OLCIAS
REBAINE Fayçal

Sports Science and Data-Driven Performance Analysis

# COMPARATIVE ANALYSIS OF PACING STRATEGIES AND PERFORMANCE IN MALE AND FEMALE 400 M ATHLETES

## Using Principal Component Analysis

REBAINE Fayçal

**Higher School of Sports Sciences and Technologies**

**Algiers, Algeria**

*Corresponding author:REBAINE Fayçal,Higher School of Sports Sciences and Technologies, Algiers, Algeria*

***Received:****December 15, 2025* ***Accepted:****December 30, 2025* ***Published:*** *January 30, 2026*

***Citation :*** *REBAINE Fayçal. COMPARATIVE ANALYSIS OF PACING STRATEGIES AND PERFORMANCE IN MALE AND FEMALE 400 M ATHLETES : Using Principal Component Analysis.* ***OLCIAS Vol.3, Issue 1***

**ABSTRACT**

This study investigates pacing strategies and performance determinants in elite male and female 400 m athletes using Principal Component Analysis (PCA). Split times over 100 m, 200 m, and 300 m were analyzed to identify key performance indicators and sex-specific differences in pacing. Results indicate that the 300 m and first 200 m splits are the most representative of overall performance in both sexes. Male athletes demonstrated more homogeneous pacing patterns, whereas female athletes showed greater variability, particularly in the final stages of the race.

***Keywords:*** *400 m sprint; pacing strategy; performance analysis; sex differences; principal component analysis (PCA)*

## INTRODUCTION

The 400 m sprint requires a balance between maximal speed, speed endurance, and fatigue management. Previous studies have shown that intermediate split times are strong predictors of final performance [1,3]. The 300 m split has been identified as a critical indicator of success, reflecting the ability to sustain high speed under fatigue [5]. Differences between male and female pacing strategies have also been reported, with males typically maintaining more consistent speed profiles [2]. Principal Component Analysis provides a powerful multivariate approach to explore these complex relationships.

## METHODOLOGY

A retrospective comparative design was used, based on official competition results of elite male and female 400 m athletes The data, collected during various international competitions, is based on results reported by the International Association of Athletics Federations. Active variables included split times for each 100 m, both 200 m segments, and the 300 m. Additional variables representing differences between segments were calculated. Data were standardized prior to PCA. Components were retained according to eigenvalues greater than 1 and scree plot analysis.

## RESULTS:

### 1. PCA Men

**Table1 : Statistical analysis**

| Variable | Observations | Minimum | Maximum | Moyenne | Ecart-type |
|---|---|---|---|---|---|
| 100m1 | 41 | 10,82 | 11,80 | 11,29 | 0,23 |
| 100m2 | 41 | 9,90 | 10,82 | 10,34 | 0,23 |
| 100m3 | 41 | 10,92 | 11,77 | 11,32 | 0,21 |
| 100m4 | 41 | 10,02 | 13,64 | 12,21 | 0,54 |
| 1er 200m | 41 | 20,85 | 22,48 | 21,62 | 0,40 |
| 2e 200m | 41 | 21,33 | 25,32 | 23,54 | 0,62 |
| 300m | 41 | 32,12 | 33,88 | 32,95 | 0,49 |
| Diff. 200m | 41 | - 0,26 | 3,69 | 1,91 | 0,71 |
| D.100m 3-2 | 41 | 0,51 | 1,39 | 0,99 | 0,23 |
| D.100m 4-3 | 41 | - 1,29 | 1,96 | 0,89 | 0,53 |

#### 1.1 Eigenvalues

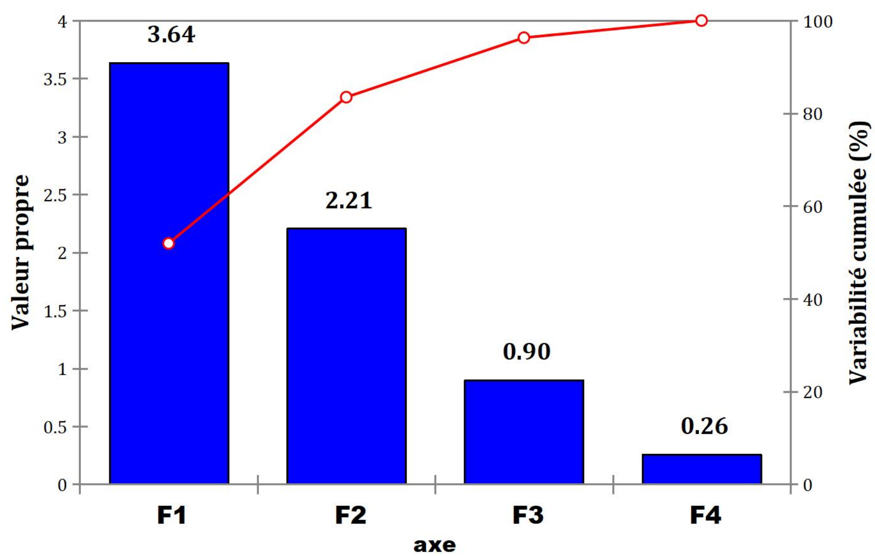
The eigenvectors represent the contribution of each stroke segment to the main factors (F1, F2, F3, F4). The values indicate the extent to which each stroke interval influences these factors.

**Table 2: Eigenvalues :**

| | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| Valeur propre | 3,64 | 2,21 | 0,90 | 0,26 |
| Variabilité (%) | 51,94 | 31,53 | 12,82 | 3,71 |
| % cumulé | 51,94 | 83,47 | 96,29 | 100,00 |

The first interesting result to analyze is the eigenvalues. The eigenvectors show the contributions of each variable to each principal component. The percentage of inertia explained by each axis gives an idea of their relative importance. Each eigenvalue (here 4) corresponds to a factor. The eigenvalues and factors are sorted in descending order of variability represented. The first axis contains the highest eigenvalue and the last, the lowest. Axis 1 explains 52% of the variance, axis 2 31.53% and axis 3 12.82%, giving a cumulative contribution from the first 3 axes of 96.29%, i.e. virtually all the information. Only the first 3 axes are of interest.

**Figure 1: Scree plot**




Determining the number of axes for the analysis There are two ways of determining the number of axes to be considered: An ‘absolute’ criterion and a ‘relative’ criterion are the two empirical criteria for selecting the number of axes

- An ‘absolute’ criterion: only include axes whose eigenvalues are greater than 1 (this is the Kaiser criterion). By convention, any factor with an initial eigenvalue greater than 1 is considered a significant factor. In our case, F3 no longer becomes significant as this axis has an eigenvalue of 0.90%. 2 of the 4 factors will be retained.

- A ‘relative’ criterion: retain the eigenvalues that ‘dominate’ the others, by referring to the bar graph of eigenvalues known as a ‘screeplot’ (Cattell, 1966). Only those factors that lie before the abrupt change in slope (F3 in this case) are selected. The ‘kink’ criterion, where there is a drop-off (kink) followed by a regular decrease. We select the axes before the drop-off, which here represents 3 factors out of 4.

In conclusion, only the first 3 axes are of interest, so the study will focus on the 1-2 and 2-3 planes.

## 1.2 Study of the variables

The fundamental result concerning the variables is the table of variable-factor correlations. This is the table of the coordinates of the variables (correlation of the variable with each dimension). These are the linear correlation coefficients between the initial variables and the factors. Consequently, the principal components must be interpreted in terms of correlation.

**Table3: Correlations between variables and factors**

| | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| 100m1 | **0,67** | -0,57 | -0,37 | -0,29 |
| 100m2 | **0,93** | -0,03 | 0,09 | 0,36 |
| 100m3 | 0,54 | 0,48 | **0,66** | -0,20 |
| 100m4 | 0,24 | **0,85** | -0,48 | 0,01 |
| 1er 200m | **0,92** | -0,35 | -0,16 | 0,04 |
| 2e 200m | 0,39 | **0,90** | -0,19 | -0,06 |
| 300m | **0,98** | -0,08 | 0,15 | -0,05 |
| Diff. 200m | -0,18 | **0,98** | -0,07 | -0,08 |
| D.100m 3-2 | -0,44 | 0,47 | 0,53 | **-0,55** |
| D.100m 4-3 | 0,02 | 0,66 | **-0,74** | 0,09 |

The results corresponding to the additional variables are shown in the second part of the table.

These values show the extent to which each race segment is correlated with the factors.

In this analysis, the first axis, associated with the first eigenvalue with a contribution of around 52%, appears to be strongly correlated with the 300m variable, followed by the 100m2 , the 1st 200m and to a lesser extent the 100m1. These correlations suggest that axis1 mainly captures performances in the first part of the race and contributes in the same direction to the formation of the axis. This indicates that F1 represents a dimension of overall performance over the total distance, through speed and endurance capacities over the 300m and the 1st 200m, or in other words, the ability of athletes to maintain a sustained speed over a long distance.

The higher a subject's score on each of the variables, the higher their score on axis 1. Axis 1 therefore represents, in a way, the overall result of all the variables considered, but it is above all strongly influenced by the variables 300m, 100m2 and the 1st 200m. These variables are very close to F1, which confirms their strong link with this component.

The significance of the second axis, which is less discriminating with a contribution of around 32%, is mainly provided by the 2nd 200m followed by the 100m4. Axis 2 reflects performance in the second half of the race, in particular endurance and the ability to maintain or increase speed in the second half of the race. The importance of maintaining speed and endurance against fatigue in the last segment appears to be the second factor in 400m performance compared with the other segments. This factor represents the ability to adjust or maintain speed over the second half of the distance.

The introduction of the additional variable, the '200m Difference', i.e. the difference between the 1st 200m and the 2nd 200m, which is strongly correlated with axis 2, confirms the management of effort and endurance in the second half of the race, as a large difference between the 200m could indicate a reduction in speed or endurance in the last segment.

This highlights that overall performance in a 400m race is mainly determined by speed in the first segments (axis 1) and the ability to maintain this speed in the second half of the race (axis 2).

The significance of the third axis, even less discriminating with a contribution of 12.82%, is mainly provided by the 3rd 100m and with the difference between intermediate segments, as D.100m 4-3 as an additional variable. This third factor reflects the consistency of performance in the intermediate segments, and the ability to manage the transitions between segments, i.e. the ability to maintain a high speed and the management of effort in the last part of the race.

The fourth factor (axis4), which will not be considered due to its very low contribution of only 3.71%, is mainly represented by the variable "Difference 100m 3-2".

### 1.3 Representation quality

The quality of the representation of variables on the PCA map is called $\cos^2$ (cosine squared). A high $\cos^2$ indicates a good representation of the variable on the principal axes

**Table 4: Squared cosines of variables**

| | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| 100m1 | **0,45** | 0,33 | 0,14 | 0,08 |
| 100m2 | **0,86** | 0,00 | 0,01 | 0,13 |
| 100m3 | 0,30 | 0,23 | **0,44** | 0,04 |
| 100m4 | 0,06 | **0,72** | 0,23 | 0,00 |
| 1er 200m | **0,85** | 0,12 | 0,03 | 0,00 |
| 2e 200m | 0,15 | **0,81** | 0,04 | 0,00 |
| 300m | **0,97** | 0,01 | 0,02 | 0,00 |
| Diff. 200m | 0,03 | **0,96** | 0,00 | 0,01 |
| D.100m 3-2 | 0,19 | 0,22 | 0,28 | **0,30** |
| D.100m 4-3 | 0,00 | 0,44 | **0,55** | 0,01 |

To confirm that a variable is strongly related to a factor, we need to consult the cosine table: the higher the cosine (in absolute value), the more closely the variable is related to the axis. In this case, the variable is positioned close to the circumference of the correlation circle. The closer the cosine is to zero, the less the variable is linked to the axis. In this case, the variable is close to the centre of the circle.

The cosine squares table confirms our analysis of the correlations between the variables and the factors. In particular, the strong representation of the 300m in F1 shows that this factor is centered on overall speed and effort management in the first 3 100m. F2 is more closely linked to endurance and the ability of athletes to maintain a good performance in the second half of the race. F3 could be interpreted as an indicator of endurance, particularly in the last 100 meters of the race. Finally, factor 4 did not provide any significant information and will not be interpreted in depth.

The principal components are used to create the correlation circle. The variables are projected onto a circle of radius 1 called the correlation circle. This corresponds to a projection of the initial variables onto a two-dimensional plane.

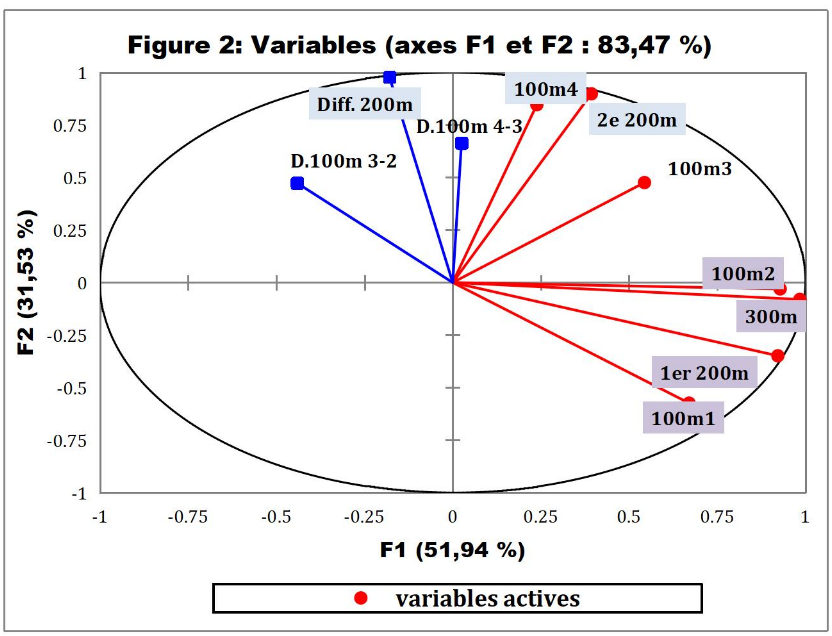



**Figure 2: Variables (axes F1 et F2 : 83,47 %)**

Variables that are well represented are those close to the circle, while those close to the origin are poorly represented. In our example, we can deduce from the plane graph 1-2 that the variables 300m, 100m2 and 1st 200m are the most strongly correlated in the first dimension. The very narrow angle formed by the points 300m and 100m2, which are closely linked, indicates that these variables are very well correlated with each other and are the explanatory variables for this axis; they are followed by 100m1.

To conclude, the results show that the 300m, the 100m2 and the 1st 200m, located in the same direction, are the variables that most influence the first principal component (F1), which means that they have the greatest impact on overall performance in the 400m. The very high correlation of the 300m on axis 1 makes it a central measure of performance. And represents the ability to maintain a high speed in the first 3 100m This makes it a central measure of performance. The importance of the initial speed (100m1) implies a good ability to start quickly.

The 2nd 200m and the 100m4 have a major influence on the second principal component (F2), underlining their importance in managing the race in the second half. The F3 axis is dominated by the variables 100m3 and D.100m 3-2. F3 could therefore be interpreted as an indicator of the variability of performance between the intermediate segments of the race, highlighting the fluctuations in speed between these segments.

## 2. PCA Women

**Table 5: Descriptive statistics :**

| Variable | Observations | Minimum | Maximum | Moyenne | Ecart-type |
|---|---|---|---|---|---|
| 100m1 | 42 | 12,08 | 13,23 | 12,53 | 0,25 |
| 100m2 | 42 | 11,08 | 12,54 | 11,82 | 0,31 |
| 100m3 | 42 | 11,98 | 13,73 | 12,91 | 0,32 |
| 100m4 | 42 | 13,24 | 16,33 | 14,22 | 0,66 |
| 1er 200m | 42 | 23,36 | 25,63 | 24,35 | 0,52 |
| 2e 200m | 42 | 25,40 | 29,50 | 27,12 | 0,89 |
| 300m | 42 | 35,34 | 39,24 | 37,25 | 0,78 |
| Diff 200m | 42 | 1,47 | 5,13 | 2,77 | 0,76 |
| D.100m 3-2 | 42 | 0,60 | 1,54 | 1,09 | 0,21 |
| D.100m 4-3 | 42 | 0,56 | 3,16 | 1,31 | 0,54 |

**Table 6 : Eigenvalues**

| | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| Valeur propre | 5,16 | 1,32 | 0,35 | 0,17 |
| Variabilité (%) | 73,67 | 18,91 | 4,98 | 2,44 |
| % cumulé | 73,67 | 92,58 | 97,56 | 100,00 |

Axis 1 explains 73.67% of the variance, axis 2 18.91%, giving a cumulative contribution from the first 2 axes of 92.58%, i.e. almost all the information. Only the first 2 axes are of interest, so the study will focus on the planes 1-2 whose eigenvalues are greater than 1 according to the Scree plot (this is the Kaiser criterion).

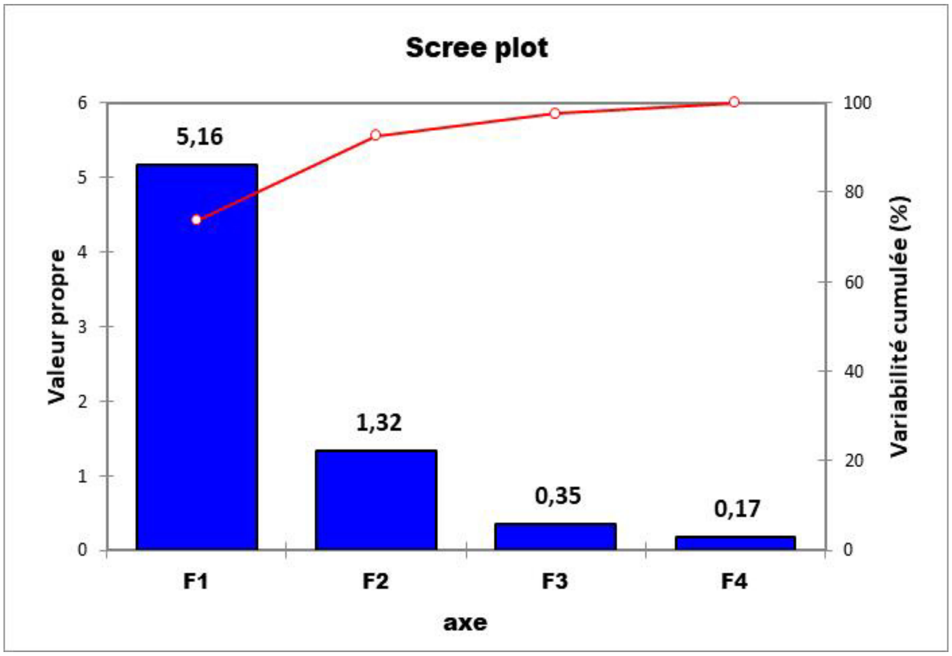



**Figure4: Screen plot eigenvalues 400m women**

**Table 6: Correlations between variables and factors**

| | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| 100m1 | **0,82** | -0,38 | -0,41 | 0,13 |
| 100m2 | **0,90** | -0,29 | 0,22 | -0,25 |
| 100m3 | **0,91** | 0,14 | 0,29 | 0,26 |
| 100m4 | 0,64 | **0,74** | -0,18 | -0,12 |
| 1er 200m | **0,93** | -0,36 | -0,07 | -0,09 |
| 2e 200m | **0,80** | 0,60 | -0,03 | 0,00 |
| 300m | **0,98** | -0,18 | 0,07 | 0,05 |
| Diff 200m | 0,30 | **0,95** | 0,01 | 0,07 |
| D.100m 3-2 | 0,05 | 0,64 | 0,12 | **0,76** |
| D.100m 4-3 | 0,25 | **0,83** | -0,40 | -0,30 |

The results corresponding to the additional variables are shown in the second part of the table.

The correlation table shows the relationships between the variables measured (race intervals) and the different axes or factors (F1, F2, F3, and F4) from the factor analysis.

The first axis, associated with the first eigenvalue with a contribution of 73.67%, appears to be strongly correlated with the 300m and 1st 200m variables, followed by the 100m3 and 100m2 and to a lesser extent the 100m1 and 2nd 200m. The first component (F1) explains most of the variables. This factor mainly represents overall speed over most of the race intervals, as it is strongly correlated with almost all the segments, in particular the 300m and the 1st 200m. This indicates that F1 captures the ability to maintain a high speed across the whole race. Considering these variables and the sign of these correlations, we observe what is known as a Size Effect, i.e. the variables are all on the same side of the axis (they all contribute in the same direction to the formation of the axis).

The second axis is represented by the 100m4 variable, which reflects speed endurance and the ability to resist fatigue. The additional variables diff 200m and D 100m4-3 indicate the athletes' ability to adjust their pace at the end of the race, by managing speed differences in the final segments.

F3 explains only 4.98% of the total variance and F4 explains even less (2.44%). This means that they contribute little to the overall description of the variations observed between performances. In a principal component analysis, we generally prefer the factors that explain most of the variance to simplify interpretation while retaining the essential information. F3 and F4 therefore do not represent an interpretable or relevant dimension in this study. It is therefore wise to limit the interpretation to F1 and F2, which capture the essential information useful for understanding the athletes' performances and strategies in this race.

**Table7: Squared cosines of variables :**

| | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| 100m1 | **0,66** | 0,15 | 0,17 | 0,02 |
| 100m2 | **0,80** | 0,09 | 0,05 | 0,06 |
| 100m3 | **0,83** | 0,02 | 0,09 | 0,07 |
| 100m4 | 0,41 | **0,55** | 0,03 | 0,01 |
| 1er 200m | **0,86** | 0,13 | 0,00 | 0,01 |
| 2e 200m | **0,64** | 0,36 | 0,00 | 0,00 |
| 300m | **0,96** | 0,03 | 0,01 | 0,00 |
| Diff 200m | 0,09 | **0,90** | 0,00 | 0,00 |
| D.100m 3-2 | 0,00 | 0,41 | 0,01 | **0,57** |
| D.100m 4-3 | 0,06 | **0,69** | 0,16 | 0,09 |

The squared cosines of the F1 and F2 factor variables are the only ones to show significant squared cosines with the active variables. They capture most of the structure of the data. The table of squared cosines confirms the fact that 300m is the variable most closely linked to axis 1, and to a lesser extent 1,200m and 100m3.

F3 has very low cosines squared for all the variables and can be ignored in the interpretation. F4 makes a weak and specific contribution to an additional variable, which limits its interest in performance analysis.

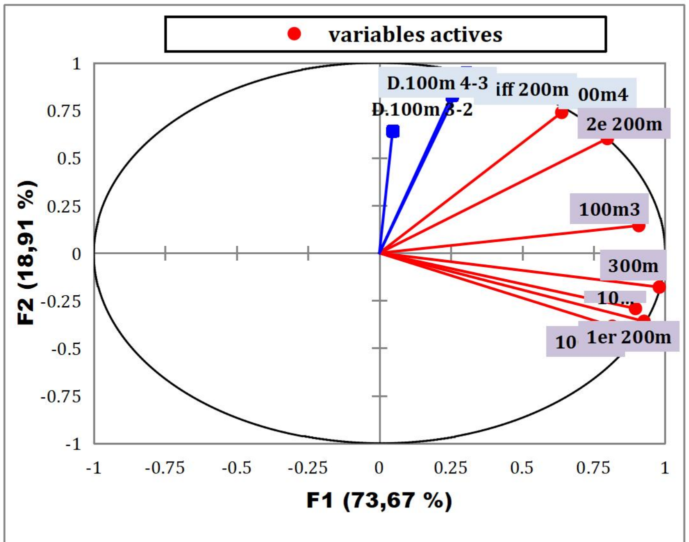



**Figure 7: Correlations circle (axes F1**

The correlation circle shows that the F1 axis is strongly correlated with the active variables. The 300m and 1st 200m variables are the most strongly correlated with the first dimension. These points correlate well with axis 1 and are the explanatory variables for this axis. Next come the 1st 200m, the 2nd 100m, and to a lesser extent the 2nd 200m and the 4th 100m. All these variables represent 3 sets of factors which explain success in the 400m for women.

The F2 axis is mainly correlated with the 100m4 and the additional variables Diff 200m and D100m4-3.

To conclude These results, highlight the importance of the 300m and 1st200m in the overall performance of female athletes, as well as the time difference between the two 200m. The additional variables, in particular Diff 200m and the difference in D.100m 4-3 segments, contribute to understanding the variations in performance between different parts of the race. This analysis of the women's 400m establishes a clear separation between the active variables, which measure direct performance, and the additional variables, which explain the differences between segments of the race.

# DISCUSSION

The present study provides a comprehensive insight into pacing strategies and performance determinants in elite male and female 400 m athletes using a multivariate statistical approach. The predominance of the 300 m split as the main contributor to overall performance confirms earlier findings indicating that this segment represents a critical transition phase between maximal speed production and fatigue resistance [1,3]. From a physiological perspective, the ability to sustain high velocity up to 300 m reflects superior anaerobic capacity, neuromuscular coordination, and efficient energy system contribution during high-intensity running [7,10].

The importance of this segment is consistent with previous descriptive and analytical studies highlighting the 300 m as a decisive indicator of success in the 400 m event [3,4]. Athletes who maintain velocity through this phase are better able to limit performance decline in the final 100 m, where fatigue-related reductions in stride length and frequency typically occur [1,7]. This finding reinforces the notion that optimal pacing in the first three race segments is a fundamental determinant of final performance.

Sex-related differences observed in the PCA structure reveal distinct performance profiles between male and female athletes. In women, performance was largely explained by a single dominant principal component, indicating a more uniform pacing pattern where global speed maintenance across the race is the primary determinant of success. Variations captured by the second component were mainly associated with pacing adjustments in the final 100 m, reflecting differences in fatigue tolerance and speed endurance capacity [5,9]. This suggests that female performance is more sensitive to late-race fatigue and pacing variability, particularly during the final acceleration–deceleration transition.

In contrast, male athletes exhibited a more complex, multidimensional performance structure, with meaningful contributions from multiple principal components. This indicates that male performance relies on a more balanced interaction between early-race speed, mid-race endurance, and late-race pacing control. Such complexity is consistent with previous reports showing smaller split-time fluctuations, higher absolute speeds, and more stable pacing profiles in male athletes [5,8]. The presence of multiple influential components suggests that elite male performance optimization requires simultaneous development of speed, endurance, and race-specific pacing regulation.

These sex-specific differences are likely influenced by physiological and biomechanical factors, including muscle mass distribution, anaerobic power, and fatigue resistance, which collectively shape pacing behavior and performance outcomes [7,10,11]. Moreover, the role of effort distribution and tactical pacing highlighted in earlier conceptual analyses of the 400 m race further supports the observed distinctions between male and female strategies [2].

From a practical standpoint, these findings underline the importance of individualized training strategies. Coaches should prioritize the development of speed endurance up to 300 m for both sexes, as this segment consistently emerges as the most influential performance determinant. For female athletes, particular emphasis should be placed on maintaining speed consistency across race segments and improving fatigue resistance in the final 100 m. Conversely, male athletes may benefit from targeted interventions aimed at optimizing transitions between segments and minimizing performance decrements associated with fatigue accumulation [8,11].

Overall, the present results demonstrate the value of Principal Component Analysis in revealing the underlying structure of 400 m performance and provide evidence-based guidance for sex-specific training and race strategy optimization.

## CONCLUSION

The results confirm the central importance of the 300 m segment in determining 400 m performance for both sexes. Female performance was largely explained by a single dominant factor, whereas male performance showed a more complex multidimensional structure. These findings support individualized training strategies focused on speed endurance and pacing control.

**REFERENCES**

[1] Duffour, S. (2009). *The influence of fatigue on running technique*. *Athletic Science Review, 12*(3), 45–52.

[2] Dlenel, P. S. (1992). *Un aspect tactique de la course du 400 m : la répartition de l'effort*. Paris, France: Institut National du Sport et de l'Éducation Physique (INSEP).

[3] Gajer, B. (2002). *400 m performance analysis*. *Track & Field Studies, 8*(2), 21–29.

[4] Gajer, B., & Reine, B. (2002). *Analyse descriptive du 400 mètres* (Research report). Paris, France: Institut National du Sport et de l'Éducation Physique (INSEP).

[5] Gajer, B., Reine, B., & Bonvin, P. (2018). *Sprint performance studies: Comparative analysis of male and female athletes*. *Athletic Journal, 24*(1), 33–41.

[6] Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

[7] Spencer, M. R., & Gastin, P. B. (2001). Energy system contribution during 200–400 m running in highly trained athletes. *Medicine & Science in Sports & Exercise, 33*(1), 157–162. https://doi.org/10.1097/00005768-200101000-00024

[8] Tucker, R., Lambert, M. I., & Noakes, T. D. (2006). An analysis of pacing strategies during men's world-record performances in track athletics. *International Journal of Sports Physiology and Performance, 1*(3), 233–245. https://doi.org/10.1123/ijspp.1.3.233

[9] Hanley, B. (2015). Pacing profiles and pack running at the IAAF World Half Marathon Championships. *Journal of Sports Sciences, 33*(11), 1189–1195. https://doi.org/10.1080/02640414.2014.988742

[10] Bundle, M. W., Hoyt, R. W., & Weyand, P. G. (2003). High-speed running performance: Assessment and prediction. *Journal of Applied Physiology, 95*(5), 1955–1962. https://doi.org/10.1152/japplphysiol.00921.2002

[11] Haugen, T., Sandbakk, Ø., Enoksen, E., & Seiler, S. (2018). Sprint conditioning of world-class athletes. *Scandinavian Journal of Medicine & Science in Sports, 28*(1), 1–10. https://doi.org/10.1111/sms.13013