Age of peak performance in professional road cycling

Leonid Kholkine[1], Steven Latré[1], Tim Verdonck[2] and Arie-Willem de Leeuw[1]

[1] Department of Computer Science, University of Antwerp – imec, Sint-Pietersvliet 7, Antwerp, 2000, Belgium;

[2] Department of Mathematics, University of Antwerp, Middelheimlaan 1, Antwerp, 2020, Belgium;

Correspondence: Leonid Kholkine, MSc
University of Antwerp
Department of Computer Science
Sint-Pietersvliet 7, 2000 Antwerp, Belgium
Email: leonid.kholkine@uantwerpen.be

Submission type: Original research
Preferred running title: Age of peak performance in professional road cycling
Disclosure of funding: No potential financial conflict of interest was reported by the authors
Word count abstract: 200
Word count manuscript: 3495
Figures and tables: 5 figures, 3 tables, 2 digital supplemental materials

Age of peak performance in professional road cycling

ABSTRACT

In this study, we investigated the relationship between age and performance in professional road cycling. We considered 1864 male riders present in the yearly top 500 ranking of ProCyclingStats (PCS) since 1993 until 2021 with more than 700 PCS Points. We applied a data-driven approach for finding natural clusters of the rider's speciality (General Classification, One Day, Sprinter or All-Rounder). For each cluster, we divided the riders into the top 50% and bottom 50% based on their total number of PCS points. The athlete's yearly performance was defined as the average number of points collected per race. Age-performance models were constructed using polynomial regression and we obtained that the top 50% of the riders in each cluster have a statistically significant ($p < 0.05$) higher peak performance age. Considering the best 50% of the riders, general classification riders peak at an older age than the other rider types ($p < 0.05$). For those top riders, we found ages of peak performance of 26.3, 26.5, 26.2 and 27.5 years for sprinters, all-rounders, one day specialists and general classification riders, respectively. Our findings can be used for scouting purposes, assisting coaches in designing long-term training programs and benchmarking the athletes' performance development.

KEYWORDS

Performance optimization; age performance modelling; machine learning; road cycling

INTRODUCTION

Sports performance is usually studied by focusing on the athlete's current level of performance (Lambert & Borresen, 2010), health status (e.g. previous injuries (Maffulli et al., 2011), fatigue (Jones et al., 2017), diet (Brotherhood, 1984), drug use (Momaya et al., 2015) or genetic endowment (Guth & Roth, 2013). Additionally, there are also possibilities to investigate the athlete's potential based on present values of performance indicators. For example, age can be used to estimate an athlete's future performance by determining the relationship between age and performance for similar athletes (Kholkine et al., 2021; Smith, 2003). Sports performance is typically enhanced as an athlete's age increases until a certain threshold value, the peak performance age, after which it starts to decrease (Baltes & Baltes, 1990). The age of peak performance is important for setting long-term goals for the athletes' performance (Allen & Hopkins, 2015).

Previous works have already established that different sports have very different age of peak performance. A systematic review by Allen et al. (Allen & Hopkins, 2015) showed that the age of peak performance can vary greatly depending on the sport type. The age of peak performance of explosive/sprint sports varies from roughly 20 years for large event durations (order of minutes) to approximately 27 years for short event durations (a few seconds). In endurance sports, there is a similar relationship between the age of peak performance and the duration of the effort. For efforts between 2 and 15 minutes in swimming, the age of peak performance is around 20 years, whereas ultra-distance cycling athletes peak at almost 39 years. This demonstrates that the age of peak performance is different across sports and varies between different disciplines of the same sport.

In cycling, there are only few studies investigating the relationship between age and performance. Ransdell et al. studied the US records in cycling per age group at time-trial races of maximally 40 km from the age of 35 years, and observed a decline in the performance of record-holders in subsequent age groups (Ransdell et al., 2009). Balmer et al. performed a controlled ramped minute power test for cyclists at different ages and observed an age of peak performance of 29 years (Balmer et al., 2005). Finally, some studies looked at ultra-distance cycling (Rüst et al., 2015; Zingg et al., 2013), and found an age of peak performance of 38 to 39 years. Interestingly, to the best of our knowledge, no study has been performed on the peak performance age of road cycling athletes and we aim to fill this gap in the scientific literature. Instead of the controlled settings with limited number of participants that are considered in the aforementioned studies, we aim to understand the relationships between age and performance in road cycling in a non-controlled setting with a large number of participants.

Of all cycling disciplines, road cycling is probably the most challenging for studying the age of peak performance. The main reason is that different types of athletes compete in road cycling races with different objectives and focuses. For example, there are sprinters who focus on races with limited altitude differences or climbers who focus on stages with more mountainous terrain. As a consequence, these riders might have different physiological development and age of peak performance. Furthermore, previous work has already shown that different types of riders already have different anthropometric characteristics (Miller & Susa, 2018; van der Zwaard et al., 2019). Therefore, we will consider different types of road cyclists, separately.

The goal of this study is to find the age of peak performance in road cycling using a data-driven approach. Based on the literature in other sports, we hypothesized that the age of peak performance of athletes will be between 25 and 30 years, with the lowest age of peak performance for the cyclists that specialize in explosive efforts. This study provides practical insights to coaches for recruiting new cyclists and setting development goals for their riders. Moreover, cyclists might have a better understanding of their future prospects.

METHODS

Experimental approach to the problem

We used a data-driven approach for investigating the relationship between the rider's age and performance. We defined the yearly performance of a rider as the total number of points collected from all races in this year, divided by the total number of races entered. In road cycling, each cyclist collects points for winning races or a stage in a multi-day race. The number of points awarded depends on how important a race is.

We collected publicly available data from the ProCyclingStats (PCS) website[1], which gathers historical race results in professional road cycling, and we considered the male riders present in the top 500 PCS yearly ranking as of 1993 until 2021. In the end, only riders with at least 700 total career PCS points are included in our analyses for reasons that will be discussed further on in this section. Note that similar to earlier studies, e.g. see (Van Erp et al., 2021) and (Miller & Susa, 2018), we considered the yearly PCS points collected as performance measurements. Since the PCS point system has been consistent throughout the years, this is preferred over using the official UCI point system that changed in 2016.

As shown in Figure 1, our approach consisted of three steps. First, we classified the race types into different categories, based on the race profiles provided by PCS and solely relying on the race's outcome. By constructing decision trees, we distinguished between sprint, mountain, and other races. Second, we clustered all riders into groups with different specialties based on the percentage of PCS points acquired in different types of races. Third, we determined the relationship between age and performance for the riders in each cluster. The details of these three steps are discussed in the remainder of this section.

(Figure 1 somewhere here)

[1] www.procyclingstats.com

Subjects

We have collected the information of 1864 male riders (mean ± SD: weight: 68.87 ± 7.07 kg and height: 1.80 ± 0.07 m). We only considered the professional careers from 18 years onwards. For those riders, we determined the age at the beginning of each calendar year and the number of PCS points at the end of that year. The career length is 12.8 ± 3.8 years (mean ± SD), with a total number of PCS points during their career ranging from 701 to 39498.

The Ethical Committee for the Social Sciences and Humanities of the University of Antwerp approved the procedures used in this study (SHW_21_118), with a waiver of the requirement for explicit informed consent of the participants.

Classification of Race Profiles

There are two types of races in road cycling: Time trials and mass-start races. In time trials, cyclists start at different times, whereas all riders start together in mass-start races. As the number of time trials and variation in their race profiles is limited, we grouped all individual time trials together. On the other hand, many mass-start races take place on different terrains. Therefore, we distinguished these races into a flat, mountain or other race.

For classifying the race profile of mass-start stages, we trained a decision tree. In essence, a decision tree is trained by recursively splitting the data using the features provided (Song & Lu, 2015). The optimal split is found by evaluating a metric. In our case, we used Entropy, a measure of uncertainty ranging from 0 to 1 in a classifier with two classes. If the Entropy is 0, the group contains only one class, and if the Entropy is 1 it means it contains a 50/50 split. Although there are other options for our task, we preferred decision trees as they provide the ability to interpret the decision-making process.

We considered all 36966 races that the riders in our dataset have participated and contains the corresponding finishing times. Of these races, 6212 contained the description of the race profile on PCS (flat, hills or mountains). The dataset consisted of 1823 flat, 1293 mountain and 3096 hilly races. With this data we trained two models: One for classifying a race as flat or not and another for identifying whether a stage falls into the category of mountain races. For each model, we split the data into a training set (70%) and a test set (30%) used for training our model and validating it's performance, respectively. We have applied stratified sampling to make sure that the training and test sets were similarly proportional in each of the classes and have also added higher weights in the Entropy for the minority classes.

To construct the decision trees, we extracted several features from the race result that might indicate the race type. Specifically, we considered the average speed of the winner, the number of groups finishing the race, the number of riders in the 1^{st}, 2^{nd} and 3^{rd} group, and the standard deviation of the time difference between the first ride and the top 10 and top 15 riders (relative finish time).

Clustering Riders

As mentioned previously, the large variation in physiological characteristics of cyclists makes it worthwhile to investigate different subgroups of cyclists separately. Therefore, we used the K-means algorithm to find natural clusters of road cyclists. Note that each rider will be placed uniquely into one of the clusters.

In this method, for a specific number of clusters K, random starting points are selected as centroids for each cluster. Hereafter, the position of the centroids as well as the size of the cluster is optimized through iterative calculations (Steinley, 2006). One disadvantage of K-means is that the number of clusters must be defined beforehand, which means several K values need to be examined before selecting the best value.

As input for the algorithm, we used the fraction of PCS points acquired through their career in the following categories: one day races, mountain races, flat races, Individual Time Trial (ITT) races and General Classification (GC). To calculate the fraction of the mountainous and flat races, we used the output from our race profile classification algorithm. Using the fractional points from different race categories makes it harder to cluster riders with a low number of points. Therefore, we introduced a cutoff level in the number of points and only considered riders that have a total number of points above this threshold. Note that this also removed riders with short career lengths from our analyses.

First, we created models for every combination of K ranging from 1 to 14 and for cutoff levels ranging from 0 to 2000 with an increment of 100. Typically, the choice of K is made by analysing the change in the SSE (Sum Squared Error), calculated by $\sum_{i=1}^{K}\sum_{p\in C_i}(p-m_i)^2$, where K represents the number of clusters, C is the set of objects in a cluster, and m is the cluster's centroid. To find the cutoff number of PCS points, we found the maximum curvature point in the graph of the SSE (Sum Squared Error) gained by a cutoff against the different cutoff levels. Using the Kneed algorithm (Satopaa et al., 2011), we obtained a cutoff value of 700 PCS points. Afterwards, we found the number of ideal clusters by observing the SSE plot against the number of clusters, using the Kneed algorithm to find the maximal curvature point and perform a silhouette analysis. The silhouette analysis consists of calculating the silhouette coefficient, a value between -1 and 1, which is a measure for the distance of a data point to the centroid of its cluster and its closest neighbouring cluster. A negative value is a sign the cluster might assigned the wrong way, 0 that there is no significance in the distance between clusters and 1 that the clusters are well apart from each other. A positive coefficient indicates that it is closer to its centroid rather than the neighbouring one (Rousseeuw, 1987).

Age-Performance Modelling

We were interested in finding the age of peak performance of road cyclists. In this context, the absolute number of PCS points per race entered that a rider obtained is irrelevant. Instead, it is the average number of points per race relative to the personal maximum obtained in a single year that specifies the development of an athlete. Therefore, we considered *relative performance* as the dependent variable. To obtain this relative performance for each cyclist, we first selected the maximal value for the number of points per races entered that the rider obtained in a single year. Hereafter, the relative performance for each year is obtained by comparing the yearly performance, i.e., the number of points divided by the total number of races entered, with this maximal value. Hence, during the cyclist's career, the relative performance runs between 0 and 1, with the value 1 indicating the age of peak performance.

Since the main aim for modelling the relationship between age and performance is to determine the peak performance age, we must be careful with the riders in our dataset that did not finish their careers yet. Especially for riders that have been professional for a relatively short time, there is a reasonable possibility that they will improve their maximal number of PCS points in future years. Therefore, we removed all riders who participated in at least one race in 2021, i.e., the final year present in our dataset.

To investigate the dependencies between age and relative performance, we applied an approach developed to study the same dependency in marathon running (Leeuw et al., 2018). The core of this method is polynomial regression, where the performance is modelled as an n^{th}-degree polynomial in age. The main advantage of this approach is that we can model the non-trivial details of the relationships between age and performance. In this approach, cross-validation is applied to minimize the risk of overfitting. For every polynomial degree n, a polynomial model is fitted on the training set. Hereafter, the mean squared error on the test set was determined. Finally, the polynomial degree with the smallest value on the test set was selected. The final model was then obtained by fitting a polynomial of this degree on the entire dataset.

Before applying this method to our setting, we had to make a small adjustment to the original method. Namely, in our case, the range of different ages is rather limited. There are approximately twenty distinct values with the precise value depending on which cluster of riders is considered. Moreover, the number of cyclists is rather small to perform too detailed analyses. More specifically, there are situations in which selecting the polynomial degree *n* that minimizes the error leads to overfitting. Alternatively, we select the most accurate polynomial model of degree 3 or lower. For all analyses performed in this study, these models appropriately describe the relationships between age and performance as the errors are within 5% of the minimal error value obtained for more complex models.

For each cluster, we modelled the relationship between age and performance for the best and bottom 50% per cent of the riders. Here, these subsets are determined by ranking the riders according to the total number of PCS points during their career. Due to a relatively small number of riders in each cluster (ranging from 159 for sprinters to 501 for all-rounders), we did not consider smaller subsets. For each of the subsets, data of a random selection of 30% of the riders was used as a test set for error estimation and assessing the generalizability. The other 70% was used to optimize the model's polynomial degree by applying 10-fold cross-validation.

Statistical Analysis

The stage profile classification was evaluated by the accuracy, which is defined by the sum of true positive and true negative decisions divided by the total number of cases. Moreover, we considered the confusion matrix with the true positive, true negative, false positive and false negative rates (Fawcett, 2006).

As our age-performance models were based on data of multiple riders and therefore included multiple values for the performance at the same age, we performed a Lack-of-fit F test to assess the statistical significance of our age-performance models (Brook & Arnold, 1985). Hereby, we tested the null hypothesis that there is no lack of fit, i.e., the error of a model is a consequence of the data variance at every age. Therefore, only if the p-value is smaller than a significance threshold we have statistical evidence that there is a lack of fit. Here, our models were considered statistically significant, i.e., there was no sufficient evidence there is lack of fit, if $p > 0.05$.

The confidence intervals of our age-performance models were determined by using the confidence intervals around the predicted values of the relative performance for each age (Dalpiaz, 2016). Moreover, the confidence intervals of the peak performance ages are the 95% confidence intervals of the age-performance models of the maxima age. By dividing the difference between the upper and lower limit of the 95% confidence intervals by 3.92, we determined the standard deviations of the peak performance age. These values were used for performing an independent t-test to compare the ages of peak performance. We considered the differences to be significant if $p < 0.05$.

Finally, effect sizes were determined via Cohen's d, including the 95% confidence intervals. Effect sizes were considered to be negligible ($|d| < 0.20$), small ($0.20 \leq |d| < 0.50$), medium ($0.50 \leq |d| < 0.80$) or large ($|d| \geq 0.80$) (Cohen, 1992).

RESULTS

In this section, we will present the results for all the steps described in the methodology section: Classification of Race Profiles, Rider Clustering and Age-Performance Modelling.

(Figure 2 somewhere here)

Classification of Race Profiles

We constructed separate models for classifying the sprint and mountain races. The decision trees are shown in Figure 2. The accuracy of the models was 0.74 for the sprint races and 0.86 for mountain races. We have also created a confusion matrix for each model (see figure, Supplemental Digital Content 1, which illustrates the confusion matrix for both models).

(Figure 3 somewhere around here)

Clustering Riders

The SSE for the different number of clusters is displayed in Figure 3. By using the Kneed algorithm, we found that the maximal curvature point of the graph was at K=4. Based on a detailed silhouette analysis (see figure, Supplemental Digital Content 2, which illustrates the full silhouette analysis), there could be either 2 or 4 clusters. In both cases, the average values of the silhouette coefficients are high, and the coefficient value for all the clusters is above average. However, there are fewer points with a negative coefficient for K=4. Therefore, we have decided to select K=4. Note that the final SSE of the clustering model is therefore 56.71.

(Figure 4 somewhere here)

In Figure 4, we plotted a heatmap with the centroid value for each feature used. Additionally, we show the average silhouette coefficient in Table 1. Based on this information, we can define different types of riders. The first cluster has a large fraction of points in one day (OD) races and corresponds to riders that focus on OD races. The riders in the second cluster do not collect the most nor the least number of points for the different race types, and have the most evenly distribution for the number of points won in the different categories. Therefore, we named this cluster *All Rounders*. The third cluster has a large fraction of GC points and can be considered riders who focus on winning multi-stage races. Finally, the fourth cluster is the sprinter cluster, as it has the highest sprinter ratio.

(Table 1 somewhere around here)

Age-Performance Modelling

We constructed age-performance models for all clusters. The results of the Lack-of-fit F tests are shown in Table 2. In all cases, the p-values are larger than 0.05, and therefore there is insufficient statistical evidence that there is lack of fit in the models. Hence, the models properly describe the relationships between age and performance.

(Table 2 somewhere around here)

The age-performance models are displayed in Figure 5. Apart from some small differences, the models are similar for all clusters. We observed a similar development for the bottom and best 50 per cent of riders in each cluster until roughly 23 years. For higher ages, we found a larger increase in the performance of the best 50 per cent of the riders compared to the bottom 50 per cent until the age of peak performance is reached. The degree of decrease in performance after reaching the peak performance age is again similar for both the best and bottom 50 per cent of the riders.

(Figure 5 somewhere around here)

In Table 3, we show the ages of peak performance. The age of peak performance of the best 50 per cent of the riders is higher compared to the bottom 50 per cent. If we considered the best 50 per cent of riders in each cluster, we found that the GC riders peak at the highest age. The sprinters, all-rounders and one day specialists peak roughly a year before GC riders ($p < 0.05$).

(Table 3 somewhere around here)

DISCUSSION

Classification of Race Profiles

To the best of our knowledge, this is the first study that uses a data-driven approach to classify road cycling races automatically into different categories. Therefore, our approach gives new and relevant insights into the important characteristics of different race types.

The most important features for the sprint race classifier were the number of separate groups finishing the race, the winner's average speed, and the number of riders finishing in the first group. These features corresponded to how sprint races usually pan out: As the cumulative elevation gain is relatively low, it is usually a faster race, but also the peloton tends to break apart less and typically, a bunch arrival is seen (Mignot, 2015). The most important features for the mountain race classifier were the standard deviation of the relative time from the top 15 riders, the number of separate groups finishing the race, and the winner's average speed. This agrees with the typical outcome of a race on mountainous terrain: It is a slower race where the peloton breaks apart more frequently and has a bigger time gap between the riders (Mignot, 2015).

Although both classifiers do not classify all races well, we believe this is a consequence of the subjectiveness of how a particular race is classified. A recent example is Tour of Flanders 2020; it is classified as flat on PCS, while it is easy to argue it is not a flat/sprint race. After manually reviewing the misclassified cases, we observed that most of these cases are related to such situations.

Clustering Riders

The clustering algorithm provides 4 different clusters related to the different types of riders. These clusters are also what is often observed in the cycling sports – Certain riders focus on flat and sprint races, others have a big focus on winning GC races and typically are good at mountain races, some riders have a bigger focus on one day, and finally, *all-rounders* have multiple skills. The main advantage of our clustering method is that it allows for a data-based categorization of riders as opposed to dividing cyclists into different groups based on subjective criteria.

Age-Performance Modelling

We have found that the age of peak performance in professional road cycling is around 27 years, with riders that focus on explosive efforts having a slightly lower peak-performance age. Note that this is close to the age of peaks in important physiological characteristics, such as VO2 max (Kim et al., 2016) and pulmonary functions (Schoenberg et al., 1978). Therefore, the results of our study suggest that the physiological measures are dominant performance indicators in professional road cycling and are more relevant than other factors, such as a rider's experience.

This age of peak performance for the best 50% of riders of roughly 27 years was higher than the results in most explosive/sprint sports (e.g. swimming, athletic events such as sprints, hurdles) and middle-distance running events (Allen & Hopkins, 2015). Compared to other cycle sports, we have found a similar age of peak performance of 27.6 in Triathlon (Malcata et al., 2014). However, in cyclocross and ultra-distance cycling events, the age of peak performance is higher with 30.2 (Anderson, 2014) and 38 (Abou Shoak et al., 2013), respectively. Moreover, other endurance events have a higher age of peak performance. For example, the peak-performance age in marathon running is around 29-30 (Hunter et al., 2011).

A difference in the distribution of the age of peak of the top and bottom 50% riders was observed. The bottom 50% tend to have a flatter curve and peak at an earlier stage. It is possible to argue that other factors next to physiological aspects are of more importance for the bottom 50% riders compared to top 50% riders. For example, athletes with lower performance might abandon the sport earlier in their career, less mental capacity and motivation to repeat their best performance year to year, or be relegated to a supporting role.

Considering the top 50% of the riders in each cluster, we have found that the performance increase before reaching peak-performance age is steeper than the decrease at later ages. This is in agreement with previous studies in 25 other Olympic sports (Berthelot et al., 2012). Moreover, we have noticed a difference amongst different clusters, with the highest peak performance age for GC riders.

Limitations

The strength of our approach is that the results are based on a large number of cyclists. However, there are also some limitations that are inherent to using large datasets for retrieving relevant information for sport and exercise sciences. (Passfield & Hopker, 2016)

In our study, we observed there was a lot of variation among the cyclists, and therefore there was relatively large uncertainty in the ages of peak performance. This indicates that it is worthwhile to investigate the dependency between age and performance on an individual level. For example, in future work, personalized models can be constructed based on the athlete's past and current performance.

Moreover, further improvements can be made in each step of our approach. The inclusion of other possibly relevant features for the model's classification of stage profiles and clustering of riders might improve the accuracies of our models. Adding features from the course profile (e.g. cumulative elevation gain) might improve our stage profile classification. The clustering of riders could be further developed by introducing features related to the relative number of points to other riders or creating separate clusters for top and bottom riders. Finally, more advanced algorithms can be applied for both the clustering algorithm and race profile classifier.

CONCLUSION

In summary, we created an approach for clustering road cycling riders and classifying races. With the aid of these approaches, we were able to model the non-trivial relationships between age and professional road cycling performance for different types of riders. We found ages of peak performance of 26.3, 26.5, 26.2 and 27.5 years for sprinters, all-rounders, one day specialists and general classification riders, respectively.

The age-performance models can have direct application in recruitment and for scouting purposes. Moreover, there are applications in the design of long-term training programs as it establishes until which age a rider is expected to improve. For the athletes, these models can be used as a point of reference to understand where they are located on the development curve compared to their peers. Another application of our study is the use of the race profile classification for optimising the race program of each road cyclist. By improving further the race profile classification (i.e. with more advanced algorithms and richer features), it is possible to label future races. Hereby, teams can improve their chances of being successful by matching the race profiles with the strengths of a cyclist (de Leeuw et al., 2020).

REFERENCES

Abou Shoak, M., Knechtle, B., Knechtle, P., Rüst, C., Rosemann, T., & Lepers, R. (2013). Participation and performance trends in ultra-cycling. *Open access journal of sports medicine*, *4*, 41-51. https://doi.org/10.2147/OAJSM.S40142

Allen, S. V., & Hopkins, W. G. (2015). Age of Peak Competitive Performance of Elite Athletes: A Systematic Review. *Sports Med*, *45*(10), 1431-1441. https://doi.org/10.1007/s40279-015-0354-3

Anderson, A. (2014). Early Identification of Talent in Cyclo-Cross by Estimating Age-Independent Ability via Probit Regression. *International Journal of Performance Analysis in Sport*, *14*, 153-161. https://doi.org/10.1080/24748668.2014.11868711

Balmer, J., Potter, C. R., Bird, S. R., & Davison, R. C. (2005). Age-related changes in maximal power and maximal heart rate recorded during a ramped test in 114 cyclists age 15-73 years. *J Aging Phys Act*, *13*(1), 75-86. https://doi.org/10.1123/japa.13.1.75

Baltes, P. B., & Baltes, M. M. (1990). *Successful aging: Perspectives from the behavioral sciences*. Cambridge University Press. https://doi.org/https://psycnet.apa.org/doi/10.1017/CBO9780511665684

Berthelot, G., Len, S., Hellard, P., Tafflet, M., Guillaume, M., Vollmer, J. C., Gager, B., Quinquis, L., Marc, A., & Toussaint, J. F. (2012). Exponential growth combined with exponential decline explains lifetime performance evolution in individual and human species. *Age*, *34*(4), 1001-1009. https://doi.org/10.1007/s11357-011-9274-9

Brook, R. J., & Arnold, G. C. (1985). *Applied Regression Analysis and Experimental Design* (C. Press, Ed. 1st ed.). https://doi.org/https://doi.org/10.1201/9781315137674

Brotherhood, J. R. (1984). Nutrition and Sports Performance. *Sports Medicine*, *1*, 350-389.

Cohen, J. (1992). Statistical Power Analysis. *Current Directions in Psychological Science*, *1*(3), 98-101. https://doi.org/10.1111/1467-8721.ep10768783

Dalpiaz, D. (2016). *Applied Statistics with R*. https://daviddalpiaz.github.io/appliedstats/

de Leeuw, A.-W., Heijboer, M., Hofmijster, M., van der Zwaard, S., & Knobbe, A. (2020). Time Series Regression in Professional Road Cycling. In A. Appice, G. Tsoumakas, Y. Manolopoulos, & S. Matwin, Discovery Science, Online.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognit. Lett.*, *27*, 861-874.

Guth, L. M., & Roth, S. M. (2013). Genetic influence on athletic performance. *Current Opinion in Pediatrics*, *25*(6), 653-658. https://doi.org/10.1097/mop.0b013e3283659087

Hunter, S. K., Stevens, A. A., Magennis, K., Skelton, K. W., & Fauth, M. (2011). Is there a sex difference in the age of elite marathon runners? *Medicine and science in sports and exercise*, *43*(4), 656-664. https://doi.org/10.1249/mss.0b013e3181fb4e00

Jones, C. M., Griffiths, P. C., & Mellalieu, S. D. (2017). Training Load and Fatigue Marker Associations with Injury and Illness: A Systematic Review of Longitudinal Studies. *Sports Medicine*, *47*(5), 943-974. https://doi.org/10.1007/s40279-016-0619-5

Kholkine, L., Servotte, T., de Leeuw, A.-W., De Schepper, T., Hellinckx, P., Verdonck, T., & Latré, S. (2021). A Learn-to-Rank Approach for Predicting Road Cycling Race Outcomes [Original Research]. *Frontiers in Sports and Active Living*, *3*(256). https://doi.org/10.3389/fspor.2021.714107

Kim, C.-H., Wheatley, C. M., Behnia, M., & Johnson, B. D. (2016). The Effect of Aging on Relationships between Lean Body Mass and VO2max in Rowers. *PLOS ONE*, *11*(8), e0160275. https://doi.org/10.1371/journal.pone.0160275

Lambert, M. I., & Borresen, J. (2010). Measuring Training Load in Sports. *International Journal of Sports Physiology and Performance*, *5*(3), 406-411. https://doi.org/10.1123/ijspp.5.3.406

Leeuw, A.-W. D., Meerhoff, L. A., & Knobbe, A. (2018). Effects of Pacing Properties on Performance in Long-Distance Running. *Big Data*, *6*(4), 248-261. https://doi.org/10.1089/big.2018.0070

Maffulli, N., Longo, U. G., Gougoulias, N., Caine, D., & Denaro, V. (2011). Sport injuries: a review of outcomes. *Br Med Bull*, *97*, 47-80. https://doi.org/10.1093/bmb/ldq026

Malcata, R., Hopkins, W., & Pearson, S. (2014). Tracking Career Performance of Successful Triathletes. *Medicine and science in sports and exercise*, *46*. https://doi.org/10.1249/MSS.0000000000000221

Mignot, J.-F. (2015). Strategic Behavior in Road Cycling Competitions. In D. V. Reeth & D. J. Larson (Eds.), *The Economics of Professional Road Cycling* (pp. 207-231). https://doi.org/10.1007/978-3-319-22312-4_10

Miller, J., & Susa, K. (2018). Comparison of anthropometric characteristics between world tour and professional continental cyclists. *Journal of Science and Cycling*, *7*(3), 3-6. https://doi.org/10.28985/181231.jsc.02

Momaya, A., Fawal, M., & Estes, R. (2015). Performance-Enhancing Substances in Sports: A Review of the Literature. *Sports Medicine*, *45*(4), 517-531. https://doi.org/10.1007/s40279-015-0308-9

Passfield, L., & Hopker, J. (2016). A Mine of Information: Can Sports Analytics Provide Wisdom From Your Data? *International Journal of Sports Physiology and Performance*, *12*, 1-17. https://doi.org/10.1123/ijspp.2016-0644

Ransdell, L. B., Vener, J., & Huberty, J. (2009). Masters Athletes: An Analysis of Running, Swimming and Cycling Performance by Age and Gender. *Journal of Exercise Science & Fitness*, *7*(2, Supplement), S61-S73. https://doi.org/https://doi.org/10.1016/S1728-869X(09)60024-1

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53-65. https://doi.org/https://doi.org/10.1016/0377-0427(87)90125-7

Rüst, C. A., Rosemann, T., Lepers, R., & Knechtle, B. (2015). Gender difference in cycling speed and age of winning performers in ultra-cycling – the 508-mile “Furnace Creek” from 1983 to 2012. *Journal of Sports Sciences*, *33*(2), 198-210. https://doi.org/10.1080/02640414.2014.934705

Satopaa, V., Albrecht, J. R., Irwin, D. E., & Raghavan, B. (2011). Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior. 31st International Conference on Distributed Computing Systems Workshops, Minneapolis, Minnesota USA.

Schoenberg, J. B., Beck, G. J., & Bouhuys, A. (1978). Growth and decay of pulmonary function in healthy blacks and whites. *Respir Physiol*, *33*(3), 367-393. https://doi.org/10.1016/0034-5687(78)90063-4

Smith, D. J. (2003). A Framework for Understanding the Training Process Leading to Elite Performance. *Sports Medicine*, *33*(15), 1103-1126. https://doi.org/10.2165/00007256-200333150-00003

Song, Y.-Y., & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, *27*(2), 130-135. https://doi.org/10.11919/j.issn.1002-0829.215044

Steinley, D. (2006). K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, *59*(1), 1-34. https://doi.org/https://doi.org/10.1348/000711005X48266

van der Zwaard, S., de Ruiter, C. J., Jaspers, R. T., & de Koning, J. J. (2019). Anthropometric Clusters of Competitive Cyclists and Their Sprint and Endurance Performance. *Front Physiol*, *10*, 1276. https://doi.org/10.3389/fphys.2019.01276

Van Erp, T., Sanders, D., & Lamberts, R. P. (2021). Maintaining Power Output with Accumulating Levels of Work Done Is a Key Determinant for Success in Professional Cycling. *Medicine and science in sports and exercise*, *53*(9), 1903-1910. https://doi.org/10.1249/mss.0000000000002656

Zingg, M., Knechtle, B., Rüst, C. A., Rosemann, T., & Lepers, R. (2013). Age and gender difference in non-drafting ultra-endurance cycling performance - the 'Swiss Cycling Marathon'. *Extreme Physiology & Medicine*, *2*(1), 18. https://doi.org/10.1186/2046-7648-2-18

Figure Captions:

Figure 1. Overview of our approach for studying the relationship between age and performance in professional road cycling. First, we trained a race profile classifier to determine the race type of all races in our data collection. Hereafter, we combined the classified race profiles with the race results to extract the number of points gained in mountain and flat races and cluster the riders into four different categories. Finally, we considered the top and the bottom 50% of the riders in each cluster and construct the age-performance models.

Figure 2. The decision tree sprint (a) and mountain (b) classifiers for race profiles based on result features. A decision tree classifier is read by following each of the criteria. If the criterium is true, the path is followed to the left and if it is false, it is followed to the right. The decisive features for sprint races are the number of separate groups finishing the race, the average speed of the winner for the whole race and the number of riders in the first group. For mountain races, the main features are the standard deviation of the relative time (the time difference with respect to the winner) from the top 15 riders, the number of separate groups finishing the race, and the winner's average speed.

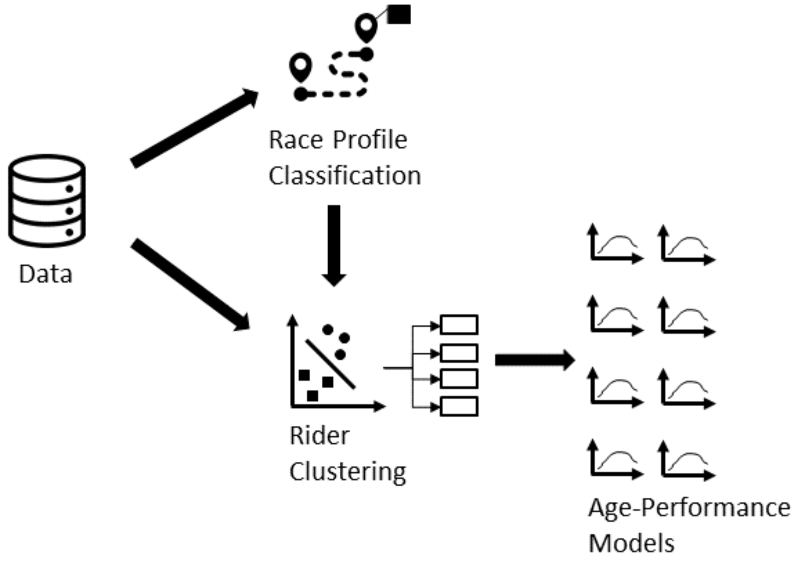
Figure 3. The Summed Squared Error (SSE) decrease for the different number of clusters. The upper bound corresponds to taking into account all the athletes, and the bottom bound corresponds to taking athletes with at least 2000 total PCS points. The final selected cutoff level of 700 was selected after analysing the SSE loss against different cutoff levels. We observe that the ideal number of clusters is between 2 and 5.

Figure 4. This heatmap represents the centroid of each cluster. Each value indicates the fraction of the total number of PCS points gained by an athlete in General Classifications (GC), One Day (OD) races,

mountain races, flat races and individual time trials (ITT). The names of the clusters are based on how the feature values are distributed in a cluster.

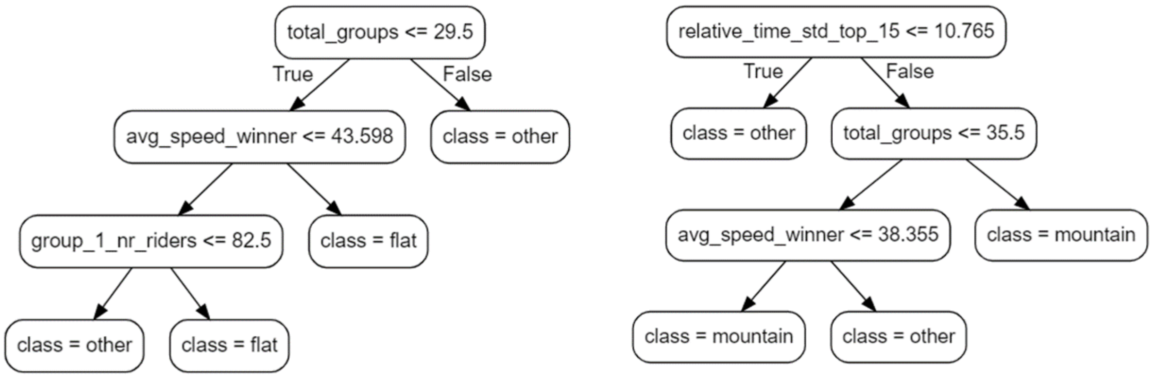
Figure 5. Age-performance models for the different clusters. We display the performance, i.e., the total number of points per race entered, with respect to the personal maximum value. We show the relationships for the best and bottom 50 per cent of each cluster with the shaded areas denoting the 95% confidence intervals.

533 **Figure 1**



534

535 **Figure 2**



(a) (b)

536

537

Figure 3

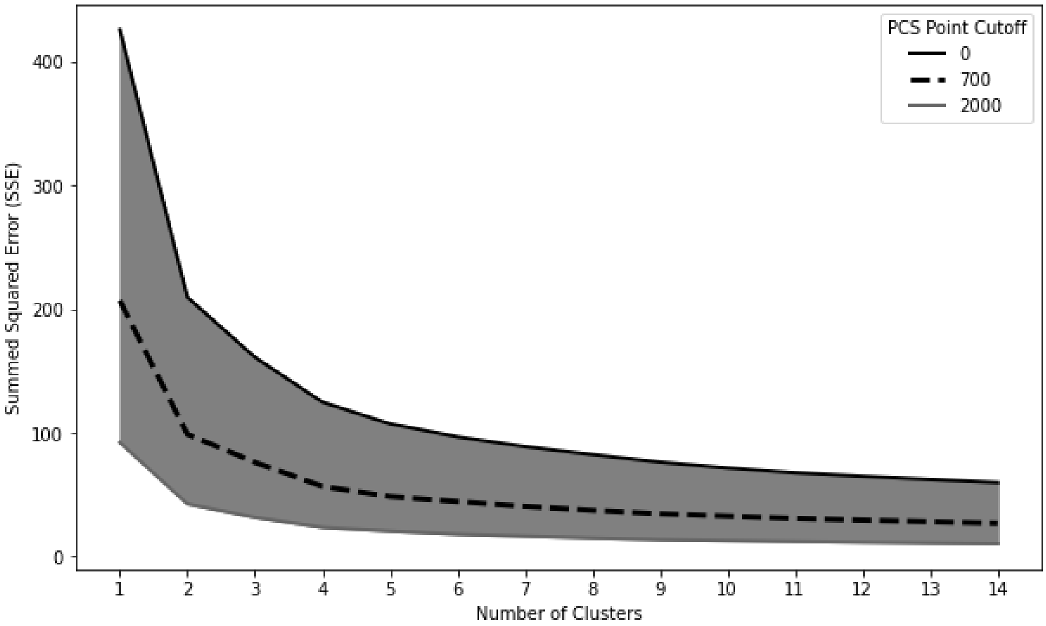
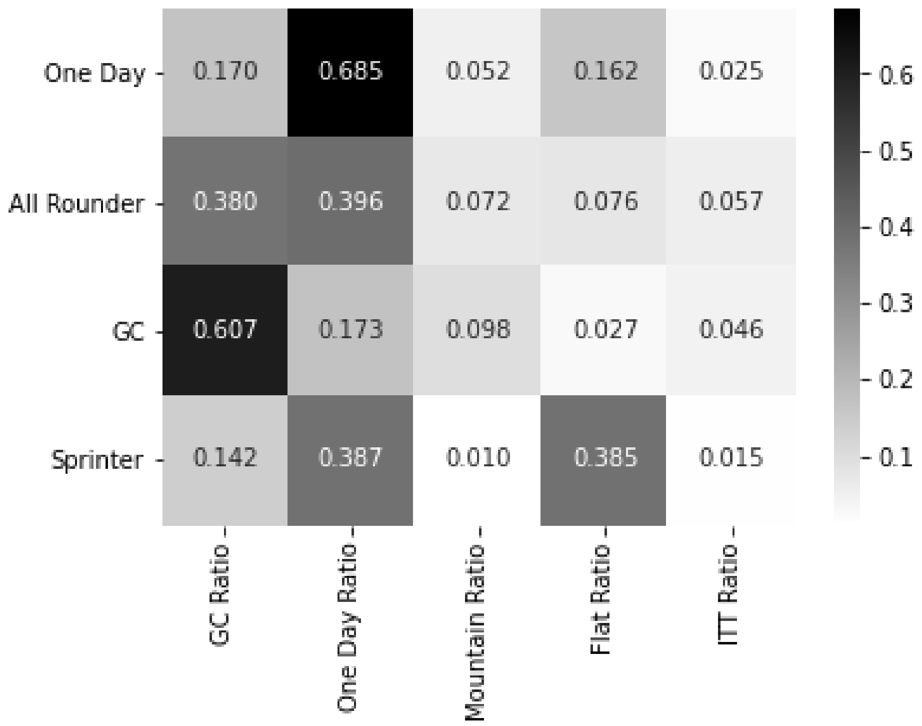
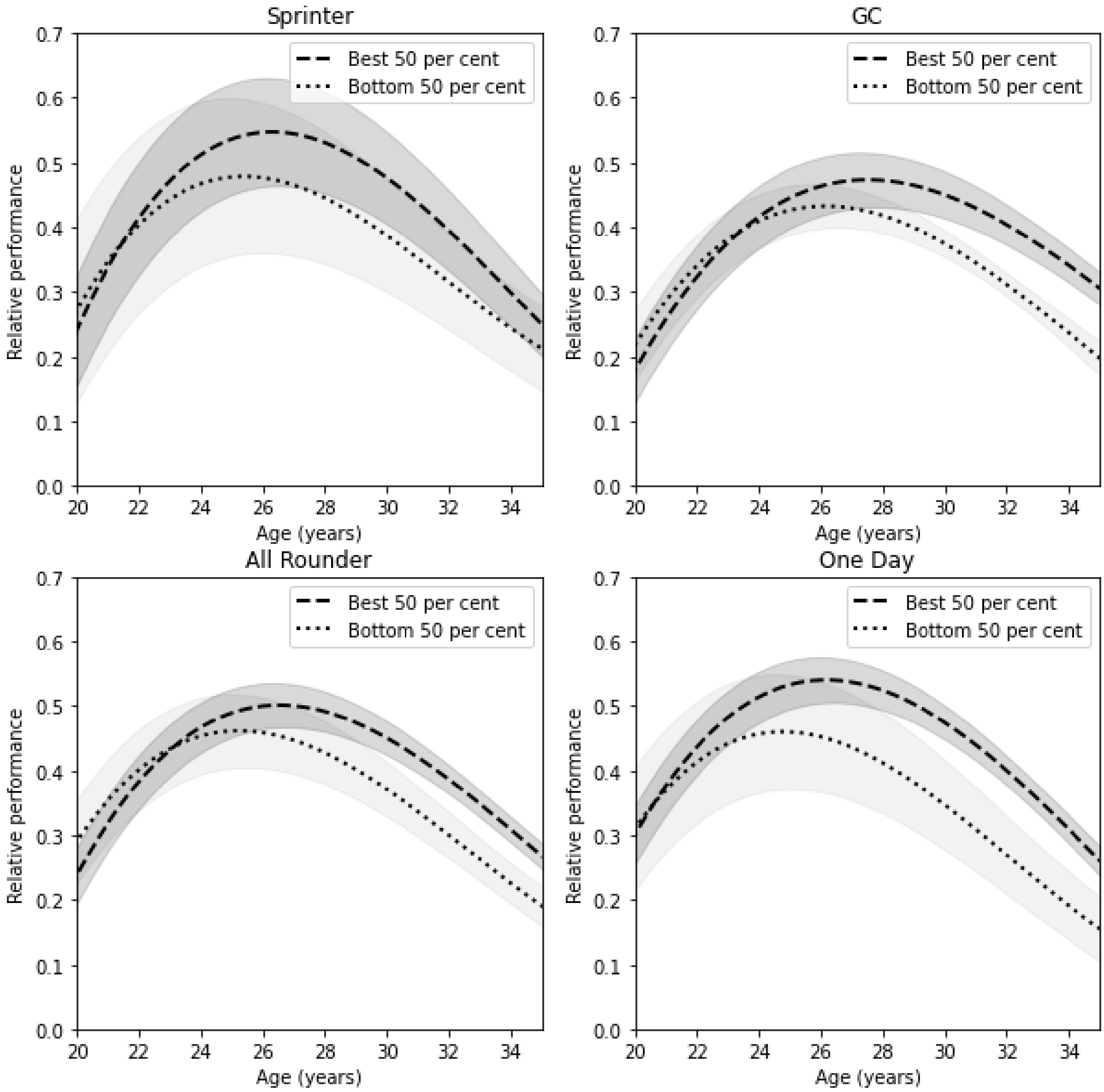

PCS Point Cutoff
0
700
2000
Summed Squared Error (SSE)
0
100
200
300
400
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Number of Clusters

Figure 4


One Day
All Rounder
GC
Sprinter
0.170
0.685
0.052
0.162
0.025
0.380
0.396
0.072
0.076
0.057
0.607
0.173
0.098
0.027
0.046
0.142
0.387
0.010
0.385
0.015
GC Ratio
One Day Ratio
Mountain Ratio
Flat Ratio
ITT Ratio
0.6
0.5
0.4
0.3
0.2
0.1

544 **Figure 5**



545

546

547

Table 1. Average silhouette coefficient per cluster. Here, a score of 0 corresponds to indifferent clusters and the larger the value of the silhouette score, the more distinct the clusters are from each other. Our cluster are well-defined as all scores are substantially larger than 0.

Rider Type	**Average Silhouette Coefficient**
Sprinter	0.354
GC	0.454
All Rounder	0.281
One Day	0.337

Table 2. P-values for the Lack-of-fit F tests for the age-performance models. In all cases $p > 0.05$, which implies there is no significant evidence for lack of fit and therefore the models are considered statistically significant.

Rider Type	**Bottom 50 per cent**	**Best 50 per cent**
Sprinter	0.17	0.39
GC	0.81	0.52
All Rounder	0.97	0.71
One Day	0.50	0.72

Table 3. Age of peak performance in years for the bottom and best 50 per cent of cyclists in the different clusters. The best 50 per cent of the riders in each cluster have a statistically significant ($p < 0.05$) higher age of peak performance than the bottom 50 per cent.

Rider Type	**Age (years) of peak performance (95% CI)**		**P-Value**	**Cohen's d (95% CI)**	**Effect size**
	Bottom 50 per cent	**Best 50 per cent**			
Sprinter	25.4 (20.9-29.9)	26.3 (22.8-30.0)	P<0.01	0.43 (0.12-0.75)	Small
GC	26.1 (23.2-28.6)	27.5 (24.3-30.6)	P<0.01	0.94 (0.75-1.13)	Large
All Rounder	25.2 (21.9-28.5)	26.5 (23.8-29.1)	P<0.01	0.86 (0.68-1.04)	Large
One Day	24.8 (20.9-28.9)	26.2 (23.5-28.7)	P<0.01	0.82 (0.59-1.05)	Large

List of Supplemental Digital Content

- Supplemental Digital Content 1. The confusion matrix for the race classifiers. pdf
- Supplemental Digital Content 2. Complete silhouette analysis of the different clusters for rider clustering. pdf