Northumbria Research Link

Citation: Woolf, Anne, Ansley, Les and Bidgood, Penelope (2007) Grouping of decathlon disciplines. Journal of Quantitative Analysis in Sports, 3 (4). ISSN 1559-0410

Published by: Walter de Gruyter

URL: http://dx.doi.org/10.2202/1559-0410.1057 <http://dx.doi.org/10.2202/1559-0410.1057>

This version was downloaded from Northumbria Research Link: https://nrl.northumbria.ac.uk/id/eprint/504/

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: http://nrl.northumbria.ac.uk/policies.html

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

Northumbria University NEWCASTLE

CUSTOMER SERVICE EXCELLENCE CSE UniversityLibrary

Journal of Quantitative Analysis in Sports

Volume 3, *Issue* 4 2007 *Article* 5

Grouping of Decathlon Disciplines

Anne Woolf, *Kingston University*
Les Ansley, *Kingston University*
Penelope Bidgood, *Kingston University*

Recommended Citation:
Woolf, Anne; Ansley, Les; and Bidgood, Penelope (2007) "Grouping of Decathlon Disciplines," *Journal of Quantitative Analysis in Sports*: Vol. 3: Iss. 4, Article 5.

DOI: 10.2202/1559-0410.1057

©2007 American Statistical Association. All rights reserved.

Brought to you by | University of Northumbria (University of Northumbria)
Authenticated | 172.16.1.226
Download Date | 7/24/12 4:05 PM

Grouping of Decathlon Disciplines

Anne Woolf, Les Ansley, and Penelope Bidgood

Abstract

The 10 disciplines in the decathlon can be broadly characterised as running, jumping and throwing. However, these simplistic characteristics may not represent the groupings defined by performances in the decathlon. The identification of groups may reveal a recondite advantage for athletes who excel in particular disciplines. Therefore this study used cluster analysis to determine the groupings inherent within the decathlon disciplines. The data set was derived from the top 173 decathletes between the years 1986 to 2005. Six clustering methods were applied to a Euclidean proximity matrix. The highest number of clusters common to all the methods was accepted as the solution. All six methods produced the same 3-cluster ([100m 400m 110H LJ PV HJ][SP DT JT][1500m]), 4-cluster ([100m 400m 110H LJ PV][SP DT JT][HJ][1500m]) and 5-cluster ([100m 400m 110mH LJ][SP DT JT][PV][HJ][1500m]) solutions. Stability tests confirmed the consistency of all the solutions. The 10 disciplines of the decathlon form into five groupings, which can be adequately explained from a physiological perspective. The clustering suggests that athletes who perform better at the sprint/track disciplines may obtain an advantage in the decathlon.

KEYWORDS: athletics, cluster analysis, personal best, classical scaling

Brought to you by | University of Northumbria (University of Northumbria)
Authenticated | 172.16.1.226
Download Date | 7/24/12 4:05 PM

Introduction

The decathlon takes place over two days and consists of a combination of ten track and field disciplines the order of which is invariant: day 1: 100 m race (100m), long jump (LJ), shot put (SP), high jump (HJ) and 400 m race (400m); day 2: 110 m hurdles (110mH), discus (DT), pole vault (PV), javelin (JT) and 1500 m race (1500m). Actual performance results in terms of time in the track disciplines and distance or height in the field disciplines are converted into point scores using scoring equations and tables formulated by the International Amateur Athletics Federation (IAAF). These equations and tables are derived from ranking lists of the best performances of both decathletes and specialised athletes in the ten disciplines that comprise the decathlon. The principle of the scoring system is an attempt to reward performances of equal merit across the different disciplines with equal points.

Simplistically the disciplines that make up the decathlon could be considered as providing three broad categories for grouping the disciplines: running (100m, 110mH, 400m and 1500m); jumping (LJ, HJ, PV); and throwing (SP, JT, DT). However, these categories are unlikely to represent the groupings based upon participants' performance in the disciplines; and indeed anecdotally decathletes are frequently classified by speed and skill levels as either a sprinter/jumper or a thrower/pole vaulter (Kenny et al, 2005).

From the analysis of five sets of World Championship decathlon data (1991 to 1999) Cox and Dunn (2002) commented that the discus throw and the shot-put throw were always grouped together while the running events (with the exception of the 1500 m) were often grouped together. They identified a 3-cluster grouping of the decathlon disciplines. Overall, these groupings were not defined by the three characteristics of running, jumping and throwing although no explanation was proffered to rationalise the groupings and the composition of the clusters were not consistent between sample sub-populations. In a subsequent analysis of decathlon data from the 1999, 2001 and 2003 World Championships and the 2000 and 2004 Olympic Games we found that better consistency of grouping and cluster content was obtained with a 5-cluster solution rather than a 3-cluster solution (unpublished data).

The clustering or grouping of the disciplines within the decathlon may have practical implications in terms of the advantages gained by athletes who are relatively better in the disciplines of one cluster than those disciplines contained within another cluster. Since decathlon data can be modelled as multivariate data, cluster analysis can be used to identify and analyse the groupings (or clusters) that may exist among the disciplines.

In the analyses by both Cox and Dunn (2001) and ourselves the sample sizes were defined by the number of athletes finishing the decathlon in any year and

Brought to you by | University of Northumbria (University of Northumbria)
Authenticated | 172.16.1.226
Download Date | 7/24/12 4:05 PM

varied between 11 and 27. These sample sizes are small when used in the context of cluster analysis and major championships do not often produce personal best performances. We therefore reasoned that an improved solution might be obtained by using a larger sample based on personal best performances of elite athletes. Accordingly, we hypothesise that the cluster analysis of the top personal best decathlon performances will provide a basis for the examination of groupings among the ten decathlon disciplines.

Methods

Data set

All personal best decathlon results that achieved over 8,000 points in international competition from 1966 to 2005 were included in the original sample set (Matthews, 2005 and Decathlon2000) (permission to use the data was obtained). However, the technical specification of the men's javelin was changed in January 1986 but this change is not reflected in the IAAF scoring tables, which were last updated in 1985. As a consequence points are harder to achieve in the javelin since 1986. For this reason only the personal best results achieved from January 1986 onwards were selected for subsequent analyses. This produced a subset of 173 personal best performances and included the results from 131 meetings. Only seven of these personal best performances were achieved in Olympic competitions and only 14 achieved in World Championship competitions.

Cluster analysis

Cluster analysis is an exploratory technique that is completely numerical using only the geometric properties of the data. The variables within the cluster analysis were defined by the ten disciplines of the decathlon event. The use of point scores rather than the actual distances, heights or times provided the necessary standardisation of the data in this analysis.

Proximity measure

In order to carry out a cluster analysis the similarity, or the distance, between every pair of variables must be defined and measured and this information stored in a proximity matrix for the sample. Weighting a variable in cluster analysis would give it greater or lesser importance than other variables when determining discipline proximity. Since the underlying premise of the IAAF scoring system is

Brought to you by | University of Northumbria (University of Northumbria)
Authenticated | 172.16.1.226
Download Date | 7/24/12 4:05 PM

that all disciplines carry equal weight, and the decathletes train to the scoring tables, we have assumed equal weighting for all disciplines in the decathlon.

Cox and Dunn (2002) used the Pearson correlation as the basis of the measure of similarity between events and defined their proximity matrix in terms of the correlation coefficients. However, there is an inherent danger in using correlation to establish general behaviour since correlation does not denote causation (Kenny *et al.* 2005). Also, we have previously used both the Pearson correlation proximity matrix and the squared Euclidean proximity matrix and found that the solutions from the clustering methods based on the squared Euclidean distance measure were more consistent than those based on the Pearson correlation similarity measure (unpublished data). Therefore this study has used only the squared Euclidean measure to define the distance between the variables.

Clustering method

Everitt *et al.* (2001, p177) state that "it is generally impossible *a priori* to anticipate what … clustering methods are likely to lead to interesting and informative classifications" and moreover different clustering methods do not always produce the same results on the same data. Therefore a solution was sought based on a consensus from six of the most widely used agglomerative clustering methods as applied to the Euclidian proximity matrix. The six methods were single linkage; complete linkage; average linkage; centroid linkage; median linkage and Ward's method. The primary difference between the clustering methods is the approach they use to measure the inter-group distance between clusters of variables or between one variable and a cluster of variables.

A graphical depiction of the clustering process is illustrated by the example dendrogram in Figure 1. With each of the variables initially represented as a separate cluster on the x-axis, the dendrogram in Figure 1 depicts how the individual variables are joined to form clusters or moved into an existing cluster at the appropriate similarity level, until all the events are amalgamated into a single cluster. At level 1, A and B are joined into a cluster resulting in 5 clusters. At level 2, C is assigned to a second cluster with A and B resulting in 4 clusters. E and F are clustered together at level 3 and D is assigned to the cluster containing A, B and C in level 4, leaving only two clusters. At level 5, the final step, the two remaining clusters are amalgamated into a single cluster. The order of the variables along the x-axis is determined by the clustering process.

Brought to you by | University of Northumbria (University of Northumbria)
Authenticated | 172.16.1.226
Download Date | 7/24/12 4:05 PM

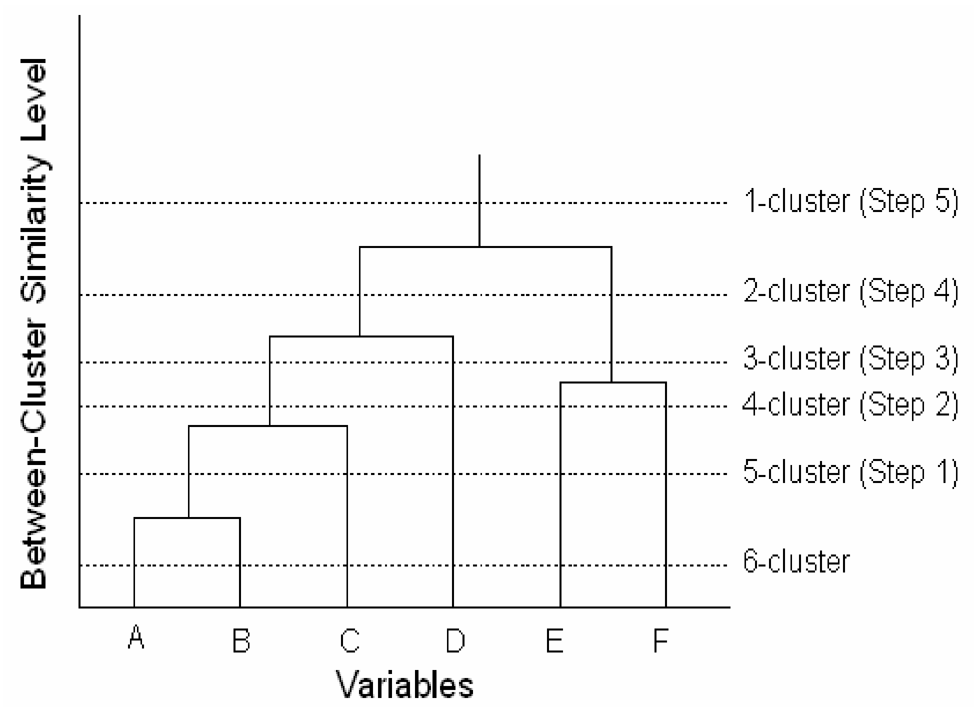

Figure 1 Depiction of the process to produce a 1-cluster dendrogram solution.

The cluster analysis presents a hierarchal clustering and it remains for the investigator to decide on the similarity level at which the number of clusters best represent the data. This decision is purely subjective since there is no methodology that can definitively determine the correct number of clusters although an examination of the classical scaling plot may help. Classical scaling is an algebraic reconstruction method for finding a configuration of points from the proximity matrix which is particularly appropriate when the distances are Euclidean.

The generation of the squared Euclidean proximity matrix and the subsequent cluster analysis using each of the six methods was carried out using Minitab version 14®. Trials runs were made using each of the six methods to produce a single-cluster solution from the data. The single-cluster solutions were examined to determine the consistency among the six clustering methods. Solutions were selected from the levels at which cluster content was common to the majority of the methods. Everitt et al. (2001, p179) caution that where several different levels are feasible, the highest number of clusters should be selected. The cluster content of the selected solutions was examined in the physiological context of decathlon disciplines.

DOI: 10.2202/1559-0410.1057

Brought to you by | University of Northumbria (University of Northumbria)
Authenticated | 172.16.1.226
Download Date | 7/24/12 4:05 PM

Cluster analysis was carried out on the decathlon data with the disciplines defining the variables and the point scores as the observations on those variables; the input data was the *173x10* Euclidean proximity matrix of discipline point scores. A classical scaling technique was applied to the data in the Euclidean proximity matrix.

Results

Consistency

Identical single-cluster solutions were produced from average linkage, centroid linkage, median linkage and Ward's method (Figure 2a); the solutions from the complete linkage (Figure 2b) and single linkage (Figure 2c) methods differed slightly. The complete linkage solution differed from the other five in assigning the LJ and the 110mH to a cluster of their own before merging that cluster with the cluster containing the 100m and 400m disciplines. All the clustering methods except the single linkage assigned the ten decathlon events between two main groups. The single linkage method retained the 1500m as a single unassigned event throughout. All six solutions assigned the first three track 'running' disciplines to a different group from the three 'throwing' disciplines and assigned the long jump to the same subgroup as the three 'running' disciplines (Figures 2a, b & c).

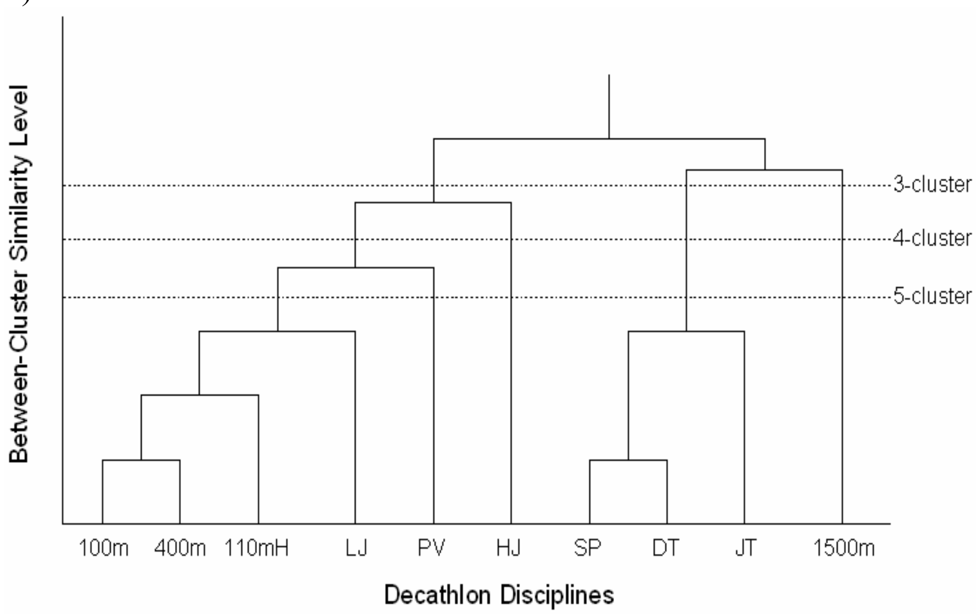

Figure 2a Representative dendrogram of Average Linkage, Centroid Linkage, Median Linkage and Ward's Method displaying the 3-, 4- and 5-cluster solutions of the decathlon disciplines for the top 173 personal best decathlon scores.

Brought to you by | University of Northumbria (University of Northumbria)
Authenticated | 172.16.1.226
Download Date | 7/24/12 4:05 PM

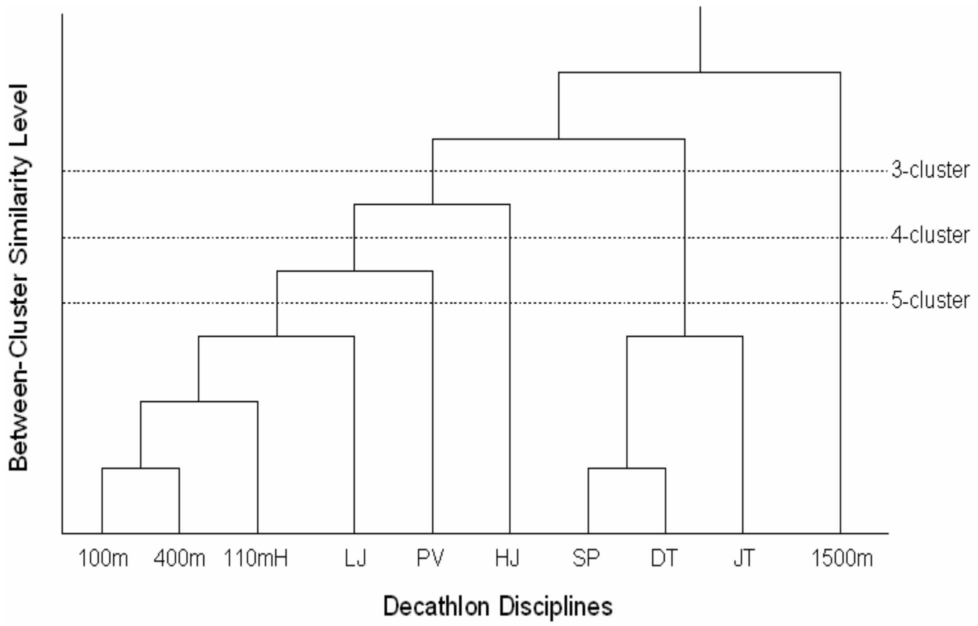

Figure 2b Dendrogram of Complete Linkage displaying the 3-, 4- and 5-cluster solutions of the decathlon disciplines for the top 173 personal best decathlon scores.

Figure 2c Dendrogram of Single Linkage displaying the 3-, 4- and 5-cluster solutions of the decathlon disciplines for the top 173 personal best decathlon scores.

Brought to you by | University of Northumbria (University of Northumbria)
Authenticated | 172.16.1.226
Download Date | 7/24/12 4:05 PM

By seeking a consensus among the six methods and bisecting the dendrograms at these levels it was observed that all six methods produced the same 3-cluster and 4-cluster and 5-cluster solutions:

[100m 400m 110mH LJ][SP DT JT][PV][HJ][1500m]	5-cluster solution
[100m 400m 110H LJ PV][SP DT JT][HJ][1500m]	4-cluster solution
[100m 400m 110H LJ PV HJ][SP DT JT][1500m]	3-cluster solution

Stability

Stability tests were carried out on all three solutions; it might have been possible to eliminate one or two of the three possible solutions if they had been found to be less stable than the others. The stability of the 3-, 4- and 5-cluster solutions was investigated with a series of stability tests carried out using the average linkage method and Ward's method. The all-time best sample was split into two subsets comprising the first 100 entries and the remaining 65 entries, respectively. A cluster solution was obtained from each subset. Both subsets reproduced the original solution with both methods.

Further stability tests were carried out by removing the point score data appropriate to the events in any of one, two or three of the clusters at random from the input data and generating a 4, 3 or 2-cluster solution from the reduced proximity matrix. All these tests produced 'reduced' solutions consistent with the original.

In a final test of stability the Euclidean proximity matrix was regenerated with the point score data rearranged as if the PV, JT and 1500m were the first three events of the decathlon. The solutions from the altered data were consistent with the original solutions indicating that the 3-, 4- and 5-cluster solutions were all stable. The solution for the average method from this test is shown in Figure 3. It was observed that the structure of the dendrogram, unlike the content of the clusters, is dependent on the sequence in which the data is stored in the Euclidean proximity matrix (Figure 2a vs. Figure 3).

Brought to you by | University of Northumbria (University of Northumbria)
Authenticated | 172.16.1.226
Download Date | 7/24/12 4:05 PM

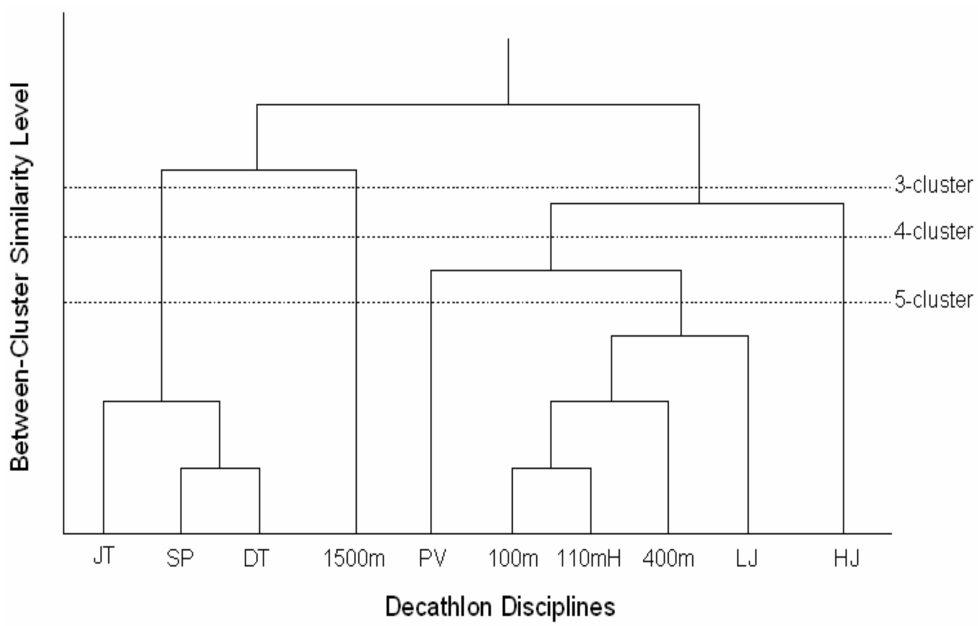

Figure 3 Representative dendrogram of Average Linkage and Ward's Method stability tests of the decathlon disciplines for the top 173 personal best decathlon scores where the data was restructured to present the PV, JT and 1500m as the first three disciplines.

Classical scaling

The classical scaling plot for the decathlon data sample depicts the grouping of the decathlon disciplines into two main clusters on either side of the plot and confirms the possible option of 3, 4 or 5 clusters (Figure 4).

Brought to you by | University of Northumbria (University of Northumbria)
Authenticated | 172.16.1.226
Download Date | 7/24/12 4:05 PM

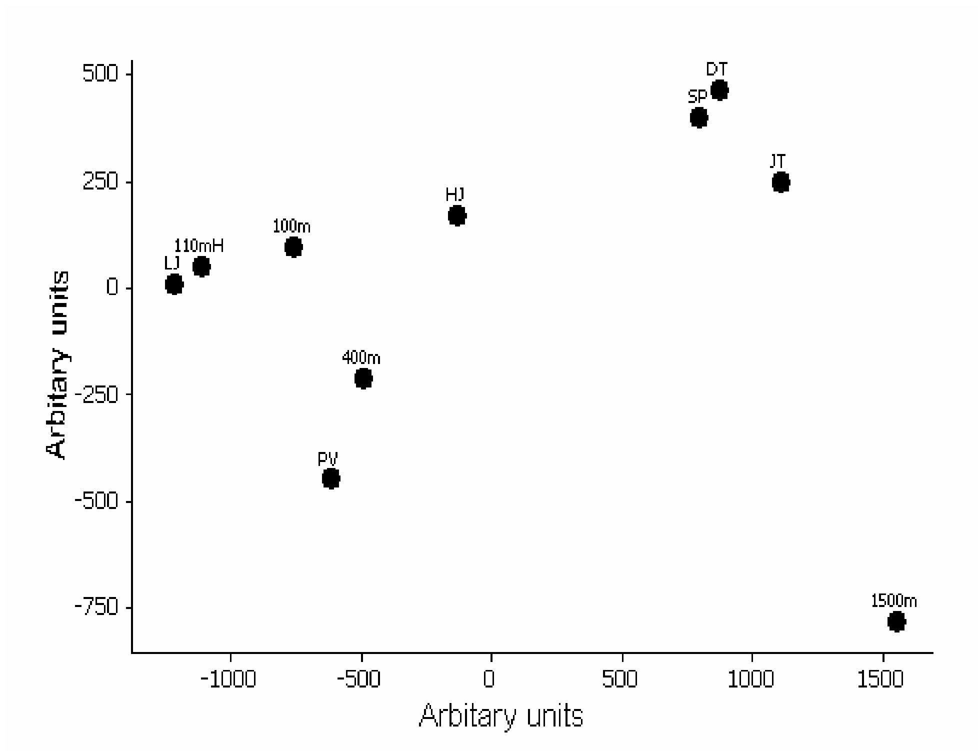

Figure 4 Classical scaling plot of the decathlon disciplines for the top 173 personal best decathlon scores.

Discussion

The cluster analysis of the all time best decathlon results presented three solutions that were all equally stable. The fundamental difference between the three solutions was whether the pole vault and the high jump were clustered with other disciplines or left in clusters of their own - as represented by the 5-cluster solution. The 5-cluster solution was chosen as the most plausible based on the principle that since it presents the highest number of clusters common to all the methods it is the solution that is least likely to force disciplines into contrived clusters (Everitt, 2001). The 5-cluster solution derived from the large data set in this study has the same cluster content as a 5-cluster solution displayed in the dendrogram presented by Cox and Dunn (2002) where the combined data from five World Championship competitions was used.

Cox and Dunn (2002) used single linkage, complete linkage and Ward's cluster methods in their analysis. Based on the smaller samples from the five individual competitions they found that complete linkage and Ward's method

Brought to you by | University of Northumbria (University of Northumbria)
Authenticated | 172.16.1.226
Download Date | 7/24/12 4:05 PM

gave similar, consistent results. Cox and Dunn (2002) also concluded that "[t]he high jump, pole vault, javelin and 1500 metre race were not so consistent with cluster membership" (p.181). Based on our previous work with smaller samples we would mostly agree with their conclusion; however, it is notable that the larger sample size from the all time best scores has produced a consistent 5-cluster solution in which the javelin is always assigned to the same cluster as the discus and the shot put.

Assuming a case can be made for the existence of groupings among the decathlon disciplines does the 5-cluster solution make physiological sense for the decathlon? Closer examination of the clustering of the disciplines provides some interesting insights. Analysis of the disciplines grouped in the first cluster i.e. [100m, 400m, 110mH, LJ] reveals that success in all four disciplines is dependent on fast running speeds on the flat (Hay 1994). Furthermore, the cross-over between success in sprints and long jump is further borne out in the athletic area where no less than three sprinters have won Olympic medals in both the 100 m and the long jump, namely Marion Jones, Carl Lewis and Jesse Owen.

The separation the 1500m from the other three running disciplines in the first cluster supports anecdotal evidence that the 1500m is "different" in the decathlon. There are a number of possible explanations for this observation. The 1500 m race is the only middle distance event in the decathlon event; all the other events are sprint and power events. Therefore the type of athlete that competes in the decathlon is probably not suited to the rigors of the 1500 m race. Furthermore, although the 1500m was initially included in the decathlon as a test of stamina, it is the final event on the second day by which time the athletes will be tired and the overall positions have usually been decided. Therefore in most instances, the athletes have little motivation to do more than just maintain their position.

The three disciplines grouped in the second cluster are the three throwing disciplines (SP, DT and JT). Although the actions of the SP, DT and JT can be described as a thrust, whip and pull, respectively (Cromwell, 1949 p.301), the principles that govern the throwing techniques are the same i.e. speed, angle of release and (in the case of the DT and JT) aerodynamics (Dyson, 1971 p.184). Furthermore the proximal-to-distal sequential order of the kinetic chain is common to the SP, DT and JT (Hay, 1994). That mastery of the fundamentals in one throwing discipline can overlap into mastery of other throwing disciplines is evident in the individual competitive arena where athletes have achieved success in two or more of the throwing events e.g. Hannes Hoply (SP and DT), Ralph Rose (SP, DT and JT), John Godina (SP and DT), Uwe Hohn (DT and JT) and Scott Russell (DT and JT).

Notably, the two events whose success is dependent on defying gravity are grouped in separate clusters. The extremely technical nature of the pole vault (Hubbard, 1980) and high jump (Dapena and Chung, 1988) possibly result in the

DOI: 10.2202/1559-0410.1057

Brought to you by | University of Northumbria (University of Northumbria)
Authenticated | 172.16.1.226
Download Date | 7/24/12 4:05 PM

scenario where the investment in the extra time in mastering the technique at the expense of time spent on other disciplines does not provide adequate returns in performance (points) improvements. Van Damme et al (2002) described this trade-off in terms of the evolutionary principle of allocation where "excellence in one task can only be attained at the expense of average performance in all other tasks" (p.755). Therefore the athletes would rather spend their time concentrating on improving performances in their "preferred" disciplines.

Cox and Dunn (2002) posited that the decathlon favours those athletes who do well in the field events. However, the composition of the clusters suggests that athletes who perform better in the sprint/track disciplines may be afforded an advantage as there are four disciplines in which they are likely to do well; whereas only three of the field disciplines are contained in the next largest cluster. This assumption is further strengthened by the observation that the average ranking of the top ten decathletes in the disciplines that comprise the first cluster [100m 400m 110mH LJ] is 11th, while their average ranking in the disciplines that comprise the second cluster [SP DT JT] is 27th. This trend continues for the top twenty positions where the average ranking for the [100m 400m 110mH LJ] cluster and [SP DT JT] cluster disciplines is 18th and 44th, respectively. This indicates that the athletes who perform better in the disciplines that comprise the first cluster are overrepresented in the top twenty personal best performances compared with those athletes who excel at the disciplines contained in the second cluster (Figure 5).

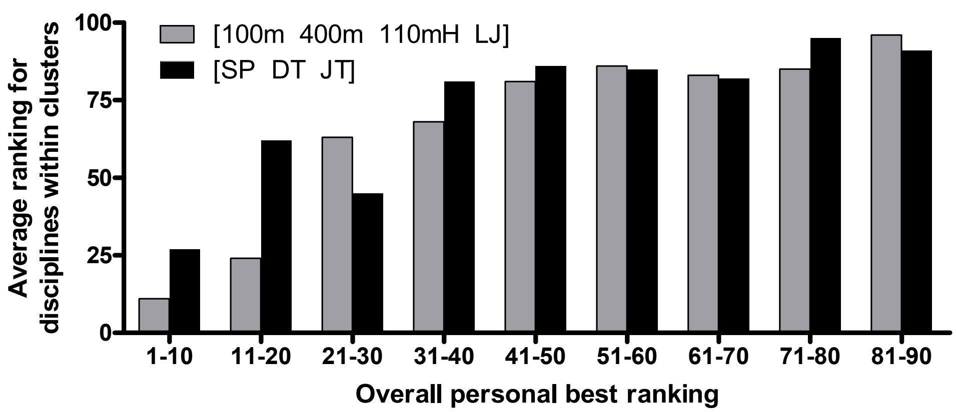

Figure 5 Comparison of the average ranking of disciplines contained within two clusters – [100m 400m 110mH LJ] and [SP DT JT] – in relation to the overall ranking of decathletes in the top 90.

Brought to you by | University of Northumbria (University of Northumbria)
Authenticated | 172.16.1.226
Download Date | 7/24/12 4:05 PM

Conclusion

Although defining the number of clusters remains subjective the results from this study support a 5-cluster solution. The clusters are not entirely defined by the three overt characteristics of the disciplines, but can be adequately explained from a teleological perspective. Furthermore, it seems that sprint track athletes may have an advantage in the decathlon event since they are more likely to do well in a greater number of events than field athletes.

References

Cox, T.F. & Dunn, R.T. (2002). An analysis of decathlon data. *Journal of the Royal Statistical Society Series D*, *51*, 179-187.

Cromwell, D.B. (1949). *Champion Technique in Track and Field* (2nd ed). New York, Whittlesey House.

Dapena, J. & Chung, C.S. (1988). Vertical and radial motions of the body during the take-off phase of high jumping. *Medicine and Science in Sports and Exercise*, *20*, 290-302.

Decathlon 2000. http// www.decathlon2000.ee/eng/statistics.php?art=834 (Last accessed 24/10/2006)

Dyson, G. (1971). *The Mechanics of Athletics* (5th ed). London, University of London Press.

Everitt, B., Landau, S. & Leese, M. (2001). *Cluster Analysis* (4th ed). London, Arnold.

Hay, J.G. (1994). *The Biomechanics of Sports Techniques* (4th ed). New Jersey, Prentice-Hall.

Hubbard, M. (1980). Dynamics of the pole vault. *Journal of Biomechanics*, *13*, 965-976.

Kenny, I.C., Sprevak, D., Sharp, C. & Boreham, C. *(*2005) *Journal of Quantitative Analysis in Sports*, *1*, Article 5.

DOI: 10.2202/1559-0410.1057

Brought to you by | University of Northumbria (University of Northumbria)
Authenticated | 172.16.1.226
Download Date | 7/24/12 4:05 PM

Matthews, P. (2005). *International Athletics Annual 1995*. Cheltenham: SportsBooks.

Van Damme, R., Wilson, R. S., Vanhooydonck B. & Aerts, P. (2002). Evolutionary biology: Performance constraints in decathletes. *Nature*, *415*, 755.

Brought to you by | University of Northumbria (University of Northumbria)
Authenticated | 172.16.1.226
Download Date | 7/24/12 4:05 PM