Visualising the complexity of the athlete monitoring cycle through principal component analysis

Dan Weaving[1,2], Clive Beggs[1], Nicholas Dalton-Barron[1,3,4], Ben Jones[1,2,4,5] and Grant Abt[6]

[1]Institute for Sport, Physical Activity and Leisure, Leeds Beckett University, Leeds, West Yorkshire, United Kingdom.
[2]Leeds Rhinos Rugby League Club, Leeds, United Kingdom.
[3]Catapult Sports, Leeds, United Kingdom.
[4]The Rugby Football League, Leeds, United Kingdom.
[5]Yorkshire Carnegie Rugby Union Club, Leeds, United Kingdom.
[6]Department of Sport, Health and Exercise Science, The University of Hull, Hull, United Kingdom.

Abstract

Purpose: The purpose of this invited commentary is to discuss the use of principal component analysis (PCA) as a dimension reduction and visualisation tool to assist in decision making and communication when analysing complex multivariate data sets associated with the training of athletes.

Conclusions: Using PCA it is possible to transform a data matrix into a set of orthogonal composite variables called principal components (PC), with each PC being a linear weighted combination of the observed variables and with all PCs uncorrelated to each other. The benefit of transforming the data using PCA is that the first few PCs generally capture the majority of the information (i.e. variance) contained in the observed data, with the first PC accounting for the highest amount of variance and each subsequent PC capturing less of the total information. Consequently, through PCA it is possible to visualise complex data sets, containing multiple variables on simple 2D scatterplots without any great loss of information, thereby making it much easier to convey complex information to coaches. In the future, athlete monitoring companies should integrate PCA into their client packages to better support practitioners trying to overcome the challenges associated with multivariate data analysis and interpretation. In the interim, we present here an overview of PCA and associated R code to assist practitioners working within the field to integrate PCA into their athlete monitoring process.

Data overload: the complexity of the athlete monitoring cycle

An important role of the coach is to prescribe an optimum training load (or dose) at the right time to enhance important responses such as physical-[1,2] and technical-tactical-performances[3] or player availability.[4] Monitoring these 'dose-response' relationships is considered an important part of decision making in sports performance.[5-7] To assist in this process, sports scientists are often tasked with collecting, processing, analysing and communicating fitness and performance data, with the aim of providing evidence-based and actionable recommendations to coaches.

In reality, for sports scientists, providing a valid representation of 'dose' and 'response' over a training period is something that is inherently difficult. This is because multiple psychological, physiological and biomechanical pathways are affected by training and competition schedules.[8-10] Therefore, the 'dose' applied, and the resulting 'response' to each pathway might vary considerably between players and across the different modes of training that are prescribed in concurrent programmes.[10-11] With continued developments in technology and the ease with which data can be collected, the sport scientist is now faced with a complex array of multivariate data, from multiple sources and in multiple units of measurement. This results in considerable ambiguity about how 'dose' and 'response' should be assessed and communicated to coaches. However, in the age of technology[12], many organisations have invested significant resources for measuring these complex constructs of 'dose' and 'response'.

The challenge of understanding, interpreting, and acting on these data within the fast-paced environment of daily training prescription has been discussed by many researchers.[12-16] Such discussions typically focus on the need for practitioners to consider: 1) how they manage large volumes of data (i.e. collecting and processing) 2) how they interpret these collective data sources (i.e. analysing) and 3) how they translate these interpretations to inform training prescription and assist stakeholder decision-making (i.e. visualisation and communication).
To mitigate some of these challenges, decision or heuristic matrices (Figure 1) have been proposed.[13,17] For example, such matrices might include plotting 'fitness' on the *x*-axis and 'fatigue' on the *y*-axis. Decision boundaries can then be created by practitioners to guide decision-making relating to when amendments (e.g. progression or regression) to training prescription should be made, depending on how 'dose' and 'response change over time.

INSERT FIGURE ONE HERE

To analyse the change in the training 'dose' and 'response' over time, previous authors[3,16] have proposed approaches such as z-scores or the *smallest worthwhile change* statistic to determine whether each individual 'dose' and 'response' variable for each athlete is deviating away from 'normal'.

Although this type of analysis can be an important tool for practitioners when determining the likelihood of a meaningful change over time, this approach suffers from the major limitation that each individual variable generates its own individual change statistic. Therefore, given the complexity of athlete monitoring, if practitioners

collect three 'dose' variables (e.g. total-distance, high-speed-distance, heart rate) and three 'response' variables (e.g. heart rate variability, perceived recovery, countermovement jump), six change statistics are generated for each athlete or 180 change statistics based on a 30-player squad, per time-point of data collection. As a result, practitioners face a major challenge of finding an efficient way to communicate these multiple change statistics to coaches without blunting the true 'signal' of each player's 'dose' and 'response'.

Multicollinearity in the data collected can also be a major problem, as multiple measurements that represent either the 'dose' and 'response' are often strongly correlated to each other. For example, with global positioning system signals, velocity and acceleration are recorded as separate variables, with the result that coaches often think that these measures are independent of each other. This however is not the case, since velocity ($m \cdot s^{-1}$) and acceleration ($m \cdot s^{-2}$) are simply the first and second differentials of distance. Consequently, the two signals are highly correlated and are not independent. Likewise, across a period of training, it is often the case that several variables used to quantify the same latent construct (i.e. quantifying 'dose' or 'response') might change positively or negatively by similar amounts thus exhibiting considerable covariance. As such, these correlated variables are to a greater or lesser extent all measuring the same thing, resulting in considerable redundancy within the data. Indeed, many commonly used measures of training load (e.g. GPS, session-RPE, etc.) exhibit inherent multicollinearity when measured across periods of training.[18-20] However, depending on the player[18] and type of training[11,19-20], other variables might change in dissimilar ways, demonstrating a degree of independence. For example, in professional rugby league, Lovell et al. (2013)[21] reported that the relationship between the session-rating-of-perceived-exertion and total distance was r = 0.80 during conditioning training, but only r = 0.37 during wrestle-based training.

Given this, sport scientists need to be able to concurrently filter out redundancy in the data that often obscures important information, but also identify unique information, so that these can be interpreted and ultimately, actionable insights communicated to coaches. In this regard, the use of dimension reduction techniques such as principal component analysis (PCA)[11,19] and single value decomposition (SVD)[20] are gaining popularity within sports performance research. For example, PCA and SVD have been used in studies examining talent identification[22], to assess the evolution of game-play[23], to develop performance indicators[24] and assess technique in athletes.[25] By orthogonalising the data, these techniques enable complex higher-dimensional systems to be represented on 2D or 3D scatter plots with minimal loss of information.[22] Unfortunately, these techniques are not commonly used within applied sport performance settings, often due to a lack of integration with the software commonly used by practitioners. As such, much of the analysis undertaken is still univariate in nature, with little account taken of covariance in the data (i.e. multicollinearity). Given this, an alternative PCA-based method for capturing the information contained in multiple variables could be advantageous, thus improving efficacy of data visualisation and communication in applied practice.

The extent to which PCA can actually be used to reduce the dimensionality in a given data set will of course depend on the degree of multicollinearity or redundancy

within the data. PCA performs better when there is considerable redundancy in the data, whereas little benefit is derived if the variables are weakly correlated or uncorrelated with each other. The extent to which redundancy is present can be evaluated using Bartletts test of sphericity, which tests the null hypothesis that the correlation matrix of the data is an identity matrix, indicating that the measured variables are uncorrelated (orthogonal), and therefore making them unsuitable for PCA. With Bartletts test, $p<0.05$ indicates that the data are not orthogonal and therefore suitable for the application of PCA.[26] A related test, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy, can also be applied when performing PCA, although strictly speaking it is a measure of how well suited the data is for factor analysis (FA). The KMO index ranges from 0 to 1, with >0.50 considered suitable for FA.[27] If this requirement is not met, it means that distinct and reliable factors (unobserved latent variables that explain the observed variables) cannot be produced.[28] When performing FA, PCA is often used to determine the number of factors that should be extracted and retained in the analysis.[29] In this context, PCA performs well when the KMO index is high (>0.50) and is more likely to identify reliable factors that explain the observed data when this threshold is exceeded.

Use of PCA to overcome data overload

Using PCA it is possible to transform a data matrix (i.e. columns of data [variables] by rows of data [observations]; Figure 2) into a set of orthogonal composite variables called principal components (PCs) (see supplementary information for a full explanation of PCA), with the following attributes:

- Each PC is a linear weighted combination of the observed variables and is derived from the eigenvectors produced when eigen-decomposition is performed on the covariance matrix of the data.
- The total number of PCs always equals the number of the observed variables and the PCs themselves are uncorrelated.
- The benefit of transforming the data using PCA is that the first few PCs generally capture the majority of the information (i.e. variance) contained in the observed data, with the first PC accounting for the highest amount of variance and each subsequent PC capturing less of the total information.

For example, Weaving et al.[19] reported the 1st PC captured 70% of the total information provided by five training load measures during small-sided-games training in professional rugby league players. However, the 2nd PC explained only 12% of the total information, with the 5th PC explaining merely 2%. This suggests that much of the complexity contained in the multiple observed training load variables can be captured in one or two PCs without losing much information. Consequently, through PCA it is possible to visualise complex data sets, containing multiple variables, on simple 2D scatterplots (e.g. 'fitness' *vs* 'fatigue'; Figure 1), without any great loss of information – making it much easier to convey complex information to coaches.

An example of the use of PCA to assess the similar and unique information provided by multiple fitness (chronic load) and fatigue (acute load) variables

Here we illustrate how PCA can be applied to training load data collected from a single professional rugby league player over two European Super League seasons. In this example, we consider the following external training load variables:

- Total distance ([TD]: m)
- High-speed-distance ([HSD]: m; 5 to 7 m·s^{-1})
- Very-high-speed-distance ([VHSD]: m; > 7 m·s^{-1})
- PlayerLoad™ ([PL]: AU)
- High metabolic power distance ([HMP]: m; > 20 W·kg^{-1})
- Number of collisions ([COL]: n)

For each of the variables above, a 7 and 28 day exponentially weighted moving average (EWMA)[30] of their daily accumulation was calculated:

$$EWMA_{today} = Load_{today} \times \lambda_a + ((1 - \lambda_a) \times EWMA_{yesterday}) \qquad (1)$$

Where λ_a is a value between 0 and 1 that represents the degree of decay, with higher values discounting older observations at a faster rate. The λ_a is calculated as:

$$\lambda_a = 2/(N + 1) \qquad (2)$$

Where N is the chosen time decay constant of either 7 or 28 days. In this example, this generated 12 EWMA training load variables across 169 training sessions.

In order to perform PCA we first examined the data for suitability, both with Bartlett's test of sphericity and the KMO measure of sampling adequacy, as described earlier. The outcome of Bartlett's test was significant ($P < 0.001$) and all training load variables had a KMO value above 0.5, suggesting that the data were suitable for PCA. The observed data were first zero mean-centred and standardized to unit variance and then stored in an n by m matrix, X, which contains m EWMA training load variables and n training session observations (Figure 2). Standardization was performed primarily to account for the different scales of the variables (e.g. number of collisions = 10 vs total distance = 3000 m).

INSERT FIGURE TWO HERE

Having compiled the matrix of standardized data, X, eigen-decomposition of the covariance matrix of the data was then performed (see supplementary information) to compute the eigenvalues and eigenvectors associated with the respective PCs (Table 1). The respective eigenvalues are key to determining the amount of information captured (i.e. variance) by each PC. From Table 1, it can be seen that in this example, the first and second PCs captured 81.4% of the total variance in the data. The respective eigenvectors associated with the eigenvalues are important in determining how the original observed variables 'relate' to each other.

INSERT TABLE ONE HERE

INSERT FIGURE THREE HERE

The PCA biplot (Figure 3) shows the scaled eigenvectors arrows (derived from Table 1) for each observed variable. From this it can be seen that the variables: Tackles (COL_7, COL_28), PlayerLoad (PL_7, PL_28), Total distance (TD_7, TD_28), and High metabolic power (HMP_7 and HMP_28) all strongly influence the first PC which accounts for 58.3% of the total variance, whereas the variables: Very-high-speed distance (VHSD_7, VHSD_28), and High-speed distance (HSD_7, HSD_28) relate more to the second PC (23.1% of the variance). As such, the first PC for this player can be broadly described as representing the 'overall load', whereas the second PC represents the 'high-speed load'. Furthermore, it can be seen from the biplot and Table 1, that many of the observed variables are actually measuring very similar phenomena (e.g. VHSD_7 and VHSD_28; and PL_7, TD_7 and HMP_28), indicating that these variables are not independent. By contrast however, variables that might be considered closely related, such as TD_7 and TD_28, point in different directions in the eigenspace.

By using PCA, practitioners can first explore the interrelationships between multiple variables, which are likely to be different for each player due to their unique characteristics. In practice, this could be used as a systematic process to evaluate which variables provide similar or unique information (via the eigenvectors) and how much of the total variance they are likely to capture (via the eigenvalues) across a training period.[10-11,19,22]

Visualising multivariate data using PCA

Through the use of the PC 'scores', PCA can also assist with the visualisation of multiple variables. The term 'score' is often used to denote the elements of each PC that relate to the individual observations. Therefore, in the context of our example, the PC 'scores' represent the combined linear weighted contribution of the original training load variables (i.e. the eigenvectors [Table 1]) of each PC for each of the individual training sessions.

To calculate the PC 'scores', the matrix of eigenvectors and the matrix of the standardised training load data are multiplied together. This applies the coefficients in the respective eigenvectors to the standardised data. In our example, the scores for the 1st PC are generated by applying the coefficients in the 1st eigenvector (corresponding to the largest eigenvalue in Table 1) to the standardised data as follows:

$$PC1 = (-0.36 \times TD\_7) + (-0.34 \times TD\_28) + (-0.19 \times HSD\_7) + (-0.26 \times HSD\_28) + (-0.36 \times PL\_7) + (-0.34 \times PL\_28) + (-0.01 \times VHSD\_7) + (0.00 \times VHSD\_28) + (-0.33 \times HMP\_7) + (-0.35 \times HMP\_28) + (-0.29 \times COL\_7) + (-0.31 \times COL\_28)$$

$$(3)$$

PCA results in the creation of an *n* by *m* matrix containing the PC 'scores', each of which relates to the respective eigenvalues.

By producing a scatterplot of the first and second PCs, it is possible to visually represent in 2D most of the information (81.4% for our example) contained in the 7 and 28 day EWMA training load data (see Figures 3 and 4). For practitioners, the PCs can then be visualised in a number of ways, such as a scatterplot (Figure 4), or alternatively across two time-series line graphs (Figure 5), the latter having the advantage that it allows the change in the respective PCs over time to be visualised.

In our example, we interpreted the first PC as the 'overall load' and the second PC as 'high-speed load'. However, because these composite variables are in reality mathematical constructs arising from eigen-decomposition of the data, it is important to recognise that individual PCs can only ever be approximations to any given ascribed attribute (i.e. 'overall' or 'high-speed' load). For example, while 'HSD 7' and 'HSD 28' are more dominant in PC2 they also make a contribution to PC1, highlighting the need to be cautious when interpreting PC scores. Regardless of this, it can be seen that the first two PCs capture the contributions of multiple aspects of the training 'dose' (e.g. total distance, high-speed-distance, collisions) and enable this to be communicated effectively to coaches and players. As such, the approach can be replicated with other variables representing fatigue (e.g. countermovement jump height, reactive strength index, heart rate variability) or multiple EWMA time periods (e.g. 2 to 30 days) to assist visualisation of the overall athlete monitoring cycle (i.e. dose and response).

In addition to allowing complex information about multiple training events to be effectively communicated, other information (e.g. injury occurrence, etc) can be overlaid on any time-series or scatter-plots produced, in order to provide a richer visualisation of the athlete monitoring cycle. For example, in Figure 4, by highlighting today's session (in green), practitioners can maximise the use of historical data by placing an athletes training load on that day within the context of every other individual training session that the athlete has completed over a longitudinal training period. Therefore, in this example, we can descriptively infer that this athlete's training load is 'much higher than normal'.

Although PCA is primarily a dimension reduction technique that is applied retrospectively to complex higher-dimensional data sets in order to see 'the wood for the trees', it is often used in a machine learning context to construct classification and regression models, which can be used prospectively to diagnose problems, make strategic decisions, and predict performance outcomes. For example, both PCA and SVD can be used in algorithms relating to performance thresholds.[22,24] Similarly, principal component regression[32], a closely related technique to PCA, can be used to build predictive regression models when multicollinearity is a major problem. As such, dimension reduction techniques such as PCA, SVD and multidimensional scaling (MDS) have considerable potential in sport performance analysis. Indeed, MDS has recently been used to evaluate the evolution of game-play in the Australian Football League.[23] Of these techniques, PCA is by far the most accessible, with many relevant articles and tutorials readily accessible on the Internet, making it an obvious choice for those wishing to explore dimension reduction when analyzing sport performance data. Furthermore, it is applicable to most sport related applications, provided of course that the data contains redundancy.

INSERT FIGURE FOUR HERE

INSERT FIGURE FIVE HERE

Conclusions and Practical Applications

Improving methods for visualising complex information from multiple sources is key to improving the efficacy of data-driven decision making in applied practice. In this invited commentary a practical solution for overcoming difficulties in visualising complex athlete monitoring datasets has been proposed through the use of PCA. Although the mathematics underpinning PCA can be complex, the method can be easily executed using open-source software packages, such as 'R' (see supplementary information for example generic PCA 'R' code), making it accessible to practitioners. In the future, athlete monitoring companies should integrate PCA into their client packages to better support practitioners trying to overcome the challenges associated with multivariate data analysis and interpretation. In the interim, an overview of PCA and associated R code to assist practitioners working within the field to integrate PCA into their athlete monitoring process has been provided.

References

1. Akubat I, Patel E, Barrett S, Abt G. Methods of monitoring the training and match load and their relationship to changes in fitness in professional youth soccer players. J Sports Sci. 2012;30(14):1473-1480.
2. Taylor RJ, Sanders D, Myers T, Abt G, Taylor CA, Akubat I. The dose-response relationship between training load and aerobic fitness in academy rugby union players. Int J Sports Physiol Perform. 2018;13:1-7.
3. Graham SR, Cormack S, Parfitt G, Eston R. Relationships between model predicted and actual match performance in professional Australian footballers during an in-season training macrocycle. Int J Sports Physiol Perform. 2017;17:1-25
4. Williams S, Trewartha G, Kemp SP, Brooks JH, Fuller CW, Taylor AE, Cross MJ, Stokes KA. Time loss injuries compromise team success in Elite Rugby Union: a 7-year prospective study. Br J Sports Med. 2016;50(11):651-656.
5. Impellizzeri FM, Rampinini E, Marcora SM. Physiological assessment of aerobic training in soccer. J Sports Sci. 2005;23(6):583-592.
6. Gabbett TJ. The training-injury prevention paradox: should athletes be training smarter and harder. Br J Sports Med. 2016;50(5):273-280.
7. Drew MK, Finch CF. The relationship between training load and injury, illness and soreness: a systematic and literature review. Sports Med. 2016;46(6):861-863.
8. Soligard T, Schwellnus M, Alonso JM, et al. How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. Br J Sports Med. 2016;50(17):1030-1041.
9. Vanrenterghem J, Nedergaard NJ, Robinson MA, Drust B. Training load monitoring in team sports: a novel framework separating physiological and biomechanical load adaptation pathways. Sports Med. 2017;47(11):2135-2142.

10. Weaving D, Jones B, Till K, Abt G, Beggs C. The case for adopting a multivariate approach to optimise training load quantification in team sports. Front Physiol. 2017;8:1024.
11. Weaving D, Jones B, Marshall P, Till K, Abt G. Multiple measures are needed to quantify training loads in professional rugby league. Int J Sports Med. 2017;38(10):735-740.
12. Coutts AJ. In the age of technology, Occam's razor still applies. Int J Sports Physiol Perform. 2014;9(5):741.
13. Gabbett TJ, Nassis GP, Oetter E, Pretorius J, Johnston N, Medina D, Rodas G, Mylinski T, Howells D, Beard A, Ryan A. The athlete monitoring cycle: a practical guide to interpreting and applying training monitoring data. Br J Sports Med. 2017;51:1451-1452.
14. Cardinale M, Varley MC. Wearable training-monitoring technology: applications, challenges and opportunities. Int J Sports Physiol Perform. 2017;12: S255-S262.
15. Robertson S, Bartlett JD, Gastin PB. Red, amber or green? Athlete monitoring in team sport: the need for decision support systems. Int J Sports Physiol Perf. 2017;12(Suppl 404 2):S273-S279.
16. Ward P, Coutts AJ, Pruna R, McCall A. Putting the "I" back in the team. Int J Sports Physiol Perform. 2018;13(8):1107-1111.
17. Jovanovic M. (2017). Uncertainty, Heuristics and Injury Prediction. Aspetar Journal 2017;6.
18. Thornton HR, Delaney JA, Duthie GM and Dascombe BJ. Importance of various training load measures on injury incidence of professional rugby league athletes. Int J Sports Physiol Perf. 2016;5:1-17.
19. Weaving D, Marshall P, Earle K, Nevill A, Abt G. Combining internal- and external training-load measures in professional rugby league. Int J Sports Physiol Perf. 2014;9:905–912.
20. McLaren SJ, Macpherson TW, Coutts AJ, Hurst C, Spears IR, Weston M. The relationships between internal and external measures of training load and intensity in team sports: a meta-analysis. Sports Med. 2018;48(3):641-658.
21. Lovell TW, Sirotic AC, Impellizzeri FM, Coutts AJ. Factors affecting perception of effort (session rating of perceived exertion) during rugby league training. Int J Sports Physiol Perform. 2013;8(1):62-69.
22. Till K, Jones BL, Cobley S, Morley D, O'Hara J, Chapman C, Cooke C, Beggs CB. Identifying talent in youth sport: a novel methodology using higher-dimensional analysis. PLoS One. 2016;11:e0155047.
23. Woods CT, Robertson S, Collier NF. Evolution of game-play in the Australian Football League from 2001 to 2015. J Sports Sci. 2017;35(19):1879-1887.
24. Parmar N, James N, Hearne G, Jones B. Using principal component analysis to develop performance indicators in professional rugby league. J Sports Sci. 2018;6:938-949.
25. Gløersen Ø, Myklebust H, Hallén J, Federolf P. Technique analysis in elite athletes using principal component analysis. J Sports Sci. 2018;36(2):229-237.bar
26. Bartlett MS. Tests of significance in factor analysis. British Journal of statistical psychology. 1950 Jun;3(2):77-85.
27. Williams B, Onsman A, Brown T. Exploratory factor analysis: A five-step guide for novices. Australasian Journal of Paramedicine. 2010 Aug 2;8(3).

28. Yong AG, Pearce S. A beginner's guide to factor analysis: Focusing on exploratory factor analysis. Tutorials in quantitative methods for psychology. 2013 Oct;9(2):79-94.29.Costello AB, Osborne JW. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Pract Assess Res Eval 2005; 10. URL http://pareonline. net/getvn. asp. 2005;10(7)

30. Williams S, West S, Cross MJ, Stokes KA. Better way to determine the acute:chronic workload ratio? Br J Sports Med. 2017;51:209-210.

31. Williams S, Trewartha G, Cross MJ, Kemp SPT, Stokes KA. Monitoring what matters: a systematic process for selecting training-load measures. Int J Sports Physiol Perfom. 2017;12:S2101-S2016.

32. James G, Witten D, Hastie T & Tibshirani R. An introduction to statistical learning with applications in R. Springer texts in statistics. Springer, New York; 2013.

Figure and Table Descriptions

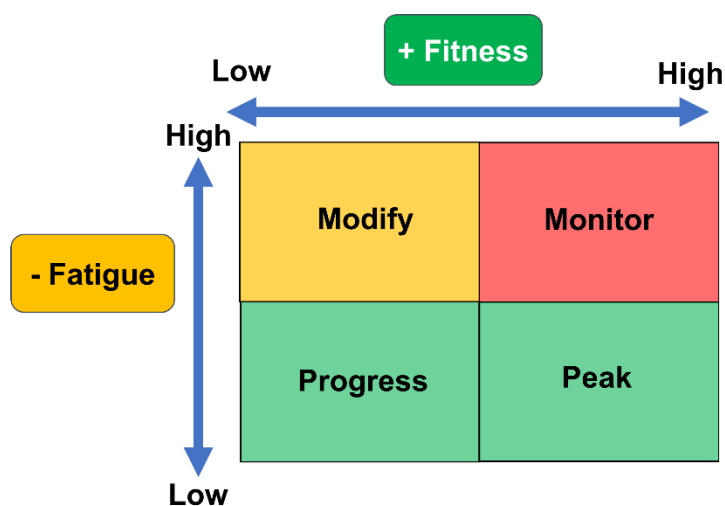Figure 1. Example athlete monitoring heuristic decision matrix.

Figure 2. Example of a typical training load matrix for a single player, showing the 7 and 28 day exponentially weighted moving averages of six training load variables (*m*) calculated across different training days across a training period (*n*).

$$m$$

Date	TD 7day	HSD 7day	VHSD 7day	COL 7day	HMP 7day	PL 7day	TD 28day	HSD 28day	VHSD 28day	COL 28day	HMP 28day	PL 28day
01-08-2018	-1.08	1.16	-0.02	-0.96	-0.99	-1.10	2.59	1.40	-0.14	-0.97	-0.52	-0.96
03-08-2018	-0.44	0.95	0.04	-0.82	-0.33	-0.88	1.77	1.24	0.43	-0.71	-0.25	-0.77
04-08-2018	-0.76	0.93	-1.07	-1.18	-0.57	-0.84	0.19	0.75	-1.55	-1.10	-0.79	-0.96

n

For each training day across two seasons (n = 169)

Matrix *X*

Figure 3. Biplot of the scaled eigenvectors (arrows) and scores (dots) for the first (PC1) and second (PC2) principal components. More horizontal arrows relate to the x-axis (i.e. PC1) while more vertical arrows relate to the y-axis (i.e. PC2).
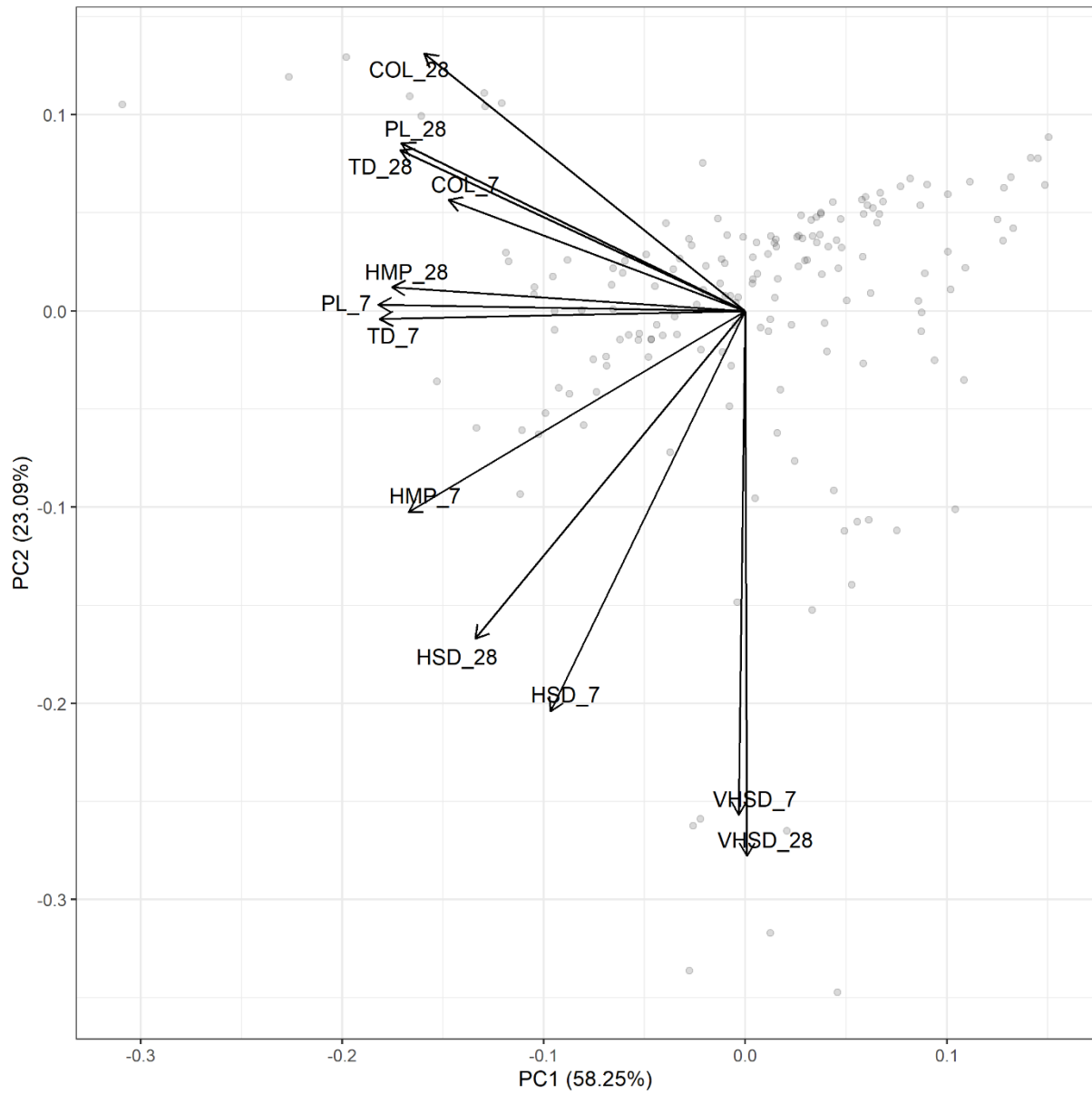
Figure 4. Scatterplot of the first (x axis) and second (y axis) principal component scores for the 169 training sessions for an individual player. The green data point highlights today's session in the context of every other training session completed.
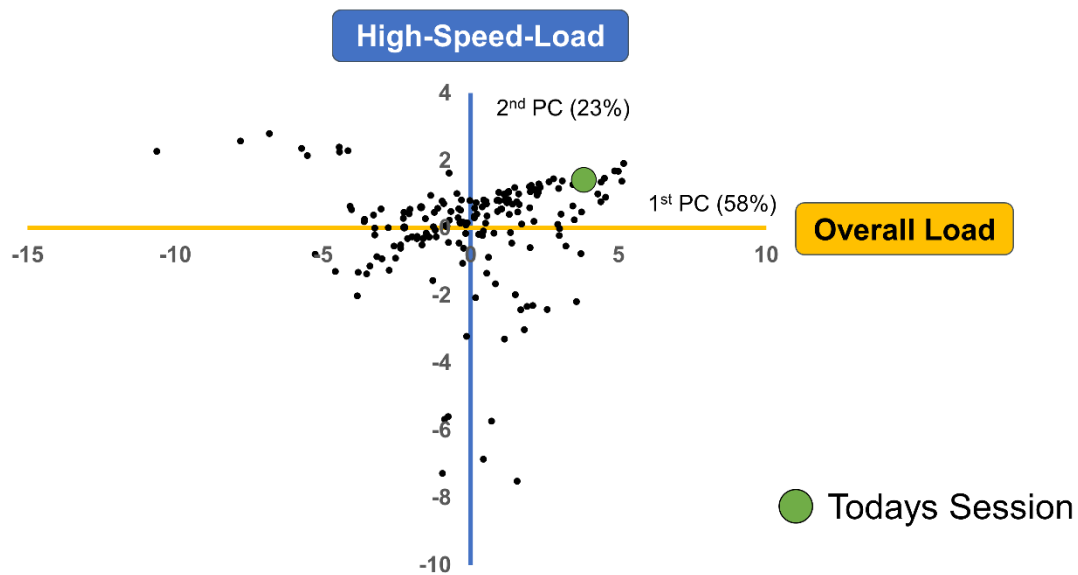

Figure 5. Time series plot of the first (overall load) and second (high-speed load) principal component scores (with signs reversed) across two seasons for a single player. Grey area represents days within 1 standard deviation of the mean, orange area within 2 standard deviations and red area within 3 standard deviations of the mean.

Table 1. Results of principal component analysis including the eigenvalues (% of variance explained) and eigenvectors for each principal component of matrix *X*.
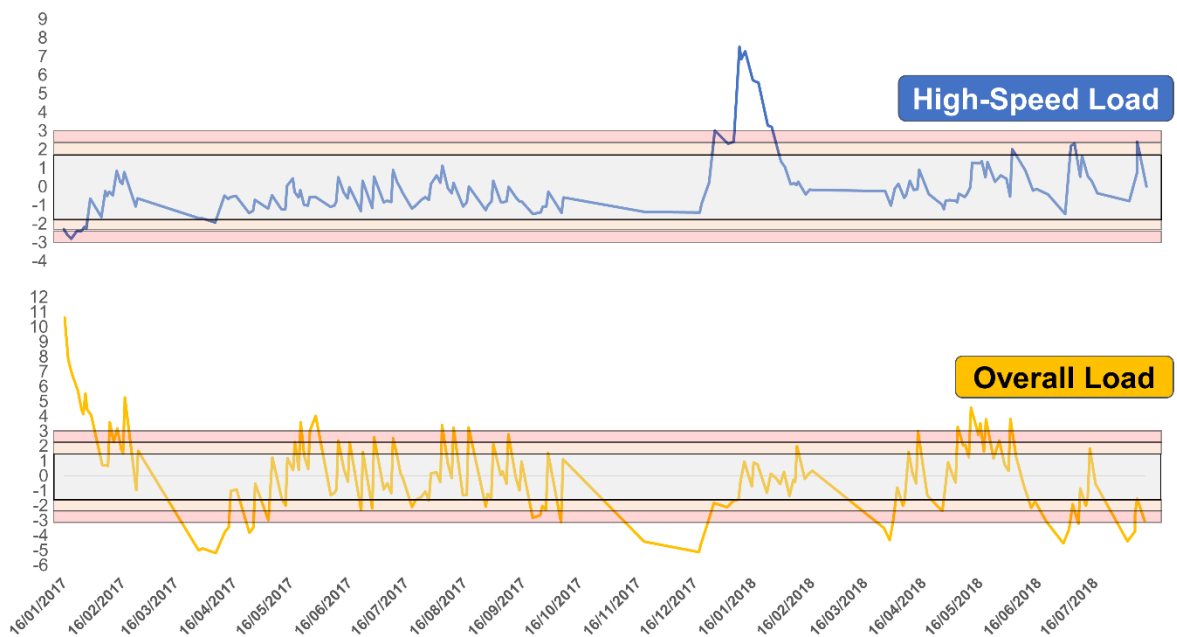
Table 1. Results of principal component analysis including the eigenvalues (% of variance explained) and eigenvectors for each principal component of matrix X

	1st PC	2nd PC	3rd PC	4th PC	5th PC	6th PC	7th PC	8th PC	9th PC	10th PC	11th PC	12th PC
Normalized eigenvalues	6.99	2.77	1.04	0.56	0.27	0.22	0.05	0.05	0.03	0.01	0.00	0.00
% of total variance explained	58.3	23.1	8.71	4.7	2	2	0.5	0.4	0.2	0.06	0.02	0.002
Eigenvectors												
TD_7	-0.36	-0.01	0.22	0.05	-0.36	0.16	-0.42	-0.14	-0.07	0.61	-0.15	0.27
TD_28	-0.34	0.16	-0.34	-0.03	-0.09	-0.13	0.05	-0.16	0.19	0.31	0.37	-0.65
HSD_7	-0.19	-0.40	0.39	-0.31	0.18	-0.56	0.08	-0.40	0.18	-0.02	-0.11	-0.02
HSD_28	-0.26	-0.33	-0.15	-0.42	0.46	0.23	-0.39	0.42	0.05	0.00	0.15	0.02
PL_7	-0.36	0.01	0.22	0.10	-0.30	0.22	-0.32	-0.07	0.16	-0.66	-0.15	-0.29
PL_28	-0.34	0.17	-0.34	0.00	-0.08	-0.13	0.06	-0.19	0.26	-0.24	0.37	0.65
VHSD_7	-0.01	-0.51	-0.22	0.59	-0.16	-0.40	-0.14	0.37	0.13	0.02	-0.01	0.00
VHSD_28	0.00	-0.55	-0.29	0.18	0.09	0.47	0.15	-0.55	-0.19	-0.02	0.01	-0.01
HMP_7	-0.33	-0.20	0.26	-0.11	-0.31	0.03	0.48	0.31	-0.48	-0.06	0.34	0.01
HMP_28	-0.35	0.02	-0.36	-0.16	-0.06	0.03	0.40	0.17	0.11	0.04	-0.72	0.02
COL_7	-0.29	0.11	0.40	0.48	0.47	0.25	0.29	0.09	0.34	0.13	0.05	0.02
COL_28	-0.31	0.26	-0.12	0.26	0.40	-0.29	-0.20	-0.12	-0.65	-0.11	-0.13	-0.02

Abbreviations: PC = principal component; TD = total distance; HSD = high-speed-distance; PL = PlayerLoad™; VHSD: very-high-speed-distance; HMP = high-metabolic-power-distance; COL = collisions. Numbers represent exponentially weighted average window (i.e. 7 or 28 days).