International Journal of Performance Analysis in Sport

ISSN: 2474-8668 (Print) 1474-8185 (Online) Journal homepage: http://www.tandfonline.com/loi/rpan20

On the search for reliable performance indicators in game sports

Martin Lames & Tim McGarry

On the search for reliable performance indicators in game sports

Martin Lames[1] and Tim McGarry[2]
[1]University of Augsburg, Germany, [2]University of New Brunswick, Canada

Abstract

This article addresses the reliability of performance indicators in game sports. In this context, reliability is invariably treated from a technical point of view as a question of observer agreement i.e. high levels of agreement between observations. That the measurement process itself should yield reliable data, as defined, for sports performance is given. Our considerations of reliability, however, extend from the process of measurement to include the trait (i.e., the performance) being measured. From these considerations, we present the argument that the performance traits, as measured, are inherently unstable and that the performance indicators are therefore necessarily unreliable (or unstable). In this light, the ongoing search for reliable (or stable) measures of sports performance indicators is questioned. Instead, alternative approaches for performance analysis are offered that recognise the dynamic interactions that characterise game sports as key features of sport performance. This notion of dynamic interactions is compatible with sporting experiences and the way that sports practitioners think about sports performance. We conclude that performance analysis for purposes of theoretical advancement should make use of mathematical modelling and simulation techniques, and that performance analysis for practical purposes should include qualitative research methods to arrive at the necessary inferences for sports practice.

Key words: reliability, stability, dynamic systems, invariant behaviour.

1. The nature of game sports

Game sports are sports with two parties (teams, doubles or singles) that interact dynamically in order to score a goal/point and simultaneously to prevent the opponent from scoring.

Lames (1991, p.33)

In almost every categorical system given for the universe of sports we find a special category for game sports (e.g., Döbler *et al*., 1989). This finding acknowledges that there is a unique structure to performance in game sports, a fundamental property of which is that game sports are always comprised of two parties, usually teams, doubles or singles, with both parties sharing mutual objectives (from the perspective of performance analysis combat sports may be treated very much in the same way as game sports). In football, for example, the basic idea is that a team strives to score a goal as well as to prevent their opponent from scoring. Of course this same objective holds for the opponent, too. Thus, the two teams pursue the same objective simultaneously, a feature that is typical of game sports. Similarly, if we consider tennis as an example of a net game, we find the same structure. Each player tries to win the rally by scoring a winning point and, at the same time, to prevent the opponent from scoring. Thus, the actions of the players on a football field, and the strokes of the players on a tennis court, are the result of the simultaneous striving towards common objectives. Importantly, because these aims are mutually exclusive, but pursued at the same time, tight interactions arise between the two parties. These interactions furthermore are dynamic, that is, the interactions among the parties change in time during a game. For this reason, we talk about an interaction process. If a certain action in a game sport is successful, the opponent has a reason to change his (her) behaviour. If the action is unsuccessful, however, then the player him/herself looks for something better to do.

This notion of game sports as the product of dynamical interaction processes would seem reasonable and, understandably, has provoked no objections in the past. On the other hand it is rather fateful when its consequences are considered. What we see when we observe a sports game is, by definition, a dynamical interaction process in which measures and countermeasures are taken in an attempt to overcome the opponent. This implies that the behaviour produced is not primarily the expression of stable properties of the individual players (e.g., technical skills or physical abilities) as is the case for other groups of sports (e.g., endurance sports such as long distance running, or power sports such as weight lifting). Instead, as mentioned previously observable behaviour in game sports emerges from the dynamic interactions between the opponents (see Figure 1), a consideration that means that game sports performances have to be considered as unique action chains. These unique chains for game sports are context dependent (i.e., they are influenced by the situation and the opponent) and time dependent and thus not repeatable (i.e., reliable). These considerations lead us to the main thrust of this article that considers the ongoing search for reliability (or stability) of performance indicators in game sports behaviour as unrewarding.

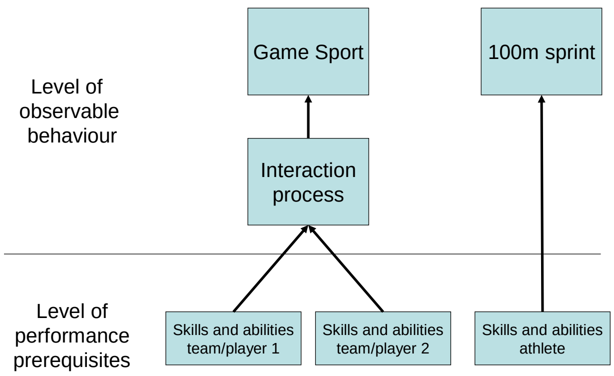

Figure 1. The differences between game sports and other sports as illustrated by the relation between observable behaviour and performance prerequisites

2. "Classical" performance analysis

Letzelter and Letzelter (1982) identified two basic aims of classical performance analysis, a theoretical aim and a practical aim. Theoretical performance analysis seeks to understand the game structure, for example, by identifying the performance behaviours that are important for a given sport. In theoretical performance analysis, the general aim is to explain sports behaviour using general models whose empirical foundations provide useful information for sports practice, such as informing on the long-term planning of training processes. The empirical data on which these models are predicated may also be used to provide statistical norms for performance measures (e.g., norms for reaction times in sprinting) or for necessary levels of performance prerequisites (e.g., norms for jumping height for volleyball players).

In contrast, practical performance analysis aims at a coupling of sports performance with individual training processes. Thus, sports performance is analysed with a view to identifying information useful for sports training. Roughly spoken, there are two primary sources of information that are relevant for training: the analysis of the player(s) and the analysis of the opponent(s). The analysis of playing performances aims at detecting the strengths and weaknesses of the player(s) in order to identify the most urgent aims for training. Analyses of the performances of the opponent are undertaken with a view to identifying an optimized strategy for the next match.

The underlying assumptions of classic performance analysis are that the observed performances can be explained by the abilities and skills of the athletes. These abilities and skills are conceived of as being stable properties, properties that may only be influenced in time by special measures taken in training. In this accounting, the stable traits of the athletes, as well as the special circumstances of the sports competition, explain the performance behaviours. The "trait paradigm", including its assumptions of stable properties for the underlying traits, therefore seemed an appropriate and scientifically accepted methodology for the analysis of sports performance. The historical root of the "trait paradigm" is found in psychological personality theory. It was a basic achievement of early sports science to apply the trait paradigm to sports behaviour.This paradigm was very successful in explaining performances in sports such as the 100m sprint, for example.

In general, the first step of classical performance analysis is to use theoretical considerations, or practical experiences, to identify a set of abilities and skills important to sports performance. In the second step, operational definitions for these traits are developed and empirical measurements taken to assess the level of these traits. Such measurements may happen on the level of performance prerequisites (e.g., motor tests, biomechanical or physiological measurements) or on the level of competition behaviour (e.g., reaction time, split time analysis, stride length and frequency). Statistical procedures like factor analysis and multiple regression of these empirical data further help to enlighten upon the structure of sports. For example, the performance prerequisites might be ranked according to their influence on sport outcome, thus providing important background information for training purposes. There are, however, some important limitations of the trait paradigm for informing upon sports performance. The first limitation is that the trait paradigm, designed to identify general laws, for example, correlations between traits and behaviour, cannot satisfactorily identify the strengths and weaknesses of individual athletes. The second limitation of the trait paradigm is found in the consideration of game sports as an interaction process (see Figure 1). If game sports are unique events, and if the game structure itself is dynamic, and if the game behaviours emerge spontaneously from the dynamic interactions among the players, then there is no expectation of stability in the observed behaviours as measured using performance indicators. And if, as suggested, the performance indicators of game sports are unstable (or variable) then the basic assumption of the trait paradigm is missed: Sports performance cannot be explained simply by an examination of the underlying traits of the athletes. Instead, the dynamic interactions among the players are key considerations in any view of game behaviour.

3. On the stability (reliability) of performance indicators in game sports

There are two notions of reliability that are common among scientists. In a narrow sense, reliability means the consistency (or stability) of a measure. With respect to observational measurements, as used in notational analysis, reliability is typically assessed using measures of intra- and inter-observer agreement. It is important for the sports scientist to establish intra- and inter-observer agreement if the data from performance analysis are to inform on sports practice as intended. In a wider sense, reliability comprises not only the consistency (or stability) of the measure, but also the consistency (or stability) of the entity that is being measured, what we will refer to as the measurement result. Thus, reliability of the measurement results require not only consistency in the measurement process (i.e., the assignment of a value to some variable), but also stability in the entity that is being measured object and stability in the conditions in which the measurements are taken (Lienert, 1969). In this wider sense, changes (or variability) within the entity itself, and/or within the conditions in which the measures were taken, present an open threat to reliability.

The consideration of reliability in its narrower or wider sense leads us to question the wisdom of looking to identify stable or reliable performance indicators in game sports, respectively. The narrow or wider consideration of reliability notwithstanding, the essential problem posed to performance analysts who seek to document game sports for purposes of information feedback remains the same, namely the issue regarding the search for invariant data.

3.1 Some conceptual problems for stability (reliability)

Unsurprisingly, performance indicators in game sports are invariably obtained by the method of observation. In considering measurement issues with regard to game observation, there is uniform consensus that the data should be independent of the observer, meaning that the observation method for data collection should be objective. Since the observer is the instrument of measurement in the observation process, it follows that objectivity on the part of the observer is a very important part of reliability in observational systems (Lames, 1994).

The usual way of reporting on performance indicators for game sports is to use frequencies, or relative frequencies, of behavioural occurrences (Hughes and Bartlett, 2004). These frequency counts provide summary statistics for parts of a match, for a single match, or for several matches aggregated. These types of descriptive statistics, however, do not contain information on the sequential context of the game (for example, what series of actions led to a specific shot on the goal), nor do they report on the situational context of the specific action (for example, was the observed behaviour produced from a fast break or from a static position). It follows, then, that the usual way of obtaining performance indicators for game sports ignores unfortunately the important dynamic interactions of which they are comprised.

The basic reason for the problems in establishing the stability (reliability) of performance indicators for game sports is simply the incorrect assumption on which such stability (reliability) is supposed. This assumption ignores the interactions between the player and opponent as important sources of variability within game sports. For example, games against different opponents usually are quite different from each other. Even within a match there are continual variations in game behaviour because of the dynamics of the interaction processes described above. Thus, the dynamic nature of game sports prevents the performance indicators from demonstrating sufficient stability (reliability). When viewed from the conceptual level there would therefore seem to be no reasonable expectation of stability (reliability) for performance indicators in game sports.

3.2 Some empirical problems for stability (reliability)

3.2.1 Variability (instability) of data between matches

Stability between matches has frequently been examined. We should distinguish between games against the same opponent and against different ones. Some common examples follow:

- We meet the same opponent in a league system in the home and away match. The demonstrated existence of home advantage in many sports (Nevill *et al*., 1996) suggests that there is a source of variability, or error, not attributed to the abilities and skills of the two teams. Performance indicators obtained when playing home and when playing away should be interpreted separately.
- Sometimes two different competitions like cup and league games provide the opportunity to analyze two matches between the same teams within a short period of time. There is no study yet that examines stability of performance indicators under these interesting conditions.
- The same individuals may play each other in an earlier and a later phase of the same tournament depending on the tournament schedule and the game results. In Germany, people often remember the two games against Hungary in the 1954 Soccer World Championships in Switzerland. Germany lost the first game 3-8 in the preliminary round and won the second (final) game 3-2.
- In a classical knockout system the next chance to meet the same opponent is only in the next tournament. Stability when playing against the same opponent should be higher than when playing against different opponents in a series of consecutive matches (McGarry and Franks, 1994; 1996a).

When considering matches against different opponents, one might expect that valid behavioural norms may be obtained given the assumption of sufficient stability. In practice, however, these behavioural norms frequently turn out to have too large variances for practical purposes. For example, in an analysis of tennis games performed on clay ground (153 games, 306 players) the first service mean error rate for men with an ATP-ranking below 80 was reported as 39.6%, with a minimum and maximum error rate of 0% and 70%, respectively (Lames, 1991). The standard deviation for these data was 16.4%, resulting in a 95%-confidence interval of 7.5% to 71.7%. These results were considered of little value for use in sports practice.

3.2.2 Variability (instability) of data within matches

The previous example demonstrated instability of a performance indicator for tennis between matches. In the next example, again from tennis, we demonstrate the instability in a performance indicator obtained from within a match. The illustration of within-match variability (see Figure 2) depicts the moving average of Boris Becker's error rate in first services in the 1989 US Open final (Lames, 1992). The error rate in first service may be interpreted as a balance between risk taking on the part of the player and pressure created from a successful first service. The data in Figure 2 demonstrates that this balance between risk and pressure varies considerably during a match. Indeed, the mean and confidence interval data fail to give a satisfactory description of within-subject stability. Inspection of the data indicates that Becker did not serve well in the second set, which he lost 1-6, but served without error towards the end of the third set.

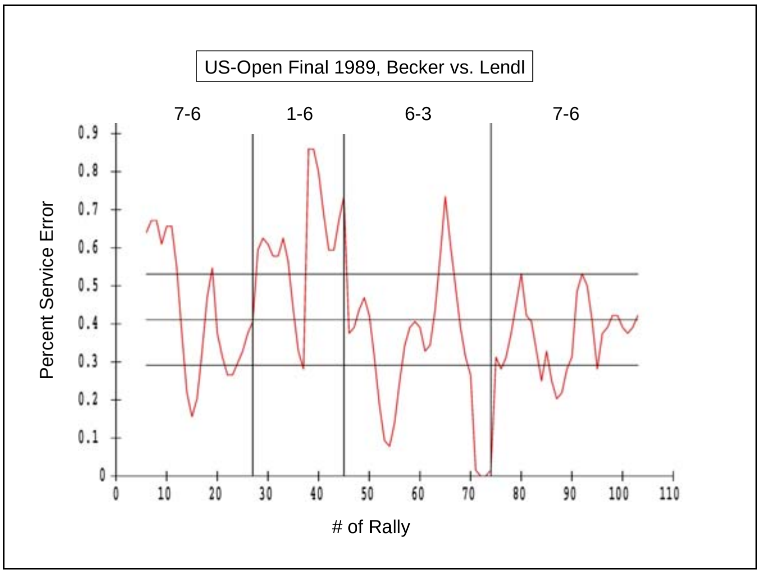

Figure 2. Moving average of Boris Becker's service errors against Ivan Lendl in the 1989 US Open final (Lames, 1992). The scores for the four sets in the match are listed in the figure.

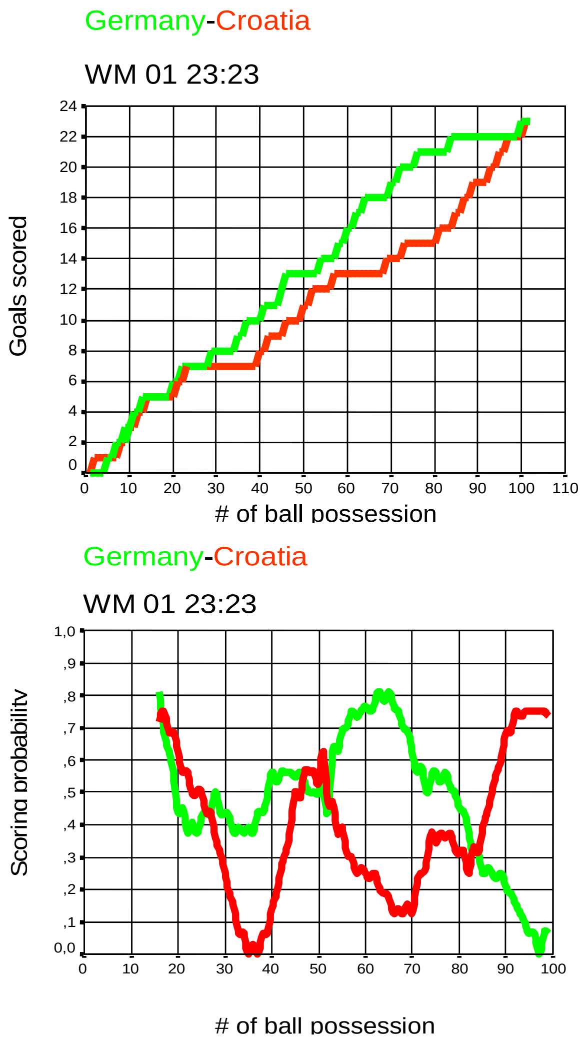

Figure 3. Scoring process (upper panel) and momentary scoring probability (lower panel) in a handball game (Lames, in press).

Figure 3 shows the development of the handball game between Germany and Croatia at the 2001 World Championships. The scoring process and the momentary attacking strength, as measured by a moving average on ball possessions coded goal/no goal, are displayed (Lames, in press). The descriptions reveal that each team demonstrated phases within the match where almost every ball possession resulted in a goal, and, in contrast each team went through phases without success. These characteristics of large within-game variabilities were found almost without exception in 40 handball games of national and international level.

The data presented in Figures 2 and 3 provide strong evidence of within-game instabilities in the performance indicators used. From a conceptual level, we reiterate the point that within-match instability is to be expected given that a sports game is the product of ongoing dynamic interactions. From a practical viewpoint, we might again predict within-match instabilities from a consideration of sporting experiences. We change tactics after scoring and taking the lead in soccer, for instance. We look for different ways to attack if our last efforts failed. Sometimes successful tactics are changed perhaps to keep the opponent guessing and prevent anticipation or, possibly, sometimes to simply add to the frustration of the opponent.

3.2.3 Stability or instability in performance indicators: Some pitfalls to watch for

In some situations the data may seem to support the presence of stability when stability is not in fact established. The use of correlations to report on the associations between performance indicators obtained between matches, or within a match (i.e., between parts of a match) in particular can mislead the unsuspecting if variance in the two data series is increased by incorrect sampling. Some practical examples include:

- The comparison of performance indicators for different positions in a team, or for different player types in singles games, may increase the correlation. For example, the rate of net attacks in tennis will show an artificially high correlation – and thus an inflated level of stability – for a sample of serve-and-volley players as compared against a sample with a high proportion of base-line players.
- Frequently, the playing times per match are different among players. If the data are not corrected for these differences, the effect will be to increase the correlations between performance indicators both within and between matches.
- That some players regularly get more playing time than others creates another pitfall. Usually playing time reflects an estimation of the player by the coach. If there is a high correlation between a performance indicator and a player ranking, then this result may arise because the playing times and the perception of the player by the coach are confounded, and not because the performance indicator is a valid measure of the quality of that player.
- Using performance indicators to assess for differences between winning and losing teams should be performed with caution. First, we should point out that trivial data contain little, if any, information for sports performance (e.g., winners and losers differ in the number of points scored). Second, performance indicators are often confounded with other information. For example, the rate of shots per goal is confounded with the number of goals scored which is necessarily higher for winners than for losers. Thus, the number of goals must be controlled for in some way perhaps by using partial correlations. In the least, it would be interesting to investigate the differences in performance indicators between winners and losers when doing so as contrasted against when not doing so.

In light of the latter point, we reiterate the suggestion of Hughes and Bartlett (2004) that performance indicators should be standardized in some way for purposes of comparison. Sometimes standardization is not a simple task as demonstrated in the example above with regard to the rate of shots per goal for winners and losers. Similarly, the common practice of standardizing volleyball statistics by the number of sets played is insufficient because of the presence of shorter and longer sets. The correct method of standardizing the data notwithstanding, the data must be standardized otherwise the results may suggest performance differences, or performance stabilities for that matter, where in fact no such differences exist.

3.2.4 Further comments on the stability (reliability) of performance profiles

In the search for a behavioural signature of playing performance, McGarry and Franks (1994, 1996a) reported that the playing profile of individual squash players, as defined from the probability of shot selections, were variant as opposed to invariant when contested against different opponents. This unexpected finding of the time challenged our presumptions of understanding of sports performance because it follows necessarily that the data from a past contest must be invariant (or reliable) if they are to inform on the next contest (as in scouting). McGarry and Franks (1996a) identified sampling error (i.e., too few data sampled) as a possible reason for their findings of variable (or unstable) data. The other possibility offered by these authors was that the frequency data used to specify the playing profiles yielded an unsatisfactory description of sports performance (McGarry and Franks, 1996a, for further details), a possibility that these authors selected to follow in further work. We will return briefly to this point in the next section.

Subsequent efforts to identify invariant behaviours by increasing the sample size have been undertaken in more recent times by some authors. For example, Hughes, *et al*. (2001) reported on the existence of "normative profiles" for performance indicators when the data were gathered from a number of games. Normative profiles are sets of performance indicators obtained by averaging performances of several games in an attempt to establish a stable profile. In their research, Hughes and colleagues determined that a performance indicator was "stable" when additional data from other games did not result in a change in the mean value estimate obtained to that time outside of what were considered to be tolerable limits for the supposed – though necessarily unknown – true mean value for that performance indicator. These interesting techniques for establishing stable performance profiles are not without criticism, however, perhaps the main one being that the Hughes *et al*. method is still subject to sampling error, a point noted by O'Donoghue and Ponting (2005).

The word stable as used in the above context means that the mean value of a performance indicator obtained from the data set is a good estimate of the true value of the trait being measured. The quality of such an estimate is usually expressed in a confidence interval for the mean, with a reasonably small confidence interval indicating a good estimate of the true value. In fact, an answer to the question of how large a sample size is required for stable performances (or normative profiles) is provided, in principle, from statistical considerations. The mean over n normally distributed measurements has a standard deviation of $s/\sqrt{n}$, that is, the standard deviation of a single measurement divided by the square of the number of aggregated trials. Thus, the required precision of the estimate can be obtained simply from taking a sufficiently large sample of data. For similar considerations using the central limit theorem to obtain the necessary number of games for establishing stable performances within specified confidence limits, see O'Donoghue and Ponting (2005).

Unfortunately, there remain some outstanding criticisms on the ongoing search for stability in performance indicators by simply increasing the sample size of data recorded. One such criticism concerns the timeliness of the some of the performance data as increasing the sample size (e.g., increasing the number of games) necessarily mean that some of the data are less recent than other data. Of course, in sports practice the interest with regard to information feedback is much more concerned with data from recent sports performances than with less recent data. The comment by O'Donoghue (2005) that more recent performances possibly be given more weight than less recent data for performance analysis is interesting, however, it would seem to reintroduce the issue of variability (or instability) in the performance data that an increased sample size is intended to negate in the first place. Of more importance, however, is the criticism that increasing the sample size to decrease variability, and thus obtain "stable" performance measures does not get around the essence of the initial problem identified by McGarry and Franks (1994, 1996a), that of variant behaviour between games when playing against different opponents. For example, while stable performances indicators identified from a number of games might predict well to the next observations taken from the same number of games at some future time, the performance indicators still contain little useful information with regard to the next (single) game. Thus, while acknowledgement of the problem of game-to-game variability, together with efforts to address this concern is to be applauded, the issue of between game variability would seem intractable. In addition, as noted before the performance indicators that invariably consist of frequency counts of game behaviours do not reflect well upon the game structure when the game is viewed as a dynamic interaction process.

4. Possible theoretical alternatives for performance analysis in game sports

4.1 Game sports as complex (dynamical) systems

A game sport is a complex system that consists of at least of two players, with each player possessing many alternatives regarding how to act. Thus a game sport is comprised of subsystems with dynamic interactions among the subsystems. Indeed, dynamical properties are essential considerations for a complex system. In this context, the shared objectives that the two parties are competing for may well be conceived as attractors for the complex system (see Figure 4). Indeed, the findings of chance, chaos, and climax in the course of a game sport are to be expected within the framework of a dynamical (complex) system. These properties of game sports, however, tend to get overlooked by classical performance analysis. There would therefore seem to be good reason to consider the merits of dynamical systems theory for the structural modelling of game sports. For the moment at least, the problem with this suggestion lies in the details.

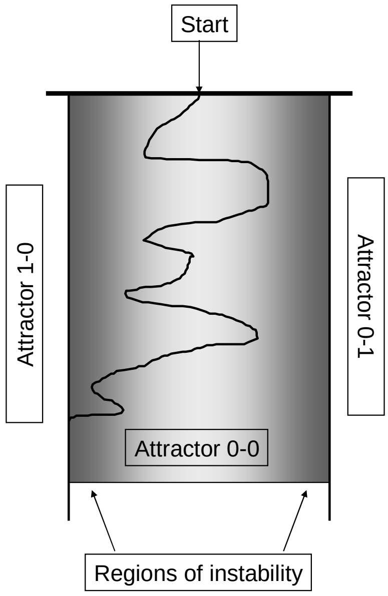

Figure 4. Illustration of football as dynamical complex system (Lames, 1998)

Figure 4 offers an illustration of a possible starting point from which to consider football as a complex system. The aims of each team are represented as two attractors with the chain of events within the game represented as the meandering black line. This line depicts the state of the interaction in the football game as a function of time. In this view, the football match might be thought of as unfolding in a type of phase plane, although the question of what variables might be appropriate for a description of the phase plane remain unanswered at present.

This new type of thinking on game sport behaviour in terms of dynamical systems theory is beginning to gain favour among some sports scientists (see, for example, McGarry *et al*, 1999, and McGarry *et al*., 2002, for earlier consideration of sports contests as dynamical systems; see also Palut and Zanone, 2005, for a demonstration of tennis as a dynamical system). In addition to these considerations, the possibilities of future contributions of neural networks analysis as a means of analyzing the dynamic data that describe sports performances was also suggested (see McGarry and Perl, 2004, for further details). For now though, many open questions remain and much work must be undertaken if a satisfactory accounting of sports performance behaviours are to be described using dynamical systems theory. Nonetheless, the general method with regards to any such accounting is evident. In order to inform upon the structure of a game sport it is first necessary to describe it using an appropriate mathematical model. Thereafter, the internal structure of the game can be investigated by analyzing the model behaviour using simulation techniques under different scenarios. In the next section, we provide an example of such an approach for advancing our understanding sports performance.

4.2 Game sports as probabilistic (Markov) processes

The use of probability analysis for investigating games of chance has a long history. Unsurprisingly, then, the structure of game sports is likewise informed using probability (stochastic) analysis, specifically finite Markov chain analysis. Some sports examples of analyzing scoring structures on the basis of probability, including identifying optimal decision-making strategies in some instances, are found for squash (e.g., Ap Simon, 1951, 1957; Clarke and Norman, 1979; Pollard, 1985; Schutz and Kinsey, 1977; and Wright, 1988), tennis (Croucher, 1982, 1986; Morris, 1977; Pollard, 1987; and Schutz, 1970), badminton (Clarke, 1979; Renick, 1977), volleyball (Pfeifer and Deutch, 1981) and baseball (Trueman, 1977). Following these authors, McGarry and Franks (1994; 1996a; 1996b) modelled the behaviour (the shots) in squash games using Markov chains. The intention of this work was to predict future behaviours from past behaviours using simulation techniques and, in so doing, to subsequently identify optimized decision-making strategies for winning performances. The approach of Lames likewise used finite Markov chains as a model for game sports, including its calculus (Kemeny and Snell, 1976). Once again, simulations were undertaken to assess the usefulness of certain tactical behaviours, as well as to assess the performance of individual players in team games. This type of work was applied to tennis (Lames, 1988; 1991), squash (McGarry and Franks, 1994; 1996a; 1996b) and volleyball (Lames and Hohmann, 1997; Lames et al., 1997), as well as table-tennis (Zhang, 2003) and handball (Pfeiffer, 2003).

	Rec. CUB	Rec. GER	Dig CUB	Dig GER	Set CUB	Set GER	Att. CUB	Att. GER	Blo. CUB	Blo GER	Point CUB	Point GER
Serve CUB		89.2									1.8	9.0
Serve GER	92.8										7.2	
Recept. CUB					94.8		1.3	2.6				1.3
Recept. GER			2.0			91.9	2.0	2.0			2.0	
Dig CUB				2.6	72.7		5.2	3.9				15.6
Dig GER			3.1			64.6	6.2	7.7	3.1		15.4	
Set CUB				0.8			97.7	0.8				0.8
Set GER			0.8				0.8	96.2			2.3	
Attack CUB				25.2						33.1	30.2	11.5
Attack GER			30.5						48.9		11.3	9.2
Block CUB			27.1	18.6						1.4	20.0	32.9
Block GER			19.1	29.8			2.1				36.2	12.8

Figure 5. Volleyball (Cuba vs. Germany 3-0, Bremen 1995) as a finite Markov chain (Lames et al., 1997).

Figure 5 shows the results of a transition matrix of a finite Markov chain model of volleyball. It depicts the match between the female national volleyball teams of Cuba and Germany at Bremen, Germany, in 1995. The model consists of distinct states and the transition probabilities between them. The transient states are the usual standard actions in volleyball (i.e. service, reception, dig, set, attack and block) and the final or absorbing states are the winning points (i.e., Point CUB and Point GER). The transition probabilities between the states are obtained from observation. For example, a transition probability (expressed as a percentage) of 89.2% from Serve CUB to Rec. GER means that 89.2% of the Cuban serves were followed with a German reception. The transition matrix itself provides for a good description of the game. For example, a comparison between the two teams of the transition probabilities between the attack and block states demonstrates that the Cuban victory was largely met without challenge.

If the properties of stationarity and independence for Markov chains are present in the sport under investigation then the calculus developed for finite Markov chains (see Kemeny and Snell, 1976) can be used to obtain useful data. Examples from tennis (see Lames, 1991, for further details) include the average and standard deviations of the rally lengths, the expected number of times the Markov chain process resides in a certain state during a rally, and the probabilities of reaching a certain final state when starting at different transient states. The latter example provides interesting results for performance analysis. For instance, if we take the starting state in volleyball as Serve CUB and the final state as Point CUB, then the combination of possible event sequences between the starting to the finishing states provides the probability of Cuba winning the point from service, a good indicator of general performance in volleyball.

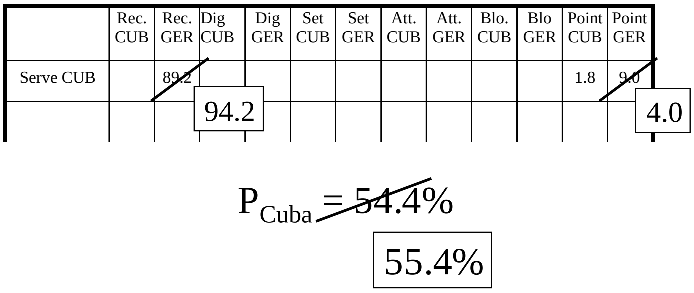

Figure 6. Simulated impact of a lower service error rate on the overall probability of scoring a point for Cuba (Lames, 1998).

The calculations suggested above may be used to investigate the validity of the model by comparing the results predicted from the model to the empirical data. Similar calculations can also be performed to allow simulations of behaviour and to investigate the expected impact of these simulations on sports performance. For example, a reduced service error rate for Cuba from 9.0% to 4.0%, together with the corresponding increase in correct services, was found to result in a predicted 1.0% increase in the likelihood of Cuba winning the point from service (see Figure 6). This finding can be interpreted as a measure of the importance of the examined tactical behaviour, an appropriate instance of the usefulness of theoretical performance analysis for sports performance. While this presented finding holds only for a single game, the simulation technique may of course be applied to a representative sample of games if desired. For example, Lames (1991) examined 153 tennis matches, Zhang (2003) examined 152 table tennis matches and Pfeiffer (2003) examined 16 junior handball matches using this technique). To further exemplify this point, the results of the impact of several tactical behaviours on volleyball performance based on 9 international female volleyball games (18 teams) are presented in Table 1a.

Action	Impact serving	Impact receiving
Service: errors	-0.63	0
Reception: errors	0	-0.64
Dig: errors	-0.43	-0.20
Set: errors	-0.25	-0.72
Attack: winners	0.29	0.72
Attack: errors	-0.39	-0.99
Block: winners	0.44	0.16
Block: errors	-0.51	-0.18

Table 1a. Impact of some tactical behaviours on individual probability for serving and receiving.

Player	Contacts	Impact
Pianka, Ines	161	9,60
Lahme, Susanne	103	2,87
Celis, Nancy	98	2,47
Naumann, Grit	95	1,52
Roll, Sylvia	100	1,25
Radfan, Constanze	1	-0,19
Wilke, Claudia	3	-0,19
Schultz, Christina	111	-1,68

Table 1b. Ball contacts and impact of in-players on overall scoring probability of the German team.

Table 1b shows another possibility for investigating the sport structure using the same method. If all the transitions of a selected player are subtracted from the initial transition matrix obtained for that team, then a transition matrix for the team without the contributions of the subtracted player is produced. Estimates of the impact of that player on team performance may then be obtained, for example by using simulation to calculate the new scoring probability of the team and subtracting the new probability from the initial probability obtained from the team matrix. Although much research is still required to provide a satisfactory method for this type of performance assessment, the point remains that simulation techniques using Markov processes may also be useful for the measurement of individual performances in team sports.

From the above considerations, it is evident that the use of appropriate models of sports performance using simulation methods permits investigations on game structure to good effect. Examples of appropriate game models include rule-based probabilistic models, not only using finite Markov chains as reported above, but also continuous Markov chains (Meyer, Forbes and Clarke, 2006), and/or models of sports behaviours that are predicated on dynamical systems theory. We suggest that a formal understanding on game structure will ultimately be developed through the methods of theoretical performance analysis as reported above.

5. Practical performance analysis

The usefulness of practical performance analysis lies in the amount of support that it provides to sports practice. In this regard, we propose two purposes of game observation to be most important; for the preparation against a future opponent and for the optimization of training. Fulfilment of these two tasks requires a coupling of information between game observation and the training process. In this regard, the coupling of competition behaviour and training was proposed by Lames and Hansen (2001) to comprise a three-step process as documented in Figure 7 below.

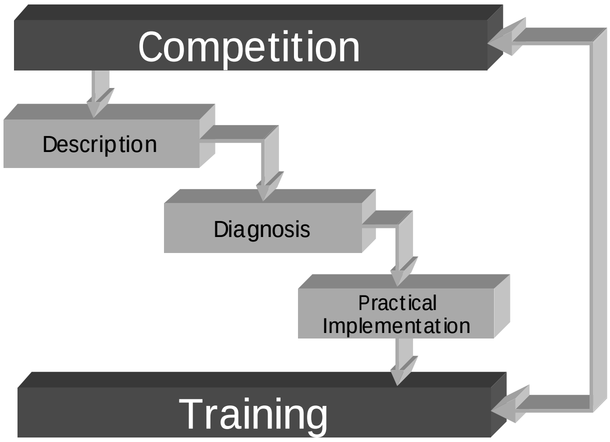

Figure 7. The coupling of competition and training by game observation as a three-step process (Lames and Hansen, 2001).

In the first step, a detailed description of competition behaviour is required using an appropriate observational system. The quality of this description depends on the reliability of the observation process as defined in the narrower sense of measurement consistency. In the second step, a diagnostic approach is used in which the information is analysed and cues detected for training purposes, particularly cues indicating weaknesses or strengths in performance. This step cannot be completed from only analysis of observational data, however, since an adequate interpretation of observational data must consider other factors. These other factors include the individual circumstances such as tactics and strategy, as well as situational aspects such as the psychological, physical and cognitive processes that occur during a game, the quality of the opponent and the level of preparation of the players.

The interpretative nature of the diagnosis step becomes more evident when attributing causes to the weaknesses and strengths observed in competition. In particular, possible causes are attributed specifically to the performance prerequisites of the individual player. For example, a high error rate in tennis volleys might be due to a detail in the execution of the specific action, or to a lack of explosive strength, or to a weak coupling between the player service and subsequent movement, or even to the length of the arm. The ambiguous, multi-causal structure of the diagnosis process requires necessarily an interpretative approach rather than an algorithmic one.

In the third step the results of the diagnosis must be transferred into practical considerations by identifying a list of possible objectives for training,. In order to do so, the issue of whether a possible objective may become a practical objective must be considered (such as arm length), as must whether the identified objective may be integrated into the ongoing training process and whether to do so on a short-term or a long-term basis. The practical implementation of the interpretation from game observation thus requires a profound knowledge in training methodology as well as a detailed involvement in the training process under consideration.

That the coupling of competition behaviour and training is an interpretive process has led some authors to apply the principles of qualitative research methodologies to this task. Qualitative research methodology exists in numerous variants, but has some features that make it well suited to practical performance analysis (Denzin and Lincoln, 1994). Such features include its applications in tackling practical problems rather than theoretical ones, its holistic and contextual approach to framing and addressing issues as opposed to the reductionist and analytic approach of quantitative methods, and its concepts of communication and intervention. Qualitative approaches adopt an interpretative and reconstructive view of social reality which is deemed appropriate for mastering the steps necessary for practical performance analysis as detailed above.

To illustrate some of the characteristics of qualitative game analysis (QGA) we report in brief on an observational method that was designed to support top-level teams in beach volleyball. The details of this approach are documented in Lames and Hansen (2001), Hansen and Lames (2001) and Hansen (2003).

The first part of QGA uses a quantitative observational system to identify the video scenes with regard to some specified parameters (e.g., service, reception, attack, result). This step is called quantitative pre-structuring and serves to prepare for the qualitative analysis to follow. In the qualitative analysis phase, the performance analyst examines the scenes selected from data base retrieval using the parameters identified in the quantitative phase. These scenes may be rallies with striking features identified from a general statistical inspection (e.g., an unusual large number of service errors), or rallies containing important tactical behaviours that are looked for as a matter of routine (e.g., the service tactics of an opposing team). At the end of this step, the performance analyst will have developed a preliminary game strategy for use against the opponent just examined. The last step of QGA is called communicative validation. In qualitative research methodology, it is typical that the scientist on one side and the participants on the other, thus the coaches and the players in this instance, communicate on equal terms. Using an iterative process, with additional support from the video scenes that are readily displayed by the observational system when necessary, an agreement on the final game strategy is subsequently reached by both sides.

Qualitative methodology provides advice for the central tasks of QGA. One idea is that the analysis of video scenes uses a method that is derived from content analysis, a common research method used in qualitative methodology (Mayring, 1993). Communicative validation is conceptualised as an iterative hermeneutic circle that is repeated until a common re-construction of the findings emerge (Guba and Lincoln, 1989). Finally, qualitative methodology provides a framework regarding how to conceptualize social interventions, which is appropriate when aiming to introduce changes in social systems.

Lames and Hansen (2001) reported a successful practical intervention with the German National Beach Volleyball teams at the 2000 Olympics at Sydney, Australia, using QGA as described. For three years one of the authors (Hansen) was involved in the preparation, qualification and, in the coaching itself during the Olympic tournament. The bronze medal was won, although the team did not finish better than 7^{th} in the world cup tournaments beforehand. Of course the authors do not claim that the intervention using QGA was the only cause for the German team success.

At this point, one might ask why the many examples of successful practical performance analysis worked in the past without the conceptual base of QGA. The answer is that every successful intervention has managed somehow to interpret observational data correctly, to communicate successfully with the coaches and players, and to intervene efficiently in this special social context. In this regard, it would seem that the principles of QGA were realised implicitly in these cases, a suggestion that is confirmed from talking with experienced game analysts. They report frequently that information has to be filtered prior to a presentation to the coaches, that there are long discussions with coaches on the meaning and consequences of the observational data, and that social acceptance of the information provided by the analyst to the coaches and players is paramount if it is to exhibit any influence on sports practice.

6. Summary

From considerations of a comprehensive notion of reliability, including not only the consistency of the measures but also the stability of the trait being measured, it was demonstrated that performance indicators in game sports would not be expected to be reliable (stable) because game sports are most appropriately conceived of as a dynamical interaction process between two opponents. In game sports, behaviour emerges as a result of this interaction process rather than it being a produced from the underlying stable properties of the athletes as is typical for some other types of sports.

There is strong empirical support for this conceptual position as well as the fact that it reflects very much the opinions held within sports practice. Speaking generally, there is much variation in performance indicators both within and between games. In short, the lack of reliability (stability) in performance indicators is compatible with the view of game sports as a dynamical interaction process, a position that encourages the search for alternative approaches to performance analysis other than that offered by classical performance analysis.

We conclude that dynamical systems theory provides a new perspective that is appropriate for advancing our understanding of the structure of game sports. While there is much research to be conducted in order to find appropriate models that are based on dynamical principles, or instead to somehow introduce dynamical considerations to existing rule-based probabilistic models of sports performance, there are encouraging signs of progress on this front. In the future, the complement of theoretical performance analysis using quantitative data, and practical performance analysis using qualitative methods, will serve to further our understanding of game structure to the benefit of sports practice.

7. References

Ap Simon, H.G. (1951) The luck of the toss in squash rackets. **Mathematical Gazette**, 35, 193-194.

Ap Simon, H.G. (1957) Squash chances. **Mathematical Gazette**, 41, 136-137.

Clarke, S.R. (1979). Tie-point strategy in American and International squash and badminton. **Research Quarterly for Exercise and Sport**, 50, 729-734.

Clarke, S.R., and Norman, J.M. (1979). Comparison of North American and International squash scoring system – analytic results. **Research Quarterly for Exercise and Sport**, 50, 723-728.

Croucher, J.S. (1982). The effect of the tennis tie-breaker. **Research Quarterly for Exercise and Sport**, 53, 336-339.

Croucher, J.S. (1986). The conditional probability of winning games in tennis. **Research Quarterly for Exercise and Sport**, 57, 23-26.

Denzin, N.K. and Lincoln, Y. (Eds.). (1994). **Handbook of Qualitative Research**. Thousand Oaks: Sage.

Döbler, H., Schnabel, G. and Thieß, G. (1989). **Grundbegriffe der Sportspiele**. Berlin: Sportverlag.
Guba, E.G. and Lincoln, Y.S. (1989). *Fourth generation* **evaluation. Newbury Park, CA: Sage.**
Hansen, G. and Lames, M. (2001). Die Qualitative Spielbeobachtung. Eine Beobachtungsvariante zur Trainings- und Wettkampfsteuerung im Spitzensport. **Leistungssport**, 31(1), 63-70.
Hansen, G. (2003). **Qualitative Spielbeobachtung im Beachvolleyball**. Köln: Strauß.
Hughes, M. and Bartlett, R. (2004). The use of performance indicators in performance analysis. In **Notational Analysis of Sport, 2nd edition** (Edited by M. Hughes and I. Franks), London: Routledge, 166-188.
Hughes, M., Evans, St. and Wells, J. (2001). Establishing normative profiles in performance analysis. **International Journal of Performance Analysis of Sport (e)**, 1, 1, 4-27.
Hughes, M., Evans, St. and Wells, J. (2004). Establishing normative profiles in performance analysis. In **Notational Analysis of Sport, 2nd edition** (Edited by M. Hughes and I. Franks), London: Routledge, 205-226.
Kemeny, J.G. and Snell, J.L. (1976). **Finite Markov Chains**. New York: Springer.
Lames, M. (1988). Techniktraining im Tennis durch Computersimulation. In **Sportspiele: animieren – trainieren** (Edited by R. Andresen and G. Hagedorn), Ahrensburg: Czwalina, 181–191.
Lames, M. (1991). **Leistungsdiagnostik durch Computersimulation: Ein Beitrag zur Theorie der Sportspiele am Beispiel Tennis**. Frankfurt, Thun: Harry Deutsch.
Lames, M. (1992). Zum Problem der Stabilität von Wettkampfverhalten im Sportspiel Tennis. In **Methodologie der Sportspielforschung** (Edited by G. Hagedorn and N. Heymen), Ahrensburg: Czwalina, 31–41.
Lames, M. (1994). **Systematische Spielbeobachtung**. Münster: Philippka.
Lames, M. (1998). Leistungsfähigkeit, Leistung und Erfolg – ein Beitrag zur Theorie der Sportspiele. **Sportwissenschaft**, 28, 137-152.
Lames, M. (in press). Modelling the interaction in game sports – relative phase and moving correlations. **Journal of Sports Science and Medicine**.
Lames, M. and Hansen, G. (2001). Designing observational systems to support top-level teams in game sports. **International Journal of Performance Analysis**(e), 1(1), 85-91.
Lames, M. and Hohmann, A. (1997). Zur Leistungsrelevanz von Spielhandlungen im Volleyball. In **Integrative Aspekte in Theorie und Praxis der Rückschlagspiele** (Edited by B. Hoffmann and P. Koch), Hamburg: Czwalina, 121-128.
Lames, M., Hohmann, A., Daum, M., Dierks, B., Fröhner, B., Seidel, I. and Wichmann, E. (1997). Top oder Flop: Die Erfassung der Spielleistung in den Mannschaftssportspielen. In **Sport-Spiel-Forschung Zwischen Trainerbank und Lehrstuhl** (Edited by E. Hossner and K. Roth), Hamburg: Czwalina, 101-117.
Letzelter, H. and Letzelter, M. (1982). Die Struktur sportlicher Leistungen als Gegenstand der Leistungsdiagnostik in der Trainingswissenschaft. **Leistungssport**, 12, 351-361.
Lienert, G.A. (1969). **Testaufbau und Testanalyse** (3. Aufl.). Weinheim: Beltz.
Mayring, P. (1993). **Qualitative Inhaltsanalyse. Grundlagen und Techniken**. Weinheim: Beltz.
McGarry, T., Anderson, D.I., Wallace, S.A., Hughes, M.D. and Franks, I.M. (2002). Sport competition as a dynamical self-organizing system. **Journal of Sport Sciences**, 20, 771-781.

McGarry, T. and Franks, I.M. (1994). A stochastic approach to predicting competition squash match play. **Journal of Sports Sciences**, 12, 573-584.

McGarry, T. and Franks, I.M. (1996a). In search for invariant athletic behaviour in competitice sports systems: An example from championship squash match-play. **Journal of Sports Sciences**, 14, 445-456.

McGarry, T. and Franks, I.M. (1996b). Development, application, and limitation of a stochastic Markov model in explaining championship squash performance. **Research Quarterly for Exercise and Sport**, 67, 406-415.

McGarry, T., Khan, M.A. and Franks, I.M. (1999). On the presence and absence of behavioural traits in sport: an example from championship squash match-play. **Journal of Sport Sciences**, 17, 297-311.

McGarry, T. and Perl, J. (2004). Models of sports contests – Markov processes, dynamical systems and neural networks. In **Notational Analysis of Sport, 2nd edition**, (Edited by M. Hughes and I. Franks), London: Routledge, 227-242.

Meyer, D., Forbes, D. and Clarke, St. (2006). Statistical Analysis of Notational AFL Data Using Continuous Time Markov Chains. In **Proceedings of the 8th Australasian Conference Mathematics and Computers in Sport** (Edited by J. Hammond and N. de Mestre), MathSport (ANZIAM), 81-90.

Morris, C. (1977). The most important points in tennis. In **Optimal strategies in sport** (Edited by S.P. Ladany and R.E. Machol), Amsterdam: North Holland, 131-140.

Nevill, A.M., Newell, S.M. and Gale, S. (1996). Factors associated with home advantage in English and Scottish soccer matches. **Journal of Sports Sciences**, 14, 181-186.

O'Donoghue, P. (2005), Normative profiles of sports performance, **International Journal of Performance Analysis of Sport (e)**, 5(1), 104-119.

O'Donoghue, P. and Ponting, R. (2005), Equations for the Number of Matches Required for Stable Performance Profiles, **International Journal of Computer Science in Sport (e)**, 4(2), 48-55.

Palut, Y., and Zanone, P.G. (2005). A dynamical analysis of tennis: Concepts and data. **Journal of Sports Sciences**, 23, 1021-1032.

Pfeifer, P.E., and Deutsch S.J. (1981). A probabilistic model for evaluation of volleyball scoring systems. **Research Quarterly for Exercise and Sport**, 52, 330-338.

Pfeiffer, M. (2003). **Leistungsdiagnostik im Handball**. Dissertationsschrift. Universität Leipzig.

Pollard, G.H. (1985) A statistical investigation of squash. **Research Quarterly for Exercise and Sport**, 56, 144-150.

Pollard, G.H. (1987) A new tennis scoring system. **Research Quarterly for Exercise and Sport**, 58, 229-233.

Renick, J. (1977). Tie point strategy in badminton and international squash. **Research Quarterly for Exercise and Sport**, 48, 492-498.

Schutz, R.W. (1970). A mathematical model for evaluating scoring systems with specific reference to tennis. **Research Quarterly for Exercise and Sport**, 41, 552-561.

Schutz, R.W., and Kinsey, W.J. (1977). Comparison of North American and International squash scoring systems. **Research Quarterly for Exercise and Sport**, 48, 248-251.

Trueman, R.E. (1977). Analysis of baseball as a Markov process. In **Optimal strategies in sport** (Edited by S.P. Ladany and R.E. Machol), Amsterdam: North Holland, 68-76.

Wright, M.B. (1988). Probabilities and decision rules for the game of squash rackets. **Journal of the Operational Research Society**, 39, 91-99.

Zhang, H. (2003). **Leistungsdiagnostik im Tischtennis**. Dissertation. Universität Potsdam.