerius, V., Pernin, J. N. and Grappe, F. (2005). Validity and Reliability of a New Mobile Cycle Power meter. *International Journal of Sports Medicine*, 26: 868-973.

Billat, V. L., Ruddy, R., Binsse, V. M., Koralsztein, J. P. and Haouzi, P. (1998). The $V0_2$ slow component for severe exercise depends on type of exercise and is not correlated with time to fatigue. *Journal of Applied Physiology*, 85: 2118-2124.

Bishop, D., Bonetti, D., & Dawson, B. (2002). The influence of pacing strategy on VO_2 and supramaximal kayak performance. *Medicine and Science in Sports and Exercise*, 34: 1041–1047.

Borreson, J. and Lambert, M. I. (2009). The Quantification of Training Load, the Training Response and the Effect on Performance. *Sports Medicine*, 39(9): 779-795.

Broker, J.P., Kyle, C.R. and Burke, E.R. (1999). Racing cyclist power output requirements in the 4000-m individual and team pursuits. *Medicine and Science in Sports and Exercise,* 31: 1677-1685.

Brooke, M. H. and Kaiser, K. K. (1970). Muscle fibre types: How many and what kind? *Arch Neurol*, 23: 369-379.

Capelli, C., Rosa, G., Butti, F., Ferretti, G., Veicsteinas, A., and di Prampero, P. E. (1993). Energy cost and efficiency of riding aerodynamic bicycles. *European Journal of Applied Physiology*, 67: 144–149.

Casella, G. and Berger, R. L. (2002) *Statistical Inference*, 2nd ed.

Cejeula-Anta, R. and Esteve-Lanao, J. (2011). Training load quantification in triathlon. *J. Hum. Sport Exercise.* 6(2): 218-232.

Chaffin, M. E., Berg, K., Zuniga, J. and Hanumanthu, V.S. (2008). Pacing pattern in a 30-minute maximal cycling test. *Journal of Strength and Conditioning Research*, 22(6): 2011-2017.

Chavarren, J. and Calbet, J. A. L. (1999). Cycling efficiency and pedalling frequency in road cyclists. *European Journal of Applied Physiology*, 50: 555-563.

Cheng, T. M., Savkin, A. V., Celler, B. G., Wang, L. and Su, S. W. (2007). Proceedings of the 29th Annual International Conference of the IEEE EMBS Cite Internationale, Lyon, France, August 23-26.

Churchill, T., Sharma, D. and Balachandran, B. M. (2009). Correlation of training load and heart rate variability indices in elite cyclists. In Mathematics in Sport II (Koning and Scarf, eds.), The Institute of Mathematics and Its Applications, UK, pp. 191-196.

Coast, J. R. and Welch, H. G. (1985). Linear increase in optimal pedal rate with increased power output in cycle ergometry. *European Journal of Applied Physiology*, 53: 339-342.

Conconi, F., Ferrari, M., Ziglio, P. G., Droghetti, P. and Codeca, L. (1982). Determination of the anaerobic threshold by a non-invasive field test in runners. *Journal of Applied Physiology*, 52(4): 869-873.

Coyle, E. F., Feltner, M. E. and Kautz, S. A. (1991). Psychological and biomechanical factors associated with elite endurance cycling performance. *Medicine and Science in Sports and Exercise*, 23: 93-107.

Coyle, E. F., Sidossis, L. S., Horowitz, J. F., and Beltz, J. D. (1992). Cycling efficiency is related to the percentage of Type I muscle fibres. *Medicine and Science in Sports and exercise*, 24(7): 782-788.

Crow, M. T. and Kushmerick, M. J. (1982). Chemical energetic of slow- and fast- twitch muscles of the mouse. *J. Gen. Physiol.* 79: 147-166.

Dahn, K., Mai, L., Poland, J. and Jenkins, C. (1991). Frictional resistance in bicycle wheel bearings, *Cycling Science*, 3(3): 28-32.

de Koning, J.J., Bobbert, M. and Foster, C. (1999). Determination of optimal pacing strategy in track cycling with an energy flow model. Journal of Science and Medicine in Sport, 2(3): 266-277.

de Koning, J.J., Foster, C., Bakkum, A., Kloppenburg, S., Thiel, C., et al. (2011). Regulation of Pacing Strategy during Athletic Competition. PLoS ONE 6(1): e15863. doi:10.1371/journal.pone.0015863

Di Prampero, P. E., Cortilli, G., Mognoni, P. and Saibene, F. (1979). *Journal of Applied Physiology*, 47: 201-206.

Eckermann, P. And Millahn, H. P. (1967). Der Einfluss der Drehzahl auf die Herzfrequenz und die Sauerstoffaufnahme bei konstanter Leistung am Fahrradergometer. *Int Z Angew Physiol* 23: 340 – 344.

Edwards, S. (1993) The heart rate monitor book. Sacramento, Fleet Feet Press.

Elder, G. C. Bradbury, B. K., and Roberts, R. (1982). Variability of fibre type distributions within human muscles. *Journal of Applied Physiology*, 53:1473-1480.

Faria, E. W., Parker, D. L., and Faria, I. E. (2005). The science of cycling: Physiology and training – Part 1. *Sports Medicine*, 35: 285 – 312.

Foss, O. and Hallen, J. (2004). The most economical cadence increases with increasing workload. *European Journal of Applied Physiology*, 92: 443 – 451.

Foster, C. (1998). Monitoring training in athletes with reference to overtraining syndrome. *Medicine and Science in Sports and Exercise*, 30(7): 1164-1168.

Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S., Doleshal, P. and Dodge, C. (2001). A new approach to monitoring exercise training. *Journal of Strength and Conditioning Research*, 15(1): 109-115.

Foster, C., Snyder, A. C., Thomson, N. N., Green, M. A., Foley, M., and Shrager, M. (1993). Effect of pacing strategy on cycle time trial performance. *Medicine and Science in Sports and Exercise*, 25: 383-388.

Gaesser, G. A. and Brooks, G. A. (1975). Muscular efficiency during steady-state exercise: effects of speed and work rate. *Journal of Applied Physiology*, 38: 1132-1139.

Grazzi, G., Alfieri, N., Borsetto, C., Casoni, I., Manfredini,F., Mazzoni, G. and Conconi, F. (1999). The power output/heart rate relationship in cycling: test standardization and repeatability. *Medicine and Science in Sports and Exercise*, 31(10): 1478-1483.

Hagberg, J. M., Mullin, J. P., Giese, M. D. and Spitznagel, E. (1981). Effect of pedalling grate on submaximal exercise responses of competitive cyclists. *Journal of Applied Physiology*, 51: 447 – 451.

Hamilton, M. T., Gonzalez-Alonso, J., Montain, S. J. and Coyle, E. F. (1991). Fluid replacement and glucose infusion during exercise prevent cardiovascular drift. *Journal of Applied Physiology*, 71: 871-877.

Hansen, E. A., Andersen, J. L., Nielsen, J. S. and Sjogaard, G. (2002). Muscle fibre type, efficiency, and mechanical optimal affect freely chosen pedal rate during cycling. *Acta Physiol Scand*, 176: 185-194.

Hickey, M.S., Costill, D.L., McConell, G.K., Widrick, J.J. and Tanaka, H. (1992). Day to day variation in time trial cycling performance. *International Journal of Sports Medicine*, 13: 467-470.

Hill, A. V. (1922). The maximum work and mechanical efficiency of human muscles, and their most economical speed. *Journal of Physiology (London)*, 56: 19-32.

Impellizzeri, F. M., Rampinini, E. and Coutts, A. J. (2004). Use of RPE-based training load in soccer. *Medicine and Science in Sports and Exercise*, 36(6): 1042-1047.

Ingjer, F. (1978). Effects of endurance training on muscle fibre ATP-ASE activity, capillary supply and mitochondrial content in man. *Journal of Physiology*, 194: 419-432.

Jeukendrupp, A. and van Diemen, A. (1998). Heart rate monitoring during training and competition in cyclists. *Journal of Sports Sciences*, 16: 91-99.

Jobson, S.A., Nevill, A.M., Palmer, G.S., Jeukendrupp, A.E., Doherty, M. and Atkinson, G. (2007). The ecological validity of laboratory cycling: Does body size explain the difference between labpratory- and field-based cycling performance? *Journal of Sports Sciences*, 25(1): 3-9.

Jobson, S.A., Passfield, L., Atkinson, G., Barton G and Scarf P.A. (2009). The analysis and utilisation of cycling training data. *Sports Medicine* 39, 833-844.

Jordan, L. and Merrill, E. G. (1979). Relative efficiency as a function of pedalling rate for racing cyclists. *Journal of Physiology (London)*, 296:49-50.

Kyle, C. R. (1988). The mechanics and aerodynamics of cycling. In: *Medical and Scientific Aspects of Cycling*, edited by Burke E. R. and Newsom M. M. Champaign, IL: Human Kinetics, 235– 251.

Lafrenz, A., Wingo, J., Ganio, M. S., and Cureton, K. J. (2008). Effect of ambient temperature on cardiovascular drift and maximal oxygen uptake. *Medicine and Science in Sports and Exercise*, 40(6): 1065-1071.

Leirdal, S. and Ettema, G. (2011). The relationship between cadence, pedalling techniques and gross efficiency. *European Journal of Applied Physiology*, 111: 2885-2893.

Lepers, R., Millet, G. Y. and Maffiuletti, N. A. (2001). Effect of cycling cadence on contractile and neural properties of knee extensors. *Medicine and Science in Sports and Exercise*, 33(11): 1882-1888.

Little, T., and Williams, A. G. (2007). Measures of exercise intensity during soccer training drills with professional soccer players. *Journal of Strength and Conditioning Research*, 21(2): 367-371.

Lucia, A., Hoyos, J., Carvajal, A. and Chicarro, J. L. (1999). Heart rate response to professional road cycling: The Tour de France. *International Journal of Sports Medicine*, 20: 167-172.

Lucia, A., Hoyos, J., Perez, M., Santalla, A. and Chicarro, J. L. (2002). Inverse relationship between $V0_{2max}$ and economy/efficiency in world-class cyclists. *Medicine and Science in Sports and Exercise*. 34(12): 2079-2084.

MacIntosh, B., R., Neptune, R. R., and Horton, J .F. (2000). Cadence, power and muscle activation in cycle ergometry. Medicine and Science in Sports and Exercise, 32(7): 1281-1287.

Martin, J. C., Milliken, D. L., Cobb, J. E., McFadden, K. L. and Coggan, A. R. (1998). Validation of a Mathematical Model for Road Cycling Power. *Journal of Applied Biomechanics*, 14: 276-291.

McCartney, N., Heigenhauser, G. J., and Jones, N. L. (1983). *Journal of Applied Physiology*, 55: 218-224.

McCole, S., Claney, K., Conte, J.-C., Anderson, R. and Hagberg, J.M. (1990). Energy expenditure during bicycling. *Journal of Applied Physiology*, 68: 748- 753.

McCulloch, C.E. and Searle, S.R.(2001). *Generalized, linear, and mixed models*. Wiley, New York.

Medbo, J. I., Mohn, A. C., Tabata, I., Bahr, R., Vaage, O. and Sejersted, O. M. (1988). Aerobic capacity determined by maximal accumulated 02 deficit. *Journal of Applied Physiology*, 64: 50-60.

Michielli, D. W. and Stricevic, M. (1977). Various pedalling frequencies at equivalent power outputs. *NY State J Med*, 77: 744-746.

Morton, R.H., Fitz-Clarke, J.R. and Banister, E. W. (1990). Modelling human performance in running. *Journal of Applied Physiology*, 69(3): 1171-1177.

Nikolopoulos, V., Arkinstall, M.J., and Hawley, J.A. (2001). Pacing strategy in simulated cycle time trials is based on perceived rather than actual distance. *Journal of Science and Medicine in Sport*, 4 (2): 212-219.

Olds, T. S., Norton, K. I., and Craig, N. P. (1993). Mathematical model of cycling performance. *Journal of Applied Physiology,* 75(2): 730-737.

Olds, T. S., Norton, K. L., Lower, E. L., Olive, S., Reay, F. and Ly, S. (1995). Modelling road cycling performance. *Journal of Applied Physiology*, 78(4): 1596-1611.

Padilla, S., Mujika, Inigo, Angulo, F. and Goiriena, J. J. (2000). Scientific approach to the 1-h cycling world record: a case study. *Journal of Applied Physiology*, 89: 1522-1527.

Palmer, G. S., Backx, K., and Hawley, J. A. (1998). Perceived rather than actual distance determines pacing strategy during simulated cycling time trials. *Medicine and Science in Sports and Exercise*, 30:305.

Paton, C. D, and Hopkins, W. G. (2001). Tests of cycling performance. *Sports Medicine*, 31: 489–496.

Pichot, V., Roche, F., Gaspoz, J. M., Enjorlas, F., Antoniadis, A., Minini, P., Costes, F., Busso, T., Lacour, J. R., and Barthelemy, J. C. (2000). Relation between heart rate variability and training load in middle-distance runners. *Medicine and Science in Sports and Exercise*. 32(10): 1729-1736.

Polgar, J., Johnson, M. A., Weightman, D. and Appleton, D. (1973). Data on the distribution of fibre types in 36 human muscles. *J. Neurol. Sci.*, 18: 111- 129.

Poole, D. C., Barstow, T. J., Gaesser, G. a., Willise, W. T. and Whipp, B. J. (1994). $V0_2$ slow component: physiological and functional significance. *Medicine and Science in Sports and Exercise*, 26: 1354-1358.

Price, D. and Donne, B. (1997). Effect of variation in seat tube angle at different heights on submaximal cycling performance in man. *Journal of Sports Sciences,* 15(4): 395-402.

Rowell, L. B., Brengelmann, G. L., Detry, J. M. R., and Wyss, C. (1971). Venomotor responses to rapid changes in skin temperature in exercising man. *Journal of Applied Physiology*, 30: 64-37.

Samozino, P., Horvais, N. and Hintzy, F. (2006). Interactions between cadence and power output effects on mechanical efficiency during sub maximal cycling exercise. *European Journal of Applied Physiology,* 97: 133 – 139.

Sargeant, A. J. (1994). Human power output and muscle fatigue. *International Journal of Sports Medicine*, 15: 116-121.

Sassi, A., Rampinini, E., Martin, D. and Morelli, A. (2009). Effects of gradient and speed on freely chosen cadence: The role of crank inertial load. *Journal of Biomechanics,* 42: 171-177.

Seabury, J. J., Adams, W. C. and Ramey, M. R. (1977). Influence of Pedalling Rate and Power Output on Energy Expenditure During Bicycle Ergometry. *Ergonomics*, 20 (5): 491-498.

Smith, M. F., Davison, R. C. R., Balmer, J. and Bird, S. R. (2001). Reliability of mean power output recorded during indoor and outdoor self-paced 40 km cycling time-trials. *International Journal of Sports Medicine,* 22: 270 – 274.

Stirling, J. R., Zakynthinaki, M., Refoyo, I. and Sampredo, J. (2008). A Model of Heart Rate Kinetics in Response to Exercise. *Journal of Nonlinear Mathematical Physics*, 15 (3): 426-436.

Stuart, A., Ord K. and Arnold, S. (1999) *Kendall's Advanced Theory of Statistics Vol 2A: Classical Inference and the Linear Model.* Arnold, London.

Swain, D. P. (1997). A model for optimizing cycling performance by varying power on hills and in wind. *Medicine and Science in Sports and Exercise,* 29: 1104 – 1108.

Swain, D. P., Coast, J. R., Clifford, P. S., Milliken, M. C., and Stray-Gunderson, J. (1987). Influence of body mass in endurance bicycling. *Medicine and science in Sports and Exercise*, 26: 58-63.

Swain, D. P. and Leutholtz, B. C. (1997). Heart rate reserve is equivalent to %V02 reserve, not to %V02max. *Medicine and Science in Sports and Exercise,* 29(3): 410-414.

Thomas, K., Stone, M. R., Thompson, K. G., Gibson, A. S. and Ansley, L. (2012). The effect of self- even- and variable-pacing strategies on the physiological and perceptual response to cycling. *European Journal of Applied Physiology*, 112: 3069-3078.

Thompson, K. G., MacLaren, D. P., Lees, A., & Atkinson, G. (2003). The effect of even, positive and negative pacing on metabolic, kinematic and temporal variables during breaststroke swimming. *European Journal of Applied Physiology*, 88: 438–44.

Ulmer, H. V. (1996). Concept of an extracellular regulation of muscular metabolic rate during heavy exercise in humans by psychophysiological feedback. *Experentia,* 52: 416–420.

Van Ingen Schenau, G. J., de Koning, J. J., & De Groot, G. (1992). The distribution of anaerobic energy in 1000 and 4000 metre cycling bouts. *International Journal of Sports Medicine,* 13: 447–451.

Weisberg, S. (2005) *Applied Linear Regression.* Wiley, New York.

Welbergen, E. and Clijsen, L. P. V. M. (1990). The influence of body position on maximal performance in cycling. *European Journal of Applied Physiology and Occupational Physiology,* 61(1): 138-142.

Wells, M., Atkinson, G., and Marwood, S. (2013). Effects of magnitude and frequency of variations in external power output on simulated cycling time-trial performance. *Journal of Sports Sciences,* May 28 [PubMed Epub ahead of print]; DOI: 10.1080/02640414.2013.794299

Whipp, B. J. and Wasserman, K. (1969). Efficiency of muscular work. *Journal of Applied Physiology*, 26: 644 – 648.

Whitt, F. R. (1971). A note on the estimation of the energy expenditure of sporting cyclists. *Ergonomics* 14: 419-424.

Wilberg, R. B., & Pratt, J. (1988). A survey of the race profiles of cyclists in the pursuit and kilo track events. *Canadian Journal of Sport Sciences,* 13: 208–213.

Wilkie, D. R. (1950). The relation between force and velocity in human muscle. *Journal of Physiology (London),* 110: 249-280.

Wingo, J. E., Lafrenz, A. J., Ganio, M. S., Edwards, G. L., and Cureton, K. J. (2005). Cardiovascular Drift Is Related to Reduced Maximal Oxygen Uptake during Heat Stress. *Medicine and Science in Sports and Exercise*, 37(2): 248-255.

Winter, E. M., and Fowler, N. (2009). Exercise defined and quantified according to the Systeme International d'Unites. *Journal of Sports Sciences,* 27 (5): 447 – 460.

Appendix

We present tables of estimates for non-linear models of power output and cadence, in table 87. We calculate and present R^2, AIC and BIC to assess the model fit, presented in table 88. In the appendix we also present fitted power cadence plots for a selection of the non-linear models of power output and cadence. For each heart rate range chosen here, we plot power output from the non-linear model of power output and heart rate (equation 10) along with power output measurements in the data, against cadence. We plot for each athlete separately (figures 24 to 28).

We next present example code for the non-linear model of power output, heart rate and cadence (in equation 7) fitted in R.

We include plots of power output against cadence from non-linear models of power output, heart rate, cadence and additional covariates. We present fitted plots for the non-linear model with extra covariates and no interactions (equation 10) in figures 29 and 30. We set heart rate, normalized power output X_{1t}, proportion of the session that has elapsed up to the current sampling point X_{2t}, mean power output over the previous seven days X_3, and mean session duration over the previous seven days (hours) X_4 to the mean values found in the data for each athlete. For the first plot, all sampled data are included, whilst for the second plot, only data in which heart rate is greater than the overall mean heart rate for each athlete are retained.

Table 88: residual standard errors of the non-linear model of power output and cadence (equation 5) and corresponding one parameter null models for different heart rate subsets, fitted in R for athletes 1 to 4.

					null model equivalent		
				Res			Res
HR	range		n	s.e.	µ	s.e.	s.e
40	-	50	136	12.10	36.072	4.214	49.150
50	-	60	187	24.08	44.378	3.995	55.350
60	-	70	340	22.71	39.942	2.062	38.030
70	-	80	915	23.28	42.876	1.383	41.830
80	-	90	2006	27.32	64.775	1.044	46.760
90	-	100	3330	30.24	95.707	0.872	50.300
100	-	110	5071	28.71	125.718	0.734	52.230
110	-	120	6094	28.38	151.552	0.730	57.010
120	-	130	7171	32.82	177.858	0.715	60.580
130	-	140	8663	33.11	203.741	0.635	59.090
140	-	150	9151	29.68	221.550	0.617	59.050
150	-	160	5065	33.98	243.288	0.973	69.270
160	-	170	2992	48.04	284.318	1.671	77.640
170	-	180	587	52.21	264.557	3.086	74.780
180	-	190	184	58.38	253.632	6.372	86.430
190	-	200	28	107.70	232.520	20.050	106.100
		ALL	51787	48.22	178.941	0.367	83.520

Table 89: AIC, BIC and explanatory power for the non-linear model of power output and cadence (equation 5), for athletes 1 to 4, with data split into sub-groups depending on heart rate.

HR range	n	AIC	BIC	R^2
40 - 50	136	532.5	555.8	80.2
50 - 60	187	749.9	775.8	69.7
60 - 70	340	1207.1	1237.7	62.6
70 - 80	915	3222.2	3260.8	64.2
80 - 90	2006	7388.1	7432.9	63.1
90 - 100	3330	12962.8	13011.7	62.5
100 - 110	5071	20617.0	20669.2	64.5
110 - 120	6094	25876.7	25930.4	66.8
120 - 130	7171	31569.2	31624.2	64.9
130 - 140	8663	38571.7	38628.3	64.1
140 - 150	9151	40318.4	40375.4	66.6
150 - 160	5065	22913.9	22966.1	67.1
160 - 170	2992	14246.5	14294.5	61.8
170 - 180	587	2599.2	2634.2	58.9
180 - 190	184	802.3	828.0	59.7
190 - 200	28	145.1	155.7	49.6
ALL	51787	232036.6	232107.5	63.4

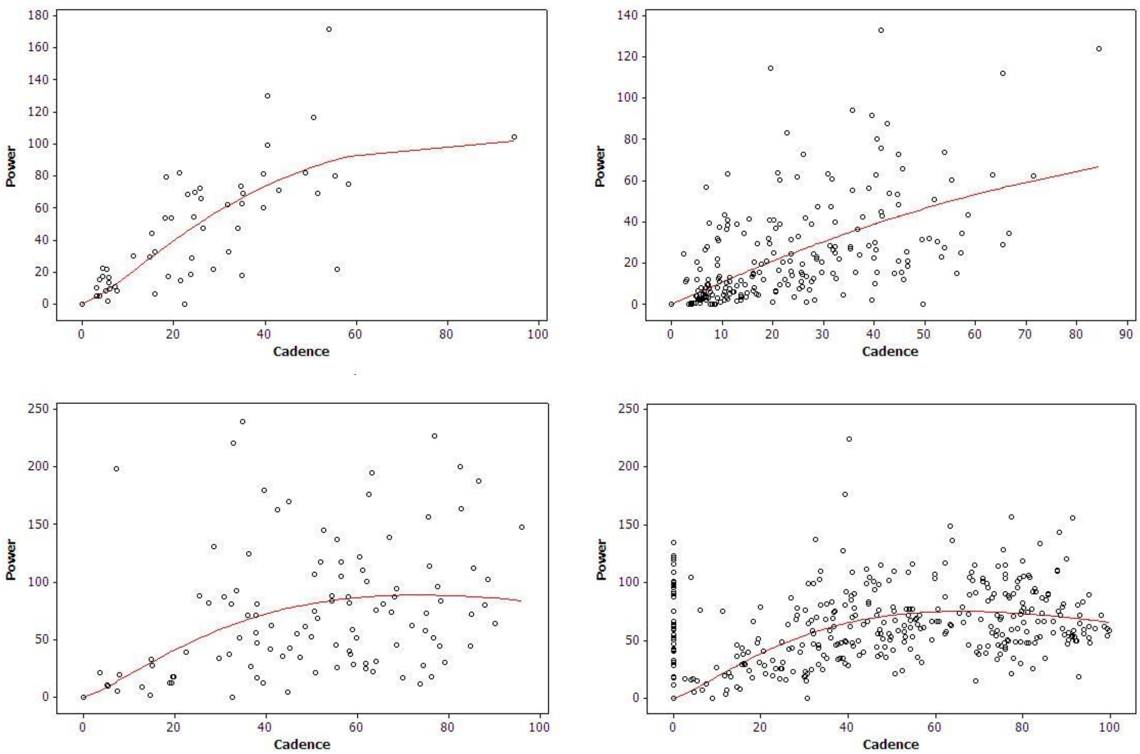


Figure 24: Power output measurements (black dots) and fitted Power output (red lines) against cadence for athletes 1 to 4 (read lexicographically), for the model in equation 5. This is for data in which smoothed heart rate is recorded to be between 70 and 80 beats per minute.

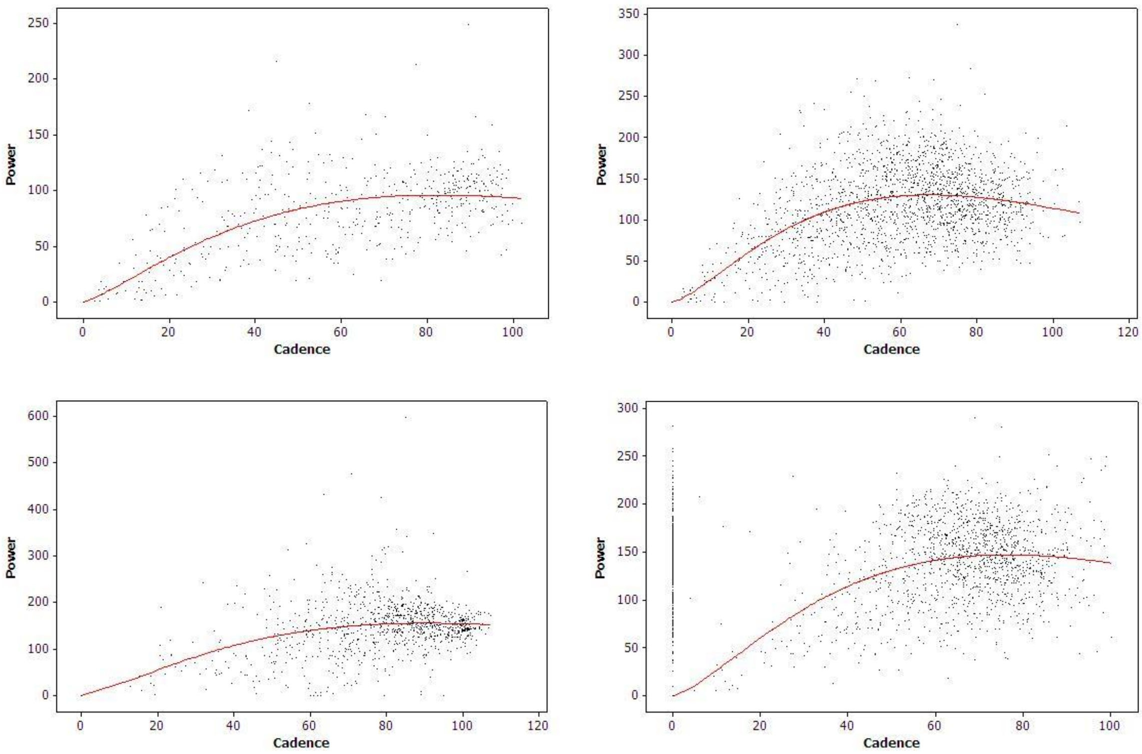


Figure 25: Power output measurements (black dots) and fitted Power output (red lines) against cadence for athletes 1 to 4 (read lexicographically), for the model in equation 5. This is for data in which smoothed heart rate is recorded to be between 100 and 110 beats per minute.

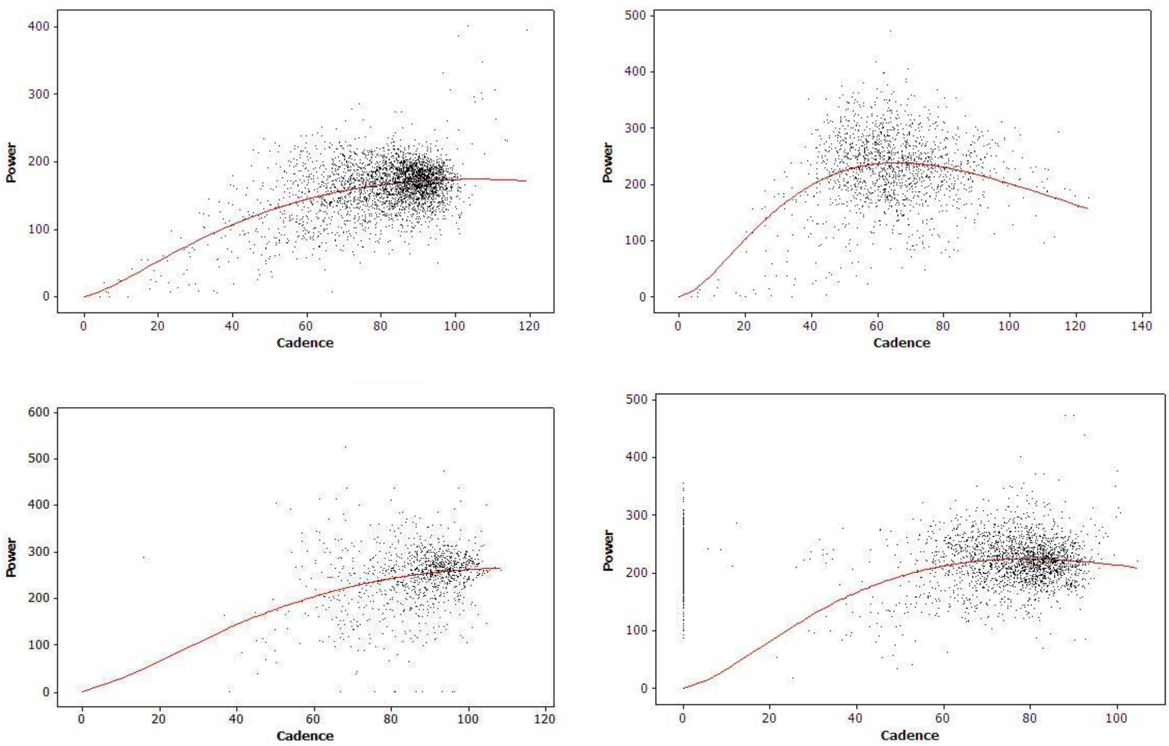


Figure 26: Power output measurements (black dots) and fitted Power output (red lines) against cadence for athletes 1 to 4 (read lexicographically), for the model in equation 5. This is for data in which smoothed heart rate is recorded to be between 130 and 140 beats per minute.

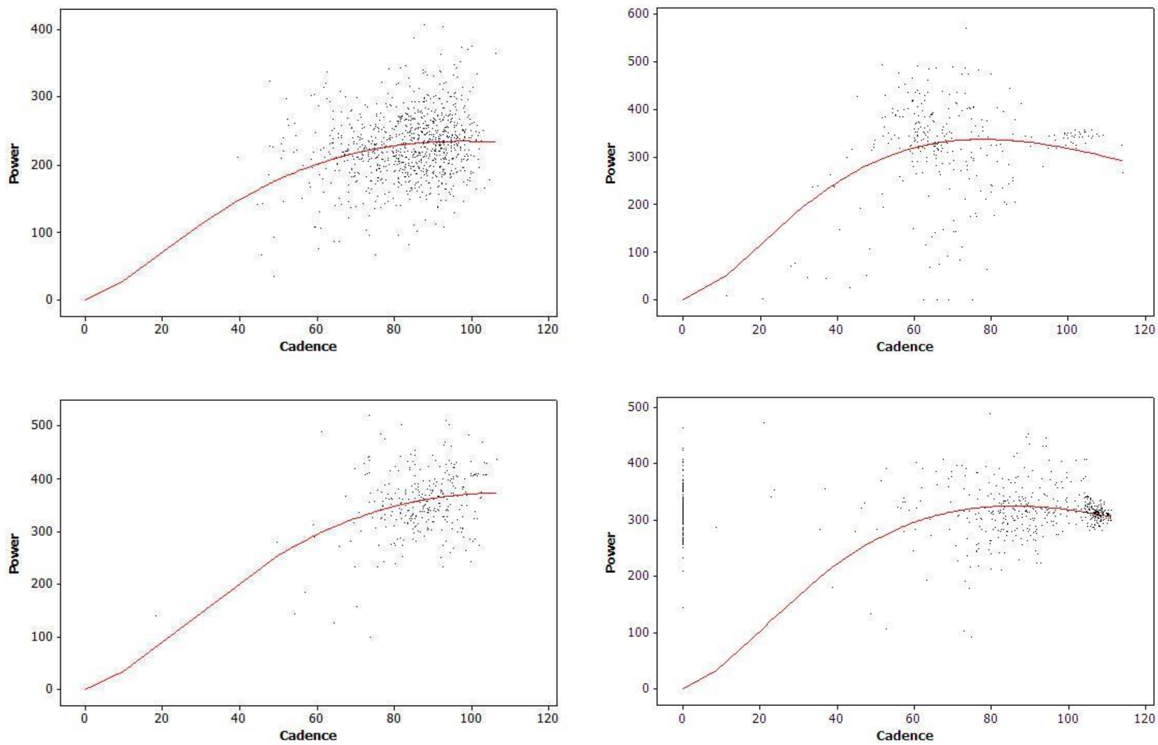


Figure 27: Power output measurements (black dots) and fitted Power output (red lines) against cadence for athletes 1 to 4 (read lexicographically), for the model in equation 5. This is for data in which smoothed heart rate is recorded to be between 160 and 170 beats per minute.

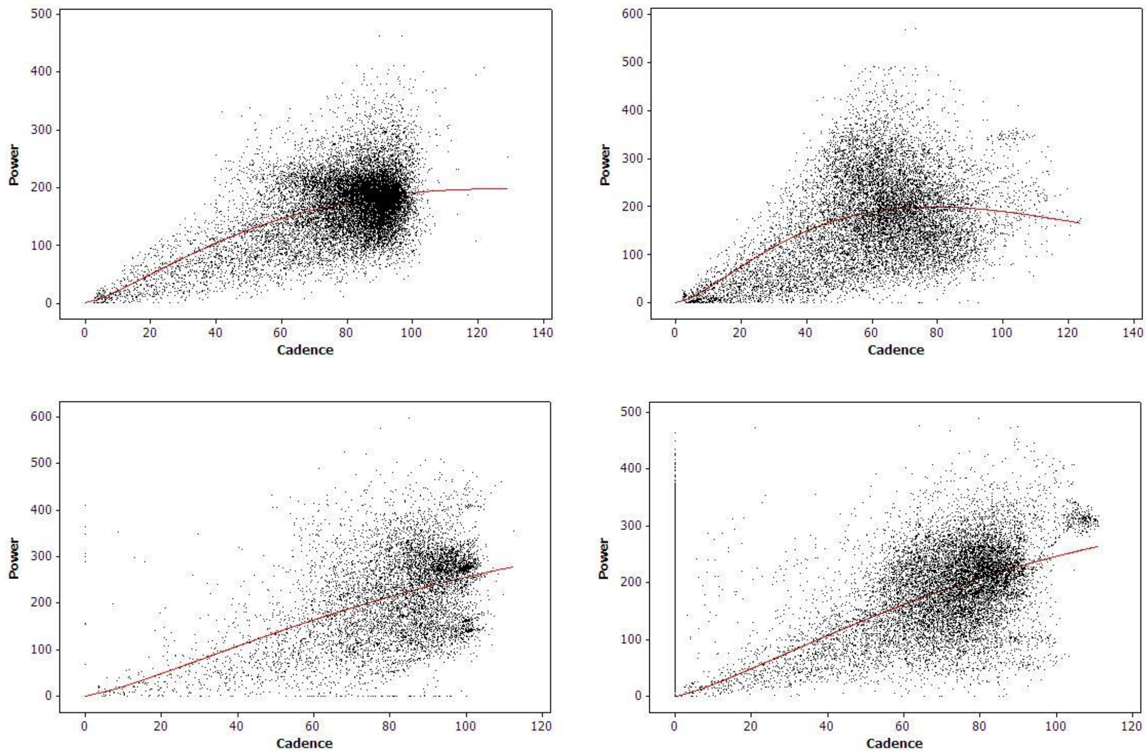


Figure 28: Power output measurements (black dots) and fitted Power output (red lines) against cadence for athletes 1 to 4 (read lexicographically), for the model in equation 5. This is for data in which all smoothed heart rate recordings are included.

Example R code for the simple non-linear model of power output, heart rate and cadence (equation 7) is given below, for athlete 1. The code “$A==’1’”indicates this is for athlete 1 in our analysis; subsequently this number is changed accordingly when we fit this code for athlete 2, 3 and 4.

```
mydata <- read.table(file.choose(), header=TRUE, sep=",")

attach(mydata)

newdata <- mydata[ which(mydata$A=='1'), ]

attach(newdata)

gammatestnew <- nls(log_P~I(logmu+(a*log_C)-(b*C)+(d*log_H_lag6)), data=newdata,
start=list(a=1, b=0.01, d=1, logmu=-5))

summary(gammatestnew)

vcov(gammatestnew)
```

Next we present power / cadence plots for the non-linear model with extra covariates (equation 10).

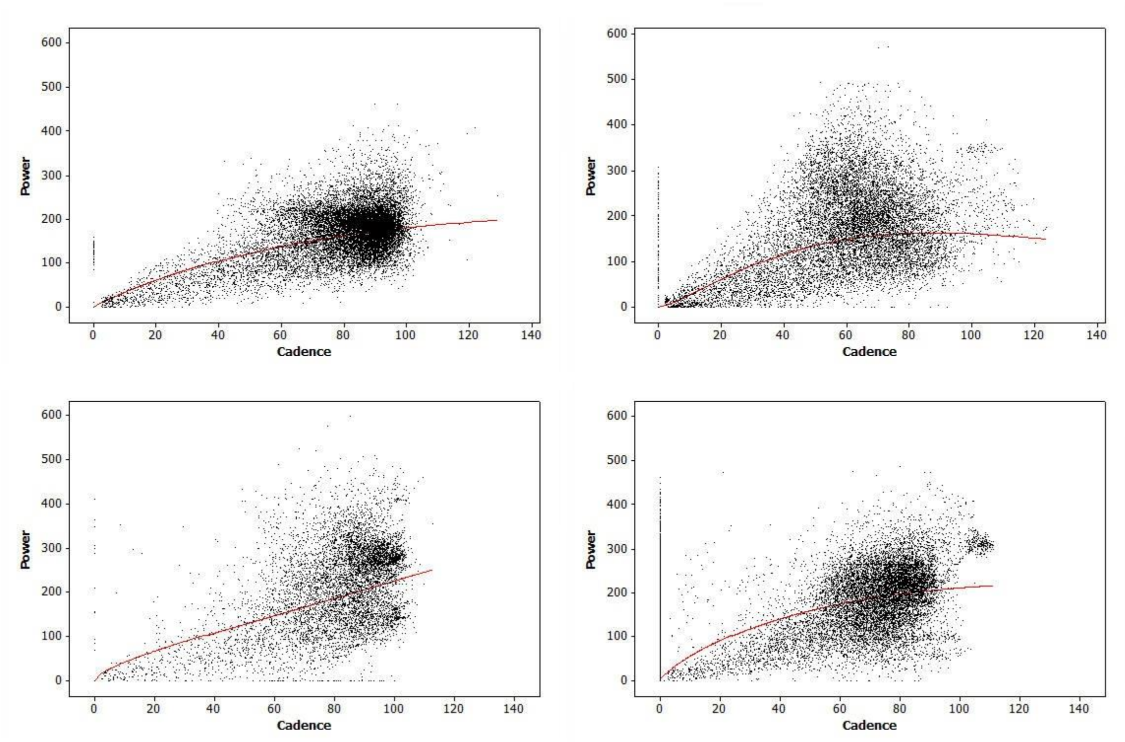


Figure 29: Fitted power output (red lines) from the non-linear model with extra covariates (equation 10) and power output measurements (black dots) against cadence, for athletes 1 to 4 (read lexicographically), for the high heart rate data subset. Heart rate is lagged at 5 seconds.

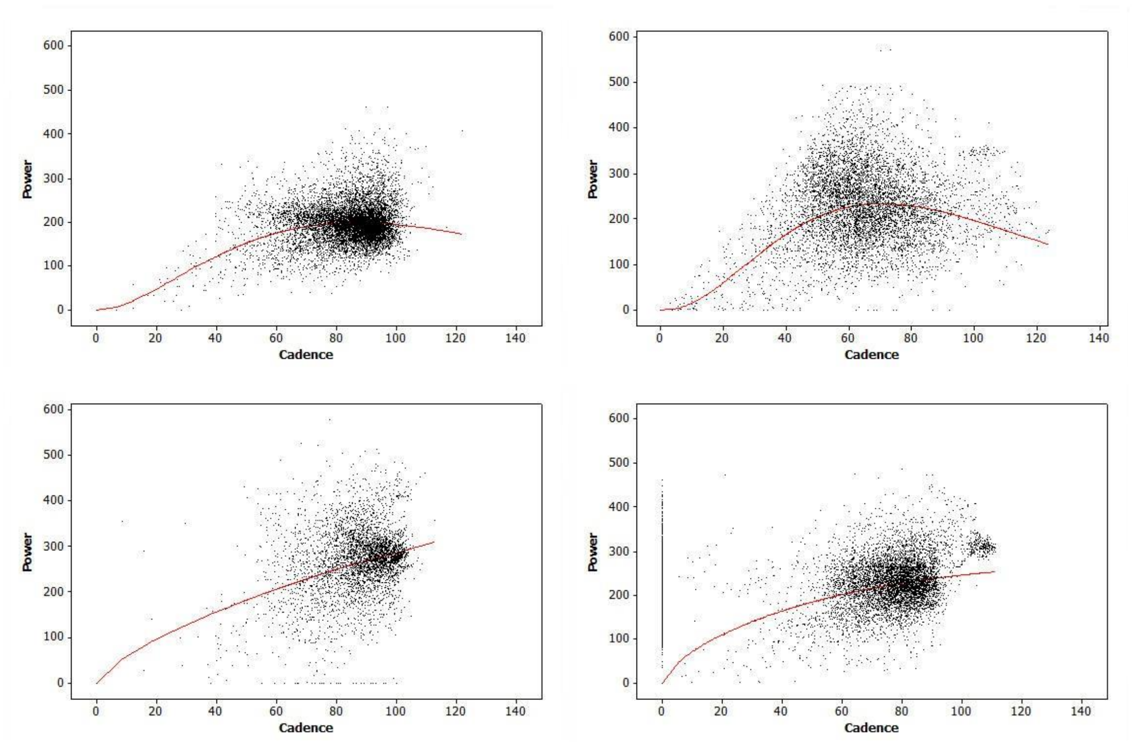


Figure 30: Fitted power output (red line) from the non-linear model with extra covariates (equation 10) and power output measurements (black dots) against cadence, for athletes 1 to 4 (read lexicographically), for the high heart rate data subset. Heart rate is lagged at 5 seconds.