THE USE OF CLASSIFICATION AND REGRESSION TREE WHEN CLASSIFYING WINNING AND LOSING BASKETBALL TEAMS

Miguel A. Gómez[1], Sergio J. Ibáñez[2], Isabel Parejo[2], and Philip Furley[3]

[1]*Faculty of Physical Activity and Sport Sciences, Polytechnic University of Madrid, Spain*
[2]*Faculty of Sport Sciences, University of Extremadura, Spain*
[3]*Institute of Cognitive and Team/Racket Sport Research,
German Sport University Cologne, Germany*

Abstract:

The aim of the present study was to identify the best predictors when classifying winning and losing teams in basketball in consideration of situational variables by the classification and regression tree (CRT) non-parametric analysis. The sample was composed of 1,404 balanced games (score-differences: 1-14 points) from the Spanish EBA Basketball League that presented high heterogeneity and a non-parametric distribution. These games were split into faster- and slower-paced games according to ball possessions per game (using a cluster *k*-means). The CRT analysis was used to predict which game-related variable/s better classified winning and losing teams during slower- and faster-paced games. In total, this approach explained 72% of the total variance in the slower- and 69.3% in the faster-paced games. The results identified importance of defensive rebounds (100%), successful free-throws (94.7%), assists (86.1%), and fouls committed (55.9%) for the classification of winning and losing teams in the fast-paced games. Conversely, in the slow-paced games the better classification of winning or losing teams was accomplished by the following variables: successful free-throws (100%), defensive rebounds (82.3%), fouls committed (68.4%), assists (66.9%), successful 2-point (62.2%) and 3-point field-goals (61.6%). The influence of situational variables was identified only for team quality in the slow-paced games. The present findings allow coaches for a better control of games and competition.

Key words: classification and regression tree, performance analysis, team sport, match analysis

Introduction

Game-related statistics that discriminate between winning and losing teams have received plenty of research attention within performance analysis in basketball (Csataljay, O'Donoghue, Hughes, & Dancs, 2009; Ibáñez, Sampaio, Sáenz-López, Giménez, & Janeira, 2003; Trninić, Dizdar, & Lukšić, 2002). Previous research has identified consistent key indicators (i.e., successful 2-point shots, defensive rebounds and assists) that discriminate between teams' performance in men's (García, Ibáñez, Gómez, & Sampaio, 2014; Gómez, Lorenzo, Barakat, Ortega, & Palao, 2008; Sampaio & Janeira, 2003) and women's basketball (Gómez, Lorenzo, Sampaio, & Ibáñez, 2006; Gómez, Lorenzo, Ortega, Sampaio, & Ibáñez, 2009) during the past decade. The pattern of results was similar for under-16 and under-18 year olds, identifying the importance

of successful 2-point field-goals, successful free-throws, defensive rebounds, turnovers, and assists (Ibáñez, et al., 2003; Lorenzo, Gómez, Ortega, Ibáñez, & Sampaio, 2010).

Of further importance for the present study, basketball research has recently pointed out the importance of situational variables that may have an interactive effect on teams' performance such as game type (i.e., the final point difference between the teams: balanced and unbalanced games), game location (i.e., playing at home or away), team quality (i.e., winning percentages: the best and worst teams), and game pace (i.e., number of ball possessions: slow- and fast-paced games) (Gómez, Lago, & Pollard, 2013; Moreno, Gómez, Lago, & Sampaio, 2012; Sampaio, Lago, & Drinkwater, 2010b). Therefore, situational variables have to be taken into account when discriminating between winning and losing teams based on game-related

statistics in team sports (Moura, Martins, & Cunha, 2014). In particular, balanced games have been shown to be better suited in identifying the key game-related statistics associated with winning and losing (Gómez, Pérez, Molik, Szyman, & Sampaio, 2014; Lupo, Condello, Capranica, & Tessitore, 2014).

Traditionally, basketball research focusing on predictive modelling through game-related statistics (i.e., relationships between team's performance and game outcome) has relied on univariate analysis (Ortega, Cárdenas, Sainz de Baranda, & Palao, 2006; Parejo, García, Antúnez, & Ibáñez, 2013), linear regression models (Moreno, et al., 2012; Sampaio, et al., 2010b), discriminant analysis (Lorenzo, et al., 2010; Sampaio & Janeira, 2003; Trninić, et al., 2002), factorial analysis using principal components (Sampaio, Drinkwater, & Leite, 2010a), Pearson's correlation analysis (Ziv, Lidor, & Arnon, 2010), and logistic regression models (Gómez, Lorenzo, Ibáñez, & Sampaio, 2013). However, these statistical approaches might be problematic considering the complexity of discriminating between winning and losing teams in basketball based on the use of a large amount of variables (i.e., game-related statistics and situational variables). In particular, the use of large amounts of data (especially non-parametric data) in previous statistical approaches in sport has shown to be problematic when identifying multivariate and complex team performance interactions (Gómez, et al., 2013b; Mateus, et al., 2015). In this respect it has, for example, been demonstrated that when the performance analysis comprised competitions with a large number of games played (e.g., competitions that included different conferences, group stages, or consecutive seasons), it proved more problematic to identify reliable key indicators that were associated with success due to non-parametric data distributions (Parejo, et al., 2013; Moura, et al., 2014). Specifically, Parejo et al. (2013) studied 231 games from the group A of the Spanish Amateur League (First Conference of the EBA League) during the 2005/2006 season. Their results pointed out that with the increasing league heterogeneity the number of game-related statistics that were able to discriminate between winning and losing teams also increased in close, balanced, and unbalanced games (i.e., 8, 10 and 15 significant indicators, respectively). This fact highlights the importance of identifying the most reliable indicators when differentiating between winning and losing teams using the best statistical approach according to the nature and type of data.

According to Moura et al. (2014), the use of multivariate techniques is useful for describing the normative profiles of game-related statistics and their association with situational variables. Therefore, this statistical approach (multivariate techniques with non-parametric data) seems better suited for identifying the game-related statistics that allow discriminating between winning and losing teams over a large sample involving high performance heterogeneity. Specifically, the classification and regression tree (CRT) multivariate technique has been suggested to be a suitable statistical tool in exploring and modelling such data (De Ath & Fabricius, 2000) as it has shown to be a powerful and robust technique in selecting the independent variables that have the greatest influence on the dependent variable. CRT is a non-parametric statistical analysis that identifies mutually exclusive and precise subgroups of a sample whose teams share similar characteristics that influence the dependent variable (e.g., game outcome). The procedure examines all possible and/or splitting variables and selects the one/s that result in binary groups that are most different with respect to the dependent variable. The CRT is presented graphically; the root node (undivided data) first branches into two descendent nodes according to the independent variables (see Figure 1). Within each branch the descending tree continues assessing the remaining independent variables to determine which variable results in the best split. At the point that no further split can be made, a terminal node is established.

Using game-related statistics CRT searches through the full range of values and finds the best combination of cut-off points (i.e., specific game-related statistic values) or categories (e.g., situational variables) according to the established relationships between the dependent and independent variables. However, to date this technique has not been used in the context of identifying the key game-related statistics that discriminate between winning and losing teams in sport.

The aim of the present study was to identify the best predictors (i.e., game-related statistics) of success in basketball games while taking the situational variables, game pace, game location, and
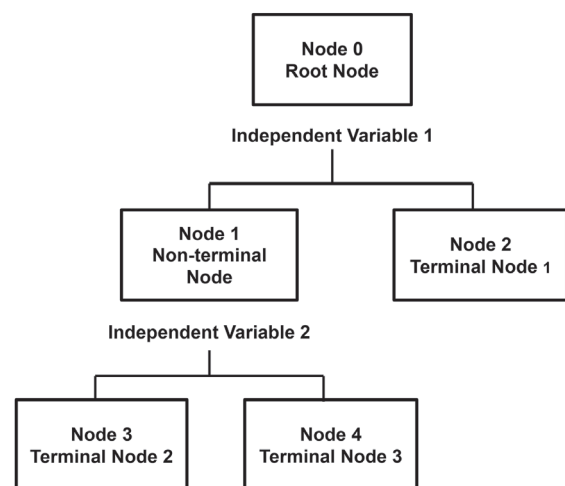

Figure 1. Graphical representation of CRT (adapted from Lemon, et al., 2003).

team quality into account using the CRT analysis. The use of the multivariate CRT technique should increase the power of data classification (i.e., into winning and losing teams) and help obtain a reliable number of key significant game-related statistics when using non-parametric large data.

Method

Sample

Researchers obtained archival data from the open-access official website of the Spanish EBA Basketball League (Spanish Amateur Basketball League) during the 2005/2006 and 2006/2007 regular seasons (N=2,074 games). The present sample was selected due to its heterogeneity and non-parametric distribution, the number of teams involved (i.e., five conferences composed of 14 teams each; n=70 teams each season), and the high number of games played during the two seasons. This assured a representative sample with great variability between teams' performance (De Ath & Fabricius, 2000; Eccles, Ward, & Woodman, 2009; Moura, et al., 2013). The local Institutional Review Board approved the study.

Variables

The absolute game-related statistics collected were: 2-point and 3-point field-goals (both successful and unsuccessful), free-throws (both successful and unsuccessful), offensive and defensive rebounds, steals, turnovers, assists, blocks (both performed and received), and personal fouls (both committed and received). Subsequently, to account for game rhythm contamination, the variables were normalized according to game ball possessions and multiplied by 100 (Gómez, et al., 2014; Ibáñez, et al., 2003). For example, the performance of a team A that makes 15 assists in a 60-possession game has to be weighed differently compared to the performance of a team B that makes 20 assists in an 80-possession game. Ball possessions equation (BP) was calculated according to Oliver (2004): BP = (field-goals attempted) – (offensive rebounds) + (turnovers) + 0.4 x (free-throw attempted).

In order to control for the situational variable effects, variables of game type, quality of opposition, game location, and game pace were analyzed (Gómez, et al., 2013a, 2013b; Moreno, et al., 2012):

Game type. The sample was split according to the final outcome, and then the game types were obtained using a *k*-means cluster analysis that allowed identifying a cut-off value for point differences in the final outcome of a given game. The results identified: balanced games (cluster 1) with the differences in the score ranging from 1 to 14 points (*n*=1,404 games), and unbalanced games (cluster 2) with the differences above 15 points (*n*=670 games). In the present study only the balanced games were included.

Game pace. The balanced games (previously defined) were divided into faster-paced and slower-paced games according to the number of ball possessions (Csataljay, Hughes, James, & Dancs, 2011; Sampaio, et al., 2010a), the *k*-means cluster analysis was performed to identify the ball possessions' cut-off value, thus two clusters were identified: the slower-paced games (cluster 1: 57.03±13.32 ball possessions; n=745 games) and the faster-paced games (cluster 2: 79.31±6.45 ball possessions, n=659 games).

Team quality. This was measured as the best and worst teams according to the winning percentage. A *k*-means cluster analysis identified two groups: the worst teams (winning%=42.25±9.32) and the best teams (winning%=61.51±12.24).

Game location. The home advantage effect was studied by analyzing the teams when playing at home or away.

Procedures

Professional statisticians from the National Federation gathered all the data; however, a sub-sample of ten games was randomly selected and observed by two experienced analysts (basketball coaches with more than 10 years of experience in performance analysis in basketball). The results showed *Kappa* coefficients of 1.0 for all the game-related statistics with the exception of assists (κ= .93) and turnovers (κ= .96) (Altman, 1991).

Statistical analyses

A classification tree analysis was used to classify winning and losing teams according to situational variables and game-related statistics of the slower- and faster-paced games. The algorithm used was the CRT technique which splits the sample into segments that are as homogeneous as possible in relation to the dependent variable (winning/losing) (Breiman, Friedman, Olshen, & Stone, 1984). The CTR is a classification and prediction tree-method. This method is robust with large samples and allows splitting the sample into different subgroups (nodes) based on the assumed impact of variables (i.e., game-related statistics and situational variables) on the game outcome. Also, this statistical procedure provides a useful visualization of the impact of each independent variable in the form of a tree model. The CTR algorithm node starts by examining the input fields to find the best split. Each split node of the root establishes two subgroups, each of which is subsequently split into two more subgroups, the growing-tree finishes when the stopping criteria is triggered (IBM, 2011).

The following statistical specifications were applied in the present study (IBM, 2011):

i)   The significance level was set at $p < .05$;

ii) To avoid overfitting, the pruning procedure was applied. The smallest tree with a cost within 1 Standard Error (minimum cost) was selected;

iii) The range of iterations was 100-50 (maximum-minimum) in order to obtain a useful tree that assured a balanced number of nodes (e.g., to avoid trees that have no nodes after the root node);

iv) The improvement measure could not be smaller than 0.001, which indicates modest differences between the nodes (i.e., higher values produce trees with a reduced number of nodes);

v)   The impurity measure tries to maximize within-node homogeneity (i.e., a terminal node that has all the cases with the same value for the dependent variable is a homogenous node that does not require further splitting and is called pure). The Twoing-method was used due to the fact that the dependent variable is grouped into two subclasses (win and lose), thus the splits should find the best differentiation between both groups;

vi) The missing independent values were excluded from the process and then surrogates were included in the descending tree. For these cases other independent variables that have a high associations with the original variable were used for the classification;

vii) The tree had a maximum of five levels in order to reduce the misclassification risk (goodness of fit);

viii) Cross-validation was conducted (the sample was randomly divided into 10 folds). For each subsample, a tree was built with the remaining cases (90%). The 10% subsample was used as a test sample, and then the 90% learning sample was applied to develop the tree with the cases of the 10% test sample. In addition, the cross-validation process showed the risk for the 10% test sample. Each of the 10% folds (which are mutually exclusive and add up to the total sample) served once as a test sample and served as part of the learning sample 9 times (for a 10-fold validation). The cross-validation process shows the risk value for the 10 test samples (IBM, 2011).

ix) Finally, each independent variable was ranked according to the importance to the model, and the risks of misclassifications were calculated as a measure of reliability of the model (IBM, 2011).

All statistical analyses were performed using IBM SPSS statistics for Windows, version 21.0 (IBM Corp., Armonk, NY, USA).

Results

For the *faster-paced games*, the CTR analysis showed five significantly influencing factors on a four-stage tree (cf. Figure 2). Fourteen nodes (6 final nodes) of contrasting groups of game outcomes in the *faster-paced games* were mainly established by assists (level 1), successful free-throws (level 2), defensive rebounds (level 3), and successful 2-point field-goals and fouls committed (level 4). Figure 1 shows the categories for the predictor variable (game outcome) and also the 14 nodes defined by the CTR analysis.

Level 1 (root node) is split by the assists. High probabilities to win the game were achieved when the obtained values for assists were higher than 13.37 (node 2) and, conversely, lower chances to win the game where achieved when the obtained values for assists were equal or lower than 13.37 (node 1).

Branch beneath node 1. The games in which the teams obtained equal or lower values than 28.69 for free-throws were split into two sub-groups (nodes); then the successful free-throws was the explanatory variable for splitting the node, and values equal or lower than 28.69 were related to losing the game (nodes 3 and 4). In this group (node 3), the defensive rebounds allowed for splitting the node and reflected lower chances to win when the teams obtained equal or lower values than 31.44 (node 7 vs. node 8). Besides, from node 7, if the teams obtained values lower than 29.17 in successful 2-point field-goals, they increased the chance to lose the game (node 11 vs. node 12). Conversely, from node 8, if the teams obtained equal or lower values than 29.75 for fouls committed, they increased the probabilities to win the game (node 13 vs. node 14).

Branch beneath node 2. The games where the teams obtained higher values than 28.69 for assists were split into two sub-groups (nodes), then the successful free-throws were the explanatory variable for splitting the node and values higher than 15.17 were related to winning the game (node 5 vs. node 6). Also, for node 6, the defensive rebounds were the explanatory variable for winning the game, with values lower or greater than 31.90 the teams increased their chance to win (node 9 vs. node 10).

This classification and regression tree model enabled explaining 72.0% of total variance after the cross validation analysis (see table 1). Also, the results showed that the explanatory variables for the classification of winning and losing in basketball, according to the importance to the model, were as follows: defensive rebounds (100%), successful free-throws (94.7%), assists (86.1%), and fouls committed (55.9%).

For the *slower-paced games*, the CRT analysis showed six significant influencing factors on a five-stage tree. Eighteen nodes (9 final nodes) of
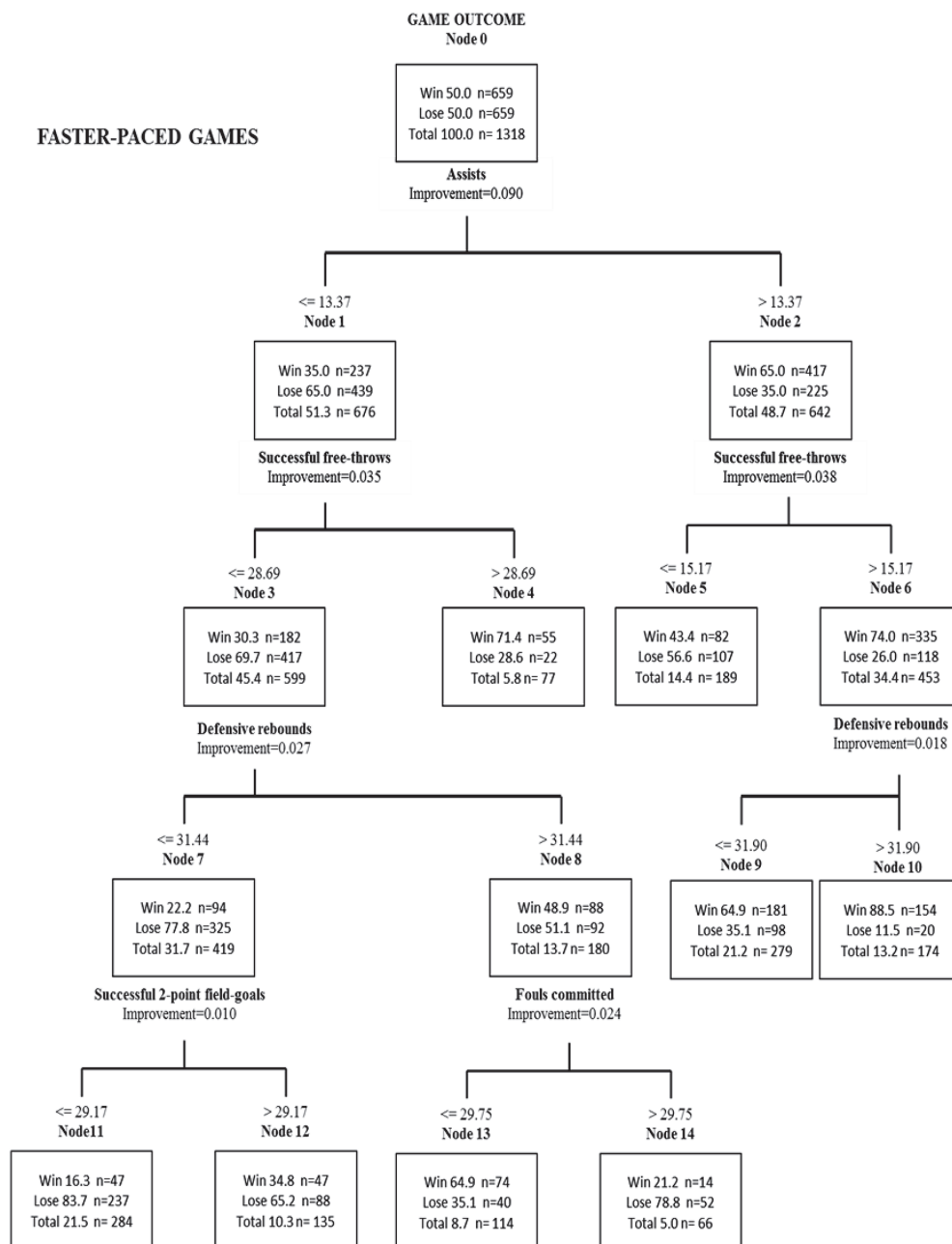
Figure 2. Classification and regression tree analysis of basketball faster-paced games (dependent variable: game outcome; independent variables: game location, team quality and game-related statistics).

contrasting groups of game outcome in the *slower-paced games* were mainly established by successful free-throws (level 1), defensive rebounds and fouls committed (level 2), successful 2-point and 3-point field-goals, and team quality (level 3), successful 3-point field-goals (level 4), and fouls committed (level 5). Figure 3 shows the categories of the predictor variable (game outcome) and also the 18 nodes defined by the classification and regression tree analysis.

Level 1 (root node) is split by the successful free-throws. High probabilities to win the game were achieved when the obtained values for free-throws were higher than 20.95 (node 2 vs. node 1).

Branch beneath node 1. The games where the teams obtained equal or lower values than 20.95 for successful free-throws were split into two sub-groups (nodes 3 and 4), then the defensive rebounds was the explanatory variable for splitting the node, and values equal or lower than 30.99 were related to losing the game (node 3 vs. node 4). In this group (node 3), the successful 3-point field-goals allowed for splitting the node, and reflected lower chances to win when the teams obtained equal or lower values than 14.05 (node 7 vs. node 8). On the other hand, node 4 was split into two sub-groups (nodes 9 and 10), the predictor variable was the successful 3-point field-goals, then an increased probability to
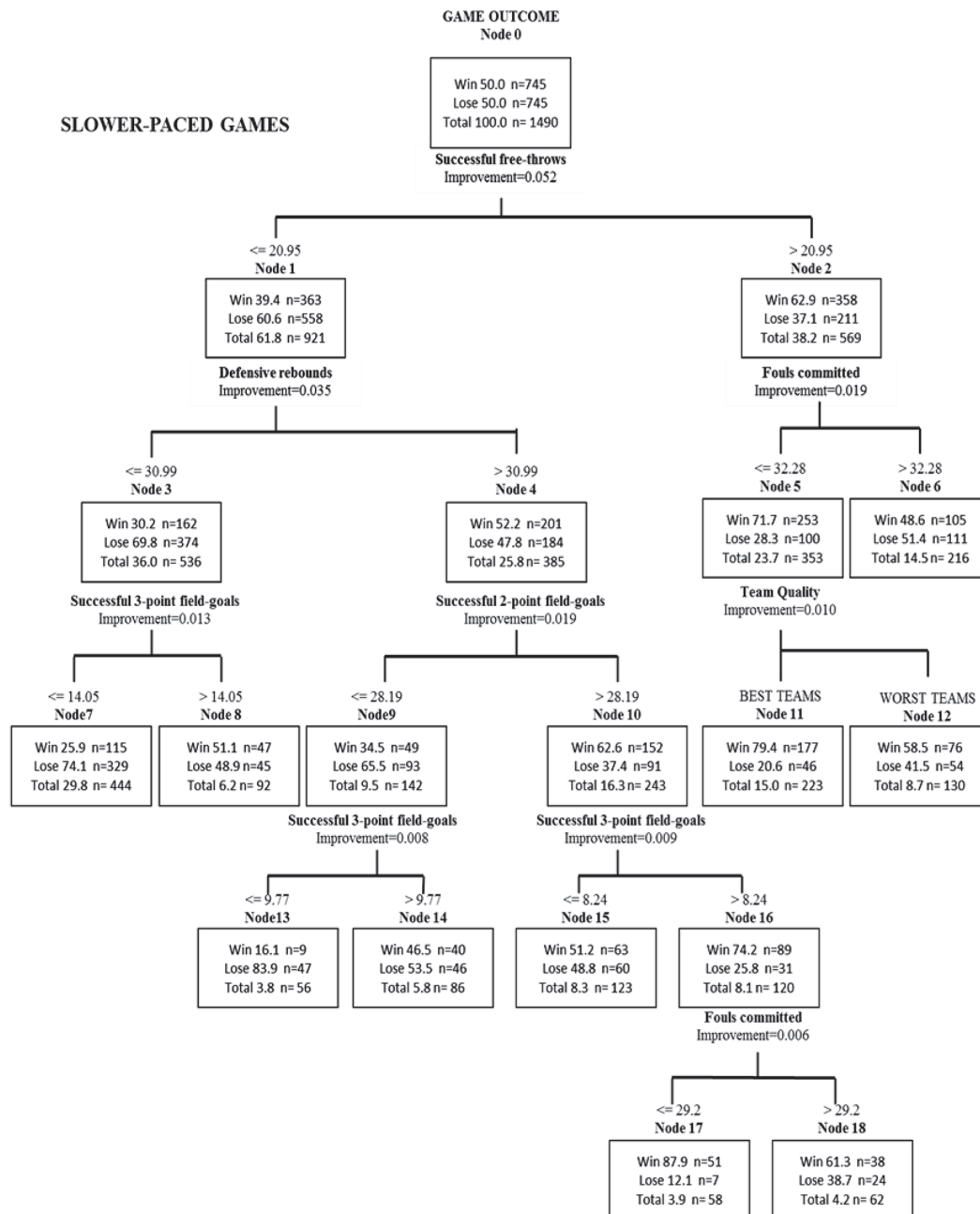
Figure 3. Classification and regression tree analysis of basketball slower-paced games (dependent variable: game outcome; independent variables: game location, team quality and game-related statistics).

win the games was obtained with higher values than 28.19 (node 10 vs. node 9). Besides, nodes 9 and 10 were divided into sub-groups according to the successful 3-point field-goals variable. The results revealed that, from node 10, values higher than 8.24 highly increased the chance to win (node 16 vs. node 15). Conversely, from node 9, values lower or equal than 9.77 increased the chance to lose the game (node 13 vs. node 14).

Branch beneath node 2. The games in which teams obtained higher values than 20.95 for successful free-throws were split into two sub-groups (nodes 5 and 6), then the fouls committed was the explanatory variable for splitting the node, and values equal or lower than 32.28 were related to

winning the game (node 5 vs. node 6). Also, for node 5, team quality was the explanatory variable for winning the game, with the best teams increasing their chance to win compared to the worst teams (node 11 vs. node 12).

The classification and regression tree model enabled explaining 69.3% of total variance after the cross validation analysis (see Table 1). Also, the results of the explanatory variables according to the importance to the model were as follow: successful free-throws (100%), defensive rebounds (82.3%), fouls committed (68.4%), assists (66.9%), successful 2-point field-goals (62.2%), and successful 3-point field-goals (61.6%).

Table 1. Cross-validation results of CRT models: estimations (standard errors of classification) and importance of the independent variables for faster- and slower-paced games

Faster-paced games			Slower-paced games		
	Importance			Importance	
	Absolute	%		Absolute	%
Defensive rebounds	0.10	100	Successful free-throws	0.06	100
Successful free-throws	0.10	94.73	Defensive rebounds	0.05	82.27
Assists	0.09	86.13	Committed fouls	0.04	68.44
Committed fouls	0.06	55.86	Assists	0.04	66.95
Team ability	0.05	46.70	Successful 2-pt field-goal	0.03	62.24
Game location	0.05	44.05	Successful 3-pt field-goal	0.03	61.59
Received fouls	0.04	34.16	Received fouls	0.02	37.93
Successful 2-pt field-goal	0.04	33.73	Unsuccessful 2-pt field-goals	0.02	35.69
Unsuccessful 3-pt field-goals	0.03	30.63	Blocks received	0.02	34.58
Unsuccessful 2-pt field-goals	0.03	26.12	Team ability	0.02	31.23
Unsuccessful free-throws	0.02	23.35	Unsuccessful 3-pt field-goals	0.02	30.26
Blocks made	0.02	21.33	Unsuccessful free-throws	0.01	22.89
Steals	0.02	19.90	Turnovers	0.01	17.92
Blocks received	0.02	17.44	Steals	0.01	9.86
Successful 3-pt field-goal	0.01	8.63	Blocks made	0.01	9.60
Turnovers	0.01	5.97	Offensive rebounds	0.00	6.95
Offensive rebounds	0.00	3.11	Game location	0.00	0.49
	Estimation (SE)			Estimation (SE)	
Cross-validation	.280 (.013)		Cross-validation	.307 (.013)	

Discussion and conclusions

The aim of the present study was to identify the game-related statistics that could best predict the game outcome (i.e., winning or losing) in balanced basketball games related to situational variables (i.e., game pace, game location, and team quality) using the CRT analysis. As was argued, the use of large samples that involve high level teams' performance variability can be problematic from a statistical approach point of view (Gómez, et al., 2013b, 2015a; Moura, et al., 2014). Therefore, the use of multivariate analysis such as CRT allows analysis of different factors and variables (i.e., game-related statistics and situational variables) within the same statistical model. Then, this procedure increases the power of data classification and allows us to measure a direct effect and interdependency of teams' performances in a visual tree (Figures 2 and 3). In fact, this analysis is not possible with other statistical tools when studying dependent variables with binary responses (i.e., winning or losing) such as discriminant analysis, Pearson's correlations, univariate analyses of variance or linear regression models (Gómez, Moral, & Lago-Penas, 2015b). In this respect, CRT analyses proved to be a powerful and effective technique in explaining winning and losing teams' performances for both the faster- (72% of total variance) and slower-paced games (69.9% of total variance).

The identified CRTs established cut-off values for key game-related statistics in an interactive exploration that differentiated winning or losing teams in basketball (i.e., five significant games statistics for the faster-paced games, and six game statistics and one situational variable for the slower-paced games). These results are important because the sample comprised balanced games and accounted for game pace rhythms, thus the classification of teams was based on representative performance variables in basketball (Csataljay, et al., 2011; Gómez, et al., 2014; Lupo, et al., 2014). The present data has important implications and can assist professionals in evaluating performance during competitions and when designing training plans.

Faster-paced games

For winning teams the CRT model showed the predictive importance of assists, successful free-throws, successful 2-point field-goals, fouls committed, and defensive rebounds. These variables were related to fast-paced rhythms that allowed for better offensive (i.e., quick attacks and high shooting accuracy) and defensive actions (i.e., defensive rebounds). This playing style generates more opportunities for fast break when securing defensive rebounds, passing and assisting to an open player in easy field-goal positions without defensive pressure (Csataljay, et al., 2011; Sampaio,

et al., 2010b). However, the significant effects of successful free-throws and fouls committed are inconsistent within available research (Sampaio, et al., 2010a). On the one hand, faster-paced games involve better performance of offensive rebounding due to the fact that this variable secures a restarted ball possession (Csataljay, et al., 2011; Oliver, 2004). On the other hand, defensive readiness may be expected during fast-paced games (e.g., steals and recovered balls instead of fouls committed). Thus, the results obtained allowed us to describe the importance of fouls committed and consequent free-throws. These variables can explain defensive actions used by losing teams when trying to stop the opponent's quick actions (e.g., fastbreaks or fast transitions).

The CRT results enhance understanding of the important key game-related statistics and their interactions that have not been previously accounted for (Ibáñez, et al., 2003; Sampaio & Janeira, 2003; Trninić, et al., 2002). In particular, nodes 6 (74% win, n=335 games) and 10 (85.5% win, n=154) accounted for high winning classification when teams obtained assists values greater than 13.37 and successful free-throws values greater than 15.17 for node 6 and, additionally, with defensive rebound values greater than 31.9 for node 10. Conversely, node 3 was closely related to losing (69.7%, n=417) when obtaining equal or lower values than 13.37 and 28.69 for assists and successful free-throws, respectively. These cut-off values are of great relevance for performance analysts and coaches during competitions (Moura, et al., 2014), but further highlight that game contexts have to be accounted for when interpreting game-related statistics (i.e., faster-paced and unbalanced games in amateur leagues).

Slower-paced games

During the slower-paced games the CRT model showed the importance of assists, successful free-throws, successful 2-point field-goals, fouls committed, and defensive rebounds. In addition, successful 3-point field-goals and team quality were included as significant predictors when classifying winning and losing teams. These key game-related statistics are associated with a high degree of the ball control playing style that is characterized by making better field-goal selection decisions, accurate passing to teammates in open positions, and drawing fouls (Sampaio, et al., 2010b). According to Oliver (2004) slower-game pace is connected with shooting efficiency, particularly from the mid- and long-range distances (e.g., 2- point and 3-point field-goals). This playing style enhances the importance of selecting good positions for field-goals. In this respect, coaches are well-advised to prepare their games and competitions accordingly (Csataljay, et al., 2011; Gómez, et al., 2009). In order to

account for game complexity during slower-paced games, the results showed powerful interaction of key game-related statistics in nodes 5 (71.7% win, n=253) and 10 (62.6% win, n=152) when winning. The cut-off values greater than 20.95 for successful free-throws and equal or lower than 32.28 for fouls committed were determinants in node 5. For node 10 the relevant values were lower or equal than 20.95 for successful free-throws, greater than 30.99 for defensive rebounds, and greater than 28.19 for successful 2-point field-goals. Conversely, node 3 was associated with losing (69.8%, n=374) when obtaining equal or lower values than 20.95 and 30.99 for successful free-throws and defensive rebounds, respectively. The identified performance profiles are different from those obtained in the faster-paced games. These may reflect the influence of key game-related statistics when the coach controls the team's performance during the slower-paced games (Moura, et al., 2014). The present results point out that game pace is a critical contextual factor that should be accounted for when training (i.e., preparing training tasks that involve fast- or slow- rhythm, modifying the number of passes, possession duration or ball possessions per minute) and during competitions (i.e., the coach has reference values that allow him/her to control for transitional or critical game periods according to the obtained cut-off values).

Situational variables

The results showed no significant effect of team quality during the faster-paced games, and only revealed its effect on node 5 during the slower-paced games (i.e., better values for best teams). This result is intriguing and may show that during balanced games (i.e., reduced point differences in the score) both teams performed similarly, suggesting a minor role of situational variables during these game contexts.

One interesting result from the present study was the absence of significant effects or interactions of game location during either faster- or slower-paced games, as this situational variable was significant in previous studies (Pollard & Gómez, 2013; Sampaio, et al, 2010b). Results may suggest that, in balanced games from leagues with high performance variability, the home advantage effect is not as important as it was found in unbalanced games (Gómez, et al., 2008).

The present findings might have some important practical implications and coaches would be well advised to especially target the identified game-related statistics (i.e., specific cut-off values related to winning or losing) identified in the present study during their training plans (Eccles, et al., 2009). Further, the use of CRT analyses allowed us to provide important information (i.e., variables that

have the greatest importance) for basketball coaches when preparing and controlling for basketball competitions.

Finally, the present study has some limitations that have to be acknowledged and should be addressed in further research. Future research might want to replicate the present study using games from other seasons, competitions/leagues or genders. For example, identifying how the critical moments of a game (i.e., last 5 minutes or overtimes) affect the game outcome in different contexts. Cut-off values for game-related statistics, performance indicators and situational variables for each quarter are also candidate actions for future studies using the CRT analysis.

References

Altman, D.G. (1991). Some common problems in medical research. In D.G. Altman (Ed.), *Practical statistics for medical research* **(**pp. 403-409). London: Chapman & Hall.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. J. (1984). *Classification and regression trees*. Belmont, CA.: Wadsworth.

Csataljay, G., Hughes, M., James, N., & Dancs, H. (2011). Pace as influencing factor in basketball. In M. Hughes, H. Dancs, K. Nagyváradi, T. Polgár, N. James, G. Sporis & G. Vuckovic (Eds.), *Research methods and performance analysis* (pp. 178-187). Szombathely: University of West Hungary.

Csataljay, G., O'Donoghue, P., Hughes, M., & Dancs, H. (2009). Performance indicators that distinguish winning and losing teams in basketball. *International Journal of Performance Analysis in Sport*, *9*(1), 60-66.

De Ath, G., & Fabricius, K.E. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology*, *81*(11), 3178-3192.

Eccles, D.W., Ward, P., & Woodman, T. (2009). Competition specific preparation and expert performance. *Psychology of Sport and Exercise, 10*(1), 96-107.

García, J., Ibáñez, S.J., Gómez, M.A., & Sampaio, J. (2014). Basketball game-related statistics discriminating ACB league teams according to game location, game outcome and final score differences. *International Journal of Performance Analysis in Sport*, *14*(2), 443-452.

Gómez, M.A., Battaglia, O., Lorenzo, A., Lorenzo, J., Jiménez, S., & Sampaio, J. (2015a). Effectiveness during ball screens in elite basketball games. *Journal of Sports Sciences*, *33*(17), 1844-1852.

Gómez, M.A., Lago, C., & Pollard, R. (2013a). Situational variables. In T. McGarry, P. O'Donoghue & J. Sampaio (Eds.), *Routledge handbook of sports performance analysis* (pp. 259-269). Oxon: Routledge.

Gómez, M.A., Lorenzo, A., Barakat, R., Ortega, E., & Palao, J.M. (2008). Differences in game-related statistics of basketball performance by game location for men's winning and losing teams. *Perceptual and Motor Skills, 106*(1), 43-50.

Gómez, M.A., Lorenzo, A., Ibáñez, S., & Sampaio, J. (2013b). Ball possession effectiveness in men's and women's elite basketball according to situational variables in different game periods. *Journal of Sports Sciences, 31*(14)*, 1578-1587.

Gómez, M.A., Lorenzo, A., Ortega, E., Sampaio, J., & Ibáñez, S.J. (2009). Game-related statistics discriminating between starters and nonstarters players in Women's National Basketball Association League (WNBA). *Journal of Sports Science and Medicine, 8*, 278-283.

Gómez, M.A., Lorenzo, A., Sampaio, J., & Ibáñez, S.J. (2006). Differences in game-related statistics between winning and losing teams in women's basketball. *Journal of Human Movement Studies 51*, 357-369.

Gómez, M.A., Moral, J., & Lago-Peñas, C. (2015b). Multivariate analysis of ball possessions effectiveness in elite futsal. *Journal of Sports Sciences*, *33*(20), 2173-2181.

Gómez, M.A., Pérez, J., Molik, B., Szyman, R.J., & Sampaio, J. (2014). Performance analysis of elite men's and women's wheelchair basketball teams. *Journal of Sports Sciences*, *32*(11), 1066-1075.

Ibáñez, S.J., Sampaio, J., Sáenz-López, P., Giménez, J., & Janeira, M.A. (2003). Game statistics discriminating the final outcome of Junior World Basketball Championship matches (Portugal 1999). *Journal of Human Movement Studies, 45*, 1-19.

IBM. (2011). *IBM SPSS Decision Trees 20*. North Castle Drive, NY: IBM Corporations.

Lemon, S.C., Roy, J., Clark, M.A., Friedmann, P.D., & Rakowski, W. (2003). Classification and regression tree analysis in public health: Methodological review and comparison with logistic regression. *Annals of Behavioral Medicine*, *26*(3), 172-181.

Lorenzo, A., Gómez, M.A., Ortega, E., Ibáñez, S.J., & Sampaio, J. (2010). Game related statistics which discriminate between winning and losing under-16 male basketball games. *Journal of Sports Science and Medicine*, *9*(4), 664.

Lupo, C., Condello, G., Capranica, L., & Tessitore, A. (2014). Women's water polo World Championships: Technical and tactical aspects of winning and losing teams in close and unbalanced games. *Journal of Strength and Conditioning Research, 28*(1)*, 210-222.

Mateus, N., Gonçalves, B., Abade, E., Liu, H., Torres-Ronda, L., Leite, N., & Sampaio, J. (2015). Game-to-game variability of technical and physical performance in NBA players. *International Journal of Performance Analysis in Sport*, *15*(3), 764-776.

Moreno, E., Gómez, M.A., Lago, C., & Sampaio, J. (2012). Effects of starting quarter score, game location, and quality of opposition in quarter score in elite women's basketball. *Kinesiology*, *45*, 48-54.

Moura, F.A., Martins, L.E.B., & Cunha, S.A. (2014). Analysis of football game-related statistics using multivariate techniques. *Journal of Sports Sciences*, *32*(20), 1881-1887.

Oliver, D. (2004). *Basketball on paper. Rules and tools for performance analysis.* Washinton, D.C.: Brassey's.

Ortega, E., Cárdenas, D., Sainz de Baranda, P, & Palao, J.M. (2006). Analysis of the final actions used in basketball during formative years according to player's position. *Journal of Human Movement Studies*, *50*, 421-437.

Parejo, I., García, A., Antúnez, A., & Ibáñez, S.J. (2013). Differences in performance indicators among winners and losers of group A of the Spanish basketball amateur league (EBA). *Revista de Psicología del Deporte*, *22*(1), 257-261.

Pollard, R., & Gómez, M.A. (2013). Variations in home advantage in the national basketball leagues of Europe. *Revista de Psicología del Deporte*, *22*(1), 263-266.

Sampaio, J., Drinkwater, E.J., & Leite, N. (2010a). Effects of season period, team quality, and playing time on basketball players' game-related statistics. *European Journal of Sport Sciences*, *10*, 141-149.

Sampaio, J., & Janeira, M. (2003). Statistical analysis of basketball team performance: Understanding teams' wins and losses according to a different index of ball possessions. *International Journal of Performance Analysis in Sport, 3*, 40-49.

Sampaio, J., Lago, C., & Drinkwater, E.J. (2010b). Explanations for the United States of America's dominance in basketball at the Beijing Olympic Games (2008). *Journal of Sports Sciences*, *28*, 147-152.

Sampaio, J., & Leite, N. (2013). Performance indicators in game sports. In T. McGarry, P. O'Donoghue & J. Sampaio (Eds.), *Routledge handbook of sports performance analysis* (pp. 115-123). Oxon: Routledge.

Trninić, S., Dizdar, D., & Lukšić, E. (2002). Differences between winning and defeated top quality basketball teams in final tournaments of European club championship. *Collegium Antropologicum, 26*, 521-531.

Ziv, G., Lidor, R., & Arnon, M. (2010). Predicting team ranking in basketball: The questionable use of on-court performance statistics. *International Journal of Performance Analysis in Sport, 10,* 103-114.

Correspondence to:
Sergio-José Ibáñez
Faculty of Sport Sciences
University of Extremadura
Av/ Universidad s/n; 10071, Cáceres, Spain
E-mail: sibanez@uex.es