Pak. j. life soc. Sci. (2025), 23(2): 900-905 E-ISSN: 2221-7630;P-ISSN: 1727-4915

Pakistan Journal of Life and Social Sciences

www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2025-23.2.00221

RESEARCH ARTICLE

Statistical Analysis of Performance Determinants among Senegalese 400- Meter Runners: A PCA and Clustering Approach

Diop Mountaga[1], Guene El Hadji Mama[2], Diene Papa Serigne[1], Diouf Daouda[1], Thiaw Ndiack[(1)], Beye Mame Ngoné[(1)], Mbengue Ndarao[(1)], Diouf Amadou[(1)], Diop El Hadji Mamouthiam[(1)], Samb Abdoulaye[(3)], Andba Abdoulaye[(3)]

[1]Staps-Jl Laboratory, INSEPS-UCAD

[2]Department, Lgl-Tpe, University of Lyon 1, Lyon, France

[3]Faculty Of Medicine, Pharmacy and Odontostomatology, UCAD

ARTICLE INFO

Received: Jul 21, 2025

Accepted: Sep 9, 2025

Keywords

400 Meters
Muscle Power
Principal Component Analysis (PCA)
Hierarchical Clustering

***Corresponding Author:**

narong@pi.ac.th

ABSTRACT

The objective of this study is to identify the determinants of performance in Senegalese 400- meter runners by combining Principal Component Analysis (PCA) and Ward's hierarchical clustering. A sample of 18 athletes, including U20 and senior athletes, was evaluated using anthropometric, physiological, biomechanical and performance variables. PCA identified axes explaining up to 76% of the variance among the best athletes, revealing the predominance of power variables (maximum power, RF Peak) and experience (training age). Clustering highlighted homogeneous athlete profiles, differentiated by their power level, lactate production and age. Contrary to popular belief, post-race lactate levels were not correlated with performance, highlighting the importance of individualized monitoring. These results argue in favor of differentiated training planning and the establishment of a national database for a better monitoring of athletes.

INTRODUCTION

Performance in athletics, particularly in extended sprint events such as the 400 meters, is the result of a complex interaction between physiological, biomechanical, psychological and contextual variables (Billat, 2003). To better understand this complexity, modern statistical tools such as hierarchical *clustering* and Principal Component Analysis (PCA) offer a promising avenue for multidimensional analysis. These techniques not only allow athletes to be grouped according to similarity profiles (Jain et al., 1999), but also reduce the dimensionality of the data while preserving its informative richness (Jolliffe & Cadima, 2016).

Ward's *clustering*, a hierarchical aggregation method based on minimizing intra-group variance (Ward, 1963), is particularly suited to identifying homogeneous subgroups of athletes who share common characteristics. This method has been widely used in sports science to understand the structure of athlete profiles (Readdy and Ebbeck, 2012; Gatti et al., 2019).

As for PCA, it allows the extraction of the main axes explaining the variance within a multivariate data set. It is often used to identify the most influential variables in performance (Hopkins, 2000), reducing the risk of redundancy due to multicollinearity between variables (Jolliffe & Cadima, 2016).

In this study, we propose to integrate these two methods to better characterize the profiles of Senegalese 400-meter runners. The aim will be to determine whether natural groupings emerge based on the data collected (morphology, power, lactate, VO_2max, etc.) and to identify the combinations of variables that best explain performance. This integrated approach aims to improve differentiated training and talent detection.

1. METHODOLOGY

1.1. Equipment

The study involved a sample of 18 Senegalese athletes specializing in the 400 meters, including both U20 runners and senior athletes. All participants were in competitive phase and had a high level of training, which allows for a relevant comparative analysis between different performance profiles.

The variables measured in these athletes fall into several categories:

- ✓ Anthropometric variables: body mass, percentage of fat mass, lean mass index.
- ✓ Physiological variables: VO_2max, lactate levels at rest, after warm-up and after running.
- ✓ Biomechanical and power variables: maximum power, force-velocity ratio (RF Peak), power density, performance over 30 meters from a standing start and a flying start.
- ✓ Experience variable: age of specific training for the 400m.
- ✓ Performance variable: time achieved over 400 metersThe data were collected under standardized conditions in a controlled environment, according to protocols validated in sports science (Billat, 2003).

Method

The analysis was conducted in two complementary stages:

1. Principal Component Analysis (PCA)

PCA was used to reduce the dimensionality of the data and identify the axes explaining the greatest proportion of variance. This method allows correlated variables to be grouped together and those contributing most to performance over 400m to be determined. PCA was applied to the entire sample, then specifically to the eight best athletes, in order to compare variance structures.

2. Hierarchical clustering (Ward's method)

Ward clustering was applied to classify athletes into homogeneous groups according to their physiological and physical characteristics. This method aims to minimize intra-group variance at each stage of aggregation. Two separate analyses were conducted:

- o The first without the performance variables (400m time) and post-race lactate levels;
- o A second analysis incorporating these two key variables in order to assess their impact on cluster structure.

II- RESULTS

II- 1. Results of the Principal Component Analysis (PCA) II- 1.1. PCA on the entire sample

Principal Component Analysis (PCA) applied to all 18 athletes identified three main axes. The first component (PCA1), which explains 67.96% of the variance, mainly includes variables related to power, such as maximum power and the force-velocity ratio (RF Peak), as well as experience (age of specific training for the 400 m) and aerobic capacity (VO_2max). These variables are negatively correlated with performance, indicating that the higher they are, the shorter the race time.

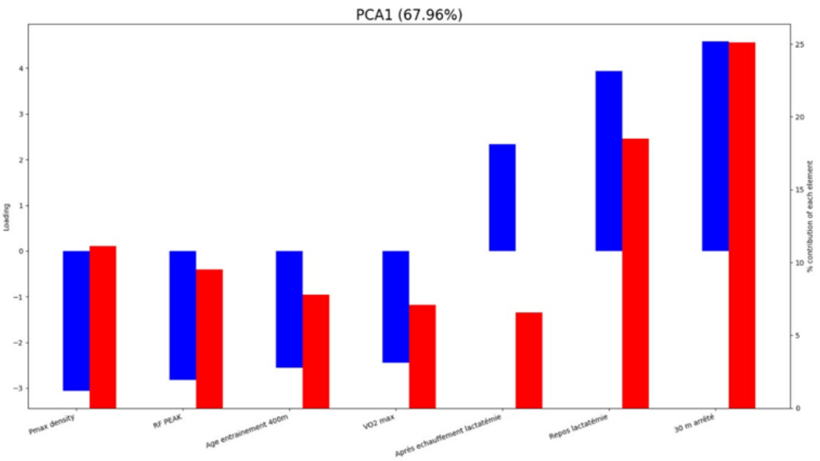

Figure 1: Principal component analysis 1

Conversely, other variables such as lactate levels at rest and performance over 30 meters from a standing start and a flying start are positively correlated and must therefore be reduced to improve results. The second component (ACP2), which explains 11.7% of the variance, is dominated by lactate levels after warm-up, which contribute 60%. However, this component remains marginal in explaining performance. Finally, the third component (ACP3), representing 9.16% of the variance, highlights a combination of bioenergetic variables such as lactate at rest and VO_2max, as well as power variables such as RF Peak.

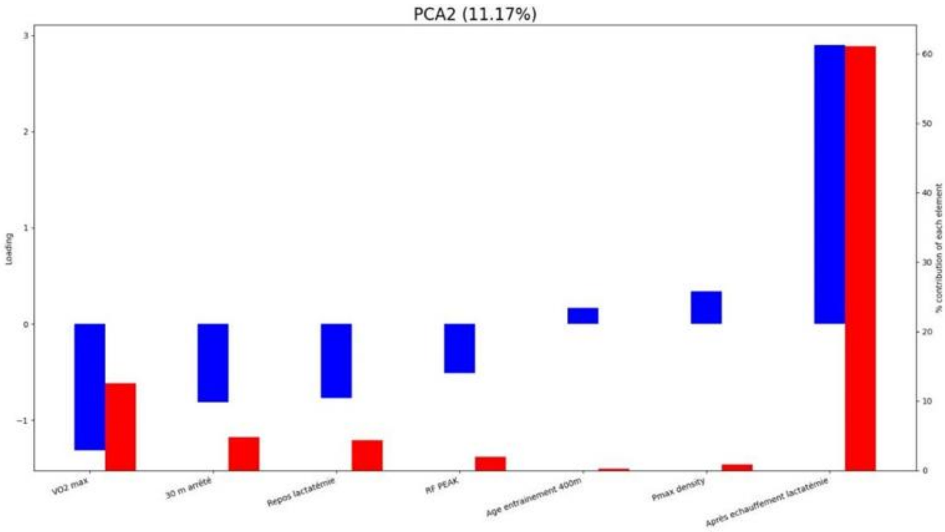

Figure 2: Principal component analysis 2

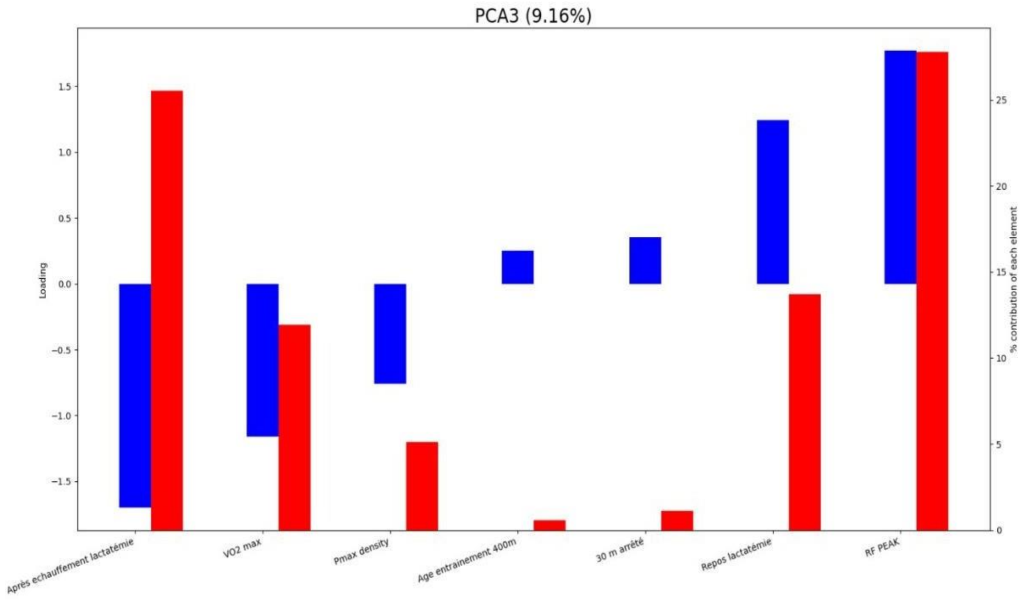

Figure 3: Principal component analysis 3

II- 1.2. PCA restricted to the eight best athletes

When PCA is applied only to the eight best runners in the sample, the proportion of variance explained by the first component (PCA1) rises to 76.24%, which is a significant improvement over the overall analysis. This component is strongly dominated by power variables, particularly maximum power and the force-velocity ratio (RF Peak), which together account for 58% of the variance. This confirms that power is a determining factor in the performance of the best athletes. The second component (ACP2), which explains 16.47% of the variance, highlights the growing importance of secondary variables such as the 30-meter standing start, lactate at rest and RF Peak. These results indicate that, among the best athletes, performance is based on a more subtle combination of power and metabolic regulation factors. In contrast, the contribution of the third component (ACP3) falls to 4.71%, suggesting that it provides little additional information in this high-performance sub-sample.

Figure 42: Principal component analysis 1 with the best athletes

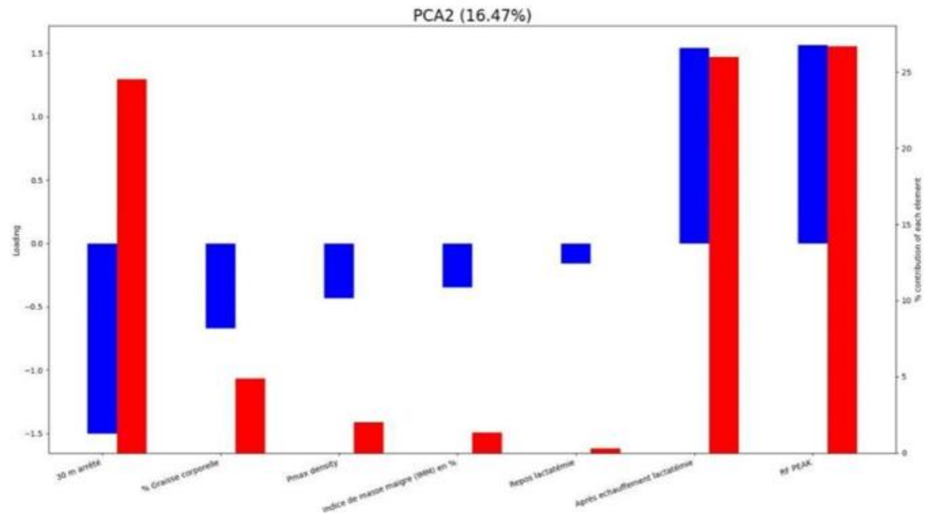

Figure 4: Principal component analysis 2 for the best athletes

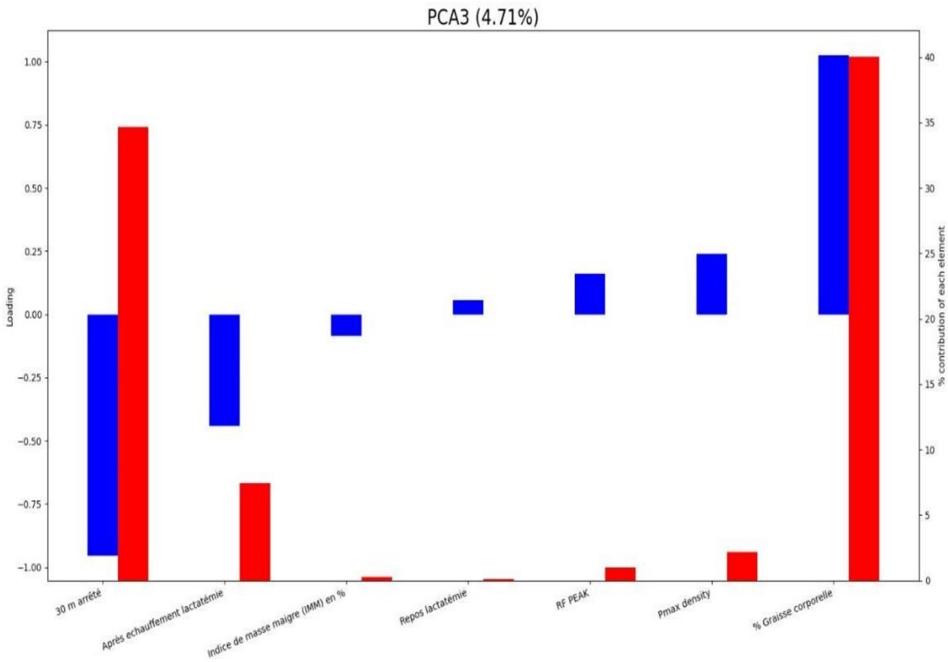

Figure 5: Principal component analysis 3 with the best athletes

II- 2. Hierarchical clustering results (Ward)

II- 2.1. Clustering on variables

Applying Ward's method to the variables helped identify two major clusters:

Cluster 1: variables to be reduced to optimize performance (30m standing start, lactate at rest, lactate after warm-up, 400m performance).

Cluster 2: variables to be increased to improve performance (maximum power, RF Peak, VO_2max, etc.).

II- 2.2. Clustering on athletes

Hierarchical clustering analysis without taking into account performance or lactate levels after 400 meters could identify five distinct groups. The first cluster, comprising athletes 1, 2 and 3, gathers the best performers, revealing a strong similarity between them, even in the absence of explicit data on their running times or metabolic response. Furthermore, age appears to be an important structuring criterion, with younger athletes more frequently grouped together in certain clusters. When performance and lactate levels are included in the analysis, the overall structure of the clusters remains relatively stable. However, a few athletes change groups, showing that these two variables refine the classification without disrupting it. This confirms the relevance of the variables initially used in the grouping process. Thus, this double clustering highlights the existence of distinct athlete profiles, determined by morpho-physiological factors, power and age. It reinforces the idea that the best runners share common characteristics, regardless of their measured performance level.

GENERAL CONCLUSION

This study has provided a better understanding of the determinants of performance among Senegalese 400-metre runners by combining two powerful statistical approaches: Principal Component Analysis (PCA) and hierarchical clustering using Ward's method. These tools revealed that variables related to muscle power (maximum power, RF Peak) and experience in the event (specific training age) are the most influential in achieving good performance.

Contrary to popular belief, post-race lactate levels, often considered an indicator of maximum exertion, have not been shown to be systematically correlated with performance. They seem to depend more on the age or metabolic efficiency of athletes, which opens up interesting prospects for individualized training.

The results from Ward clustering have highlighted distinct athlete profiles, structured around factors such as power, age, lactic endurance and mechanical efficiency. These typologies underscore the value of differentiated physical and physiological training tailored to each runner's characteristics.

Practical recommendations

To improve the performance of 400-meter runners, several practical recommendations can be made. First, it is essential to increase the maximum power of young athletes, particularly those under 20, by incorporating targeted strength-speed, plyometric and short sprint exercises into their training. Furthermore, as experience in the event has proven to be decisive, it is important to emphasize consistency and progressive specialization among 400-meter runners, ensuring longitudinal monitoring of their training. Lactate measurements should be interpreted with caution, cross-referencing them with other indicators such as running time, VO_2 max or RF Peak, while taking into account the athlete's age and physiological profile. The use of profiles derived from clustering allows training to be adapted to the specific characteristics of each group of athletes, whether they are focused on power, lactic endurance or aerobic capacity. Finally, it would be useful to set up a national database of runners, allowing for the centralization of morpho-physiological and performance information in order to optimize their monitoring, detection and progress with a view to sustainable performance.

BIBLIOGRAPHY

Aubert, S., Ratel, S., & Doré, E. (2020). *Physiologie du sport : de la théorie à la pratique*. Éditions De Boeck Supérieur.

Billat, V. (2003). *Physiology and methodology of training. From theory to practice*. De Boeck Supérieur.

Campbell, B. I. (2017). Power as a critical component of sprint performance. *Journal of Strength and Conditioning Research*, 31(4), 1101–1107.

Gatti, C., Chiodera, P., Salvioni, A., et al. (2019). Cluster analysis for evaluating training responses in athletes. *European Journal of Sport Science*, 19(4), 500–508.

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, 30(1), 1–15.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys*, 31(3), 264–323.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202.

Pavei, G., Cazzola, D., La Torre, A., & Minetti, A. E. (2019). Sprint running: Power, force and velocity relationships and the efficiency of energy conversion. *European Journal of Applied Physiology*, 119(6), 1423–1432.

Readdy, T., & Ebbeck, V. (2012). Athlete identity and disordered eating: A cluster analysis. *Journal of Clinical Sport Psychology*, 6(2), 166–182.

Smith, D. J., Norris, S. R., & Hogg, J. M. (2016). Performance evaluation of swimmers: scientific tools. *Sports Medicine*, 26(6), 317–328.

Takei, N., Sugiyama, K., & Inoue, T. (2018). Blood lactate concentration is not related to performance in a 400-m sprint. *International Journal of Sports Physiology and Performance*, 13(10), 1343–1349.

Tzankoff, S. P., & Norris, A. H. (1978). Effect of muscle mass decrease on age-related BMR changes. *Journal of Applied Physiology*, 45(4), 536–539.

Ward, J. H. (1963). Hierarchical Grouping to Optimise an Objective Function. *Journal of the American Statistical Association*, 58(301), 236–244.