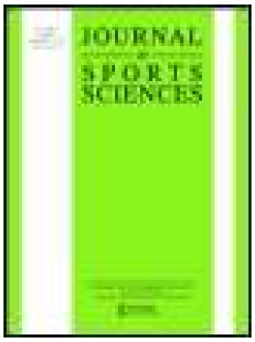

Journal of Sports Sciences

Routledge
Taylor & Francis Group

ISSN: 0264-0414 (Print) 1466-447X (Online) Journal homepage: http://www.tandfonline.com/loi/rjsp20

Knowledge is power: Issues of measuring training and performance in cycling

L. Passfield, JG. Hopker, S. Jobson, D. Friel & M. Zabala

JOURNAL OF SPORTS SCIENCES, 2016
http://dx.doi.org/10.1080/02640414.2016.1215504

Routledge
Taylor & Francis Group

Knowledge is power: Issues of measuring training and performance in cycling

L. Passfield[a], JG. Hopker[a], S. Jobson [b], D. Friel[c] and M. Zabala[d,e]

[a]Endurance Research Group, School of Sport and Exercise Sciences, University of Kent, Chatham Maritime, UK; [b]Poligono Industrial de Egües, Egües (NAVARRA), Spain; [c]TrainingPeaks, Peaksware, Boulder, CO, USA; [d]Faculty of Sport Sciences, University of Granada, Granada, Spain; [e]Movistar pro-Cycling Team, Spain

ABSTRACT

Mobile power meters provide a valid means of measuring cyclists' power output in the field. These field measurements can be performed with very good accuracy and reliability making the power meter a useful tool for monitoring and evaluating training and race demands. This review presents power meter data from a Grand Tour cyclist's training and racing and explores the inherent complications created by its stochastic nature. Simple summary methods cannot reflect a session's variable distribution of power output or indicate its likely metabolic stress. Binning power output data, into training zones for example, provides information on the detail but not the length of efforts within a session. An alternative approach is to track changes in cyclists' modelled training and racing performances. Both critical power and record power profiles have been used for monitoring training-induced changes in this manner. Due to the inadequacy of current methods, the review highlights the need for new methods to be established which quantify the effects of training loads and models their implications for performance.

ARTICLE HISTORY
Accepted 15 July 2016

KEYWORDS
Modelling; endurance; training; cycling; performance; power output

Introduction

Mobile power meters are devices that can be fitted to a bicycle to measure cyclists' power output in the field. The detailed data obtained from power meters can then be used to monitor and evaluate cyclists' training and race performances. This power output data can be gathered in a range of field conditions including cycling on the road, track, off-road or even indoors. The data obtained can also be used in different way depending on the cycling discipline to inform decisions relating to cycling position and technique (i.e., the effect of position or technique change on physiological parameters at a set power output), competition demands, and team and equipment selection. Power meters were first developed in the 1980s with SRM (Schoberer Rad Messtecnik, Jülich, Welldorf, Germany) generally being acknowledged as the first to produce a commercially available system. Early adopters of the SRM system included the East German national cycling team, and Greg Lemond in the European professional peloton. Since its inception the SRM power meter has established itself as the standard against which others are compared. In recent years, the market for power meters has developed considerably and there are now a number of manufacturers producing devices (e.g., Cycleops Powertap, Stages Cycling Powermeter, Garmin Vectors). Their technological approaches to measuring power output vary, but the most common method is to use strain gauges to measure the torque generated by the cyclist. Power output can be measured from a number of locations in the propulsive transmission system of a bicycle. Thus power meters can derive their measurement from the shoe (e.g., Zone DPMX), pedal (e.g., Garmin Vector), crank (e.g., Stages Powermeter), bottom bracket axle (e.g., Rotor INpower), chain (e.g., Wattbike), or hub (e.g., Cyclops Powertap). The utility and success of these approaches depend upon the particular power meter's measurement method and location. The majority of commercially available power meters measure torque directly at the pedal, crank or rear wheel. The specific position of the power meter on the bicycle can be important for some cyclists. For example, track sprinters may be more interested in monitoring torque produced, i.e., at the pedal or crank, rather than power output delivered to the wheel (at the hub). However, the primary concern for most power meter users is their validity sensitivity, reproducibility and, repeatability of measurement.

Validity

The validity of the power meter can be high where power output is measured directly and calculated from its derivatives, angular velocity multiplied by torque Abbiss et al. (2009). For example, at the rear hub angular velocity is calculated from wheel rotation, and torque from the force transmitted by the chain to the hub. The principle is similar at the pedal or crank, except angular velocity is given by cadence. The use of strain gauges allows accurate measurement of torque, but they are sensitive to changes in ambient temperature (Gardner et al., 2004; Wooles, Robinson, & Keen, 2005). Therefore, care is needed in calibration, especially at the start of the ride, if the bicycle is moved from a warm to a cold location, for example. The placement of the strain gauges dictates whether measured torque is separate for each leg, combined across both legs or measured for only one leg (and doubled). Instrumenting the pedals allows the torque pattern of left

CONTACT Louis Passfield l.passfield@kent.ac.uk Endurance Research Group, School of Sport and Exercise Sciences, University of Kent, Canterbury, UK

and right legs to be measured separately. This makes possible analysis of negative forces, generated as the pedal rises between bottom and top dead centre, and any bilateral asymmetry in pedalling style. Measurement of the combined torque of both legs occurs where the bicycle is instrumented anywhere in its propulsive transmission after the bottom bracket axle. This method cannot quantify ineffective torque, although some gross pedalling asymmetry may still be detectable. Moreover, although some power meters purport to examine negative forces, this requires a constant measurement of angular velocity, which most devices do not measure, instead calculating average angular velocity at every revolution. A simple approach to determining power output is to bond strain gauges to a single crank and measure the torque from one leg only. Total power output is calculated as double the measured value, by assuming an equal and symmetrical contribution for the unmeasured leg. The validity of this assumption for pedalling symmetry remains unclear. Smak, Neptune, and Hull (1999) found that asymmetry is related to limb dominance, and reported asymmetry ranging from 0.5% to 2.0%. Carpes, Mota, and Faria (2010) reviewed a number of studies with asymmetry values ranging from 5% to 20%. They also noted that increasing cadence and power output tend to improve indices of symmetry. Therefore, where an overall measure of work rate in the field is required, power meters relying on a single crank measurement may be sufficient. For careful comparison between cyclists and work rates, stable bilateral symmetry should not be assumed though.

The principle of the power meter is valid, but the expected power output and its accuracy can vary according to the measurement conditions. The location of the power meter on the bicycle affects the expected power output. Frictional losses especially from the drive train dissipate some of the energy input. Therefore, a difference in simultaneous torque measurements should be found where these are made before and after the drive train, e.g., from the pedal and hub, respectively. Drive train frictional losses are thought to be proportional to the total power output and have been suggested to amount to ~2.4% (Kyle, 1988; Martin, Milliken, Cobb, McFadden, & Coggan, 1988). Regardless of where they are located, most commercially available power meters measure angular velocity simply by detecting complete hub or crank rotations. As a consequence when angular velocity is low or changes notably within a single revolution, the power meter's sensitivity may be affected. Most power meters are unable to evaluate power output until its angular velocity is well above zero. Even once a minimum angular velocity threshold is exceeded, changes within a single revolution cannot be detected. For both of these reasons power output measurement may not be accurate under conditions involving low angular velocity or marked acceleration, such as when evaluating standing starts (Bertucci, Crequy, & Chiementin, 2013; Martin, Gardner, Barras, & Martin, 2006). Under these conditions of low or variable cadence and high torque, it may be preferable to evaluate torque separately.

Accuracy and reliability

The high accuracy and reliability of commercially available power meters have been demonstrated repeatedly (Bertucci, Duc, Villerius, Pernin, & Grappe, 2005; Gardner et al., 2004; Jones & Passfield, 1998; Martin, Milliken, Cobb, McFadden, & Coggan, 1998; Wooles et al., 2005). The early studies (Jones & Passfield, 1998; Martin et al., 1998) mounted SRM power meters onto a laboratory friction-braked ergometer for comparison. Both studies found an $R^2 > 0.99$, and Jones & Passfield reported 95% limits of agreement to be as low as 0.3% between ergometer and power meter. But the assumption that a rope-braked laboratory ergometer provides an accurate reference calibration has been questioned (Franklin, Gordon, Baker, & Davies, 2006; Gardner et al., 2004). Gardner et al. (2004) examined 26 power meters from two different manufacturers (SRM and Powertap), re-testing 15 power meters after 11 months' use. They found that both manufacturers' power meters had similar reproducibility (~2.5% error), with good long-term reliability over 11 months' of use. Wooles et al. (2005) performed repeat calibrations on 185 SRM devices across a period of 18 months. Their reported mean percentage drift in the calibration factor was only −0.15 once three devices with mechanical problems were excluded. Gardner et al. (2004) noted that some discrepancy in power measurement between SRM and Powertap devices was evident at high power outputs when used in the field. Bertucci et al. (2005) reported similarly high agreement when comparing the same manufacturers' power meters, and the same exception for the highest power outputs. Indeed, it is noted that most validity and reliability studies have been conducted across power outputs typical of elite endurance riders. Therefore for starts and sprints such as in the studies of Martin et al. (2006), and Bertucci et al. (2013), it may be worth checking that the linearity of response is maintained across the expected range of measurement. Furthermore, fastidious attention to routine maintenance, e.g., checking tightness of crank and chain ring bolts can be critical to achieving replicable results. In more recent studies not all power meter manufacturers have compared favourably with criterion devices (Bertucci et al., 2013 [G-Cog], Duc, Villerius, Bertucci, & Grappe, 2007 [ErgomoPro], Hurst & Atkins, 2006 [Polar S710], Kirkland, Coleman, Wiles, & Hopker, 2008 [ErgomoPro], Millet, Tronche, Fuster, Bentley, & Candau, 2003 [Polar S710]). Consequently, it appears that the reasonable accuracy of commercial power meters should not be assumed until verified. Once established though, power meters can be used for monitoring training and performance with a long-term accuracy and reproducibility of 2.5% or less. Gardner et al. (2004) point out that this level of accuracy may still present an issue in detecting changes important to competitive cyclists.

Analysing power output data from training and races

Cyclists from recreational to elite use power meters to examine in detail the power output profile for their training or race performances. There are several studies characterising the power output of notable competitive events (Abbiss, Straker, Quod, Martin, & Laursen, 2010; Ebert et al., 2005; Vogt et al., 2006, 2007). In flat road races mean power output for elite men was found to be 220 ± 22 W or 3.1 ± 0.2 W·kg^{-1}, and for a hilly time trial 392 ± 60 W or 5.5 ± 0.4 W·kg^{-1} (Vogt et al., 2006). Mean power output for elite women in flat road races was 192 ± 21 W or 3.3 ± 0.3 W·kg^{-1} (Ebert et al., 2005). In contrast to racing, however, there is relatively little information or analysis of power meter training data, especially for elite cyclists over the course of a season.

To assist in exemplification of how power data from training and racing can be analysed, we present power meter data from the 2011 season of a prolific Grand Tour cyclist in the form of a case study. To enable use to present this data within the review we obtained local university ethics committee approval and informed consent from the cyclist for the use of his data. During the year the Grand Tour cyclist completed 1143 h of training and covered a total of 35,622 km. He competed regularly throughout the 2011 season most notably in the Tirreno-Adriatico, the Spring Classics, the Criterium du Dauphine, the Tour de France, the Eneco Tour and the World Road Championships. In this review, we have restricted our discussion to consider only methods of data interpretation that have been published in peer-reviewed journal articles. There are further proprietary methods, such as Normalised Power™ and Training Stress Score™ that we do not review here as they have not been validated in scientific studies published in peer-reviewed journals despite their common use by coaches and cyclists.

Interpreting mean power output

Figure 1(a,b) illustrates the 30-s rolling mean power output from two training sessions. Analysis for many scientists, athletes and coaches may consist of simple visual inspection to identify characteristics of interest, such as the highest power output, the number of intervals completed or the extent of variation in power output. The mean power output for a training session provides one method of summarising or "smoothing" the variation seen in Figure 1. Reducing a training session to a single number is attractive. The mean power output calculated for sessions in Figure 1(a,b) are 125 W and 269 W, respectively. However, these mean values provide no indication of the degree of variability in power output evident in Figure 1.

Reflecting the implications of such variability usefully presents a major challenge for power meter data analysis. Often the mean power output will not be commensurate with the physiological strain a cyclist experiences unless the training session is constant power in nature. Coggan (2003) proposed the use of an exponentially weighted mean or "normalized power" output to reflect the added stress a cyclist perceives during variable intensity sessions. Using the "normalized power" approach, data are smoothed using a 30-s moving average (as this is the approximate time constant for many physiological processes [e.g., heart rate] to respond to a change in exercise intensity), before being raised to the fourth power (derived from a regression of blood lactate concentration against exercise intensity). The transformed values are then averaged with the fourth root taken to provide the "normalized power". Constant intensity sessions result in this weighted mean remaining unchanged from the actual mean, but for variable intensity sessions it increases as a function of the proportion of higher intensity training completed. As an example the weighted means of the two sessions in Figure 1 (a,b) are increased by their variability from 125 W to 158 W and from 269 W to 307 W, respectively. Although widely used by cyclists to summarise their training sessions and races, the use of a "normalized power" or weighted mean has received limited scientific evaluation (Skiba, 2007). It is important to note that training sessions with very different physiological and metabolic characteristics can still result in the same weighted mean power output. Consequently, a more detailed analysis of power meter data is required where it is important to determine how the volume and intensity of training (and racing) has been distributed. In the sections below we will propose some alternative methods to address the limitations of using averaged or weighted mean power outputs.

Binning training data

The mean and weighted mean provide helpful summary statistics, but cannot convey the power output distribution where a session is variable in nature. Instead, the power output distribution within a session can be described by the amount of time spent within designated training "zones" or data bins. To present the data visually the bins can be plotted to produce a session histogram. Indeed previous studies have used a data binning approach to investigate physiological responses during training and cycling competitions (Lucía, Hoyos, Carvajal, & Chicharro, 1999; Palmer, Hawley, Dennis, & Noakes, 1994). This histogram approach to describing training data is illustrated below with data obtained from a Grand Tour

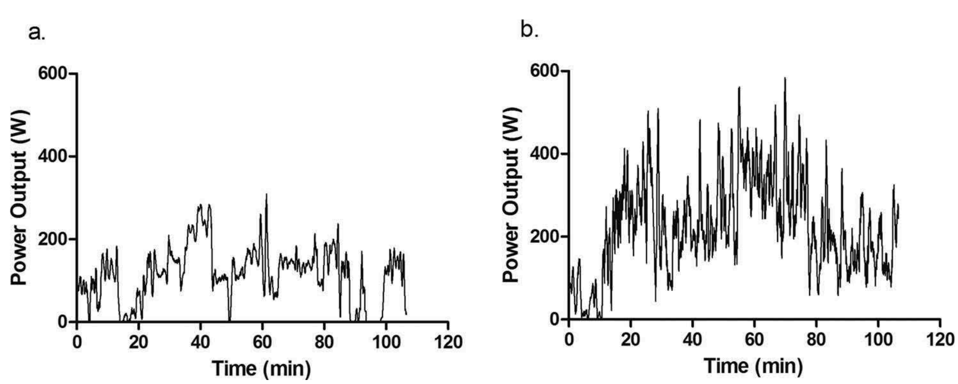

Figure 1. Power output for two training sessions from a professional Grand Tour cyclist. Power output is 30-sec rolling mean. See text for further details.

Cyclist. The histogram illustrated in Figure 2 shows the two training sessions from Figure 1(a) and (b) separated into time bins. Ebert et al. (2005) used a similar comparison for two types of women's World Cup cycle road races. They calculated the percentage of total race time spent within four data bins (0–100 W, 100–300 W, 300–500 W and >500 W). Although simple, this method is excellent for the purpose of overall session comparisons (Jobson, Nevill, & Jeukendrup, 2005).

The use of data binning transposes the complex stochastic power meter data into a simple, easy to interpret output. A further method for analysing power meter data is to calculate the maximum mean power output. This method sub-divides the power meter data into efforts of varying durations or epochs (typically from 5 to 600 s) rather than intensities. The maximum mean power output produced for each of these epochs is then identified (Quod, Martin, Martin, & Laursen, 2010). Changes in the power output associated with each epoch may better reflect specific training effects. However, as the data are collected during training and racing, changes in cadence, gear ratio, drafting, road gradient, environmental conditions and the tactical nature of mass start road races will all affect the power output that is recorded in each epoch. Consequently, it may be more appropriate to examine the maximum mean power output across a period of training or series of races rather than for individual sessions (Quod et al., 2010). Figure 2 demonstrates the maximum mean power output over two periods of the Grand Tour cyclist's season.

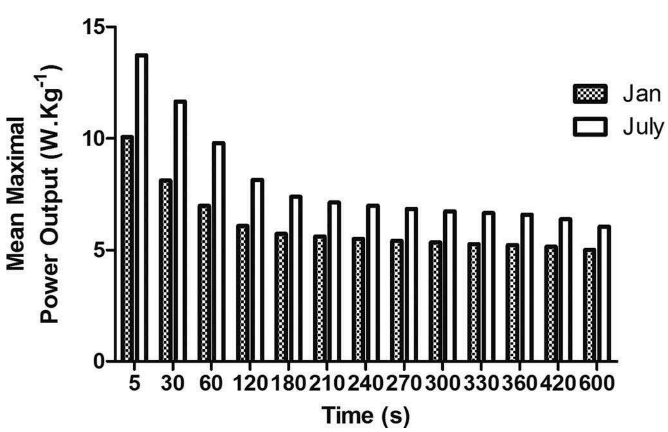

Figure 2. Mean maximal power output for two training sessions from a professional Grand Tour cyclist. Data are the same as used in Figure 1.

Although simple and clear in use, the histograms depicting training zones or maximum mean power output have some limitations. The values used to define each bin largely remain arbitrary and as such may not capture an important aspect of the data. However, some research has attempted to address this limitation by defining the data bin according to certain physiological landmarks such as the ventilatory or anaerobic thresholds (Munoz, Cejuela, Seiler, Larumbe, & Esteve-Lanao, 2014). However, the use of these physiological landmarks as a method to stratify training stress has yet to be fully validated. As training changes fitness, bin values may also need altering, but comparison between differently binned data becomes problematic. Furthermore, the number or duration of efforts within a given data bin is not apparent. For example, a session that requires a single 4-min effort at 400 W cannot be differentiated from one with four 1-min efforts at 400 W. The subsequent training effects of these two sessions may be very different (Theurel & Lepers, 2008). In this regard, Figure 3 illustrates data from two different races for the Grand Tour cyclist. Both races in Figure 3 have exactly the same mean (236 W), but the variability in power output differs notably (SD 138 W vs. 205 W). Consequently, it would be anticipated that the resultant physiological stress from these two races would be very different. Using a binning method to analyse the power data would not necessarily be capable of identifying the difference in the variability of the two races.

Mathiassen and Winkel (1991) proposed exposure variation analysis as a method to examine activity that is stochastic in nature. Exposure variation analysis is a versatile data reduction method that can be used to analyse numerical data which is recorded continuously over time. Subsequently, exposure variation analysis method has been used to examine not only how power meter data is distributed between training zones, but also the duration of sustained bouts too (Abbiss et al., 2010; Passfield, Dietz, Hopker, & Jobson, 2013). Thus exposure variation analysis is performed by defining a fixed number of power bins which represent specific, non-overlapping power output intervals (in Watts), and a fixed number of acute time bins that represent specific, non-overlapping intervals of the time spent (in seconds) in a given power bin. Abbiss et al. (2010) used exposure variation analysis to compare variations in the amplitude and time distribution of power meter data for different cycling events. They found that exposure variation analysis was able to detect differences in the distribution of

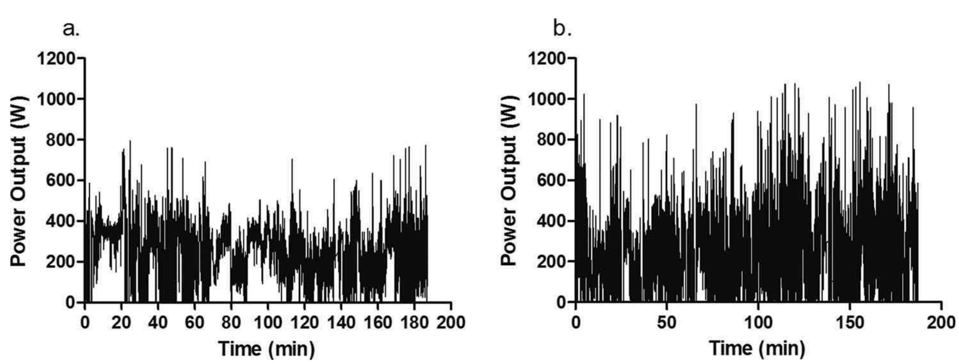

Figure 3. Power output for two races from a professional Grand Tour cyclist. Mean power output in both races is identical but SD varies notably (138 W vs. 205 W).

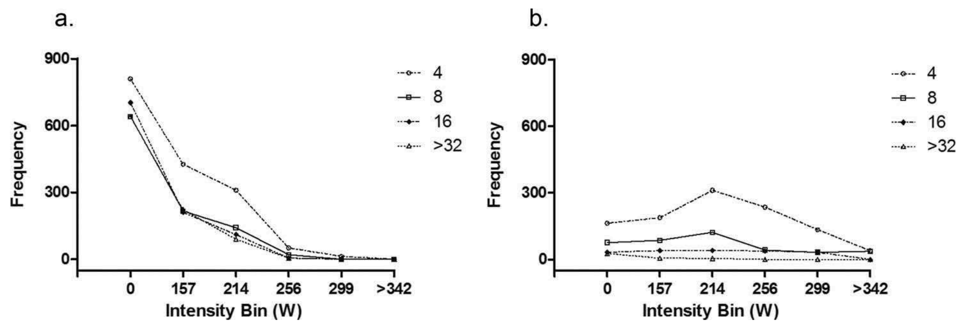

Figure 4. Exposure variation analysis for two races from a professional Grand Tour cyclist. The frequency of data observed between the different intensities (W) is shown. Different symbols are used to show the effort duration (seconds). Data are the same as used in Figure 3.

power output for different race formats. Moreover, exposure variation analysis has previously been used to examine the influence of fatigue and pacing on cycling performance (Peiffer & Abbiss, 2011). In Figure 4 we use exposure variation analysis to further examine the two races with similar means but differing variation in power output from Figure 3. After exposure variation analysis, Figure 4 shows the distribution of power output measures across training zones, but also classified according to the duration of each effort. The effect of the greater variation in Race B can be seen as longer efforts are sustained at the higher exercise intensities. However, whilst this method can differentiate between different race characteristics, it is has yet to be established whether it is sensitive to training-induced changes (Passfield et al., 2013).

Critical power

An alternative approach to assigning power meter data to bins or training zones is to model it instead. In recent years probably the most popular method for modelling endurance performance has been the critical power model. The critical power model is based upon the hyperbolic relation between power output (P) and time to exhaustion (t) originally described by Monod and Scherrer (1965) for bouts of repetitive lifting exercises performed using isolated muscle groups. A simple two-parameter model provides the mathematical representation of this relation:

$$(P - CP)t = W' \quad (1)$$

where P is sustainable power output, CP is critical power, t is time and W' is anaerobic capacity.

To determine critical power a cyclist must typically complete 3–5 bouts of exhaustive exercise lasting between 3 and 20 min (Vandewalle, Vautier, Kachouri, LeChevalier, & Monod, 1997). Mean power output from each bout is then modelled using Equation 1 to construct a power output–duration curve. Thus the critical power is a relevant parameter for cyclists to consider as a significant period of time during both road race and time trial competitions is spent within the severe-intensity exercise domain (Vogt et al., 2006). Consequently, a significant proportion of the total energetic contribution must be derived from the predominantly "anaerobic" parameter of W'. The critical power model can also be used to inform training and predict performance, such as monitoring changes in endurance fitness; assessing the effectiveness of training on specific points on the curve and determining a cyclist's relative strengths and weaknesses.

The traditional method of critical power determination required cyclists to complete exhaustive exercise bouts on separate days in a laboratory (Hill, 1993). Recent studies have proposed two alternative methods for estimating critical power output from a single testing session; a 3 min test (Vanhatalo, Doust, & Burnley, 2007) and a field test (Karsten, Jobson, Hopker, Jimenez, & Beedie, 2014a). Vanhatalo et al. (2007) proposed that the power output sustained during the final 30 s of a 3 min all-out test corresponds to critical power. In a follow up study (Vanhatalo, Doust, & Burnley, 2008), these researchers also found the 3 min test to track training-induced changes in critical power. However, recent studies indicate that the interpretation of the 3 min test is controversial. Dekerle, Barstow, Regan, and Carter (2014) found high intra-subject variability in the agreement between 3 min test and critical power, whilst Karsten, Jobson, Hopker, Passfield, and Beedie (2014b) suggest that the ergometer used may also affect agreement. As an alternative single visit protocol, Karsten, Jobson, Hopker, Stevens, and Beedie (2015) found a field test comprising of three all-out trials of 3, 7 and 12 min, with 30-min recovery, provides a measure of critical power (Karsten et al., 2014a, 2015). Indeed, Karsten (2014) has shown that critical power can be estimated reasonably from the peak 3-, 7- and 12-min power output values observed during training (i.e., without employing a specific test protocol). Figure 5 illustrates critical power calculated in this manner from the combined training and racing data obtained from the Grand Tour cyclist over the course of a season. Both training and race data are used to construct the critical power profile so as to capture the absolute peak 3-, 7- and 12-min efforts that the cyclist was capable of during the period of observation. It can be seen that the Grand Tour cyclist's critical power and W' were highest during his main competitive phase of the season (Dauphine, National Championships, Tour de France, Eneco Tour). The obvious double peak in critical power suggests this method of analysis may reflect changes in fitness. Interestingly, the second peak in the cyclist's critical power, and his highest W', is seen in October which was associated with his preparation for and competition in Paris-Bourges and Paris-Tours races. There are, however, obvious limitations with

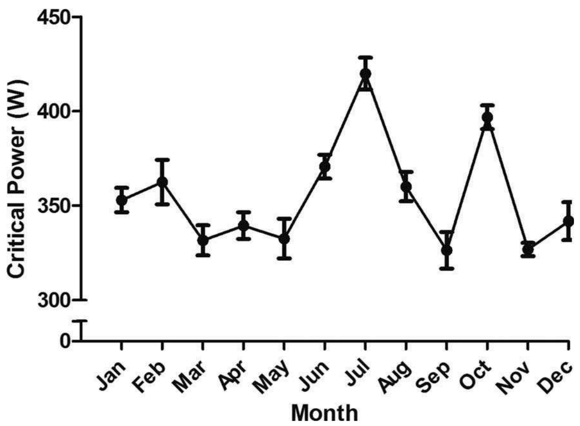

Figure 5. Critical power modelled from power meter data of a professional Grand Tour cyclist. Critical power is calculated from all training and racing data each month. Error bars show SD.

the critical power model in that it is asymptotic in nature, and typically restricted to efforts between 3 and 20 min (Vandewalle et al., 1997).

Record power profile

It has long been recognised that human performances are not asymptotic but tend follow an exponential curve (Kennelly, 1906). The record power profile (Pinot & Grappe, 2011) acknowledges this by using maximum power output for different durations to generate a power output–duration curve that is much more extensive than the 3–20 min used to calculate critical power (Vandewalle et al., 1997; Vanhatalo et al., 2007). Thus, the record power profile extends the previously mentioned MMP and CP methods of analysis by establishing the relationship between different sequential records of power output and the corresponding time training/race durations during a whole race season.

Figure 6 shows the record power profile for the Grand Tour cyclist over different phases of the cycling season. The record power profile is constructed from time intervals of 5 s to 5 min, and then over 5 to 240 min. The record power profile presents the exponential curve that reflects mean record power output of 12 W·kg^{-1} (5 s) and 3 W·kg^{-1} (4 h). In Figure 6 the average of all training and racing data for the specified time period are presented. Figure 6 shows power output for the May–August period is higher than for any other time point of the season. It is also apparent that 5 s to 5 min power output is higher in September–December than January–April. In contrast, 5 to 240 min power output is lower in September–December than January–April. The record power profile can be divided into sections; from 5 s to 5 min the profile decreases by ~50% regardless of time of the season. From 5 min to 60 min the profile decreases by 30% in January–April and October–December respectively, but by less (27%) in May–August. From 60 min to 240 min a decline of 20% in January–April and October–December, is slightly less (19%) than in May–August.

Variability in power output

As with many other behavioural and physiological processes, cycling power output is highly irregular or stochastic, even during apparently steady-state exercise. The variance or standard deviation of the data set provides an indication of the extent to which power output varies during training and racing. In Figure 3, we presented data from two races for the Grand Tour cyclist with exactly the same mean power output of 236 W, but where the standard deviation was quite different (Figure 3(a) = 138 W vs. Figure 3(b) = 205 W). Despite the identical mean power output, the higher variation in power output is likely to be indicative of a more stressful race and therefore could be useful to monitor and evaluate. Tucker et al. (2006) noted that during time-trial type efforts, the large variability in power output between and within a group of 11 cyclists, also exhibited a high degree of self-similarity. This observation suggests that the standard deviation is not the best index for monitoring power output variability during training and racing. Instead, methods that provide a calculation of long-range correlations in time series data such as detrended fluctuation analysis (DFA) may be more appropriate. Within DFA analysis stronger correlations suggest a more predictable, regular time series, whereas weaker correlations indicate a less predictable time series (Peng, Havlin, Stanley, & Goldberger, 1995). The main advantage of using DFA as opposed to other analytical methods (such as spectral analysis) is that it is robust in regard to non-stationary, or unpredictable, data in the time series (Chen, Ivavnov, Hu, & Stanley, 2002). A DFA was performed on the race data presented in Figure 3 (Figure 3(a) DFA = 1.07 and

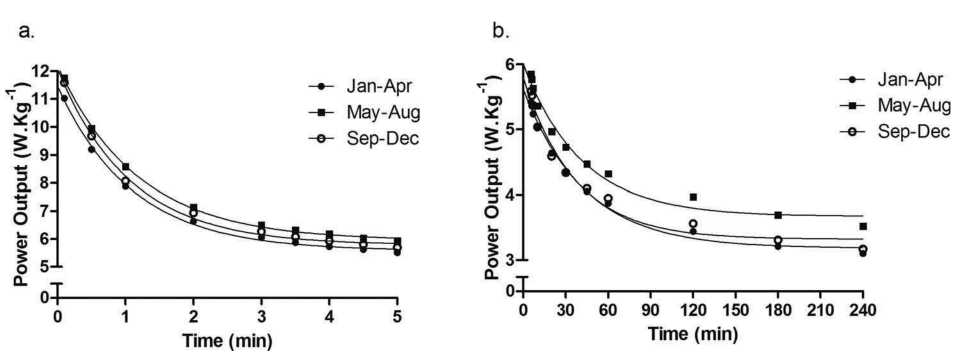

Figure 6. Record power profile for a professional Grand Tour cyclist over three different phases of the cycling season (January–April, May–August and September–December). Figure 6(a) shows the record power profile for efforts of 5 s to 5 min. Figure 6(b) shows the record power profile for efforts more than 5 to 240 min.

Figure 3(b) DFA = 0.87, respectively). These results are consistent with the anticipated physiological stress of the different races (Theurel & Lepers, 2008). Further research is required to establish whether this method reflects real physiological phenomena, or the wider applicability of fractals.

Modelling training and performance

Monitoring training sessions and race performances with a power meter provides an opportunity for the relation between them to be modelled. Power meter data could be used to form the input for a model used to predict future performance and to prescribe and optimise training. Banister, Calvert, Savage, and Bach (1975) proposed a systems theory approach to modelling the responses to endurance training. Subsequently developed by others (Busso, 2003; Morton, 1997) their approach attempted to abstract the training process into an impulse-response-based mathematical model. The model was characterised by a training impulse and a performance response linked by a mathematical "transfer function" (Busso & Thomas, 2006). This modelled function follows the general form:

$$\text{Performance} = (\text{fitness from training}) - (\text{fatigue from training})$$

Calvert, Banister, and Savage (1976) suggested that training data could be used to calculate an elicited fatigue response (that decreases performance), and two fitness responses (that increase performance). Hellard et al. (2006) suggested that modelling-based research could provide information about inter-individual differences and inform the construction of individualised training programmes. However, Taha and Thomas (2003) observe that current models (e.g., Busso, 2003; Jobson, Passfield, Atkinson, Barton, & Scarf, 2009; Morton, 1997) do not correspond with contemporary understanding of physiological mechanisms and are unable to distinguish the specific effects of different training impulses. Furthermore, inter-study and inter-subject variability in model parameter estimates limit the ability to develop and apply a generalisable model. Addressing the latter issue, some of the present authors examined whether individualised parameter values can be determined from the relation between power output and heart rate data (unpublished study). However, this method was successful, the resulting model cannot determine an individual's capacity for fatigue. Consequently, impulse-response models might inform training planning theory, but alternative models are required to produce acceptable accuracy (Busso & Thomas, 2006).

Training adaptation is a complex non-linear problem because the biological system changes itself (Pfeiffer & Hohmann, 2012). Recognising this, Edelmann-Nusser, Hohmann, and Henneberg (2002) and Pfeiffer and Hohmann (2012) used a non-linear multi-layer perceptron neural network to model the performance of an Olympic-level swimmer. In both cases, the model produced a "prediction error" of less than 1%. But whilst the predictive power of neural networks is impressive, they function as a "black box" and cannot explicitly identify causal relationships (Hellard et al., 2006). A further problem is that "training" neural network models requires a large amount of training data to be collected from athletes over a prolonged period of time. In predicting the performance of a single swimmer, Edelmann-Nusser et al. (2002) and Pfeiffer and Hohmann (2012) overcame this problem by training the model with data from a second swimmer. This method proved to be successful but, as noted by the authors, it may have been fortuitous that the adaptive response of both athletes was similar.

Future directions and considerations

Since the introduction of the first commercially available power some 30 years ago, the availability and use of power meters has changed considerably. From current trends, it seems likely that the cost and specification of commercially available power meters will continue to improve. These developments will facilitate our ability to monitor cyclists' training and racing with the accuracy necessary to detect meaningful changes in performance. This in turn will require an improvement in our current methods for visualising and analysing large volumes of training data such as that proposed by Kosmidis and Passfield (2015). Particularly challenging is the development of novel methods and metrics for quantifying the training load given the stochastic nature of cyclists' training and racing. A further challenge is to develop useful and valid models linking training and performance. An exciting prospect for the future is to be able to model the effects of individual cyclist's training on performance. This would mean that cyclists' training and consequent performance could be optimised with the appropriate analysis of their power meter data. Perhaps the most significant issue of all, however, is that despite so many different ways to analyse power output, there is not a single reference measurement of performance. It is difficult to evaluate the implications of different methods of analysis of power meter data without being able to benchmark against corresponding changes in performance. Consequently, the biggest issue with many of the methods of analysis discussed is that they have not been able to use a model that has clear input and output variables. In this regard a promising approach may be to develop new ways of analysing large amounts of training and race data that links time spent in training to a flexible model of performance (Kosmidis & Passfield, 2015).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

No funding was received for this work.

ORCID

S. Jobson http://orcid.org/0000-0002-1377-2128

References

Abbiss, C. R., Quod, M. J., Levin, G., Martin, D. T., & Laursen, P. B. (2009). Accuracy of the velotron cycle ergometer and SRM power meter. *International Journal of Sports Medicine*, *30*(2), 107–112. doi:10.1055/s-0028-1103285

Abbiss, C. R., Straker, L., Quod, M. J., Martin, D. T., & Laursen, P. B. (2010). Examining pacing profiles in elite female road cyclists using exposure variation analysis. *British Journal of Sports Medicine*, *44*(6), 437–442. doi:10.1136/bjsm.2008.047787

Banister, E. W., Calvert, T. W., Savage, M. V., & Bach, T. M. (1975). A system model of training for athletic performance. *Australian Journal of Sports Medicine*, *7*, 57–61.

Bertucci, W., Crequy, S., & Chiementin, X. (2013). Validity and reliability of the G-Cog BMX Powermeter. *International Journal of Sports Medicine*, *34* (6), 538–543.

Bertucci, W., Duc, S., Villerius, V., Pernin, J. N., & Grappe, F. (2005). Validity and reliability of the powertap mobile cycling powermeter when compared with the SRM device. *International Journal of Sports Medicine*, *26* (10), 868–873. doi:10.1055/s-2005-837463

Busso, T. (2003). Variable dose-response relationship between exercise training and performance. *Medicine & Science in Sports & Exercise*, *35*, 1188–1195. doi:10.1249/01.MSS.0000074465.13621.37

Busso, T., & Thomas, L. (2006). Using mathematical modelling in training planning. *International Journal of Sports Physiology and Performance*, *1*, 400–405.

Calvert, T. W., Banister, E. W., & Savage, M. V. (1976). A systems model of the effects of training on physical performance. *IEEE Transactions on Systems, Man and Cybernetics*, *6*(2), 94–102. doi:10.1109/TSMC.1976.5409179

Carpes, F. P., Mota, C. B., & Faria, I. E. (2010). On the bilateral asymmetry during running and cycling - a review considering leg preference. *Physical Therapy in Sport*, *11*(4), 136–142. doi:10.1016/j.ptsp.2010.06.005

Chen, Z., Ivavnov, P. C., Hu, K., & Stanley, H. E. (2002). Effect of nonstationarities on detrended fluctuation analysis. *Physical Reviews E, Statistical, Nonlinear Soft Matter Physics*, *65*, 041107. doi:10.1103/PhysRevE.65.041107

Coggan, A. R. (2003). Training and racing using a power meter: an introduction. Retrieved December 10, 2014, from *http://www.ipmultisports.com/ref_lib/Coggan_Power_Meter.pdf*

Dekerle, J., Barstow, T. J., Regan, L., & Carter, H. (2014). The critical power concept in all-out isokinetic exercise. *Journal of Science and Medicine in Sport*, *17*(6), 640–644. doi:10.1016/j.jsams.2013.09.003

Duc, S., Villerius, V., Bertucci, W., & Grappe, F. (2007). Validity and reproducibility of the Ergomopro power meter compared with the SRM and powertap power meters. *International Journal of Sports Physiology and Performance*, *2*(3), 270–281.

Ebert, T. R., Martin, D. T., McDonald, W., Victor, J., Plummer, J., & Withers, R. T. (2005). Power output during women's World Cup road cycle racing. *European Journal of Applied Physiology*, *95*(5–6), 529–536.

Edelmann-Nusser, J., Hohmann, A., & Henneberg, B. (2002). Modeling and prediction of competitive performance in swimming upon neural networks. *European Journal of Sport Sciences*, *2*, 1–10. doi:10.1080/17461390200072201

Franklin, K. L., Gordon, R. S., Baker, J. S., & Davies, B. (2006). Comparison of methods for determining power generated on a rope-braked cycle ergometer during low-intensity exercise. *Sports Engineering*, *9*, 29–38. doi:10.1007/BF02844260

Gardner, A. S., Stephens, S., Martin, D. T., Lawton, E., Lee, H., & Jenkins, D. (2004). Accuracy of SRM and power tap power monitoring systems for bicycling. *Medicine & Science in Sports & Exercise*, *36*(7), 1252–1258. doi:10.1249/01.MSS.0000132380.21785.03

Hellard, P., Avalos, M., Lacoste, L., Barale, F., Chatard, J.-C., & Millet, G. P. (2006). Assessing the limitations of the Banister model in monitoring training. *Journal of Sports Sciences*, *24*(5), 509–520. doi:10.1080/02640410500244697

Hill, D. W. (1993). The critical power concept. A review. *Sports Medicine*, *16*, 237–254. doi:10.2165/00007256-199316040-00003

Hurst, H. T., & Atkins, S. (2006). Agreement between polar and SRM mobile ergometer systems during laboratory-based high-intensity, intermittent cycling activity. *Journal of Sports Sciences*, *24*(8), 863–868. doi:10.1080/02640410500245678

Jobson, S. A., Nevill, A. M., & Jeukendrup, A. (2005). The efficacy of power output measurement during a professional cycle stage race: A case study. *Journal of Sports Sciences*, *23*(11–12), S1292.

Jobson, S. A., Passfield, L., Atkinson, G., Barton, G., & Scarf, P. (2009). The analysis and utilization of cycling training data. *Sports Medicine*, *39*, 833–844. doi:10.2165/11317840-000000000-00000

Jones, S. M., & Passfield, L. (1998). The dynamic calibration of bicycle power measuring cranks. In S. J. Haake (Ed.), *The engineering of sport* (pp. 265–275). Oxford: Blackwell Science.

Karsten, B. (2014). Analysis of reliability and validity of critical power testing in the field (PhD thesis). University of Greenwich, Chatham Maritime.

Karsten, B., Jobson, S. A., Hopker, J. G., Jimenez, A., & Beedie, C. (2014). High agreement between laboratory and field estimates of critical power in cycling. *International Journal of Sports Medicine*, *35*(04), 298–303.

Karsten, B., Jobson, S. A., Hopker, J., Passfield, L., & Beedie, C. (2014b). The 3-min test does not provide a valid measure of critical power using the SRM isokinetic mode. *International Journal of Sports Medicine*, *35*(04), 304–309.

Karsten, B., Jobson, S. A., Hopker, J., Stevens, L., & Beedie, C. (2015). Validity and reliability of critical power field testing. *European Journal of Applied Physiology*, *115*(1), 197–204.

Kennelly, A. E. (1906). An approximate law of fatigue in the speeds of racing animals. *Proceedings of the American Academy of Arts and Sciences*, *12*(15), 275–331. doi:10.2307/20022230

Kirkland, A., Coleman, D. A., Wiles, J. D., & Hopker, J. G. (2008). Validity and reliability of the Ergomopro powermeter. *International Journal of Sports Medicine*, *29*(11), 913–916.

Kosmidis, I., & Passfield, L. (2015). Linking the performance of endurance runners to training and physiological effects via multi-resolution elastic net. Retrieved from http://arxiv.org/abs/1506.01388

Kyle, C. R. (1988). *The mechanics and aerodynamics of cycling*. Champaign, IL: Human Kinetics Press.

Lucía, A., Hoyos, J., Carvajal, A., & Chicharro, J. L. (1999). Heart rate response to professional road cycling: The tour de France. *International Journal of Sports Medicine*, *20*, 167–172.

Martin, J. C., Gardner, A. S., Barras, M., & Martin, D. T. (2006). Modeling sprint cycling using field-derived parameters and forward integration. *Medicine and Science in Sports and Exercise*, *38*(3), 592–597.

Martin, J. C., Milliken, D. L., Cobb, J. E., McFadden, K. L., & Coggan, A. R. (1988). Validation of a mathematical model for road cycling power. *Journal of Applied Biomechanics*, *14*, 276–291.

Mathiassen, S. E., & Winkel, J. (1991). Quantifying variation in physical load using exposure-vs-time data. *Ergonomics*, *34*(12), 1455–1468.

Millet, G. P., Tronche, C., Fuster, N., Bentley, D. J., & Candau, R. (2003). Validity and reliability of the Polar S710 mobile cycling powermeter. *International Journal of Sports Medicine*, *24*(3), 156–161.

Monod, H., & Scherrer, J. (1965). The work capacity of a synergic muscular group. *Ergonomics*, *8*, 329–338.

Morton, R. H. (1997). Modelling training and overtraining. *Journal of Sports Sciences*, *15*, 335–340.

Munoz, I., Cejuela, R., Seiler, S., Larumbe, E., & Esteve-Lanao, J. (2014). Training-intensity distribution during an Ironman season: Relationship with competition performance. *International Journal of Sports Physiology and Performance*, *9*, 332–339.

Palmer, G. S., Hawley, J. A., Dennis, S. C., & Noakes, T. D. (1994). Heart rate responses during a 4-d cycle stage race. *Medicine and Science in Sports Exercise*, *26*, 1278–1283.

Passfield, L., Dietz, K. C., Hopker, J. G., & Jobson, S. A. (2013). Objective time-binning in exposure variation analysis. *IMA Journal of Management Mathematics*, *24*(3), 269–282.

Peiffer, J. J., & Abbiss, C. R. (2011). Influence of environmental temperature on 40km cycling time-trial performance. *International Journal of Sports Physiology and Performance*, *6*, 208–220.

Peng, C. K., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in non-stationary heartbeat time series. *Chaos*, *5*(1), 82–87.

Pfeiffer, M., & Hohmann, A. (2012). Applications of neural networks in training science. *Human Movement Science*, *31*, 344–359.

Pinot, J., & Grappe, F. (2011). The record power profile to assess performance in elite cyclists. *International Journal of Sports Medicine*, *32*, 839–844.

Quod, M. J., Martin, D. T., Martin, J. C., & Laursen, P. B. (2010). The power profiles predicts road cycling MMP. *International Journal of Sports Medicine*, *31*, 397–401.

Skiba, P. (2007). Evaluation of a novel training metric in trained cyclists. *Medicine and Science in Sports and Exercise*, *39*(5), S448.

Smak, W., Neptune, R. R., & Hull, M. L. (1999). The influence of pedaling rate on bilateral asymmetry in cycling. *Journal of Biomechanics*, *32*(9), 899–906.

Taha, T., & Thomas, S. G. (2003). Systems modelling of the relationship between training and performance. *Sports Medicine*, *33*(14), 1061–1073.

Theurel, J., & Lepers, R. (2008). Neuromuscular fatigue is greater following highly variable versus constant intensity endurance cycling. *European Journal of Applied Physiology*, *103*(4), 461–468.

Tucker, R., Bester, A., Lambert, E. V., Noakes, T. D., Vaughan, C. L., & St Clair, G. A. (2006). Non-random fluctuations in power output during self-paced exercise. *British Journal of Sports Medicine*, *40*(11), 912–917.

Vandewalle, H., Vautier, J. F., Kachouri, M., LeChevalier, J.-M., & Monod, H. (1997). Work-exhaustion time relationships and the critical power concept. A critical review. *Journal of Sports Medicine and Physical Fitness*, *37*(2), 89–102.

Vanhatalo, A., Doust, J. H., & Burnley, M. (2007). Determination of critical power using a 3-min all-out cycling test. *Medicine and Science in Sports and Exercise*, *39*(3), 548–555.

Vanhatalo, A., Doust, J. H., & Burnley, M. (2008). A 3-min all-out cycling test is sensitive to a change in critical power. *Medicine and Science in Sports and Exercise*, *40*(9), 1693–1699.

Vogt, S., Heinrich, L., Schumacher, Y. O., Blum, A., Roecker, K., Dickhuth, -H.-H., & Schmid, A. (2006). Power output during stage racing in professional road cycling. *Medicine and Science in Sports and Exercise*, *38*(1), 147–151.

Vogt, S., Schumacher, Y. O., Roecker, K., Dickhuth, H. H., Schoberer, U., Schmid, A., & Heinrich, L. (2007). Power output during the tour de France. *International Journal of Sports Medicine*, *28*(9), 756–761.

Wooles, A. L., Robinson, A. J., & Keen, P. S. (2005). A static method for obtaining a calibration factor for SRM bicycle power cranks. *Sports Engineering*, *8*, 137–144.