The variable and chaotic nature of professional golf performance

Michael Stöckl and Peter F. Lamb

Department of Sports Science, University of Vienna, Vienna, Austria; School of Physical Education, Sport and Exercise Sciences, University of Otago, Dunedin, New Zealand

ABSTRACT
In golf, unlike most other sports, individual performance is not the result of direct interactions between players. Instead decision-making and performance is influenced by numerous constraining factors affecting each shot. This study looked at the performance of PGA TOUR golfers in 2011 in terms of stability and variability on a shot-by-shot basis. Stability and variability were assessed using Recurrence Quantification Analysis (RQA) and standard deviation, respectively. About 10% of all shots comprised short stable phases of performance (3.7 ± 1.1 shots per stable phase). Stable phases tended to consist of shots of typical performance, rather than poor or exceptional shots; this finding was consistent for all shot categories. Overall, stability measures were not correlated with tournament performance. Variability across all shots was not related to tournament performance; however, variability in tee shots and short approach shots was higher than for other shot categories. Furthermore, tee shot variability was related to tournament standing: decreased variability was associated with better tournament ranking. The findings in this study showed that PGA TOUR golf performance is chaotic. Further research on amateur golf performance is required to determine whether the structure of amateur golf performance is universal.

Introduction
Generally, as sports matches unfold, teams and individuals perform actions that are meant to help achieve certain objectives, given the context of the match. Each sport is made unique by the differing objectives and/or different ways of achieving those objectives. Team sports rely on an interaction between several teammates, opponents and the objectives of the match (e.g. Dutt-Mazumder, Button, Robins, & Bartlett, 2011). Individual sports can involve a direct interaction with an opponent, as in squash and tennis; these interactions, although less complex than in team sports, have been modelled as self-organising dynamical systems (McGarry, Anderson, Wallace, Hughes, & Franks, 2002; Palut & Zanone, 2005). Golf is unique to these sports in that the players do not react directly to the actions of the opponent. Golfers do, however, react to numerous constraints affecting each shot, such as: accessibility of pin locations, lie of the ball, wind conditions, and their position in the tournament. All of these factors influence the golfer’s decision making (see Glazier, in press for a theoretical overview). Subsequently, individual skills of golfers (e.g. Robertson, Burnett, Newton, & Knight, 2012), although valuable, are unlikely to predict PGA TOUR tournament outcomes because the situations golfers find themselves in, even following well planned and executed shots, are unpredictable. The ability of the golfer to adapt to the situation is likely to be another important characteristic of the top-level golfer. To understand this characteristic in better context, a better understanding of the structure of golf performance is necessary (Lames & McGarry, 2007).

Until recently, performance patterns shown during professional golf tournaments have mostly been analysed indirectly while investigating the existence of the psychological concepts of momentum or streakiness. Round scores (Clark, 2004; James, 2009) and hole scores (Clark, 2005a, 2005b; Livingston, 2012; Rees & James, 2006) have been looked at on the PGA TOUR to determine whether there was a tendency for good rounds or holes to follow other good rounds or holes. Clark (2004) found that there was a tendency for good as well as bad rounds to cluster together, whereas James (2009) found that round scores were rather independent from each other. With respect to hole scores, the support for the idea of streakiness has been weak (e.g. Clark, 2005a, 2005b; James, 2009; Livingston, 2012; Rees & James, 2006). Since round scores are comprised of hole scores and hole scores are the result of shot sequences, streaks on a shot-by-shot level – the highest level of behaviour and, thus, of performance – has not yet been investigated.

Since 2003, nearly every shot taken on the PGA TOUR has been recorded and stored in the ShotLink™ database. The existence of data describing individual shots led to the development of the ISOPAR Method (Stöckl, Lamb, & Lames, 2011), a model which allows the assessment of the quality of individual shots. The ISOPAR Method provides the performance indicator, Shots Saved, which quantifies the quality of a shot by comparing its performance to typical performance by the players in the field (field refers to players in the tournament; see next section for more details). Streakiness or momentum are difficult to define as they depend critically on the choice of
an arbitrary threshold, above or below which determines the shot’s classification. Instead we used Recurrence Quantification Analysis (Marwan, Romano, Thiel, & Kurths, 2007), a rigorous method for analysing time-series data from dynamical systems, which allows the stability of the sequence to be characterised. This study aims to analyse performance patterns in golf – the pattern of stroke sequences according to their quality – in terms of stability and variability, to better understand the nature of professional golf performance.

**Methods**

**Dataset**

The current study is based on data from the ShotLink™ database provided by the PGA TOUR, which contains information describing every shot taken on PGA TOUR sanctioned tournaments (these exclude majors and match play events) since 2003. We analysed every shot played in 32 PGA TOUR tournaments (by players who made the tournament cut) in 2011. We included only four-round stroke play tournaments, resulting in 9,152 rounds played by 308 different golfers. The shots (N = 563,354) were divided into five shot types: Tee shots (n = 164,994), Long approach shots (n = 107,023), Short approach shots (n = 14,658), Around the green shots (n = 64,621), and Putts (n = 212,058). Holed putts less than 0.46 m (1.5 ft) were excluded from the analysis because they precluded exceptional performance and could have the effect of making stable phases of exceptional performance appear more typical. The study was approved by the University of Otago Human Ethics Committee.

**The ISOPAR method**

The analysis in this paper is based on assigning ISOPAR values to each ball location in the dataset. ISOPAR values approximately represent an average number of shots remaining to hole out from a given x,y location on any hole in the above-mentioned PGA TOUR tournaments. The ISOPAR values are calculated in three main steps. The algorithm is explained in detail in Stöckl, Lamb, and Lames (2012), but a short summary is provided below:

1. A two-dimensional grid is assigned to the hole. In this study we used a mesh size of two inches.
2. At the grid nodes ISOPAR values are calculated using an exponential smoothing algorithm (Hamilton, 1994). In this study we used a smoothing parameter of 0.17.
3. Based on the ISOPAR values from the previous step a three-dimensional, continuous ISOPAR surface is generated through a cubic spline interpolation (Fahrmeir, Kneib, & Lang, 2009).

The algorithm was programmed with MATLAB 2014a using built-in procedures for some steps. Using the ISOPAR values we have defined a measure of performance for individual shots – Shot Quality (SQ). SQ describes the performance of a shot compared to the performance of the field and is defined as the difference between ISOPAR value (IPV) before and after a shot (Stöckl et al., 2012):

\[
SQ = IPV_{before} - IPV_{after}.
\]

From SQ the performance indicator Shots Saved is derived which is the difference between the SQ of a shot and the average SQ of the field (Stöckl et al., 2012)

\[
Shots\ Saved = SQ - SQ_{avg}
\]

and describes how a player gains performance advantage on the field. In this study we calculated the Shots Saved values for each of the five different shot types mentioned above. All strokes were divided into these categories and the Shots Saved values were calculated with respect to the average SQs of the field for the respective shot type.

**Recurrence plots**

A golfer’s performance in a tournament is represented by the (number of) shots played. The quality of each shot can be expressed by a Shots Saved value. Since the order in which the shots are played forms a shot sequence a golfer’s performance is represented by a Shots Saved sequence.

For this analysis, stability captures the fluctuating changes in performance throughout a tournament. We used Recurrence Plots (RP) to identify performance stability in this study. RPs visualise the similarity of states of a dynamical system using a binary system – two states are either similar or not, based on a predefined threshold. In our case RPs highlight when players perform similarly in terms of Shots Saved. If the shots in a sequence are similar to each other they constitute a stable phase.

In general, the behaviour of a dynamical system is represented by a trajectory in phase space. In our case the behaviour of the system is represented by the one-dimensional stroke series of a player’s Shots Saved values from the whole tournament or from all shots within a shot category. According to Takens (1981) the phase space needs to be reconstructed by embedding the measured trajectory into a higher dimensional phase space which “guarantee[s] the existence of a diffeomorphism between the original and the reconstructed” (Marwan et al., 2007, p. 246) phase space. The calculation of an RP is based on the respective dynamical system’s trajectory in the reconstructed phase space. To achieve the phase space reconstruction we used the time delay embedding method suggested by Marwan et al. (2007).

This method roughly consists of three steps:

1. A time delay is specified for the embedding; using the method of mutual information (Cao, 1997) we determined a time delay of 2.
2. The embedding dimension of the new phase space is determined based on the time delay from step 1. We used a false nearest neighbour algorithm (Kennel, Brown, & Abarbanel, 1992) to find the embedding dimension m = 2.
3. The recurrence matrix is calculated as
where \( \varepsilon \) is the threshold distance for assessing the similarity between two states and \( \theta \) is the Heaviside function \( (\theta(x) = 0 \text{ if } x < 0, \text{ otherwise } \theta(x) = 1) \); Marwan et al., 2007. \( \varepsilon \) was determined as suggested by Mindlin and Gilmore (1992); \( \varepsilon \) should be a few percent of the maximum phase space diameter, but not greater than ten percent of the maximum phase space diameter (Koebe & Mayer-Kress, 1992).

In order to find an appropriate \( \varepsilon \) which fulfils these constraints and is valid for all players’ stroke series, we determined \( \varepsilon \) as the mean of the different thresholds of 74 golfers from THE PLAYERS Championship in 2011. Each player’s threshold was determined as 10% of the diameter of the phase space trajectory excluding extreme performances, which resulted in \( \varepsilon = 0.14 \). The RP visualises the recurrence matrix colour-coded – a black point represents the similarity between two states and is called a recurrence point, while a white point represents non-similarity between two states.

To find stable phases of performance an RP was calculated for each player in each tournament. The recurrence point structure in an RP allows one to draw inferences on the investigated golfer’s behaviour (see Figure 1). Vertical lines indicate that the investigated system remains in similar states for a time period corresponding to the line length. The presence of vertical lines in the current analysis indicates that the Shots Saved values achieved stability for a time period indicated by the length of the line. Therefore, vertical lines were used as the measure of stability. We chose a minimum vertical line length of three to reduce the likelihood of accepting chance recurrences of only two strokes. Intuitively, stable performance over a time period might suggest good performance; however, the RP does not discriminate between good and poor performance. Therefore, determining the quality of the performance constitutes a follow-up step. The following measurements on vertical lines were considered in our analysis:

- Proportion of shots composing vertical lines
- Average length of the vertical lines (referred to as trapping time in Marwan et al., 2007)
- Quality of stable phases was determined as the mean Shots Saved value of the strokes forming the respective vertical line.

Finally to assess the structure of professional golf performance as deterministic or chaotic, we used the determinism (DET) measure. Determinism describes the proportion of recurrence points that form diagonal lines in an RP. By investigating the diagonal lines (with a minimum length of two) and their average length \( \langle L \rangle \) the predictability of the system can be inferred. Webber and Zbilut (2007) state that short diagonal lines with the determinism measure tending to 0% indicate a chaotic system, while determinism of 100% and long diagonal lines would indicate a periodic system, which at any point is predictable. Determinism does not relate to stability, but the underlying structure of golf performance.

For the RP analyses, the shot categories short approach shots and around the green shots were combined to a category called short game, so that there were enough shots for meaningful RPs. All RP calculations were conducted using the CRP toolbox for MATLAB (Marwan et al., 2007).

**Statistical analysis**

Statistical analyses were conducted with respect to stability as well as the variability of the golfers’ performances. Descriptive statistics, mean and standard deviations, of the RQA measurements and Spearman correlations between the tournament ranks and the RQA measurements were calculated to explain the stability of performance.

Performance variability was measured as the standard deviation of the Shots Saved values. Variability for each golfer was calculated as the average of the standard deviation for each tournament played. Because the proportional odds assumption required for regular ordered logistic regression was violated, a generalised ordered logistic regression was calculated to analyse whether the variability in a certain shot type had significant influence on a player’s ranking in the tournament.

The alpha level was set to 0.05 for all statistical test procedures. Statistical procedures were conducted using SPSS (IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp).

**Results**

**Stability**

Figure 1 shows the RP for a golfer’s performance at the Arnold Palmer Invitational in 2011. The RP is fairly typical for PGA TOUR golfers; it contains many isolated points with very few vertical and diagonal lines. His play from the 10th to 13th holes in round one (starting at around shot 35), however, shows a block with a higher recurrence point density. In this section there is also a greater number of vertical lines, compared to the rest of the plot. He made pars on each of those four holes and showed fairly typical performance (according to the Shots Saved values).

![Figure 1. Recurrence plot for a golfer's performance at the Arnold Palmer Invitational in 2011.](image-url)
The mean length of a stable phase was roughly the same for all shot categories (range 3.3–3.7) and only slightly longer than the minimum cut-off, standard deviations were also quite similar across shot categories (range 0.7–1.2; see Table 1). The longest stable phase was 16 shots, which occurred, both, in the Putts category, as well as the All Shots category. These long stable phases did not represent particularly good or poor performance, with average Shots Saved values of 0.05 and –0.01, respectively. The stable phases described in Table 1 account for nearly 10% of all shots played. In general, the length of stable phases was not correlated with the tournament rank ($p = .00, P = .665$).

Table 1 also shows the descriptive statistics for the quality of stable phases represented by the mean Shots Saved value of the strokes that form stable phases. On average the quality of stable phases is close to zero for all shot categories. However, the stable phases in Putts and Long approach differ from zero slightly, being positive and negative, respectively. The distribution of Shots Saved values of stable phases (Figure 2) shows that most of the stable phases tend to consist of typical performance. The quality of the stable phases is positively correlated with tournament performance (low tournament rank indicates good performance e.g. the leader is rank 1) with the exception of Putts (Table 1). In particular, for Tee shots ($p = -.17, P < .001$) and Long approach shots ($p = -.14, P < .001$) there is a weak but significant trend indicating that better ranked players had stable phases of higher quality. None of the shot categories show a significant correlation between stable phase length and quality.

About 9% of the strokes played were part of stable phases for the categories All shots, Tee shots and Putts; fewer strokes comprised stable phases for the categories Long approach (3.0%) and Short game (2.4%) (Table 2). With the exception of Tee shots ($p = -.08, P < .001$), the proportion of shots that were part of stable phases was not correlated with tournament rank. As more Tee shots were part of stable phases, players tended to rank better in the tournament standing, although the correlation was very weak.

The proportion of all recurrence points that form diagonal lines is relatively small (range 9.0%–13.8%; Table 3). Further, the existing diagonal lines are short (range 2.0–2.4), which, according to Webber and Zbilut (2007), indicates that the Shots Saved values from the PGA TOUR in 2011 are quite chaotic.

### Variability

Variability in performance across all shots was of 0.32 (Table 4). The Spearman correlation between the tournament rank and the variability value showed that as variability increases tournament rank becomes worse ($p = .13, P < .001$), the correlation, however, is quite weak.

Performance variability across different shot types was similar (Table 4). Variability was roughly the same across all shot types (∼0.3 Shots Saved) and very close to the variability based on All shots. The generalised ordered logistic regression model was not able to explain much of the variance in the tournament ranks because the Nagelkerke-$R^2 = .054$ was so small.

<table>
<thead>
<tr>
<th>Table 1: Descriptive statistics for the length of stable phases across all tournaments.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All shots</strong></td>
</tr>
<tr>
<td><strong>Mean (SD) stable phase length</strong></td>
</tr>
<tr>
<td><strong>Spearman (p) correlation</strong></td>
</tr>
<tr>
<td><strong>Shots Saved</strong></td>
</tr>
<tr>
<td><strong>Mean (SD) stable phase quality</strong></td>
</tr>
<tr>
<td><strong>Spearman (p) correlation</strong></td>
</tr>
<tr>
<td><strong>CI</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>CI</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

CI = 95% confidence interval.
The analysis of stability showed that relatively short, stable phases of performance occurred fairly infrequently. Stable phases refer to a series of shots, which is relatively robust to perturbations and, thus, do not necessarily imply good performance. The mean Shots Saved values for shots that occurred during stable phases was close to zero, meaning stable phases tended to consist of fairly typical performance, although the Shots Saved distribution is skewed slightly toward negative values. This could suggest that players are more likely to find themselves in stable phases of poor performance than good performance. Although the negative Shots Saved values are still very close to zero, we speculate that a professional golfer’s expectations of performance are much higher than ‘average performance’ and therefore, a stable phase of slightly lower than average performance could be perceived as a bout of very poor performance. Data on golfers’ perceptions of their performance would be helpful in substantiating these interpretations.

We should emphasise that stability analysis is not the same as hot hand, momentum or streakiness, as found in the literature (Clark, 2005a, 2005b; Hughes et al., 2015; James, 2009; Livingston, 2012; Rees & James, 2006); although extended stable phases of exceptional performance could be compared to the so-called hot hand phenomenon. With this in mind we found very little evidence of extended stable phases of exceptional performance from any players in any tournaments in the 2011 PGA TOUR season. For example, considering only stable phases lasting at least nine shots (there were 87 in 2011) average Shots Saved was 0.01 ± 0.03, with maximum and minimum mean Shots Saved value for the stable phase of 0.1 and –0.06, respectively. Here the best performance was by a player at the Valero Texas Open in round 2; the stable phase included shots on holes 11–14 and lasted for ten shots. He went par – par – par – birdie on those holes, which, as the best performance in a stable phase, does not suggest that “hot streaks” on a shot-by-shot basis are common on the PGA TOUR.

This analysis also shows that shot series’ performed by PGA TOUR players are not particularly robust to the perturbations

### Table 2. Determinism for different shot categories.

<table>
<thead>
<tr>
<th>Mean</th>
<th>Short Game (n = 6808 rounds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DET (%)</td>
<td>All shots (n = 6808 rounds)</td>
</tr>
<tr>
<td>L</td>
<td>13.8</td>
</tr>
<tr>
<td>L</td>
<td>13.0</td>
</tr>
<tr>
<td>L</td>
<td>9.0</td>
</tr>
<tr>
<td>L</td>
<td>12.8</td>
</tr>
<tr>
<td>L</td>
<td>8.5</td>
</tr>
</tbody>
</table>

### Table 3. Proportion of shots comprising stable phases.

<table>
<thead>
<tr>
<th>Mean % (SD, max)</th>
<th>Spearman Correlation (p) (proportion vs. rank)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All shots</td>
<td>9.0 (3.8, 23)</td>
</tr>
<tr>
<td>Tee shots</td>
<td>8.7 (7.0, 36)</td>
</tr>
<tr>
<td>Long approach</td>
<td>3.0 (5.0, 40)</td>
</tr>
<tr>
<td>Putts</td>
<td>8.2 (6.0, 32)</td>
</tr>
<tr>
<td>Short game</td>
<td>2.4 (5.2, 50)</td>
</tr>
</tbody>
</table>

CI = 95% confidence interval.
imposed by the changing constraints from shot to shot. This finding is not unique to players low on the leaderboard or players at the peak of their performance; rather the lack of stability found should be considered part of the structure of golf performance. The diagonal lines on the recurrence plots, quantified by their proportion of points (determinism) and their length, shows us that golf performance is quite chaotic as opposed to being predictable. This can potentially have important implications for players and golf psychologists – if a player finds him or herself in what they consider a stretch of poor performance they should feel assured that it may be that their perceptions are incorrect since stable phases of poor performance are quite rare. The shots they consider to be poor are not likely as poor as they perceive, relative to the field. By improving their outlook they may be able to improve their performance on average, accepting the fact that the stability of their performance is likely to be unaffected.

The chaotic nature of golf performance may be unique to golf because of all the factors unique to the sport. For example, McGarry, Khan, and Franks (1999) suggested that in squash a steady state of performance could be perturbed by exceptional or poor performance by one of the competitors. Golfers do not react to the actions of their opponents in nearly the same way as in net/wall games such as squash. Accordingly, golfers are often instructed to play their own game (e.g. Stockton & Rudy, 2014) and not take notice of the performance of their opponents. In golf there is also more time in between shots, compared to squash, for the player to recover from the excitement of a good shot or the disappointment of a poor shot. Furthermore, the ISOPAR method of quantifying golf performance helps standardize performance, so that, a difficult shot that gives the player little chance to get the ball close to the hole does not preclude exceptional performance. Rather, relative to the average, a difficult shot does not have to end up as close to the hole as an easy shot to be given the same quality. In other sports where the outcome depends on other factors than the performance itself, it can be difficult to determine whether an exceptional performance was the result of a player’s good performance or the poor performance of the opponent.

The average variability for each player’s Shots Saved values for each tournament was fairly consistent across all shots and between the shot types. However, variability was slightly more variable for Tee shots and much more variable for Short approach shots, which indicates a greater spread around the mean variability score for Tee shots and Short approach shots. Importantly, variable performance for the Short approach category did not correlate with tournament ranking. Variability in Tee shots performance, on the other hand, was strongly related to tournament rank, showing that less variable tee shot performance tends to be rewarded with better tournament standing. In other words the leaders are less variable, shot to shot, on Tee shots than players who finish lower in the rankings, whereas variability for Short approach shots is consistently higher compared to other shot types for all players, regardless of their position in the tournament. For instance, an exceptional tee shot can gain the player about 0.5 shots on a risk-reward hole (see 18th hole at Pebble Beach Golf Links in Stöckl et al., 2012); however, a poor drive finding a hazard or out of bounds costs the player much more. The best players tend to make small gains on the field for most tee shots and according the variability analysis here, they limit the number of very poor shots, compared to players who finish further down in the standings. Moreover, we have previously shown that driving performance, according to mean Shots Saved, was most highly correlated with money earnings in 2011 on the PGA TOUR ($p = .61$) and in THE PLAYERS tournament ($p = .57$; Stöckl, Lamb, & Lames, 2013) compared to other shot types. Similarly, Broadie (2012) also showed that long-game performance, which consists of shots longer than 90 m to the hole, explained most of the variability in PGA TOUR scoring from 2003 to 2010. Again, the implications of these findings relate to how players perceive their performance. A player may perceive his or her performance on short approaches as being inconsistent; however, inconsistency for this shot type is typical and has a minor effect on their place in the tournament.

**Limitations**

We can only generalise our results to professional golfers – specifically, those on the PGA TOUR. We expect amateur golf performance to be much more variable than professional golf performance, however, it remains to be seen whether the chaotic nature of golf performance is inherent to the sport. In future work we look to gain amateur golf performance data, which will allow us to better understand characteristics of amateur golf performance.
Conclusions

This study is the first to have looked at performance stability in golf on a shot by shot basis. By looking specifically at stable phases of exceptional performance we found very little evidence for the “hot hand” phenomenon in golf. We have also shown further support for the importance of long game performance in PGA TOUR golf; specifically that low variability in tee shot performance is related to better tournament performance. In future research we look to collect performance data from amateur golfers so that comparisons with PGA TOUR golfers can be made. In particular we would like to know whether golf performance, in general, is chaotic or if this is specific to professional golf. Future work should also look at whether golfers’ abilities to adapt to certain constraints can be effectively quantified.

Acknowledgements

We thank the PGA TOUR for providing us with access to the ShotLink™ data used in this study.

Disclosure statement

No potential conflict of interest was reported by the authors.

References


