"Engine matters": a first large scale data driven study on cyclists' performance

Paolo Cintia
KDD Lab, ISTI-CNR
via G. Moruzzi 1, 56124 Pisa, Italy
Email: cintia@di.unipi.it

Luca Pappalardo
KDD Lab, ISTI-CNR
via G. Moruzzi 1, 56124 Pisa, Italy
Email: lpappalardo@di.unipi.it

Dino Pedreschi
KDD Lab, University of Pisa
Largo B. Pontecorvo 3, 56127 Pisa, Italy
Email: pedre@di.unipi.it

***Abstract*—The recent emergence of the so called online social fitness constitutes a good proxy to study the patterns underlying success in sport. Through these platforms, users can collect, monitor and share with friends their sport performance, diet, and even burned calories, giving an unprecedented opportunity to answer very fascinating questions: What are the main factors that shape sport performance? What are the characteristics that distinguish successful sportsmen? Can we characterize the role of social influence on fitness behavior?**

In the current work, we present the results of a study conducted on a sample of $29,284$ cyclists downloaded via APIs from the social fitness platform Strava.com. We defined two basic metrics: a measure of *training effort*, that is how much a cyclist struggled during the workout; and a measure of *training performance* indicating the results achieved during the training. Analyzing the relationship between these two metrics, an interesting result immediately emerges: at a global level, there is no correlation between effort and performance. This means that, in general, the performance is not simply a function of training: two athletes with the same level of training have different performance. However, by deeply investigating workouts time evolution and cyclists' training characteristics, we found that athletes that better improve their performance follow precise training patterns usually referred as overcompensation theory, with alternation of stress peaks and rest periods. Studies and experiments related to such theory, up to now, have always been conducted by sports doctors on a few dozen professionals athletes. To the best of our knowledge, our study is the first corroboration on large scale of this theory, mainly confirming that "engine matters", but tuning is fundamental.

I. Introduction

The striking proliferation of data that characterizes our modern era is now affecting even another interesting aspect of our society: the complexity underlying *sports performance* is starting to be unveiled through the powerful tools of data science. Nowadays, the perspective of securing a competitive advantage versus their peers is driving major sports organization (including baseball's Boston Red Sox and Italy National Soccer Team) to collect and analyze more and more data on their athletes: individual player performance, coaching or managerial decisions, game-based events, and the list goes on. Obviously, since such data represent a great wealth in terms of competitiveness for sport organizations, they are rarely made public. Fortunately, the recent emergence of the so called online social fitness constitutes a good proxy to study the patterns underlying success in sports. Indeed, through these platforms users can collect, monitor and share with friends their sports performance, diet, and even burned calories. Since such data are generally made available by public APIs, this allows researchers to download and analyze information about thousands of professional and amateur sportsmen, giving an unprecedented opportunity to answer very fascinating questions: What are the main factors that shape sports performance? What are the characteristics that distinguish successful sportsmen? Can we characterize the role of social influence on fitness behavior?

In the current work, we present the results of a study conducted on a sample of $29,284$ cyclists downloaded via APIs from the social fitness platform Strava.com. This platform makes fitness a social experience: cyclists and runners all over the world can share, compare and compete with each other's personal fitness data via mobile and online apps. By using the available information about cycling workouts (such as average speed, duration of ride, cyclists' heart bit rate and power), we derived two basic metrics: a measure of *training effort*, that is how much a cyclist struggled during the workout; and a measure of *training performance* indicating the results achieved during the training. Analyzing the relationship between these two metrics, an interesting result immediately emerges: at a global level, there is no correlation between effort and performance. This means that, in general, employing greater effort does not necessarily produce better results. In peloton slang, cyclists say that "the engine matters": an important component of success is talent, expressed in terms of aerobic capacity and other power indicators.

However, this does not necessarily mean that if you were born with the right characteristics, you will be a top cyclist. Indeed, a dynamic temporal analysis of workouts characteristics revealed that talent is not everything: to be a successful sportsman you need to train in the right way. Using clustering techniques, we split the dataset in three subpopulations of cyclists presenting similar training behaviors. By analyzing the way people trained in the clusters we discovered that athletes that better improve their performance follow precise training patterns, with alternation of high stress peaks and rest periods. This is a confirmation of the overcompensation theory [6] (Figure 1), the main medical sports theory applied to aerobic sports. The interesting and new result is that overcompensation seems the only way to reach good performances, since users who do not follow such workouts plan do not achieve good

training results. Studies and experiments related to such theory, up to now, have always been conducted by sports doctors, in specialized (and expensive) laboratories on a few dozen professionals athletes. To the best of our knowledge, our study is the first corroboration on large scale of this theory.
The paper is organized as follows. Section II provides a literature review on sports data mining. In Section III we describe the dataset, define used metrics and present outcomes of experiments. Finally, Section IV concludes the paper providing possible future research scenarios.

Fig. 1. The Overcompensation theory

II. Related works

Sports data mining is a relatively recent field, with the larger diffusion on sports where tactics have a big influence on results: baseball, basketball, football, soccer or volleyball. In these sports, the sequence of moves made by players and teams produces a huge observable mass of data, making data mining essential to support coaches and players in their decision process. In [2], authors give a complete survey on state-of-the-art techniques for extracting patterns from sports data. They describe the sports data mining process in the classical terms of statistics, artificial intelligence and machine learning, defining the methods and techniques involved in a framework for sports data mining. An example of this kind of approach is depicted in [4], where a pattern discovery exploration has been made to find common winning tactics in tennis matches. Authors of [3] propose a Bayesian classifier for predicting baseball awards, prizes assigned to the best pitchers in the Major League Baseball. Since winners are elected by members of the Baseball Writers Association of America, criteria are not static, making the prediction a challenging task. Results of the proposed model show that predictions are correct in the 80% of the cases, highlighting the usefulness of underlying data on describing sports results and performances.
Our work falls into an even newer field than classic sports mining research. Cycling is an individual sport where tactics are not crucial, since the athlete's performance mainly depends on physiological characteristics. To give an example, VO_2max [5] indicates the maximal volume of oxygen the body can process to produce movements: it has been proved that the higher this value, the better the cyclist's performance. Research in this field, to the best of our knowledge, is usually pursued in specialized lab by sports medicine researchers on small sets of athletes. However, thanks to the increasing diffusion of training assistant devices, nowadays non-professional cyclists can easily track their workouts in spatial terms (speed, elevation gain), aerobic indexes (heartrate) or physics measures (watts). The last edition of Tour de France, the most important cycling race in the world, clearly showed that data are becoming more and more important for cycling. The team of the yellow jersey Chris Froome released all the data regarding the performance of the winner of Tour de France, in order to clean away all the suspects about doping. An interesting introduction to the measures used in cycling training sciences, such as VO2max, is given in [5]. Another important metric we use in this paper is Training Stress Score (TSS), introduced in [8]: based on watts expressed by the athlete, it gives a measure of the intensity of the workout. The *Mean Ascent Velocity* (VAM) [9], defined as the elevation gain over time, is used to compare cyclists' performance across different uphills climbing. The importance of the Heart rate parameter is explained in [7], where authors show how heartrate is a good index for expressing the lactate threshold. In the next section a review of these measures connected to sports science is provided. As stated in [6], periodization is the base for building a correct training plan. Athletes aim to reach one or more performance peaks across the season: in order to do this, the training year is sub-divided in more cycles, alternating high and low intensity training sessions. One of the main results of our work is learning the characteristics of the training season from the observation of a high number of athletes' workouts: data mining is used to confirm and prove some well known training theories, on a large scale population of amateur cyclists.

III. Results

A. Data

Strava.com[1] is a social fitness platform where cyclists and runners all over the world can share, compare and compete with each other's personal fitness data via mobile and online apps. It makes fitness a social experience, providing motivation and camaraderie even if athletes are exercising alone. When a user signs up for an account, she can download a free app for tracking her rides or runs. Once she completed a ride or run, data are automatically sent to Strava.com. By the app, a user is able to find popular and competitive segments nearby the place where she is located, and participate in virtual rides or races with other users. Moreover, the platform also includes a social dimension to the experience, allowing users to follow friends and their activities, join clubs and create new ones.

Using Strava.com APIs, we downloaded a set of features regarding a sample of the $29,284$ users from around the world (Table I briefly describes our dataset). We selected, from the total 29k riders, a subset of 1,868 users with the following characteristics: i) they have more than 30 training sessions in the period November 2012 - May 2013 (25 total weeks), in order to select users that are active throughout the period of observation; ii) they live and perform workouts in the northern hemisphere, in order to obtain similar seasonal

[1] http://www.strava.com/

Total users	29,284
Users in period Nov 2012 - May 2013	1,868
Total Km traveled	$4.8023 * 10^9$ km
Total elevation gained	$4.58612 * 10^7$ m
Total training time	195,625h 53m 43s
Estimated power production	11.796472 MW
Total training session analyzed	88,632
Average training session per user per week	1.89

TABLE I
DESCRIPTION OF THE STRAVA DATASET

AVAILABLE INFORMATION	
Training sessions	moving time
	traveled distance
	elevation gain
	heart rate stream
	estimated average watts
Crossed segments	elapsed time
	elevation gain
	mean ascent velocity (VAM)
	estimated average watts

TABLE II
DESCRIPTION OF AVAILABLE TRAINING SESSION AND SEGMENT INFORMATION

weather conditions and include countries with the stronger cycling diffusion and tradition (for instance France, Italy, and USA). As Figure 2 shows, the number of training sessions per week is not constant during the period of observation: in cold seasons (from November to February) we observe a law training activity, but as the weather conditions start to improve the number of active users grows more and more.

For each training session of each rider we have the following information: moving time, traveled distance, elevation gain, heart rate stream with a sample rate of 3 seconds, estimated average watts produced. Furthermore, for each session we have the crossed competitive segments. In particular, every segment has its elapsed time, elevation gain, VAM (mean ascent velocity), and estimated average watts (Table II summarizes the available information). We use data from segments to evaluate the performance of the rider, while, as explained in the following section, data from training session – i.e. heartrate stream – has been used to define the intensity of the workouts.

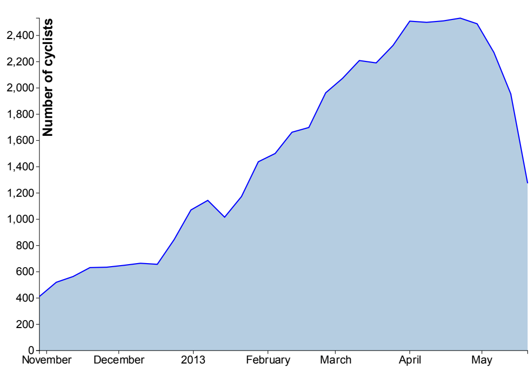

Fig. 2. Number per week of cyclists performing workouts in the period of observation.

B. Metrics

Science of training defines the fitness performance as a combination of aerobic capacity, lactate threshold and economy. The *aerobic capacity* indicates the amount of oxygen the body can process to produce movement. It is referred with the acronym VO_2max and measured in terms of milliliters of oxygen per kilogram of body weight per minute [5]. A world-class male rider usually produces numbers in the 70 to 90 ml/kg/min range. A normal non-professional rider typically tests in the range of 40 to 50 ml/kg/min. Although it is trainable to a certain extent, aerobic capacity is largely determined by genetics and limited by physiological factors: size of heart, heart rate, heart stroke volume, blood hemoglobin content, aerobic enzyme concentrations, mithochondrial density, and muscle fiber type. Knowing the aerobic capacity of each athlete is not enough to predict the result of a race: the rider with the highest value of VO_2max could be not the actual winner.

Indeed, another important fitness index is crucial for a good performance: the *lactate threshold*. Sometimes called "anaerobic capacity", it is the level of effort intensity above which lactate begins to rapidly accumulate in the blood [5]. At this point, in the production of energy metabolism rapidly shifts from dependence on the combustion of fat to dependence on glycogen, that is the storage form of carbohydrates. The higher this threshold as percentage of aerobic capacity, the faster the athlete is able to ride for an extended period of time. Usually, this measure is approximated in terms of heart rate: under a certain value, the body can handle the lactate production; over that value, lactate starts to circulates into cardiovascular system, giving muscle failures. In contrast to aerobic capacity, lactate threshold is a highly trainable characteristic, to such an extent that in general the goal of a training period is mainly intended to elevate the lactate threshold.

Economy can be viewed as the capacity of a machine to use less fuel to produce the same amount of power. This is obviously an advantage in competition. In a rider, economy is given by a combination of genetic and physiological factors, like the percentage of slow-twitch muscle fibers, the body mass or the psychological stress and the aerodynamic of the bike-rider system [5].

Aerobic capacity, lactate threshold and economy depict together a clear picture of the state of fitness of an individual, describing both effort and performance performed by the athlete. Clearly, due to the difficulty of measuring the needed values, the above described measures can be obtained only in laboratories with the necessary equipment. Unfortunately, Strava.com does not provide enough information to compute the complete set of fitness indexes. Therefore, a crucial issue has to be addressed: how to extract from our data reliable estimations of training effort and performance?

Modern sports science suggests that *power* is an adequate value to concisely summarize the fitness behavior of an individual. Expressed in watt as the ratio between workload and time, it is the more accurate index to evaluate both workout intensity and performance: all the complex system represented by the athlete (training, goals, and underlying physiological parameters) is representable by the watts produced during the workouts. Chris Froome, for example, based his training method on the measuring of power values, who presumably played a key role in his victory of 2013 edition of Tour de France. Typical watts values achieved by Lance Armstrong in famous climbing races are shown in Table III.
The *Training Stress Score* (**TSS**) measure is a power-based effort measure, as defined in [8]. For each athlete, five different power zones are identified and the score of each zone is obtained with the following formula:

$$\text{TSS} = \sum_{i \in Z} t_i * c_i \tag{1}$$

where Z is the set of zones; t_i is the time spent running in zone i; and c_i is a score directly proportional to the watts range and, subsequently, to the physics stress amount of the zone. Since power meters are expensive and hence not yet widely diffused, only a few users in our dataset have reliable information about watts produced during the workouts. Conversely, heart rate monitors are largely diffused between non-professional cyclists. For this reason, we approximated TSS using HeartTSS, that is the training stress score measured using heart rate based zones. Although it is a little less precise parameter, Coggan showed a direct correlation between watts and heart pace [8]. For each cyclist in the dataset, we computed the corresponding HeartTSS measure, by considering her heart rate extension through the minimum and maximum values, referred respectively as HR_{min} and HR_{max}. Then, according to Table IV and equation 1, we computed the scores in the corresponding zones.

Once defined a reliable index of *training effort*, we need a measure of *training performance*. To this purpose, we used the *Mean Ascent Velocity* (**VAM**), defined by Ferrari (former trainer of Lance Armstrong) as the elevation gain over time:

$$\text{VAM} = (M * 3600)/T \tag{2}$$

where M indicates the meters ascended and T the time (in seconds) it took to ascend. This value, calculated in meters/hour, is used to compare cyclists' performance across different uphills climbing, in order to have a capability estimation of riders not influenced by other factors such as wind or aerodynamics. Although climbing is usually done at low speed, it is considered the hardest and most important part of a cycling race: "When road goes up", said great cyclist Eddy Merckx, "you can't hide yourself". Figure 3 shows the typical VAM values achieved by professional cyclists during climbing races of Tour de France, the most important cycling race in the world.

In addition to VAM, we also evaluated performance by an estimation of watts produced in workouts, calculated via software by Strava through the physics formulae explained in [1]. Although this measure is only an approximation (as stated above we do not have actual measurements of power in workouts), we use it to have another clue about the quality of the patterns we extract from our dataset.

Year	Climb	Power (W)
1999	Sestrieres	430
	Alpe d'Huez	410
	Piau Engaly	387
	Soulor	436
2000	Hautacam	457
	Mont Ventoux	425
	Col d'Izoard	424
	Courchevel	430
	Joux Plane	406
2001	Alpe d'Huez	452
	Chamrousse	445
	Plateau de Bonascre	475
	Pla d'Adet	439
	Luz Ardiden	411
2004	Alpe d'Huez	495

TABLE III
LANCE ARMSTRONG'S POWER OUTPUTS IN TOUR DE FRANCE CLIMBS

Zone	Heartrate interval	Score
Z1	$< HR_{min} + (HR_{max} - HR_{min}) * 0.6$	30
Z2	$\geq Z1 and < HR_{min} + (HR_{max} - HR_{min}) * 0.7$	60
Z3	$\geq Z2 and < HR_{min} + (HR_{max} - HR_{min}) * 0.8$	80
Z4	$\geq Z3 and < HR_{min} + (HR_{max} - HR_{min}) * 0.9$	100
Z5	$\geq Z4$	120

TABLE IV
HEARTTSS ZONES

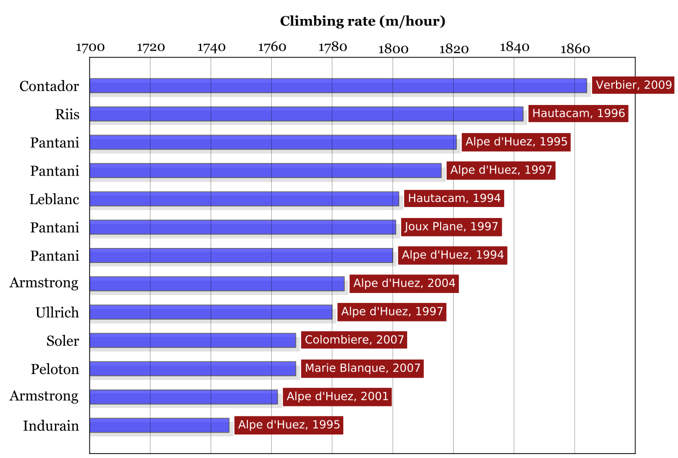

Fig. 3. Highest VAM values recorded on Tour de France climbs

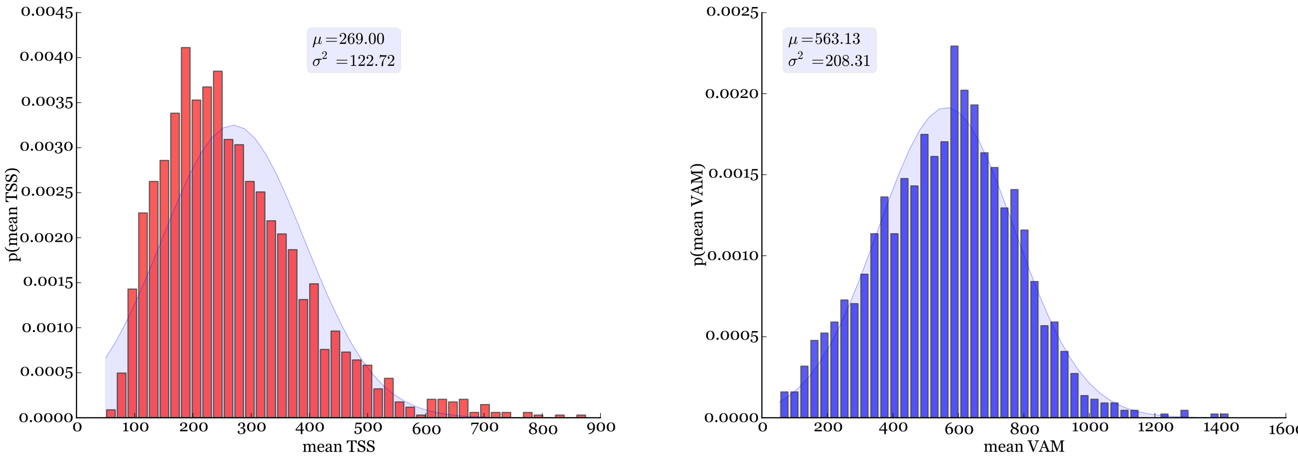

Fig. 4. Probability distributions of mean TSS and mean VAM for users in our dataset. μ and σ^2 indicates respectively mean and variance of the fitted curve.

C. Experiments

The first interesting aspect we investigated in our study is the distribution of endeavors and results accomplished by the amateur cyclists in our dataset. In order to do this, we computed for each user the mean values of TSS and VAM she performed in her workouts during the period of observation. As Figure 4 shows, the distributions are very well fitted by Gaussians, with mean and variance values of $\mu = 269$, $\sigma^2 = 122.72$ (TSS) and $\mu = 563.13$, $\sigma^2 = 208.31$ (VAM). Therefore, in the world of amateur bikers a typical profile of cyclist clearly emerges, showing performance considerably lower than the ones accomplished by professional cyclists of Figure 3. Included in the distribution, those cycling champions would appear as outliers, with performance about three times our average Strava users.
As the scatter plot in Figure 5 suggests, at a global level TSS and VAM are weakly correlated, presenting a very low Pearson's correlation coefficient $\rho = 0.16$. This means that employing greater effort does not necessarily produce better results. Indeed, while some users seem to achieve very good results with very little effort, others fail to achieve good performance, independently from the effort employed. Moreover, most of the points are concentrated around the means of the normal distributions, bringing out a typical workout behavior: most of the users train with an average effort getting an average performance (heat map in Figure 6). Such results, however, concern the aggregated behavior of users. As the overcompensation theory suggests [6], the time evolution of workouts strongly influences the performance improvement: the intensity of your workouts and the way you distribute them over time will determine, in large part, your future sports performance.

In order to study the workouts time evolution and detect those producing the best benefits, we used the metrics described before to extract, for each rider and each week of training, the following information:

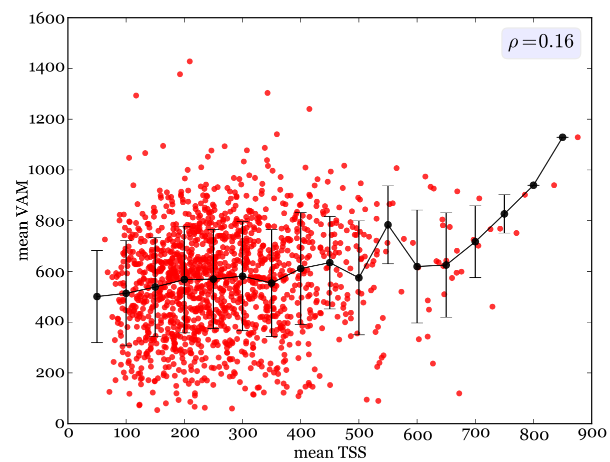

Fig. 5. Correlation between mean TSS and mean VAM. ρ indicates the Pearson's correlation coefficient.

- Sum of TSS of every training session performed during the week;
- VAM variation achieved in the week, computed as the difference between the average VAM of all the segments faced during the week and the same value of the previous week;
- Estimated Watts variation achieved in the week, calculated using the average estimated watts for all the segments faced during the week. Though the Strava.com watts index is an estimation from other data (rider's weight, road climb factor, etc), we use it as a comparison with results obtained with VAM index.

We considered a period of observation of 25 weeks, from November 2012 to the end of April 2013, for two main reasons. First of all, since Strava's APIs have been closed to public access on June 1st, this did not give us enough time to

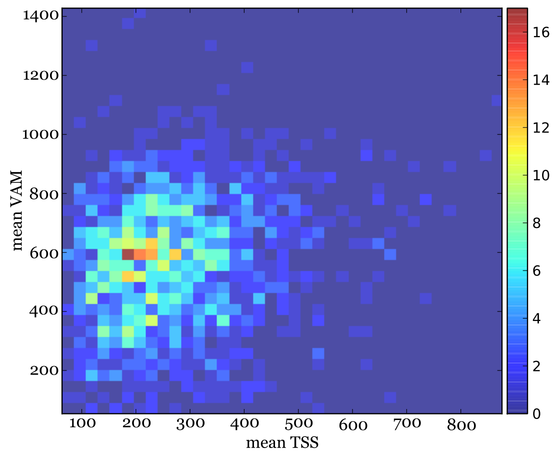

Fig. 6. Heat map of mean TSS and mean VAM values.

CLUSTERING	CLUSTER	#USERS	PROFILE
TSS	T_1	929	*low trained*
	T_2	552	*normal*
	T_3	387	*would-be professional*
VAM variation	V_1	418	*low trained*
	V_2	543	*normal*
	V_3	481	*would-be professional*
Watts variation	W_1	314	*low trained*
	W_2	643	*normal*
	W_3	480	*would-be professional*

TABLE V
DESCRIPTION OF RESULTING CLUSTERS

collect data regarding rides performed in May. Secondly, the chosen period is the first and important half part of a cycling season, that usually starts with the winter initial workouts and it is oriented to the first days of May, when the most important amateur races take place. The same, with different scale factor, applies for professional cyclists: Giro d'Italia usually begins in the first week of May.

Once we characterized each athlete with her weekly TSS, VAM variation and Watts variation, we split the whole population in clusters of similar users with respect to each single metric. To perform this task, we used the K-MEANS clustering algorithm [10] with $k = 3$ as the number of clusters parameter, chosen as the number of clusters produced by the density-based DBSCAN clustering algorithm [10]. Table V summarizes the characteristics of the resulting clusters for TSS clustering, VAM variation and Watts variation clustering.

The TSS based clustering (Figure 7, left) split the riders in a realistic manner. The three behaviors highlighted are easy to find across the "peloton", as the population of bike riders is called between the domain experts. Cluster T_1 (blue dashed curve) identifies the "low trained" rider, who does not have so much time (or will) to perform long training sessions; Cluster T_3 (green dashed curve) represents the opposite "would-be professional" rider, which dedicates a big part of her life to cycling. Between them, in cluster T_2 (red solid curve) there is the "normal" rider, with training workload increasing as well as spring arises and weather conditions improve.

VAM and Watts variation clustering show similar trends and seem to be coherent with the previous results, with an important evidence that confirms the main motivation of this work. In these plots, the three clusters show an initial striking difference: a starting condition where riders' "engines" seem to have different "capacity". Going forward in time, the values almost start to converge, indicating us the efficiency of the three different approaches to training. With respect to the VAM variation clustering (Figure 7, center), the most interesting is the behavior shown by cyclists in cluster V_2: even though in the beginning the average increment is lower than cluster V_3, it finally reaches the highest peaks during the important part of the season. In a virtual battle of the clusters, cluster V_2 would win, highlighting how fundamental are quality of training and workout planning for a training season. In an individual aerobic sports like cycling, where the focus is on the individual physiological parameters, not only the "engine" matters, but type and quality of the training plan play a fundamental role. In Figure 8 (left) we show the TSS time evolution of two users extracted from the most interesting and competitive clusters V_2 and V_3 (VAM variation clustering). Such users are the closest to the centroid of the respective clusters, and present very different training characteristics: while the V_3 user starts with high intensity since the beginning of the winter, user V_2 seems to adopt a more focused periodization of the training. Indeed, she starts with a low-stress winter preparation and increases the intensity during the season, with two resting periods where the intensity decreases (red solid curve in Figure 8 left). This allows the user to tolerate harder training when needed, that is in spring season, as we mentioned before. The benefits of such kind of "periodized" training plan are evident from Figure 8 (center), where the VAMs of the athletes are shown.

Figure 8 (right) shows the relationship between TSS and VAM for the clusters, introducing a refined "intensity-of-training" index. In fact, the ratio TSS/VAM could be viewed as a measure of energy consumption, a way to enrich the information about the intensity of training given by TSS. The V_3 user has a lower average consumption ($\mu_{V_3} = 0.60$) than the V_2 user ($\mu_{V2} = 1$), highlighting the trend of the V_2 user to train harder. In fact, the higher V_2 consumption is in contrast with the higher average TSS of cluster V_3, making the TSS/VAM an accurate parameter to evaluate the quality of training. Furthermore, the TSS/VAM curve of V_2 user has the highest peak around the 15th week, followed by a low-consumption period. Looking at the VAM plot of V_2 (Figure 8, center), user's performance starts to significantly grow just around the same period. High stress peak, resting, performance increasing: this is the exact physiological process known as *overcompensation*.

The main result obtained is an evidence: quality of training influences performance. To the best of our knowledge, this is the first experiment on this field made on this large number of

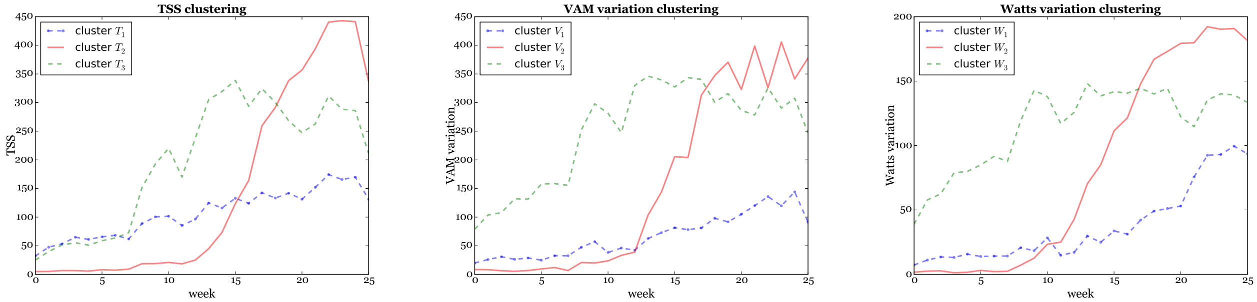

Fig. 7. Values performed by cyclists in the clusters during the period of observation for TSS clustering (left), VAM variation clustering (center) and Watts variation clustering (right).

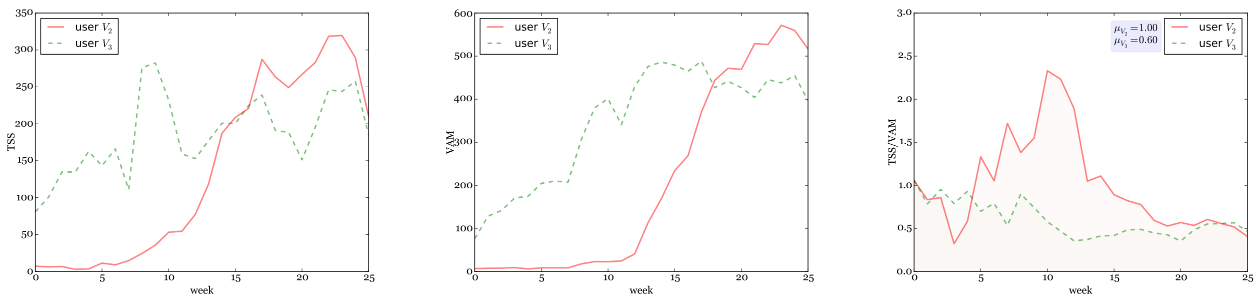

Fig. 8. Comparison of two users from clusters V_2 and V_3 in terms of TSS (left), VAM (center) and TSS/VAM ratio (right).

athletes, mainly confirming that "engine" matters but "tuning" is fundamental.

IV. Conclusion and future works

In this paper, we proposed a first large scale study on sports performance by analyzing a big dataset of cyclists' training sessions, with the aim to understand the secrets of best riders. We first observed that training effort and training performance are weakly correlated, meaning that employing greater effort does not necessarily produce better results. Then, we used more sophisticated data mining techniques to cluster cyclists according to their effort and performance profiles regarding the first half part of the sporting season (from November to May). We discovered that, in an individual sports like cycling, not only "the engine matters", but type and quality of training also play a key role: only users following the well known overcompensation theory reach good training performance.

Data are becoming more and more important in sports understanding. During last Tour de France, the team of the winner Chris Froome gave full access to his personal and historical data, in order to answer to the doping accusations. Data showed that the amazing performance of Froome comes from a precise training planning, that are the base of the incredible Froome results' grown. In the near future, data analysis could help antidoping agencies to detect suspicious athletes by checking whether, for example, an unlikely performance arises from a given training plan.

Future will give us even more chances to collect and study cycling data, due the increasing diffusion of monitoring devices. Such a big amount of data could be exploited by several software tools like, for example, a mobile application that is able to help athletes in their training plans. Starting from individual ride data collected by the cyclist, such "smart" software trainer assistant could use data mining and machine learning techniques to detect whether the athlete is training correctly or has a high risk to fall into over- or under-training, appropriately suggesting a new workout plan.

References

[1] *www.strava.com*
[2] R. Schumaker, O.Solieman, H. Chen, *Sports Data Mining*, Springer, 2010.
[3] L. Smith, B. Lipcomb, A. Smikins, *Data Mining in Sports: Predicting CY Young Award Winners*, Journal of Computing Sciences in Colleges archive, Volume 22 Issue 4, April 2007.
[4] A.Terroba, W.A. Kosters, J.K. Vis, *Tactical Analysis Modeling Through Data Mining*, Proceedings of the International Conference on Knowledge Discovery and Information Retrieval, 2010.
[5] J. Friel, *The cyclists' training bible*, Velopress, 2003.
[6] Boris Blumenstein, Ronnie Lidor, Gershon Tenenbaum, *Psychology of Sport Training*, Meyer & Meyer Verlag, 2007
[7] A. Luca, J. Hoyos, M. Perez, J.L. Chicharro, *Heart rate and performance parameters in elite cyclists: a longitudinal study*, Medicine & Science in Sports & Exercise, 2000
[8] A.R.Coggan, *Training and racing with a power meter: an introduction*, Revised 25 March 2003 [online]. Available from URL: http://www.midweekclub.ca/articles/coggan.pdf
[9] M. Ferrari, *The riders' performance: how to measure it?*, Published 1 mar 2003. Available form URL: http://www.53x12.com/do/show?page=article&id=21
[10] Pang-Ning Tan, M. Steinbach, V. Kumar, *Introduction to Data Mining*, Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 2005.