Middlesex University

Applied Performance Analysis in Canoe Slalom

A thesis submitted to Middlesex University

In partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Julia Wells

Faculty of Science and Technology

Middlesex University

October 2018

Abstract

A needs analysis of canoe slalom (coach and athlete led) resulted in a reliable performance analysis system for training and competition. This was achieved using a procedure recommended by O'Donoghue & Longville (2004). Reliability tests concluded that errors for split-times and upstream analyses were unacceptably high if coaches or less trained analysts collected data due to the inconsistent application of operational definitions.

Study 2 used percentage times off the K1M and class winners as a measure of performance. Variability was high due to athlete, course and class differences. Race percentages from the 2013-16 Olympic cycle were used to test the probabilities associated with winning, medalling or reaching a final calculated from the 2009-12 Olympic cycle. Signal detection theory determined an appropriate balance between the risk of misses and false alarms (inversely proportional) with results supporting the use of race percentages off the class winner, rather than the K1M, and a 50% level of probability for predicting gaining a medal i.e. a low rate of false alarms (maximum 6%) and a high hit rate (over 70% of medals correctly identified).

Study 3 tracked athletes' performances over time using exponentially smoothed ICF race points. Performance funnels were created for winning (previously won a major championship) and winless athletes using the median and 95% confidence intervals for the median. Time series plots for an athlete (from the start of their International career) were synchronised with the performance funnels to allow easy visualisation of performance. Nearly all athletes' time series depicted a period of initial improvement

i

followed by a plateau and then deterioration in performance over a 7½ year period. Athletes were also classified into probable, possible and unlikely to win a future major.

This thesis provided coaches and athletes with academically rigorous methodologies to aid their understanding of canoe slalom performance.

Acknowledgements

Peer Reviews

Wells, J., James, N., Hughes, M. & Caudrelier, T. (2012, July). A needs analysis of canoe and kayak slalom to identify a coherent performance analysis strategy. Poster session presented at the World Congress of Performance Analysis of Sport IX, Worcester, UK.

Wells, J. (2013, December). Measuring performance in training – a case report in canoe slalom. Poster session presented at the English Institute of Sport Conference, Hinckley, UK.

Hughes, M., Hughes, M.T., James, N., Wells, J & Murray, S. (2015). Performance profiling. In M. Hughes & I.M. Franks (Eds), Essentials of Performance Analysis in Sport, 2nd Edition (pp. 327-331). London, UK: Routledge.

Wells, J. & James, N. (2017, October). The likelihood of an athlete medalling or not in canoe slalom based on predicted race percentages. Presented at the ISPASA Conference, Langkawi, Malaysia.

Ethics Approval

Middlesex University London

School of Health and
Social Sciences
The Burroughs
London NW4 4BT

www.mdx.ac.uk
Main switchboard:
020 8411 5000

To: Julia Wells

Date: Thursday, 27 September 2012

Dear Julia

Re: **902** " Performance Analysis in Canoe Slalom " Supervisor: Nic James Category:
A1

The ethics subcommittee (Health Studies) considered your application on Monday 14th of May 2012. On behalf of the committee, I am pleased to inform you that your application has been approved. However, please note that the committee must be informed if any changes in the protocol need to be made at any stage.

I wish you all the very best with your project. The committee will be delighted to receive a copy of the final report.

Yours sincerely

Prof. Gordon Weller
Chair of Ethics Sub-committee (Health Studies)

Contents

List of Tables

List of Figures

9

10

List of Appendices

11

16

Chapter One: Introduction

1.1 The scope of the study

At elite levels of sports performance, the coaches have a wide access to performance analysis (PA) support systems, for example through the English, Scottish, Welsh and Irish Institutes. Carling, Reilly, & Williams (2009) stated that if athletes are to attain world-class levels of performance, information from a continuous assessment of training and competition must be made available to aid in the evaluation of how athletes are performing and progressing. Many of the analysts within the English Institute of Sport (EIS) are full-time within a sport which allows the practitioner to fully develop their role within a National Governing Body and benefit from a multitude of expert analysts from other sports within the network. An example of the role of PA service provision within the EIS:

- Capturing video from competition and training.
- Developing performance indicators (PI's) for the sport.
- Analysing performance based on these PI's in competition and training.
- Feedback of analysis within the coaching process.
- Overseeing and managing the development of performance analysis within their sport.
- Developing relationships and partnerships with external companies working with technology companies and Universities for research.
- Delivering workshops and seminars on PA.
- Working in partnership with other sport science & medical disciplines.

The EIS performance analysts are working with a number of key Olympic & Paralympic sports. There has been a large increase in PA staff over the past 15 years

since the establishment of the EIS in 2002. There has also been an increase in posts within professional clubs such as football and rugby union. All of this is a credit to the development of the PA discipline, which is continuing to grow. In 2004 the British Canoe Slalom team invested in a full-time PA, and as a consequence of this, the practitioner was able to pursue a PhD alongside her work programme to ensure the methodologies and processes were current and progressing with academic rigour. Hughes (2004) stated that if we consider the role of a performance analyst in its general sense in relation to the data that the analyst is collecting, processing and analysing, then there are a number of mathematical skills that will be required to facilitate the steps in the processes:- i) defining performance indicators, ii) establishing the reliability of the data collected, iii) ensuring that enough data have been collected to define stable performance profiles, iv) determining which are important, v) comparing sets of data, vi) modelling performances and vii) prediction (Hughes, 2004).

Canoe slalom is an extremely challenging white water-sport, demanding skill, stamina and courage. There are 5 classes of competition: men's and women's kayak class (K1M, K1W), men's and women's Canadian (singles) class (C1M, C1W) and men's Canadian doubles class (C2M). The aim of all five classes is to run a white-water course marked by "gates" fast, and ideally without touching the gates. A "gate" consists of two poles, suspended over the water. There are upstream gates and downstream gates which are distinguished by the following colours: Green and white gates are negotiated in a downstream direction; red and white gates are upstream. The gates are placed so that the athlete must manoeuvre through cross-current waters, eddy's and waves. A touch is penalised with 2 seconds and added to the athlete's time. Therefore, if an athlete touches

a pole with anything - paddle, boat, buoyancy aid, helmet or any part of the body - a 2 second penalty is imposed. Missing a gate or going through the gate in the wrong direction costs an athlete 50 seconds. This usually means "game over" in serious competition. Therefore, the task is a trade-off between a fast run and a clean run. The gates are positioned to test the competitor's skill in using the water and coping with the demands of the water. There will be an upstream gate to test the ability to break out into the eddy behind a rock, then a downstream gate the far side so that the competitor must surf a wave to reach it before the river pushes the competitor past. It takes skill, as well as speed. The athletes must select and paddle a line that turns the current to their advantage. Therefore, the ability to read the water is essential.

At International competitions there are qualifying heats (x 2) and finals (semi-final and final). The slalom courses around the world are either natural or artificial. The majority of International and Olympic racing is competed on artificial slalom courses. The rules have changed dramatically over the past 10-20 years with regards to race format and equipment requirements. The running time for International canoe slalom has dropped by around 50% from an average 200s to approximately 90-120s. This has had an impact on the physical components of slalom racing and has also resulted in a much more attractive event for athletes and spectators. At many venues now, spectators can see most of the race from start to finish from the viewing gallery. This has also had a positive impact on delivering performance analysis support. Capturing video of full runs of a performance (60s intervals = 2 paddlers on the water at one time) is challenging in the outdoor environment. However, this is possible as a result in more staffing, cable length issues, reduction in video file size and ease of turnaround of video for coaches. Coaches

and performers can use video recordings to discuss reasons for certain behaviours, reinforce positive actions and identify areas that need improvement (Robertson, 1999). Video recordings can also be a useful tool in establishing and checking intra- and inter-observer reliability. Treadwell and Lyons (1997) also stated that video enables researchers and sport performers to share an understanding of performance. Data is also rich in information and combined with video can be extremely powerful in sport. Understanding what the sport needs, when and why is key to impacting performance and performance analysis can play a crucial role in this process in sports such as canoe slalom.

From the perspective of this thesis a number of issues related to canoe slalom were deemed influential in regard to how the research could develop. Races take place on courses that are significantly different from all others and each can offer a different challenge on different days due to the flow and currents in the water, gate placement etc. Substantial between race variation, in terms of race time, was therefore expected and meant that evaluation of race performance problematical. World records cannot exist in this sport, as a consequence of these course differences, meaning that performance evaluation is typically determined as the relative race time measured against the other competitors in the class or even different classes. As the K1 men's class is the fastest class many coaches use the winning time in this event as the benchmark against which race times are compared. However, the validity of this has not been questioned although some coaches suggest comparing race times against the same class winner is a better approach. The method for evaluating the overall race times was therefore a key objective of this thesis although it was also recognised that this is an outcome measure and the

processes undertaken to achieve this time are probably more valuable from a coaching perspective. However, the ability to measure these processes accurately has been a challenge due to inadequate equipment and manpower. For example, one process suggested as being very important for successful canoe slalom performance is the type of stroke used. However, it is not a simple task to measure or indeed categorise this because much of what happens during a race takes place under water and to the naked eye or even under examination of video some subjectivity is inevitable. Factors such as blade velocity, force and direction require sophisticated technology, which is currently unavailable, to measure. Consequently, a primary aim of this thesis is to determine what process variables can be measured with sufficient accuracy, given technology limitations, to provide meaningful performance information to aid the coaching process with regard performance improvement. Finally, world rankings were introduced in 2006 to provide a measure of performance based on a rolling two-year cycle. These rankings provide coaches and athletes with a relatively insensitive rating as poor performance over a period of time may not affect the World ranking because only the five best race results contribute to the ranking. With lottery funding based on the number of medals a sport achieves at Olympic and World championships, some understanding of potential to medal has become an important factor for British canoe slalom. Hence some examination of the usefulness of the world ranking, and the variability of race performances, was deemed an important goal of this thesis.

1.2 Aim of the research

Given the relative novelty of scientific research in canoe slalom at the outset of the study, the main aim of this PhD was to determine robust methods of performance analysis to better inform the coaching process in terms of reliable data capture, race strategy and team selection.

1.3 Objectives of the research

At the initialisation phase of this study canoe slalom had limited research within performance analysis and therefore a clear gap and opportunity to explore new and novel ideas to apply performance analysis techniques that could be returned back into the applied front line of duty. 3 main objectives were identified:

1. Conduct a needs analysis on the sport of canoe slalom and develop a reliable performance analysis system for coaches and analysts to adopt in their day to day delivery.

2. Explore the use of winning times in all classes for assessing other performances.

3. Develop a methodology that can track individual performance over time and determine success or failure.

Chapter Two: Review of Literature

2.1 Performance analysis applications to canoe slalom

Messias, Masselli dos Reis, Ferrari & Manchado-Gobatto (2014) identified just 21 publications involving canoe slalom (period analysed: 1971- July 2013) within 7 different scientific disciplines (Figure 2.1).
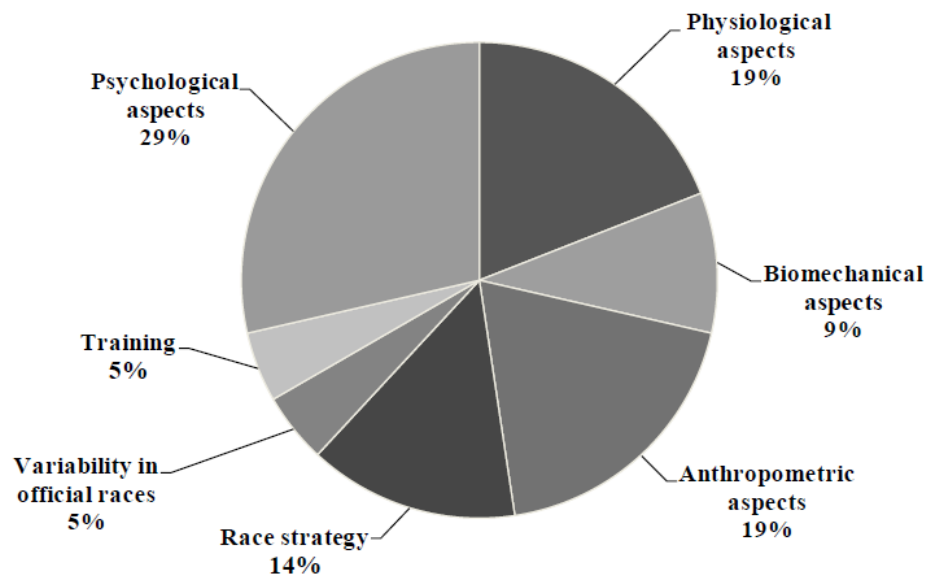
Figure 2.1   Percentage distribution of scientific articles involving canoe slalom indexed by the main topics investigated (Messias et al., 2014)

The number of publications of direct relevance to this thesis is much smaller however. Lyons (2005) mentioned some canoe slalom analysis projects at the Australian Institute of Sport (AIS) when describing brief examples of performance analysis in a variety of sporting applied contexts. This research aimed to address (1) The quantity of each type of paddle stroke being used, (2) The time spent working upstream and

downstream and (3) Effective and ineffective paddle strokes of winners and losers. An unspecified software system was developed to analyse canoe slalom both in training and competition at the International course in Penrith, Australia. It seems likely the work mentioned in Lyon's paper was related to the Hunter et al. research projects discussed later.

Hunter, Cochrane & Sachlikidis (2007) assessed intra-observer and inter-observer reliability of data gathered from a lapsed-time time-motion analysis system for canoe / kayak slalom (discussed further in section 2.3 of the literature review). Hunter, Cochrane & Sachlikidis (2008) also published an analysis of this competition which quantified the differences between groups of elite canoe slalom athletes based on the class they paddled in and the strategies used in the 2005 World Championships in Penrith, Australia. Three cameras captured the 10 fastest competition runs for men's and women's kayak and the men's canoe. Gate split-times, turn times, total stroke information, left and right stroke information and gate errors were extracted for analysis. Unsurprisingly the top ten athletes in the women's kayak were significantly ($p < 0.05$) slower than the top ten men's kayak and men's canoe who were not significantly different to each other. This was easily determined by total run time although there were some clear differences within certain gate sequences which can be determined by strategy, but also the different physical and technical demands of different classes. The splits were determined by times between each gate which has the danger of losing some context of the move when determining a time loss / gain versus another athlete. This could lead to the possibility of combining splits based on tactical and technical decision-making to account for the whole move sequence. In Hunter et al.'s paper the strategy chosen was determined by 'sided'

moves, which can be determined by left & right handedness and can be viewed as an advantage over the other. To keep left and right biases to a minimum the International Canoe Federation identify the following course criteria for International competition:

20.1 The course must be entirely navigable throughout its length and provide the same conditions for right-handed and left-handed C1 and C2 paddlers. The ideal course should include:

1. Minimum one gate-combination, which offers the competitor several options.

2. Constant direction changes and flowing movements using the technical difficulties of the water (eddies, waves and rapids).

Extract from ICF rules 2015, p29

http://www.canoeicf.com/sites/default/files/icf_canoe_sprint_2015_0.pdf

Within Hunter et al.'s (2008) study they stated that the technical sections caused the greatest differences between athletes even within the top ten. This is a combination of technical and physical capabilities to perform such moves and it was highlighted that the athlete should spend most of their time on these technical gate sequences on white water. There were also examples of a gate sequence which incorporated four possible moves, all with varying pro's and con's on time loss / gains and it was stated that further investigation would be needed to truly determine the exact cause of these time differences. A strategy decision is based on, what are the fastest moves and whether an athlete is capable of performing the move. This is an area where performance analysis

techniques could be applied in competition to inform an athlete's strategic plan combined with coaching intelligence from training on whether the athlete can perform the moves.

Within canoe slalom racing a course design has 6 upstream gates (new rule 2013) that the athlete has to negotiate with the rest of the gates being downstream gates (up to 25 gates in total). In Hunter et al. (2008) elite canoe slalom coaches highlighted the upstream gate as a key component to analyse winning. This was identified as a 'turn' and was divided into 4 sections (Figure 2.2). There was no clear reason for why the 'turn' in the upstream gate was separated into the respective sections.
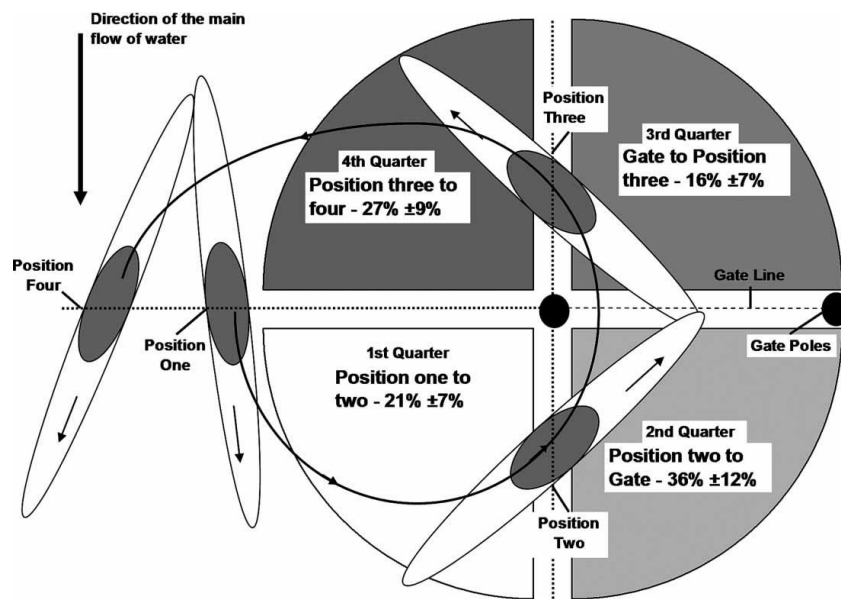
Figure 2.2   Division of time around an upstream gate (from Hunter, Cochrane & Sachlikidis, 2008)

The major limitation of this method was that each upstream gate is unique with different approaches (1st quarter) and exits (4th quarter) from the gate determined by the position of previous gates and the next gate required to negotiate the course. This means that it is impossible for the researcher to reliably distinguish when an upstream has started and ended (definition of terms), this problem is exacerbated when the other moves are incorporated into the timing splits. The 2nd and 3rd quarter events were the most repeatable due to the location of the gates in relation to the boat. However, the accuracy of the analysis of splits would have been severely compromised by using a 3-camera view of the race course. More cameras would improve the situation although the complexity of this type of analysis suggests that more accurate measuring devices such as timing gates or local positioning systems are needed for this type of analysis. Whilst the accuracy of Hunter et al.'s (2008) study has to be questioned their finding that athletes spent 16%, 36%, 21% and 27% in quarters 1-4 of the turn was the first attempt at distinguishing a potentially important measure of canoe slalom performance but, as the authors suggested, further research is necessary, using more accurate equipment.

Hunter (2009) went on to study the upstream gate further by determining how the path chosen by the elite slalom athletes influenced the time taken to negotiate the upstream gate with the aim to provide critical information on technical characteristics that were beneficial to performance. This study focused on kayak (n=11) and canoe (n=6; 5 right handed, 1 left handed) men athletes and interestingly the left handed paddler's data was included because analysis revealed no differences between the strategies used with respect to the variables analysed. Total time was measured as the time taken for the head to travel from positions 1 to 4 and boat trajectory was defined as the mean distance

between the paddler's head and the inside pole between positions 1 and 4. The research found that the absolute variability of a paddler decreased as their level of skill increased (as determined by total time taken), but the percentile variation remained constant. There didn't appear to be different strategies performed across the two classes. Even within this elite population the lines of the two fastest athletes compared with those of the two slowest kayak (K1M) and canoe (C1M) were significantly different. It was proposed that athletes could produce faster times if they focused on minimizing the distance between their head and the inside pole combined with an appreciation of the risk of touching a gate. The use of the head as an indicator of performance was assessed using an overhead camera, portable calibration rig and markers attached to the body, only possible during training but not competition. However, the validity of this approach is questionable due to the small sample used, as the conclusion that the head being close to the pole was associated with better performance could have been due to the sample of paddlers being unrepresentative of paddlers in general. This small sample may also have been unduly influenced by the one left handed paddler included in this study. This paddler would have used a different technique, in comparison to the right handers even though the authors suggested his strategy was similar.

Hunter et al. (2008) also investigated the number, length and types of canoe slalom strokes used. 67-71% of strokes were forward across all classes. There was no correlation between the percentage of time that athletes spent with the blade in the water and run time. There was evidence of dominance of stroke sidedness, for example there were more right sided strokes and it was suggested that this may be an indication of underlying limb dominance or a result of the course design. They suggested there was a

need for further research into profiling courses and individual athletes. This paper acknowledged the difficulty associated with categorising a stroke type as in canoe slalom many stroke types are either a combination of stroke types, take place under water and are therefore not possible to discern or the stroke type changes during the stroke. The authors suggested defining the stroke type by the most predominant action of the stroke but in practice this is subjective and therefore unreliable.

Penalties were also investigated in Hunter et al.'s (2008) work and due to the low number of penalties in the top ten it was difficult to determine if any relationship existed between penalties and performance. This is a limitation of analysing one competition and there is scope to transfer this analysis across a number of major championships to understand trends and patterns. In a paper that did conduct analysis across competitions they stated that, clearly, it is always important for an athlete to have a ''clean run'' (Nibali, Hopkins & Drinkwater, 2011). In D'Angelo's (2013) applied work he observed athletes he coached over the race seasons and there were examples where young developing athletes easily performed clean runs and others found this quite hard to achieve. Also some athletes performed clean runs in their 1st run and then hit penalties in their 2nd run and vice versa. D'Angelo (2013) stated that in a competition where there were a high number of clean runs this might mean that the athletes were very good technically or the course was less technical in design. However D'Angelo (2013) believed this data should not be limited to physical and technical aspects of performance but should be combined with understanding the individual's psychology and philosophy.

Green (2012), in an unpublished undergraduate dissertation, applied Hunter's research to a different population of premier division canoe slalom athletes competing at

the 2011 UK Premier race being held at Cardiff's (Wales, UK) International white water course (CIWW). There were some comparisons with Hunters work in run time percentages and it could be suggested that the use of percentages could inform differences between International and National levels. Turn time produced a medium correlation to overall run time in general, however when considering individual categories, female athletes did not show this trend and Green (2012) suggested that at a National level, fast turn times did not distinguish the best athletes. However, it is important to know that these claims were based on 1 competition and limited in-depth analysis. Questions could be asked whether there were International athletes competing in this National competition? Similarly, would specific strengths and weaknesses within a nation's athletes and between the respective classes determine differences not found in this research?

Nibali et al. (2011) statistically analysed the variability and predictability of elite competitive slalom canoe and kayak performance. They also examined home advantage which was found to be small. They concluded that the variability of performance and smallest worthwhile enhancements in canoe slalom were larger than those of comparable sports and that race outcomes were largely unpredictable. They stated that one implication of poor predictability was that athletes with a low true ranking still had a reasonable chance of winning a medal; however, this could be quantified in future research. The variability was larger for the bottom ranked athletes resulting in the mean run time being longer in the final compared to the semi-final. This was a consequence of an uncharacteristically fast run in the semi-final which qualified the athlete for the final, but they were unable to repeat or improve on this run time in the final. The study

31

examined races from 2000-2007 although since 2008 the rules have changed from aggregate times from semi-final and final to a single race where the final run time determines a winner. There is scope, therefore, to investigate whether variability has changed as a consequence of the current competition format i.e. 2 Olympic cycles 2008-2016 have taken place. There is also a new class, in 2010 the ICF introduced rules allowing women's canoe singles (C1W) competitors at World Championships and in 2020 they will be performing in their first Olympics.

D'Angelo (2013) investigated 'fast' runs by determining the percentage off the winning kayak man including and excluding penalties. Excluding penalties allows the coach to monitor raw speed and the use of percentages allows a comparison of performances from competition to competition.

> "Collecting data and verifying it in detail with more objective
> methods of analysis allows me to confirm or disprove what I
> observe during training sessions".

D'Angelo, (2013)

Percentages have been largely used by coaches and National governing bodies (NGB's) such as British & Australian Canoeing to measure the performance of an individual and have been used in team selection processes. For example British Canoeing (2015) stated in their published policy "For 2015 selection qualification, appropriate performance percentages relevant to age (Senior, U23 or U20) must be achieved as per

the published performance criteria". Senior performance standards in all classes were worked on scores (accuracy 0.01 seconds) as a percentage off the winning K1M. For automatic selection an athlete needed to achieve at least three runs inside the performance standard (Table 2.1). For example, percentage calculation on a single run K1M winning time was 94.44s total and the 5th place K1W run time was 109.93 secs total, therefore the 5th place K1W percentage was 116.4%. The Australian performance standards were included in Table 2.1 for a comparison where most standards are similar except for the K1M class.

Table 2.1   Standard percentages for national team selection races

Class	K1M	K1W	C1M	C2M	C1W
GBR Senior Standard	<107%	<126%	<115%	<126%	<145%
AUS Senior Standard	<104%	<125%	<114%	<124%	<142%

With a large weighting on the use of percentages in NGB's selection criteria's research is needed to examine the variability or in applied terms the consistency of the percentages as a measure of performance from race to race. Are they valid percentages to determine performance standards? Consideration for class percentages should also be explored as there could be less variability within-class performances to establish more accurate performance standards.

2.2 Learning from similar sports

Unlike many individual sports (e.g. running and jumping events in athletics) canoe slalom does not have world records to aim for or to be measured against. This brings challenges to training sessions when determining what a good performance was, and how this could be compared with competition performance. Canoe Slalom does have some similar characteristics to the following sporting examples: show-jumping, alpine slalom skiing, cross-country skiing, motor track racing, and skeleton. These similarities span across:

- Courses venue & design changes

- Time based moves & splits to determine time loss & gains

- Turns & line choice

- Tactics & techniques for moves

- Penalties for hitting and / or missing moves within the sports rules

Arundel & Holmes (2013) compared strategies of rider / horse combinations that achieved clear rounds with those that didn't. The indicator of strategy was determined by riding time between fences and the number of strides taken directly towards fences. Working with ground times between fences they found a mean absolute error of 0.18s which was deemed sufficiently reliable. The study did not find a speed-accuracy trade off as the timings between the two groups of performers were very similar. Interestingly they highlighted the timing of a scheduled jump as a potential limitation, i.e. competitors in the second half of the show jumping schedule may hold an advantage by being able to watch opponents. The rider could therefore revise their strategy based on the visual

feedback. This is a strategy that canoe slalom competition schedules allow as athletes perform one after another and are able to watch live on the riverbank or through the broadcast signal or a recorded video clip.

In alpine skiing Kirby (2009) stated that feedback usually comes in the form of video, coach's comments and timing systems. The video and coach's comments are subjective, and the timing system only provides one or a few data points. Kirby's research focused on the development of the vLink racing computer which measures forward and lateral displacement of alpine skis and converts the displacement into real-time audible feedback. Alpine skiing has similar challenges to canoe slalom, with changing environments and a highly technical sport. Not all courses will have the use of highly sophisticated systems such as vLink and therefore it is understandable that video analysis plays a primary role for providing athletes with feedback related to technique and tactics (Kirby, 2009). From the video the ability to compare and time performance enhancements is also achieved to aid coaching feedback. The challenge is how quickly this feedback can be provided, of which 83% of subjects in Kirby's study stated that real-time feedback definitely or probably helped improve their technical skills.

In 2014 Spencer, Losnegard, Hallen & Hopkins estimated variability and predictability in cross country skiing using a mixed-linear modelling procedure. Variation was 1.1% for men and 1.3% for women suggesting a greater competition depth in the men's events. Predictability was high when all athletes were considered, however it was hard to predict placing among the top 10 due to the small spread in ability. These techniques were applied to many sports such as cycling (Paton & Hopkins, 2006), track and field (Hopkins, 2005), swimming (Pyne, Trewin & Hopkins, 2004), rowing (Smith,

& Hopkins, 2011), sprint canoeing (Bonetti & Hopkins, 2010), Skeleton (Bullock, Hopkins, Martin & Marino, 2009), and also included canoe slalom (Nibali et al., 2011) as previously discussed. The focus of the research was based on practical applications that encouraged coaches and sport scientists to focus on improvements as little as the smallest worthwhile enhancements. As with canoe slalom environmental factors have an impact on performance outcomes in skiing. Some of these studies, where applicable, were able to factor in wind direction, speed, snow conditions, race terrain and altitude. However they stated that findings were inconsistent and this could be due to subjective ratings, for example snow conditions.

Motor car racing also has similarities to canoe slalom racing, the best driver (paddler) is the one that is able to drive (paddle) on a given track (course) in the shortest possible time (Braghin, Cheli, Melzi & Sabbioni, 2008). There is a large focus on the trajectory of the car with a balance between the path and speed, and the ability of the driver to navigate this optimal plan. Determining a speed profile can result in an estimation of the lap time achievable by a specific vehicle on a given track. Given the high profile of motor sport there is no surprise that there has been a large investment into research involving numerous engineering technologies to measure performance and develop models that identify the optimal trajectory (Braghin et al., 2008). Canoe slalom would benefit from such feedback however as a public funded sport in the UK, resources for such projects could be minimal. Therefore realistically coaches can determine the best possible trajectory through video, observation and timing systems of their athletes and others on the course.

As Bullock & Hopkins (2009) identified in skeleton the tracking of race times can be problematic due to the variation in race time between races arising from differences in venues and weather rather than the athlete's true ability. They stated that sports similar to skeleton could track an athlete by using their race placing or percent behind the winning time. It was highlighted that caution would need to be taken when using percentages as differences in technical demands could increase the spread in performance times and this measure could suffer from errors such as the assumption that the fastest run time is actually the fastest or even an exceptional winning performance. Bullock & Hopkins (2009) used the percentages off the winning time to predict an athlete's outcome at the 2006 Winter Olympic Games. The main aim of the study was to compare this to the coaches' method of using placing. The results concluded that the percentage time behind the winner was superior for the men but inferior for the women. It was unclear why there was a difference between men and women and it was concluded that future research was needed as it would be unreasonable to say to coaches in the meantime that they can use the ranking method with women but not with the men. Furthermore this type of analysis using percentages off the winning time to predict an athlete's outcome during a competition could be applied to canoe slalom as a method to support decisions on race plans and outcome goals.

2.3 Reliability issues

2.3.1 Reliability in canoe slalom

In the area of applied performance analysis it is vital to ensure that data obtained is accurate and reliable within both data capture and the provision of statistics (O'Donoghue

& Longville, 2004). The importance of this is paramount as the data is used to impact on the coach's feedback to the athletes performing. The Hunter, Cochrane & Sachlikidis, (2007) study analysed gate split-times (time taken between gates), touched and missed gates (penalties), turn times, major and minor avoidance, rolls, paddle in and out of water times, and stroke categorisation. This analysis was conducted by viewing video from 2 camera positions, one covering the top section and the other covering the bottom section with an overlap between cameras to link the two video clips together as one whole run. For all time-based variables, the mean, minimum, maximum, range, standard deviation, error of measurement, and limits of agreement were calculated and used to indicate the variability and reliability of the data. For intra-observer analysis, gate split-times and turn times demonstrated a similar variation of 0.21s (limits of agreement). However, for inter-observer analysis this increased to 0.39 and 0.37s. Stroke identification, which links to technique used (a combination of strokes make up a technique), produced 78% accuracy levels that increased to 81% with the inclusion of half-correct strokes. The reason for the intra-observer analysis being more reliable than inter-observer analysis was suggested to be due to slightly different interpretations of the operational definitions and differences in the knowledge of canoe slalom among the observers. All three observers in this study were relatively inexperienced in the use of the system and the definitions used because they had only recently been developed. It is possible, therefore, that learning was still taking place and the reliability results overstated the errors for fully trained observers. Hunter et al. (2007) also suggested that the water caused visibility problems, which affected the reliability, and made direct comparisons of reliability values with other sports, and previous research, unfair. This may be true, but irrespective of circumstances,

reasonable reliability is essential if the analysis data is to be used effectively. Methods for increasing accuracy of measurement are therefore imperative, for example, by increasing the number of camera positions and adjusting the camera angles. Given all the concerns raised above, caution must be given to the accuracy of the measurements derived for this study, in particular, split times and identification of strokes. Future studies will therefore need to consider the feasibility of collecting this type of data in relation to the equipment available.

2.3.2 Reliability – learning from other research methodologies

Previous research has utilised expert coaches to validate the variables collected for analysis (Wells, Robertson, Hughes & Howe, 2004; O'Donoghue & Longville, 2004; James, Mellalieu & Jones, 2005 and Choi, O'Donoghue & Hughes, 2006). For applied research this should be the starting point for any researcher and could expand this to gaining feedback from the multi-disciplinary team and the athletes.

The majority of performance analysis research has been based on categorical data, sets of which are 'tagged' or 'coded' (notated) through computerised video analysis software programmes such as Sportscode, Dartfish and Nacsport. There are different levels of sophistication of the 'tagging' panel however there is always room for human error and therefore requires a high level of reliability testing. This is vital when the information is being used in coaching contexts to make important decisions (O'Donoghue, 2007) for instance around team selection. Intra- and inter-analyst tests should be conducted to determine the accuracy and summed data should be avoided to prevent invalid results (Hughes, Cooper & Nevill, 2002). Kappa (Cohen, 1960) has

39

become a useful reliability statistic to use with nominal variables because Kappa determines the proportion of cases where the independent observers agree excluding the proportion where they could have agreed by chance (Robinson & O'Donoghue, 2007).

For ratio data Cooper, Hughes, O'Donoghue and Nevill (2007) presented a method for assessing reliability based upon Bland and Altman's (1999) suggestions for the non-parametric treatment of comparison data and Nevill, Lane, Kilgour, Bowes and Whyte (2001) who recommended that 95% of differences should be recorded within a reference value thought to be of 'no practical importance'. This study resulted in a method that any applied analyst could adopt to assess their reliability of individual variables. The added benefit is that these Bland & Altman plots are clearly presentable on a simple *XY* graph and easily understood by the user. It was concluded that decisions about the reference value should depend upon the type and frequency of data being analysed and upon the context in which the data is to be used in practice.

2.4 Profiling performance in sport

There are clearly strengths and weaknesses to profiling which need to be considered prior to producing profiles. The dangers of having too little or too much data can prove challenging when attempting to produce profiles. Potter and Hughes (2001) reinforced the theory that the greater the database, the more accurate is the model against which to compare future performances. However, as Hughes, Evans & Wells (2004) stated, even this safeguard will have inherent disadvantages: it can be argued that as a database increases in size, it will become less sensitive to changes in playing patterns (Hughes et

al., 2004). O'Donoghue (2005) stated the following strengths could be highlighted in Hughes et al., (2004) work:

- It reduces variability due to individual match effects by basing performance indicators on multiple match data.
- It provides a systematic means of determining the number of matches (performances) required to produce stable value for a performance indicator.
- There is flexibility in terms of the limits of error that performance indicators need to stabilise within.

(O'Donoghue, 2005)

On the flip side O'Donoghue (2005) highlighted some criticisms to the Hughes et al., (2004) work, these are:

- As more matches (performances) are added, the limits of error relate the evolving mean to the eventual mean using all of the matches. By expressing the difference between these 2 values as a percentage of the eventual mean, the technique risks interpreting a meaningful difference as being tolerable.

- The terminology that has been used can be challenged. The word "normative" suggests that subject data is related to normative percentiles for the performance indicators. This is not the case with the method proposed by Hughes et al. (2001).

- Where we are studying the subject to determine a performance profile, we should recognise that some subjects will be inconsistent or erratic with respect to some performance indicators. The instability in such subjects is an important aspect to investigate and understand.

(O'Donoghue, 2005)

Moving from the initial profiling studies in 1999-2001 into work presented by O'Dongohue (2005) and James et al., (2005) it was clear to see that early methods had provided motivation to discover new methods to overcome some of the problems when profiling performance. O'Donoghue's (2005) method is applicable in sports such as athletics, swimming, cycling, canoeing, walking and triathlon where good performance is associated with low times or high distances achieved. Furthermore, in these types of sports, performance indicators are not affected by the quality of opponents as much as performance indicators in formal games would be. O'Donoghue's (2005) paper proposed an alternative method to Hughes et al., (2001) that represents not only the typical performance of a team or individual but also the spread of performances. The technique also relates the set of performance indicators for a team or individual to normative data for a relevant population of teams or individuals. There are three stages to O'Donoghue's proposed method; (1) determining normative data, (2) describing the performances of the subjects of interest and (3) relating the performances of the subjects of interest to the normative data. When determining normative data O'Donoghue (2005) used percentiles (19 percentiles from 5% to 95%) and stated that this is a powerful means of interpreting the value, however be aware that there are some variables where the lower the value the better, some variables where higher values are better and some variables where there is an optimal range of values (O'Donoghue, 2005). The study involved analysing a sample of elite level tennis players (female and male) in stage one and a number of individual players to complete stages 2.

O'Donoghue (2005) stated there are four key advantages over the technique proposed by Hughes et al., (2001), these are:

1. The technique is concise allowing all of the performance indicators to be displayed on a single radar chart.
2. The technique represents the mean as well as the variability for each performance indicator.
3. The performance indicators are related to performance norms for the population of interest.
4. The technique can be tailored to compare the typical performances of different performers or to compare different types of performance by the same performer.

James at al.'s (2005) work on profiling in Rugby Union also suggested an alternative approach to Hughes et al.'s work (2001). The method involved the specific estimates of population medians to be calculated from the sample data through confidence limits. They believed that the use of confidence limits was the most applicable methodology, particularly to the applied practitioner, in that performance profiles of individual and team behaviours could be established after collection of relatively few data sets. Interestingly James et al., (2005) stated that not all profiles would stabilize due to the variability of the data, an area considered in Hughes, Wells & Matthews, (2000) work when analysing different standards of play (in this case recreational levels produced limited stabilizing of profiles).

O'Donoghue et al. (2008) stated that the methods of O'Donoghue (2005), Hughes et al. (2001) and James et al. (2005) could be criticised for not addressing the quality of the opposition when interpreting performance indicator value. This is a reasonable point and proves that performance profiling has limitations which still need to be explored within

43

the academic world of performance analysis to move the study of profiling further so that more appropriate techniques can be adopted based on the sports own individual 'make-up'. Quality of opposition is not a new phenomenon, it has been on the 'back of the minds' of many a researcher / analyst for a long time when considering their methods. In the notational analysis literature there are a number of studies that have considered quality of opposition and also quality of performance in methods. Such studies include Hughes et al. (2000) that analysed the different standards so that profiles for each level could be distinguished; Hughes and Franks (1994) also studied squash analysing different levels of standards and decided to analyse the first three games only, even though a number of the matches contained four or five games within the match. With particular focus on level of opposition research (Hughes and Franks, 1994; Hughes et al., 2000) these papers had the potential to use these 'profiles' when analysing individual performances against these profiles but this was not the aim of the studies.

In a performance profiling review paper of James et al.'s (2005) and O'Donoghue's (2005) techniques (Butterworth, O'Donoghue & Cropley, 2013) highlighted that several different methods of profiling have been identified, each with their own qualities, however, one enduring adverse trait runs in parallel throughout each; a lack of relation to the coaching process. Interestingly the work in Squash (Hughes et al. 2000) had this at its forefront and the interaction of this research to the coaching process was the catalyst. Looking at current day this is also the case for this research thesis within canoe slalom.

In O'Donoghue et al.'s (2008) work the purpose of their investigation was to determine percentiles for British National Super League netball performances. Teams were classified as being in the top or bottom half of the league based on their finishing

position in the league table for the season of the given match and appropriate analyses were conducted based on these classification (Top v Top, Top v Bottom, Bottom v Top, Bottom v Bottom). A question arises, how would this application be used by the working analyst that would require profiles through the season instead of at the end of the season? Is there any flexibility in this approach? Do you have to update the normative profiles to ensure they are up to date and relevant to the group? As O'Donoghue et al. (2008) recommended future research with this approach on other sports needs to be explored.

The work of James et al. (2006) analysed Tiger Woods' outstanding record in the sport of golf by comparing his performances on eight performance indicators and one performance outcome, weighted scoring average. This was achieved by converting the player's performance scores into standard scores (z scores) relative to the other professionals playing on the PGA tour during a particular year. Thus z scores values of 0 represent tour average performance; values of 1 and -1 represent performance one standard deviation better or worse than tour average. If a normal distribution is assumed then 95% of players will score between -2 and +2 and 99% between -3 and +3 (James et al, 2006). They hypothesised that this form of analysis would enable a true relationship between variables to be ascertained since each value would be relative to the tour average and standard deviation for that variable. In conclusion it was suggested that analysing performance indicators relative to the other tour players provided more information regarding a player's strengths and weaknesses than correlation analysis alone. This approach can assess performance indicators and potentially be used to profile performance to identify a number of areas of interest (refer to areas highlighted below).

To summarise from the literature there are a number of areas to consider prior to producing performance profiles:

-        Is it an individual profile?
-        Is it a team profile?
-        Is it a group (standard) profile?
-        What are the levels of standards?
-        What are the criteria for the level of standards?
-        What performance indicators are being used in the profile?
-        Which profiling methods meet the needs of the data being analysed?

In the applied world once profiles are produced the analyst is suggesting that a team or individual is likely to perform based on the profile but is this ever proven and should we be stating with such commitment that this outcome will happen. With performance's being so variable in many sports, including canoe slalom should performance analysts be leaning towards probability as a method to calculate the chance of something happening. James et al., (2005) suggested that a performance profile should offer some indication of future performance. In an attempt to progress performance profiling techniques O'Donoghue & Cullinane's (2011) approach included the evaluation of players and teams, in particular quality of opponents. The functions in their analyses allowed them to determine how much better or worse a player had done than would have been expected in a match against an opponent of the given World ranking. The main difficulty within their study was obtaining enough data to do a meaningful regression modelling exercise, a common issue in some sports analysis. In canoe slalom there are limited events per year and limited athletes per nation which results in low data sets to profile performance over time.

2.5 Predicting and forecasting performance techniques

Historically, the prediction of sports performance has been a concept usually reserved for those associated with the betting culture (Hughes, 2004) however this can be widened to business, economics, politics, geology and medicine (O'Donoghue, Ball, Eustace, McFarlan & Nisotaki, 2016). Within Hughes' review it rightly stated that for athletes and coaches predictions are often made about forthcoming opponents based upon previous encounters and known traits. This human prediction is however sensitive to subjective bias and even though the expertise of the coach is valid it is unable to compete with computer computations of chance and uncertainty. There has been some research that has tested human predictions with computer-based models (O'Donoghue & Williams, 2004; O'Donoghue, Dubitzky, Lopes, Berrar, Lagan, Hassan & Bairner & Darby, 2003). In fact O'Donoghue & Williams (2004) found a far greater range of accuracies for the human predictors with the most successful human correctly predicting the outcomes of 46 of the 48 matches in the 2003 Rugby World Cup. Overall, the study provided evidence that computer-based methods are more successful at predicting the outcomes of International rugby union matches than the average human, but is not as successful as human experts.

More recently, O'Donoghue et al. (2016) found that match outcomes in International rugby union performance are more difficult to predict than in previous years. This could be a result of the competition structure. Their model also considered the quality of opposition and revealed that higher ranked teams did not do as well as predicted. Sport continues to be a challenge to predict due to its variability in nature. It was also concluded that larger data sets produced more accurate predictive models than smaller sets of more recent data. As stated previously not all sports have large data sets

that are relevant to current performance standards and therefore ineligible for statistical analysis.

Suitability of time series analysis has the potential to be explored in canoe slalom. The analysis of time series is based on the assumption that successive values in the data file represent consecutive measurements taken at equally spaced time intervals (StatSoft, Inc., (2013). De Smith (2015) however stated that the data are often, but not always, measured or defined for times that have equal intervals between them. This is applicable to canoe slalom data as the competition dates differ each year. There are a number of useful examples of why time series analysis is conducted such as the ability to visually present data, (averages, peaks and troughs, critical turning points) and to predict and forecast. Most time series patterns can be described in terms of two classes of components, trend and seasonality. Within canoe slalom there could be an interest in both. The literature stated that there are no proven 'automatic' techniques to identify trend components in the time series data however if the trend is not consistently increasing or decreasing then the suggested route is using smoothing techniques (StatSoft, Inc, 2013).

Smoothing techniques comprise of simple moving averages, exponentially weighted moving averages, double and triple exponential smoothing. Moving averages simply use more recent data and ignore the older data. For example you could use the previous 5 observations and keep rolling them as new data comes in. This does not predict peaks and troughs in the data series. The next level is to weight more recent data over past data (exponential smoothing). The application of this to sports performance is a valid one as decision-makers are interested in current form and potential for improvement

based on historical time series data. Another exponential smoothing technique devised by Holt, (1957) included an adjustment value to account for a trend in the data. The method was extended by Winters, (1960) to double exponential smoothing (Holt-Winters method) which incorporated a seasonal component to the analyses.

In sports research there are examples in athletics research that have been challenged (James, 2012) for their use of linear regression to predict that women marathon runners would run as fast as men by 1998 (Whipp & Ward, 1992). Of course, progress in world records cannot be considered to be a continuous process (Kuper & Sterken, 2007). Nevill & Whyte, (2005) suggested a flattened S shape logistic curve was best for analysing world record times in m/s, showing a slow improvement at the beginning followed by an improvement and then back to a slow rate of improvement. This model clearly makes more sense than a simple linear one since there is bound to be some limit to human performance. James's (2012) criticism of Nevill & Whyte's (2005) model was that it did not consider the limits of human performance and therefore their predictions would become more inaccurate over time. Increased professionalism in sport leads to the saturation in the sense that recent improvements are generally smaller however this may not be true in sports where technology plays an important role (Kuper & Sterken, 2007).

James' (2012) predictions in real tennis concluded that to include inherent error related to the uncertainty of future performance suggested the use of upper and lower confidence limits for any predictions. Confidence intervals can be specified to be relatively certain of successfully predicting future performance by using a high percentage certainty but this will result in a large difference between lower and upper

49

limits. It was suggested that future studies should determine the optimum trade-off between certainty of prediction and acceptable range for lower and upper limits of performance.

When presenting forecasts with control limits, funnel plots were recommended by Spiegelhalter (2005) as a graphical aid. Spiegelhalter (2005) used the funnel for institutional comparisons that were public funded services and required accountability. They argued that a suitable form of such a control chart is the 'funnel plot' in which the observed indicator is plotted against a measure of its precision, so that the control limits form a 'funnel' around the target outcome. It was stated that a funnel plot has four components, (1) An indicator, (2) a target, (3) a precision parameter and (4) control limits. These funnels had some advantages highlighted in the research of which could be adapted to sports analysis, such as canoe slalom rankings:

- The axes are readily interpretable, so that additional observations can be added by hand if desired.

- The eye is naturally drawn to important points that lie outside the funnels. • There is no spurious ranking of institutions.

- There is clear allowance for additional variability in institutions with small volume.

- The relationship of outcome with volume can be both informally and formally assessed.

- Over-dispersion can be taken into account.

- If repeated observations are made over time then it would be possible to plot repeated points and join them up to show progress

- They are easy to produce within popular spreadsheet programmes.

(Spiegelhalter, 2005)

More recently Rakow et al. (2015) used funnel plots to present quantitative risk information for major medical intervention that involved patients and family to use this information to make informed decisions around treatment. Initially the use of funnel plots was to inform clinical governance to assist managers and senior clinicians in making decisions about service delivery. With involving the public in decision making this prompted their research to examine how accurately funnel plots could be interpreted by the general public and how they use this information to make decisions about treatment. Interestingly in sport the British system which is publicly funded works closer than ever with the individual sports to ensure transparency on funding decision making and processes. This Tokyo Olympic cycle, UKSport has a dedicated Sports Intelligence team that works solely on tracking performance (Slot, 2017) to aid better decisions around the probability of winning. Funnel plots based on historical data could provide the winning edge and ensure smarter decisions are taken in investment.

Probability is an important and complex field of study (Lane & Osherson, Online statistics education). Nearly all reasoning and decision making takes place in the presence of some uncertainty (Heeger & Landy, 2010). In sport historical data can be used to understand the chances of something happening. Signal detection theory provides a precise language and graphic notation for analysing decision making in the presence of uncertainty (Heeger & Landy, 2010). Early research in the 1960s has used signal detection theory to analyse operant behaviour in Psychology (Nevin, 1969). There is also evidence of signal detection theory being used in medical decision-making and Heeger

and Landy (2010) explained its use when detecting tumour's in a patient. There are four possible outcomes, two of these are correct, a 'correct rejection' (the tumour is absent and the radiologist says "no") or a 'hit' (the tumour is present and the radiologist says "yes"). The other two are incorrect responses, a 'miss' (the tumour is present and the radiologist says "no") or a 'false alarm' (the tumour is absent yet the radiologist says "yes"). Applying historical data and testing it against future performances could be achieved by using signal detection outcomes, for example sports such as canoe slalom work in cycles and could adopt Olympic cycles. This could also help coping with small data samples and could be tested yearly for changes, meaning that current data combined with historical data could be used to inform decision-making.

2.6 Summary

The literature review has presented previous research in many different areas, each of which is relevant to this thesis. Initially it was pointed out that there has been very little research in canoe slalom and the majority of that not relevant to this thesis. Hence the need to seek out techniques and methodologies that are applicable to analysing performance in canoe slalom. This review presented the need to be reliable in data collection followed by various factors of concern in profiling performance. In essence this is what this thesis is about, collecting reliable data and analysing it to make informed decisions about performance. The conundrum is that sporting performance is unreliable, mistakes are made, but distinguishing the errors made by athletes within an environment that includes a multitude of obstacles preventing perfection, is difficult but less so if we at least measure things properly.

Chapter Three: Study One

CREATING A RELIABLE PERFORMANCE ANALYSIS SYSTEM FOR ELITE CANOE SLALOM

3.1 Introduction

Canoe slalom analysis typically involves identifying time losses or gains by an athlete in comparison to their competitors. At competition, and in training, it is common to see a number of coaching staff on the riverbank or behind a live video feed on a laptop with stopwatches and clip boards collecting time-based information to identify moves that were quicker and to provide objective feedback to their athlete on their performance. At competitions there are official timing gates, but typically these only record start, half-way and finish times. The use of more timing gates is not currently in place although these may not satisfy all coaches as different coaches want specific splits based on their own feedback needs. Hence coaches tend to have support staff recording times, using stop watches, at their specified positions with additional information such as movement type e.g. spin versus forward moves on gates sequences. This timing technique has obvious issues in relation to reliability with different people potentially providing times that are not recorded in exactly the same way e.g. when to stop the timer. Clear operational definitions for this minimises the error but individual bias and lack of precision e.g. viewing angle, contribute to the variability of the measurement. In an attempt to quantify and minimise these errors Hunter, Cochrane & Sachlikidis (2007) measured gate split-times and turn times in canoe slalom competition using a post-event time-motion analysis

system. Video from 2 camera positions, one covering the top section of the course, the other the bottom, had overlapping images so that the two video clips could be spliced together to view the whole run. Intra-observer analysis for gate split-times and turn times had similar limits of agreement (0.21s) whilst inter-observer agreements were 0.39 and 0.37s respectively. The three observers were relatively inexperienced in using the system and the operational definitions however as they had only been developed recently. Hunter et al. (2007) also suggested that the water caused visibility problems although the camera views would also have been an issue. Clearly an effort to improve the reliability of measurements in canoe slalom is imperative, possible for example, by increasing the number of cameras and adjusting the camera angles.

The management team of a National canoe slalom team decided to utilise the services of a full-time performance analyst in 2004. This thesis is a consequence of that decision. This chapter focusses on the preliminary steps taken to achieve the initial stages of providing relevant performance analysis support to satisfy the needs of both the coaches and athletes.

3.2 Methodology

3.2.1 Participants

Elite canoe slalom coaches (n=10) and athletes (n=10) who were coaching or competing in European, Olympic & World Championships provided specific objectives for the performance analyst. The coaches all worked full-time for the British canoe slalom team and had varying years' experience (1 to 10 years) with most being ex-International

54

athletes. Each coach had been employed on the basis of a World-wide reputation at the peak of International canoe slalom performance.

3.2.2 What the coaches and athletes originally did

All coaches had a clear view of how performance analysis (PA) could assist their role. They used video in training and competition to replay performances to athletes both during sessions and post-event. Some used video analysis software (www.dartfish.com) and/or Windows media player (www.microsoft.com) to aid this process e.g. slow motion, split screens. They also recorded section split times, again both in training and competition. At competitions coaches identified which sections of the course to produce split times for, based on specific moves and gate formations, and timed the top ranked athletes for analysis. These were reviewed in Microsoft Excel whilst viewing the relevant video which was manually located in Dartfish so that performances could be compared in the Analyser module. A penalty and technique analysis received little attention but was an area they wanted to develop. Athletes, however, considered only the use of video to analyse technical aspects of their performance and were not familiar with other performance analysis techniques such as the assessment of tactics. Only 30% of them used video software on their own with the reason being a lack of access to a personal computer of their own.

3.2.3 Needs analysis and the implementation of a PA data capture system for canoe slalom

3.2.3.1 A developmental approach

O'Donoghue & Longville (2004) documented a step by step procedure for developing a PA data capture (notation) system for netball, involving the collaboration between sports scientists, coaches and athletes (Figure 3.1).
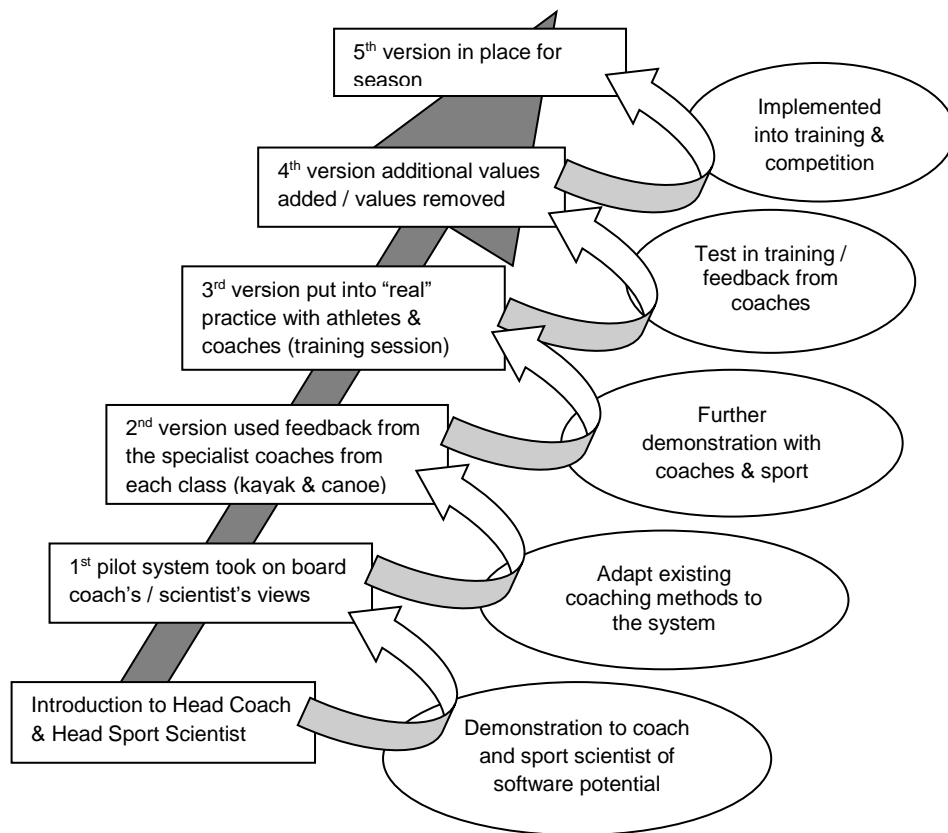
Figure 3.1 The development of a PA data capture system analysis (adapted from O'Donoghue & Longville, 2004)

O'Donoghue and Longville (2004) highlighted the importance of communication between stakeholders, sharing of expertise and demonstrations of practice. To ensure this

process produced a system of relevance and practical use to coaches and athletes, key objectives were reliability within the data capture and processing and the provision of relevant statistics. Using O'Donoghue and Longville's model (Figure 3.1) as a guide, the system developed for canoe slalom had to evolve through the testing of prototypes, systems tested, re-tested, changes implemented and further tests conducted until results and end-user feedback deemed satisfactory. These objectives were also refined to enable the delivery of feedback sessions during International competitions, a key requirement of the coaches. To facilitate the development of a feedback system O'Donoghue & Longville's (2004) model was not adhered to with regard the number of iterations of the system, rather multiple planning meetings, system tests and output modifications were implemented until a satisfactory system had been developed. O'Donoghue and Longville (2004) highlighted the importance of reliability within both data capture, processing and the provision of statistics within feedback sessions during International competitions. Of relevance to this process was the published finding of poor reliability for timing data capture (Hunter, Cochrane & Sachlikidis, 2007) and the need to produce data of practical value to an elite team.

3.2.3.2 Needs analysis

One-on-one (coach/ Sport Science Officer and analyst) and group semi-structured meetings (multiple coaches) were held between the performance analyst, the Sport Science Officer, Head Coach, other coaches and athletes to determine a performance analysis solution. At the outset the coaches perceived the benefits of PA support would be:

- Enhance the coaching process

- Assist in improving athletes' performance

- Help prepare athletes for training and competition

- Enhance key coaching messages

- Visually demonstrate key aspects of a performance

- Review a performance

The athletes also perceived the following benefits:

- Review a run and identify time losses

- Compare their performances with other athletes (split screen).

To deliver relevant feedback, specific information was also required regarding the demands of the sport with coaches identifying the feedback required for three types of analysis:

- Split times analysis

- Penalty analysis (when an athlete hits a gate pole or misses a gate)

- Technique analysis (choice of technique for different gates)

More specific questions were also provided to the analyst as desirable outcomes of an analysis system (Table 3.1).

Table 3.1 Questions posed by coaches to influence PA data capture design

Analysis type	Specific information
Split times	Who has the fastest splits on the course
	Where do time losses occur on the course
	How do individuals compare on upstream and downstream sections, gates, techniques
Penalty	Compare penalty scores individually and between athletes
Technique	What techniques/moves have other athletes chosen on specific sections
	What are the best approach trajectories for specific moves or sequences
	What are the fastest techniques for paddling through water features
Race	Can performances in different classes be used to assess performance in a particular class
	Can a criterion gold standard be used to judge individual performance (using medallists or virtual runs)
	Can individual performance trends be measured over a time period e.g. a season
	Can specific courses be prepared for based on individual characteristics

3.2.3.3 Identifying performance variables

The final tagging panel (to record individual data events) allowed the capture of nine performance variables deemed to be performance indicators by the coaches (Figure 3.2). Whilst the coaches were instrumental in determining which variables were collected a final validation event was held prior to implementing the collection phase. This involved coach's justifying why each variable was an important measure within a canoe slalom event, discussions took place regarding the use of each variable, why it was deemed important and any arguments against its inclusion. This process agreed the variables to be collected (Figure 3.2) during the testing phase for the reliability of the measures. The

total individual run time (seconds) did not include penalties which were recorded separately (2 seconds for touching a gate and 50 seconds if a gate was missed. Split times were recorded based on coaches' determination of the start and end points which was a consequence of the course design. There was no standardised methodology for this. Upstream and downstream techniques were defined (see Figure 3.3). Water features are combinations of white water identified as waves, stoppers and rapids.
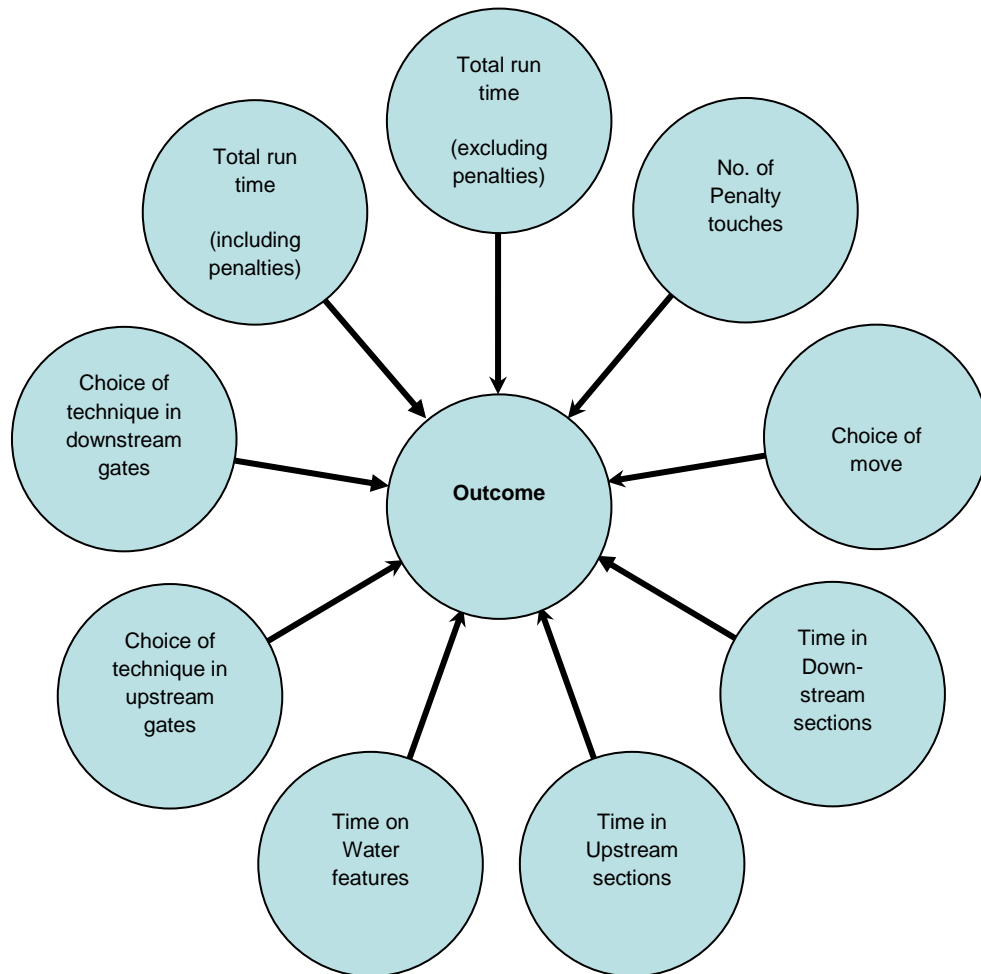
Figure 3.2 Performance variables notated for canoe slalom

3.2.4 PA data capture system development

Dartfish video analysis software (Team Pro) was selected as the software platform as it met the coach demands i.e. allowed split screen playback, slow motion as well as a user defined tagging panel.

3.2.4.1 Design

A flowchart of the data capture system (Figure 3.3) was created to determine the tagging requirements for subsequent data analysis. Since split times were required each section had to be initially coded as a split section. Event buttons were used (in Dartfish) to "tag" the section position on the video, the duration of the event and other related values. For example, when selecting an upstream event, the operator needed to add directional information (left or right) the technique used and if it was situated on a water feature or not. Additional persistent competition information was also tagged once, e.g. athlete name, event, and class (Appendix 3.3).
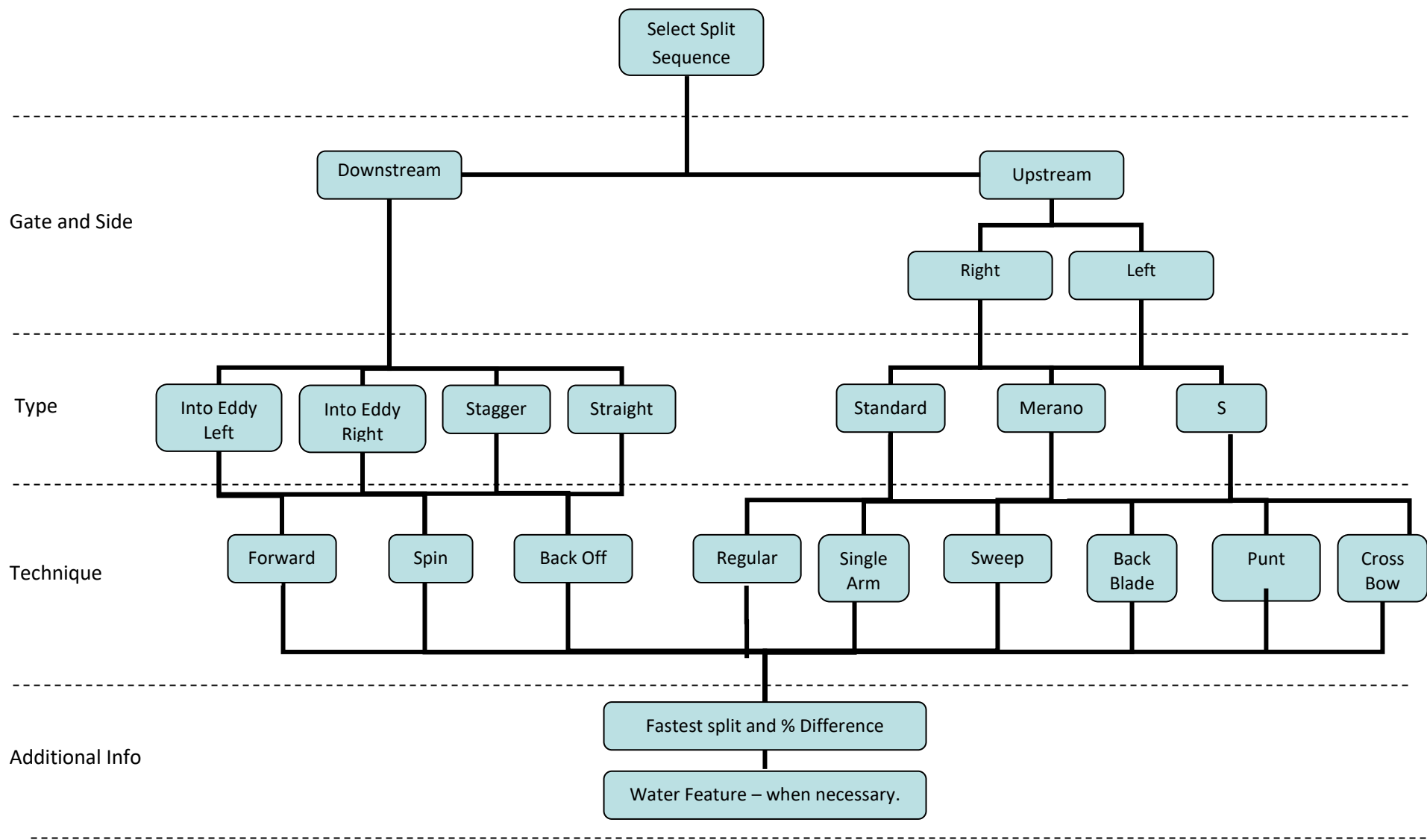
Figure 3.3 Data capture for canoe slalom competitions

Similarly, the logic required to code penalties was identified via a flowchart (Figure 3.4).

Choose Penalty Gate Type
Upstream left / Upstream right / Downstream

Choose Gate Number
1 / 2 / 3 / 4/ etc

Choose Penalty
2 sec / 50 sec / Clean (separate event)

Choose Pole Side
Left / Right

Choose How
Paddle / Head / Arm / Bow / Stern / Body
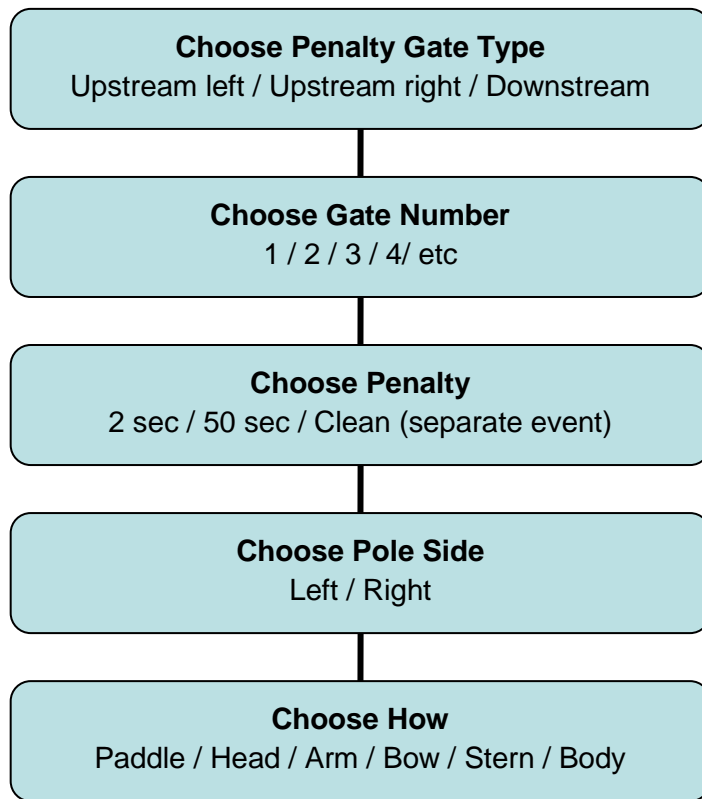
Figure 3.4 Breakdown of the penalty race analysis

The final data capture system was designed to allow both real-time tagging but also other event data could be added post event (Appendix 3.1). Operator training identified which variables could be recorded real-time and which needed to be lapsed-time due to time and operator constraints (penalty and technique analysis).

3.2.4.2 Equipment and operators

Due to the location of video cameras 400m analogue video cables were used. Waterproof camcorder covers, mini dv recorders, network video storage, video switch box (Kramer

Electronics), network hub station (Netgear) along with a variety of network and 4 pin fire-wire cables and BNC connectors made up the equipment. This required 5 operators (3 on cameras, 1 capturing and 1 analysing) during competitions.

Course venues varied in size and hence the logistics for capturing full runs had to be organised on an individual venue basis (Figure 3.5 is an example of a man-made venue used to host the Sydney Olympic and 2005 World Championships).

Figure 3.5 Example capture set-up an at International competition

Camera 1 filmed the athlete from the start pool and followed the athlete down the course until in view of camera 2. Camera 2 and 3 followed a similar procedure to ensure the whole course was recorded. In the qualifying rounds one-minute intervals between starts meant that two athletes were on the course at the same time, semi-finals and final depended on the television broadcaster's requirements resulting in intervals between one and five minutes. This setup meant that all athletes could be recorded over the entire course except in rare occasions when a race was badly compromised, and the athlete took

over two and a half minutes to complete the course. In this situation, the camera operators decided on which athlete to record based on the coach's preferences. The positions adopted for the video capture was always at the discretion of the event organisers which meant that each team was always provided with an area although these could be brick buildings, scaffolding tents or even the team minibus. Some teams used these areas to both capture video and for coaches to provide video feedback to the athletes but the use of a capture base (Figure 3.6) where the three videos signals were sent via cable meant that video files could be edited and copied to the coaches' laptops for athlete feedback sessions.

Figure 3.6 Live capture base set-up at slalom competitions

The analogue video signals were converted to digital via Sony Mini DV decks and connected using DV ports (fire-wire) to the laptops. The capture base operator used a video switch box to switch camera feeds to ensure the complete run of an athlete was recorded on one of the two laptops. Hence two video streams were recorded semi-simultaneously on the two laptops using Dartfish video analysis software (dv import module) to ensure all athletes were recorded in avi format. The laptops simultaneously

recorded to a local and a networked external hard-drive so that the video footage could be accessed within 35 seconds of being captured (average video clip was 300mb). The split times were exported to Microsoft Excel for analysis (Appendix 3.2).

3.2.5 Reliability of data

3.2.5.1 Introduction

Reliability in performance analysis relates to the extent to which the data collected reflect what happened in the event (James, Taylor & Stanley, 2007). Similarly, Atkinson and Nevill (1998) suggested that reliability refers to the amount of measurement error deemed acceptable for the effective practical use of an analysis system. In canoe slalom data is captured as different events where times are recorded via stopwatches (ratio data) and classified into different categories (nominal). These two data types require different approaches for determining reliability issues (James, Taylor & Stanley, 2007). For ratio data Bland and Altman (1986) presented a method of plotting the difference between two times against the average of them (named Bland and Altman plots). Mean biases and 95% limits of agreement for the between recording differences can then be calculated.

For categorical data, James et al. (2002) suggested three error types 1) Operational errors: where the observer presses the wrong button to label an event, 2) Observational errors: the observer fails to code an event and 3) Definitional errors: the observer labels an event inappropriately. These types of error are more common when lots of events need to be coded relatively quickly as the analyst either misses some game events hence the record is incomplete or assigns wrong codes to events. Hughes, Cooper and Nevill (2002) presented a method for calculating the percentage error between two operators' recordings

of event categories. They suggested the reliability test should be examined at the same depth as the subsequent data processing. This recommendation ensures that data is not summed for reliability calculations as this can produce invalid results. Cooper et al. (2007) discussed the number of performances required to generate enough data upon which to perform a realistic and worthwhile 'test-re-test' reliability study. This, of course, is determined by the frequency that a variable occurs in a performance as if insufficient data is tested the possibility exists that if one discrepancy exists between two operators the reliability result can be erroneously large.

To calculate the accuracy of an analyst more than one notation is needed and undertaken in two ways. One person can code the same event twice (known as an intra-analyst test) or two people can code the same event independently (inter-analyst). Both tests can give a good indication of analyst accuracy although the inter-analyst test has the advantage of detecting operational definition misinterpretations more fully since if only one analyst is used, and that person consistently misapplies an operational definition the test will suggest good accuracy.

3.2.5.2 Methods for reliability tests

Four observers (2 analysts & 2 elite coaches) recorded split times (ratio data) for canoe slalom runs of five International standard athletes (canoe & kayak classes) in different competitions and training scenarios. Mean absolute errors were calculated for real-time (the video of the run was watched either live or post-event using a stopwatch to time the events with no use of the pause function), lapsed time (events were tagged in Dartfish with the use of the pause function) and between real- and lapsed-time. Bland and Altman plots (Bland and Altman, 1986) were then used to assess for patterns where the greatest errors

were found. This type of plot was selected because the data was ratio and the expected distribution of errors (difference between observations) was random. Hence, the Bland and Altman plots provided a simple visual depiction of the pattern in the errors along with their magnitudes.

The analysts categorised each run for gate penalties, gate type, technique, water features and general race information (Figures 3.3 and 3.4). However, these were always input into Dartfish post event after the official timings, penalties and gate types had been released. The subsequent tagging was always checked against the official record to ensure accuracy. Intra- and inter-analyst reliability tests were calculated using the chance-corrected measure of agreement named Kappa (Cohen, 1960). This approach calculates expected frequencies for each cell (in the same way that Chi-square does) using the formula:

$$\kappa = \frac{\sum f_O - \sum f_E}{N - \sum f_E}$$

where f_O represents the observed frequencies on the table diagonal (concordant responses) and f_E the corresponding expected frequencies.

The chance corrected Kappa statistic was selected for the categorical data on the basis that this statistic adjusts to account for guessing. Some researchers have suggested the weighted Kappa is better for performance analysis as this gives some credit for close but not exact agreements i.e. partial agreement (Robinson & O'Donoghue, 2007). However, this is most suitable for ordinal data where the difference between saying there were 3 passes instead of the correct 2 passes is less of an error than saying there were 4. In this classification scheme the data was nominal and hence each error was the same irrespective of whether the categories were listed next to each other or not.

3.3 Reliability results

The mean difference (bias delta) for the intra-analyst reliability test for the real-time analysis of split times was 0.01s with 95% limits of agreement (-0.1s and 0.2s) and coefficient of repeatability (0.07s, Figure 3.7). No discernible pattern in the errors were evident with 51.4% of times larger for the second measurement.
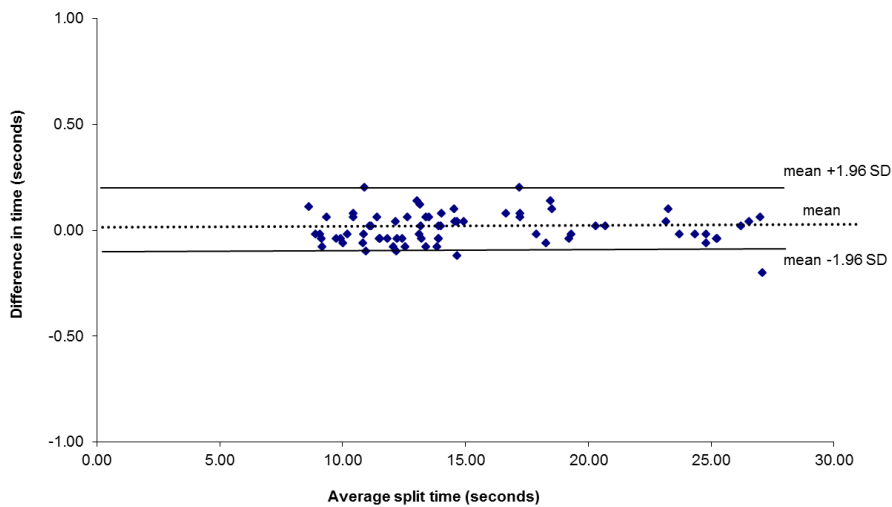
Figure 3.7 Intra-analyst reliability test for real-time analysis of split times

The mean difference for the intra-analyst reliability test for the lapsed-time analysis of split times was 0.00s with 95% limits of agreement (-0.2s and 0.2s) and coefficient of repeatability (0.08s, Figure 3.8). No discernible pattern in the errors were evident.
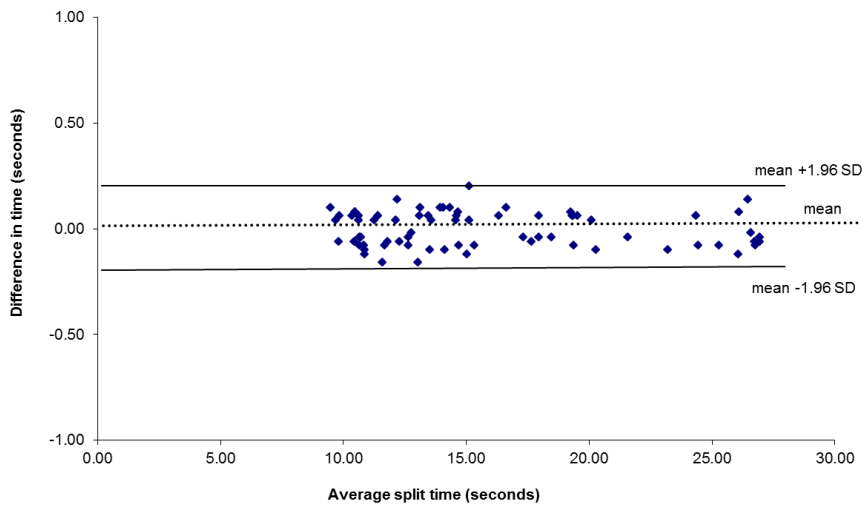
Figure 3.8 Intra-analyst reliability test for lapsed-time analysis of split times

The mean difference for the intra-analyst reliability test for the real- versus lapsed-time analysis of split times was 0.00s with 95% limits of agreement (-0.3s and 0.3s) and coefficient of repeatability (0.15s, Figure 3.9). No discernible pattern in the errors were evident.
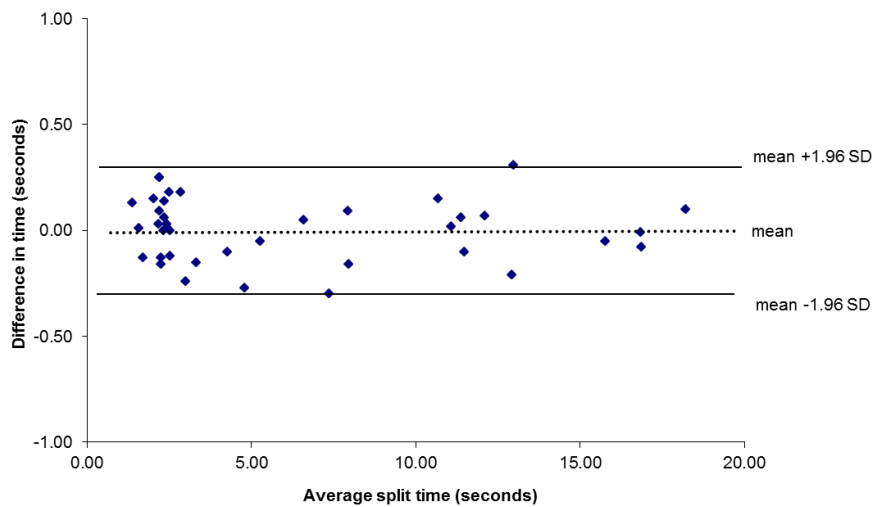
Figure 3.9 Intra-analyst reliability test for real- versus lapsed-time analysis of split times

The mean difference for the inter-analyst reliability test for the real-time analysis of split times was 0.05s with 95% limits of agreement (-0.5s and 0.6s) and coefficient of repeatability (0.27s, Figure 3.10). No discernible pattern in the errors were evident with 0.62s the maximum difference in times. 41.7% of times were larger for the second measurement.
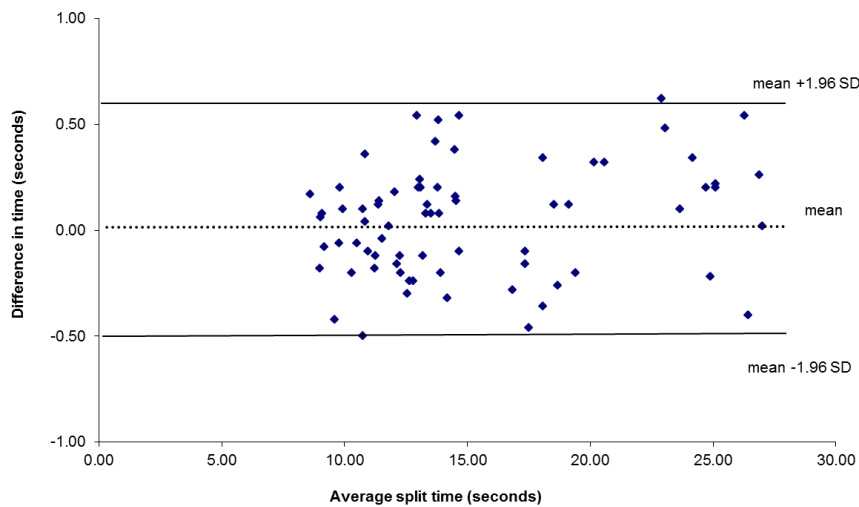
Figure 3.10 Inter-analyst reliability test for real-time analysis of split times

The mean difference for the inter-analyst reliability test for the real-time analysis of split times was 0.02s with 95% limits of agreement (-0.4s and 0.5s) and coefficient of repeatability (0.23s, Figure 3.11). No discernible pattern in the errors were evident with three differences greater than 0.65s.
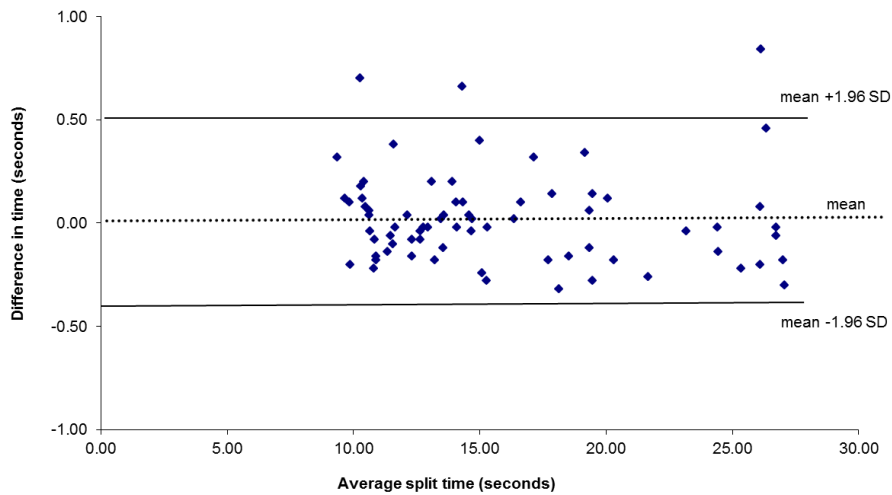
Figure 3.11 Inter-analyst reliability test for lapsed-time analysis of split times

The mean difference for the inter-analyst reliability test for the real- versus lapsed-time analysis of split times was -0.01s with 95% limits of agreement (-0.4s and 0.4s) and coefficient of repeatability (0.21s, Figure 3.12). No discernible pattern in the errors were evident with 0.66s the maximum difference in times. 55.4% of times were larger for the second measurement.
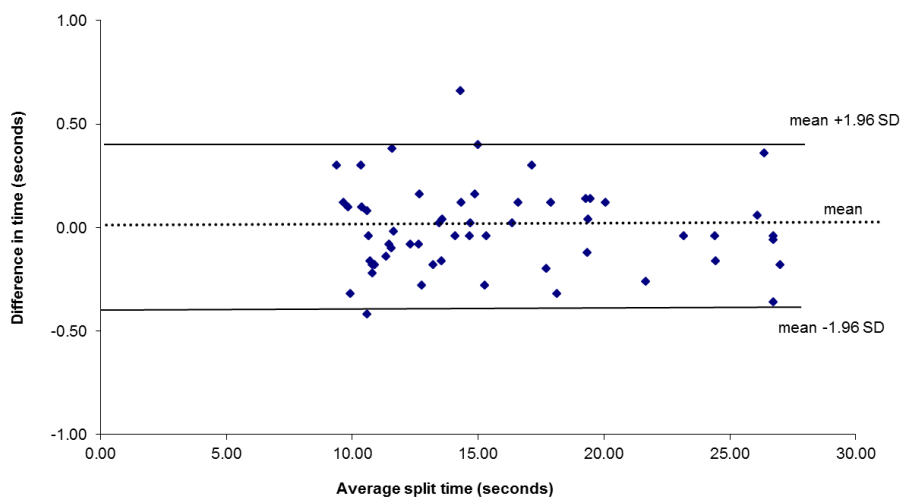
Figure 3.12 Inter-analyst reliability test for real- versus lapsed-time analysis of split times

Gate penalties, gate type, techniques for down-stream gates, water features and general race information were all 100% reliably coded for both inter- and intra-analyst tests. The 6 techniques associated with the upstream gates had a 95.1% Kappa agreement for the intra-analyst test (Table 3.2).

Table 3.2 Contingency table for intra-analyst reliability test of upstream techniques

		Observer 2						
		Back Blade	Cross Bow	Punt	Regular	Single Arm	Sweep	Total
Observer 1	Back Blade	58	0	0	2	0	0	60
	Cross Bow	0	120	0	0	0	0	120
	Punt	0	0	28	0	0	0	28
	Regular	0	0	0	290	5	0	295
	Single Arm	0	0	0	17	13	0	30
	Sweep	0	0	0	10	0	57	67
	Total	58	120	28	319	18	57	600

The inter-analyst reliability test of upstream techniques had a 91.6% Kappa agreement (Table 3.3).

Table 3.3 Contingency table for inter-analyst reliability test of upstream techniques

		Observer 2						
		Back Blade	Cross Bow	Punt	Regular	Single Arm	Sweep	Total
Observer 1	Back Blade	58	0	0	2	0	0	60
	Cross Bow	0	120	0	0	0	0	120
	Punt	0	0	28	0	0	0	28
	Regular	0	0	0	296	2	0	298
	Single Arm	0	0	0	11	17	0	28
	Sweep	0	0	0	5	0	61	66
	Total	58	120	28	314	19	61	600

3.3.1 Reliability discussion

The limits of agreement identified 2.3% errors (n = 9/383) over 0.5 seconds, unacceptably high for canoe slalom split times. However intra-analyst's tests using the sport's performance analyst found that 96.2% of errors were less than 0.2s. At competition athletes and coaches view differences in split times of less than 0.3s as trivial whereas differences around and above 0.5s are deemed large. The inter-analyst tests found that 20% of errors were greater than 0.3s which would suggest that relying on coaches or less trained analysts to time splits would lead to unacceptable precision. This was in concordance with Hunter et al. (2007) who suggested a single analyst should complete all analyses to obtain the greatest accuracy and repeatability. This study demonstrated better reliability values than Hunter et al. (2007) whose limits of agreement for between gate

times were ≤0.21s for intra- and ≤0.39s for inter-analyst measurements. Hunter et al. only used two different camera positions to analyse split times compared to three in this study which would affect the visibility of start and end points and hence reliability. Within the constraints of the British canoe slalom team it was impossible to have a coding system which did not use coaches or less trained analysts to time splits. On this basis further analysis of split times was deemed unreliable and hence not continued during the lifespan of this thesis.

The technique analysis found some discrepancies between and within analysts for the upstream gates which were reviewed and deemed caused by inconsistent application of the operational definitions. The coach deemed this error rate to be unacceptable for competition analysis and more training was provided by the coaches to improve the analyst's interpretation of the operational definitions. It was also decided that a new approach to measuring and categorising upstream gates should be part of a future development plan which determined that this form of analysis was not to be continued within this thesis.

3.4 General discussion

Previous research has utilised expert coaches to validate the variables collected for analysis (Wells et al, 2004; O'Donoghue & Longville, 2004; James et al., 2005 and Choi et al., 2006) and this study adopted the same process. This study was a first step to providing a reliable system for collecting the performance variables requested by the coaches in canoe slalom. To a large extent this was successfully achieved although the reliability tests identified areas for improved learning and future development. The system

has been demonstrated as suitable for collection of competition data within strict conditions, namely well-trained analysts collect data in a consistent manner.

Since the British team only had one analyst this compromised data collection, at this point, to the extent that the only reliable data suitable for analysis within this PhD were the whole run times. The second study will aim in answering some of the specific questions posed by the coaches (Table 3.1) around race run time percentages and whether this can assess performance across classes and within class. This type of analysis could play a role during competition, for example using historical data to assist preparation for a final performance.

3.5 Conclusions

Canoe slalom takes place on different courses which are of different length and difficulty with limited access to the course for camera placement. This means that a performance analysis solution needs to operate within the limitations experienced at major International competitions. The solution developed in this study fits these constraints, albeit with more analysts available, and thus did not provide a reliable method for collecting some of the performance variables desired by the coaches. The selection of total run times for analysis in study 2 have the potential to determine between and within athlete differences that may discriminate performance at the level for developing improvements in athletes.

Chapter Four: Study Two

THE USE OF WINNING RACE TIMES FOR ASSESSING OTHER PERFORMANCES

4.1 Introduction

In canoe slalom, there are no world records because races take place on courses of varying distance and conditions. It is therefore problematical for athletes and coaches to accurately identify the level of a performance, both in training and competition. In competition, the winning time of a race may be the best possible time although it is highly likely that this is not always the case. Similarly, the performance of the top 10 athlete's will also vary between races due to the distribution of both mistakes and course features. Nibali, Hopkins and Drinkwater (2011) found the run-to-run variability for the top ranked athletes (men's canoe, men's kayak and women's kayak) at different courses (World Cups, World Championships and Olympics, 2000 to 2007) was between 0.8% and 3.2%. A race winner's time can therefore only be used as a proxy for the world's best performance, although a virtual best performance using best split times could be used. A common approach in the applied world is to use race percentages i.e. the percentage time off the winner's time, either K1 men (fastest class) or class winner, as a benchmark for best performance. National Governing Bodies have used percentages as a performance measure for National team selection without any published evidence that these standards are valid.

D'Angelo (2013) calculated the percentage time off the K1 men (K1M) winner's time for the 1991 and 2013 World Championships (Table 4.1). The reasons for the relative improvement in 10th position for all classes was considered due to rule changes, improved boats, more athletes or technical improvements. These factors could apply to all athletes of

course, and consequently, when considering race times in general, two factors can be considered to influence race times. First the race conditions, which include all technical factors such as equipment and training, these will vary between races, influence different race classes differently and may result in improvement in performances over time. Secondly, individual performances will vary due to a presence or absence of errors, skilful moves etc., with the top placed finishers likely to exhibit less errors and more skilful moves than the lower placed finishers. D'Angelo (2013) also showed that the number of penalties (both missed and hit gates) had reduced although absolute values were presented for different numbers of athletes making this comparison invalid.

Table 4.1 Percentage times off K1 men's winner's time (D'Angelo, 2013)

	1991 World Championships		2013 World Championships	
	1st position	10th position	1st position	10th position
Men's kayak (K1M)	0%	6%	0%	3%
Men's canoe singles (C1M)	11%	21%	7%	11%
Men's canoe doubles (C2M)	21%	31%	19%	26%
Women's kayak (K1W)	26%	39%	16%	23%

D'Angelo (2013) suggested the analysis of times could include and exclude penalties. Excluding penalties allows the coach to monitor raw speed whereas the inclusion of penalties reflects the true performance level. Penalty analysis was suggested to be of value to coaches in Study 1 (Table 3.1) with a common perception that a "clean" run is either desirable or a necessity to win a race. Similarly, penalties can prevent an athlete making it through the heats or semi-finals into the final. Hunter et al. (2008) stated that due

to the low numbers of penalties in the top ten finishers it was difficult to determine if any relationships existed between penalties and performance.

Nibali et al (2011) found that performances of the top athletes (finished in the top 5 of the final) varied substantially less than athletes finishing in 6th to 10th places. Substantial between race variations (which is the same as within athlete variation) were deemed to be largely related to course difficulty differences (gate placement, water flow and depth of waves). Their analysis of semi-final and final times suggested very poor predictability (correlations between 0.1 and 0.5) between the two, purported to be due to the small between athlete variability (top 10 performances similar) in comparison to the large within athlete variability (race to race). These results led the authors to conclude that a poor semi-final run would not preclude an athlete from performing well in the final and vice versa. However, the rules for competition for the races analysed in their study meant that the semi-final and final times were aggregated to determine finishing positions. These rules have subsequently changed to the current system, where the semi-final times are merely qualifying times for the final (top 10 generally proceed to final). It is unknown whether this rule change has affected times in either the semi-final or final. Nibali et al. (2011) also concluded that the effect of a 2 second penalty increased the variability of performances but the penalty offset any gains in speed. Finally, to improve a medal prospect, improvements of at least 0.4-0.6% were suggested.

International Canoe Federation (ICF) rules state that a canoe slalom course should be navigable in around 95 seconds by K1 men. However, conditions on a given day may mean that the course takes longer or shorter to navigate than the course designers anticipated. Bullock & Hopkins (2009) used the percentages off the winning time to predict an athlete's outcome at the 2006 Winter Olympic Games. Using race percentages has the effect of normalising between race and event differences although the underlying

assumption that the winning performance is the best possible time is an approximation. Therefore, caution is needed (Bullock & Hopkins, 2009). An analysis of race times between the different classes, men's kayak (K1M), men's canoe singles (C1M), men's canoe doubles (C2M), women's kayak (K1W) women's canoe singles (C1W) may help to discern performance levels for the different classes although improvements in individual classes are likely to confuse the analysis.

Race percentages may also be more appropriate when calculated from the specific class rather than using the K1M time for all classes as between class variation may be different between events. Competitions on the circuit are also limited in numbers (1 x European, 1 x World Championships and 3 World Cups, increased to 5 in 2014) which may lack the power to enable statistical analyses. An option of combining data over a number of years in an attempt to model the race percentages should be explored. This study will use the percentage time off the winner's time, either K1 men or class winner, to assess which approach is more accurate for assessing performance. Times from heats were excluded from analyses due to different tactics being employed i.e. the goal is to progress rather than produce a best time. Race finals were used to determine what it took to achieve a medal in 3rd place, 5th place and semi-finals used to determine what it took to qualify for the final (10th place). Signal detection methods were also used to estimate the likelihood of an athlete medalling or not based on predicted race percentages.

4.2 Methodology

4.2.1 Participants

Elite level canoe slalom athletes competing at the highest level within a governing body in canoe slalom from 2009 to 2016 in World Championship, World Cup & European events (n=49 K1M, C1M & K1W; n=45 C2M & C1W). Olympics were excluded as the rule that only one boat per nation for each class potentially omitted some top athletes and fewer boats progressed in each round compared to other races.

4.2.2 Procedure

Competition run times were the official results downloaded from each competition organiser's website. Each competition had an official timing beam system (http://www.siwidata.com) that recorded the time from the official start to the official finish for each competitor. These were imported into Microsoft Excel and percentages off K1M & class winning times were calculated. Both SPSS and Excel were used for subsequent analyses to determine:

- Characteristics of race performances;
- Variability in race percentages using 1 standard deviation;
- Proportional change or not between the 2012 Olympic cycle and the 2016 Olympic cycle using chi square analysis for comparison between K1M winner percentages and class percentages;
- Predictability of race percentages using percentiles and signal detection theory.

Signal detection theory provides a precise language and graphic notation for analysing decision making in the presence of uncertainty (Heeger & Landy, 2010). In this study data from the 2012 Olympic cycle (events between 2009 and 2012) were used to determine the probability of winning a medal or not i.e. a prediction using percentages off K1M and class winners. Data from the 2016 Olympic cycle (events between 2013 and 2016) were then used to evaluate the accuracy of the predictions (Figure 4.1).
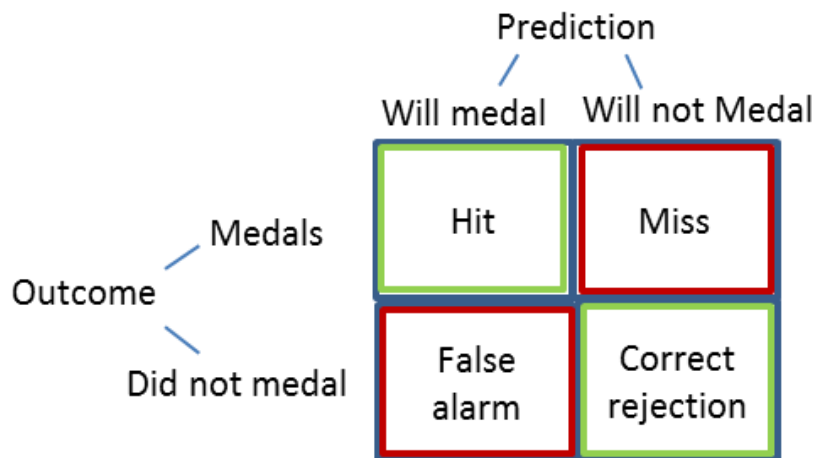
Key: Hit – 2012 data predicted an athlete's percentage time was good enough to medal and the athlete did gain a medal
False alarm – 2012 data predicted a medal but the athlete didn't
Miss – 2012 data predicted no medal but the athlete did medal
Correct rejection – 2012 data predicted no medal and athlete didn't

Figure 4.1 Signal detection methodology applied to medal prediction and outcome in canoe slalom

Whilst both misses and false alarms were erroneous predictions it was thought that false alarms (specifying a target percentage to an athlete who goes on to achieve it but failed to gain a medal) were worse outcomes than misses (specifying a target percentage to an athlete who didn't achieve it but gained a medal). In the latter case, the athlete has gained an unexpected medal and is very happy although this may lead to the athlete

considering future targets as being unnecessarily tough, Athletes who were disappointed at not medalling could consider false alarms from the perspective that bad advice was given and in an extreme situation with funding dependent on medals the analyst's role becomes at risk.

To determine whether K1M or class, winner percentages were useful indicators of race performance some understanding of consistency was required. For example, does achieving 103% off the K1M winner percentage equate to a 3rd place finish in a different class, all of the time, some of the time or never. If a consistent answer to this question was found, then the usefulness of the winner percentage was confirmed. To determine if this relationship existed some rules needed to be determined and then tested against different data. Hence, all race times from the 2009-12 Olympic cycle were compared to the K1M and class winner percentages from the same competition by calculating the percentage time off the respective winning time. To answer the specific question above, all 3rd place percentages were used to calculate 25th, 50th and 75th percentiles. Thus, the frequency that the 3rd place race percentages fell into each of the four percentile categories (0-25, 25-50, 50-75, 75-100) could be ascertained. These percentile boundaries were then used to calculate the frequency into which the 2013-16 Olympic cycle 3rd place race percentages fell. The two sets of frequencies were then compared using a Chi square analysis with the effect size calculated using Cramer's V.

4.3 Results

This section has been split into four to highlight the different approaches used to try to determine whether performances in canoe slalom could be measured in a valid and reliable manner, and if so, could coaches accurately comment on these performances in terms of overall time and whether they were of sufficient quality to represent a chance of making a final, medalling or even winning a future race.

Section 4.3.1 compared race times for all classes to see whether classes did perform differently and whether the variability in race time was related to that time. As the pattern fitted the prevailing view, and that of coaches, historical patterns in race times were ascertained (Section 4.3.2). Coaches used percentages off race winners in the belief that these were valid judgements of performance although opinions were divided as to whether using K1 Men (the fastest class) or the class winner was the best solution, neither of which had been subjected to rigorous scrutiny. Hence the variability of performances in relation to race winners were investigated to see whether race performances were changing over time e.g. larger improvements in some classes compared to others would suggest using class winners may be more appropriate than using K1 Men's. The results suggested that variability was evident and hence some form of longitudinal analysis necessary to account for this.

The analysis then considered races over time (Section 4.3.3), for both one year and Olympic cycles (4 years). These analyses demonstrated more consistency in the variability, as would be expected with more data, with the two Olympic cycles (2009-12 and 2013-16) used to test for differences using percentiles (25th, 50th and 75th). Relatively small differences, which exhibited no clear pattern, were found for 23 out of the 31 comparisons.

Section 4.3.4 utilised the percentile approach, expanded to 95%, 75%, 50%, 25% and 5%, using the Olympic cycles to assess their predictability (accuracy). Signal detection theory was used to compare the theorised likelihood of a race percentage gaining a medal, determined using the 2009-12 data, and tested to see if the percentile boundaries equated to success or failure during the 2013-16 cycle. To assess the best percentage to use receiver operating characteristic (ROC) curves (false alarms plotted against hits) and detection error trade-off (DET) graphs (false alarms plotted against misses) were produced. These analyses suggested the use of race percentages off the class winner, rather than the K1M, with a 50% level of probability of medalling resulting in a low rate of false alarms (maximum 6%) and a high hit rate (over 70% of medals correctly identified).

4.3.1 Canoe slalom race characteristics

The K1M winning times for 49 International race finals ranged from 69.9 to 108.0s (median = 92.6s, IQR = 7.6, Figure 4.2) with winning times in the other race classes typically exhibiting slower times but similar IQRs, excluding C1W who were introduced to ICF World Championship status in 2009, (C1M = 7.2, C2M = 9.1, K1W = 7.9, C1W = 18.5).
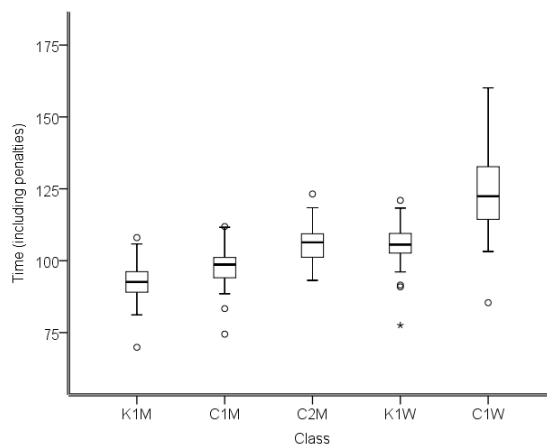
Figure 4.2 Canoe slalom winning times for finals in International competitions (2009-2016)

4.3.2 Canoe slalom race percentages off K1M winning times between 2009 and 2016

4.3.2.1 Average performances

In the K1M class the average percentage time off the winner varied more for the lower places compared to the higher ones (Figure 4.3; Figure 4.4).
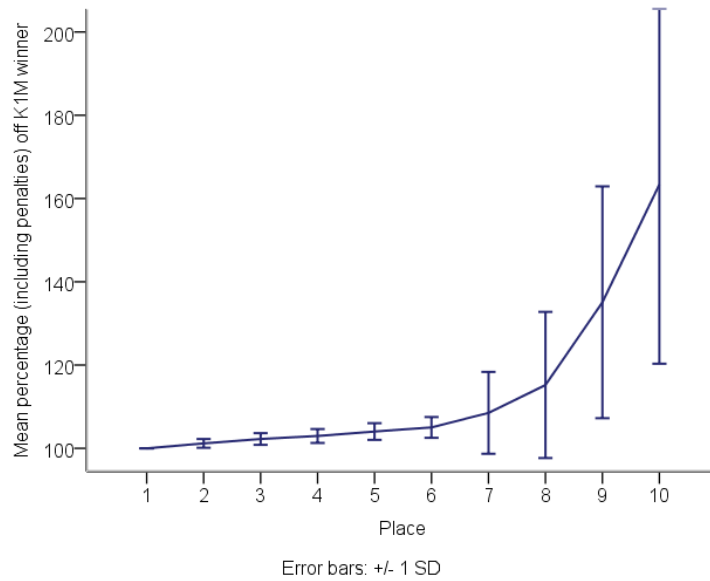
Figure 4.3 K1M percentage times for each final place off the K1M winning time (2009-2016)
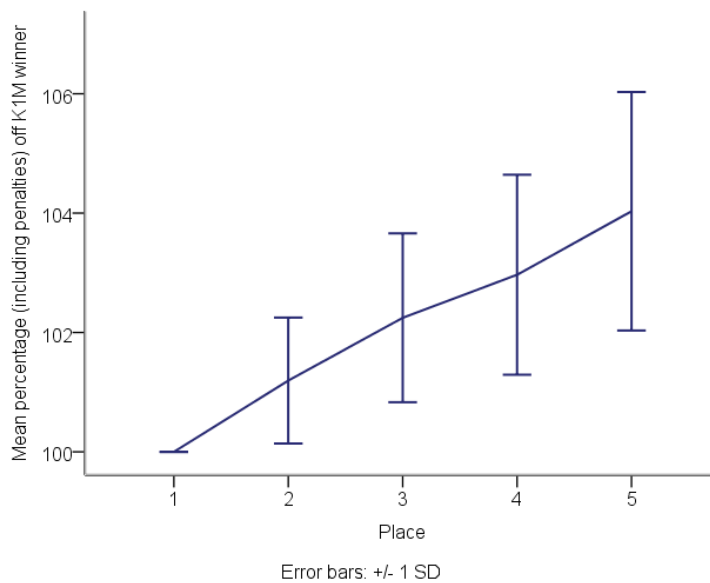
Figure 4.4 Percentage times for top 5 final places off the K1M winning time (2009-2016)

Since the 10th place in a final varied so much (due to 50 second penalties) subsequent analyses will only compare 1st, 3rd and 5th place race times from finals, for the different classes, against the class and K1M winner (at the same event). The 10th place in the semi-final was compared against the semi-final class and K1M semi-final winning times to determine what it took to make a final.

In relation to the K1M winner (final except for 10th place which was against K1M semi-final winner) the percentages tended to vary more for the slower classes and lower places (Figure 4.5).
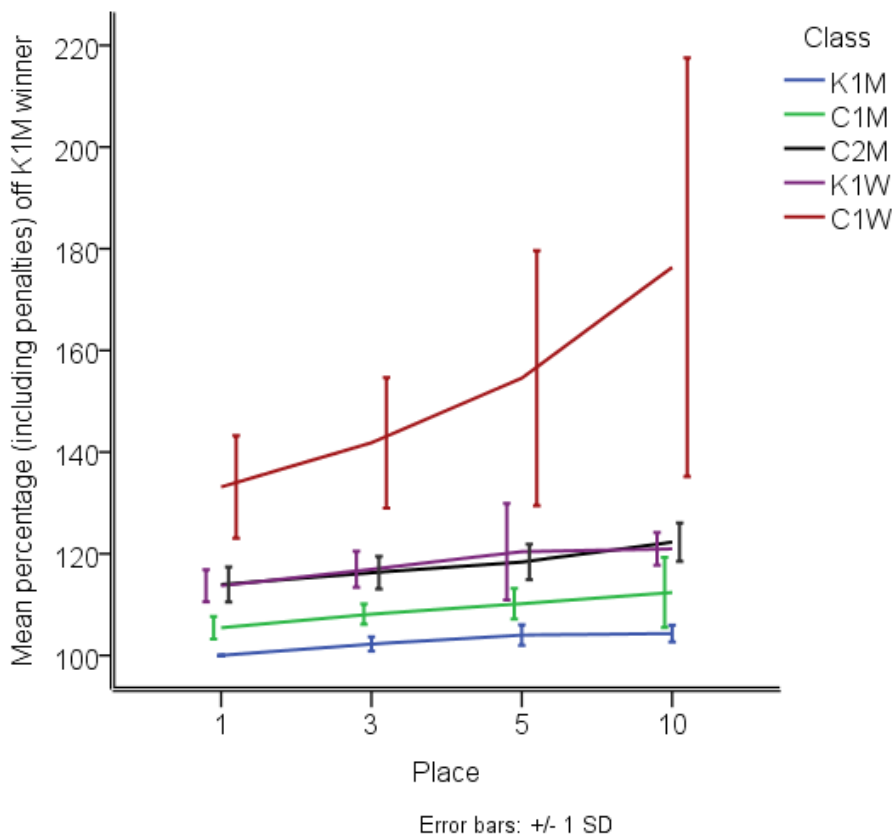
Figure 4.5 Mean percentages for each class off the K1M winning time (2009-2016)

4.3.2.2 Yearly K1M performances

There was no clear improvement in performances for 3rd, 5th or 10th places over the 8 year period relative to the K1M winning times (Figure 4.6). However, the variability of percentage times did change between years for all three positions. For example, in 2016 the 5th place percentages off the K1M winning time averaged 102.9% (minimum 101.3% and maximum 107.6%). In 2014 the 10th place averaged 103.1% off the K1M semi-final winning time (minimum 102.2% and maximum 104.3%).
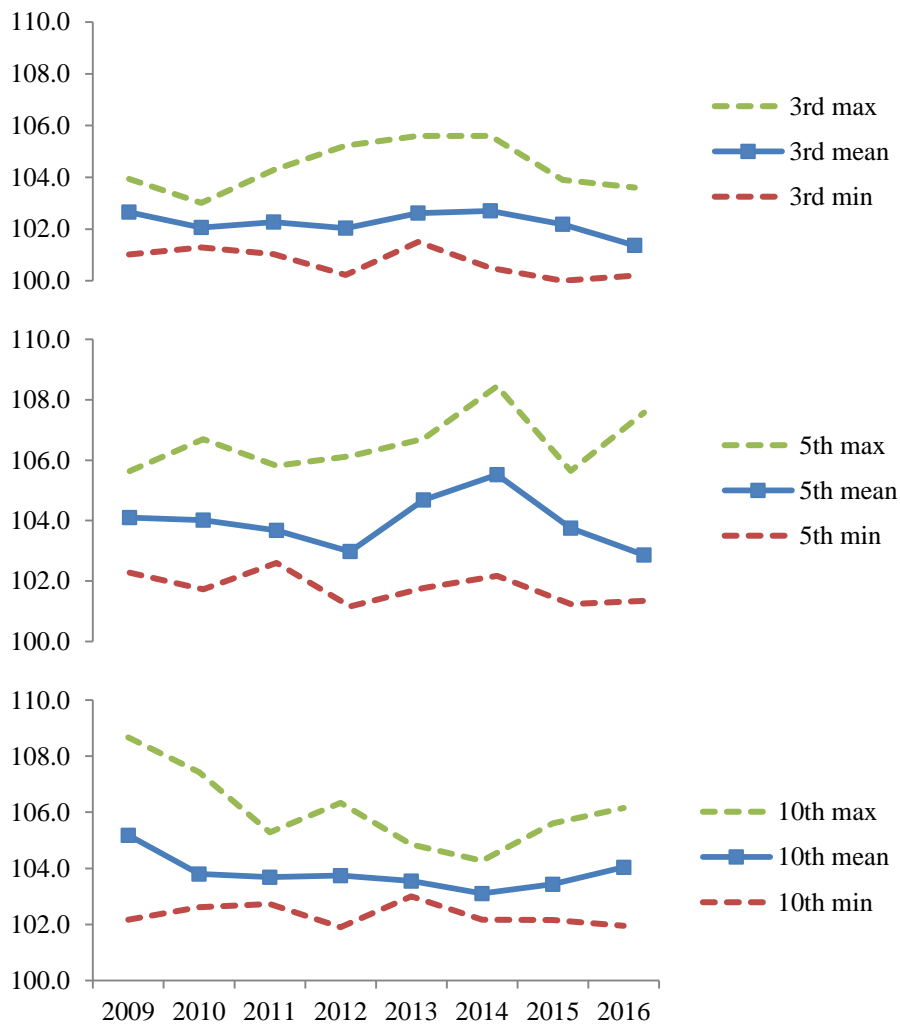
Figure 4.6   Mean K1M percentage times off the K1M winner by place & year (10th place calculated from semi-final winner)

4.3.3 Canoe slalom race percentages off K1M and class winning times (2009 – 2016)

4.3.3.1 Yearly percentages

To assess the veracity of using the K1M or the class winning times the percentages for 1st, 3rd, 5th and 10th places were plotted by year. C1M mean percentage time off K1M and class winner remained fairly constant through the 8 years (Figure 4.7) with no obvious improvement or decrement over time. Similarly, the variability in percentages did not indicate any overall pattern to suggest a trend.
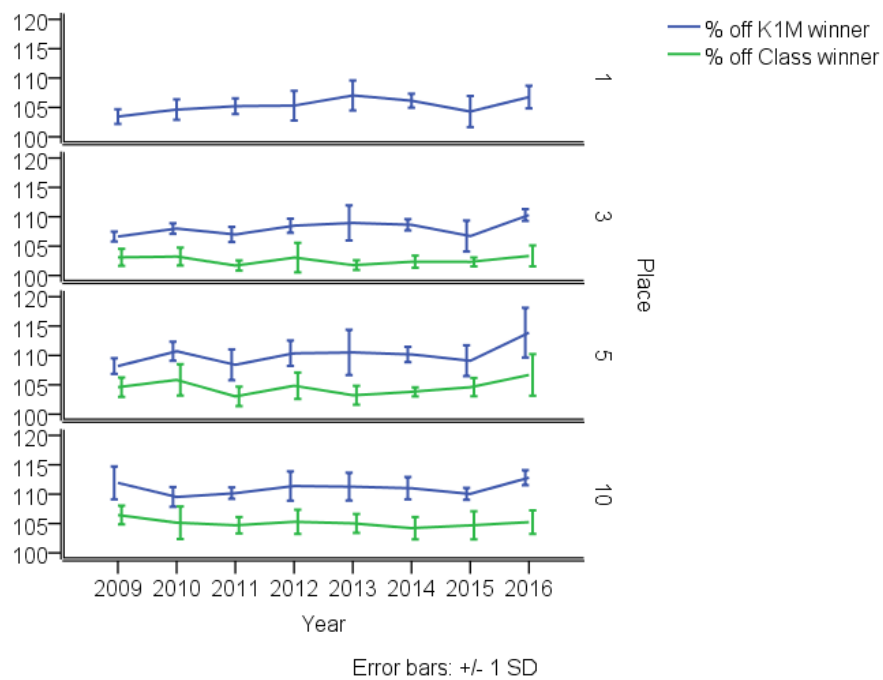
Figure 4.7   Mean C1M percentage times off the K1M winner and C1M winner by place & year (10th place calculated from semi-final winners)

Similarly, random patterns for percentage times were found for C2M (Figure 4.8) and K1W (Figure 4.9).
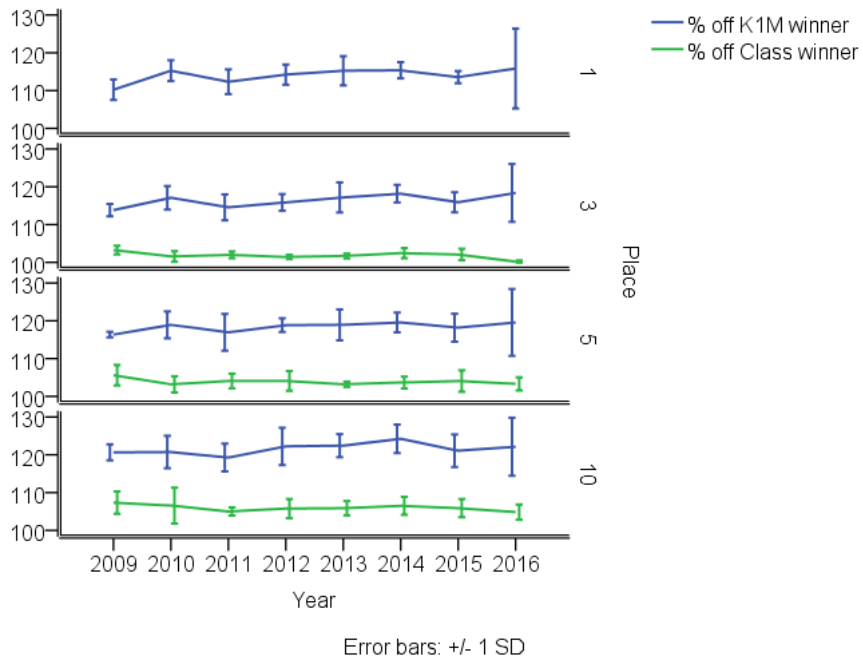
Figure 4.8   Mean C2M percentage times off the K1M winner and C2M winner by place & year (10th place calculated from semi-final winners)
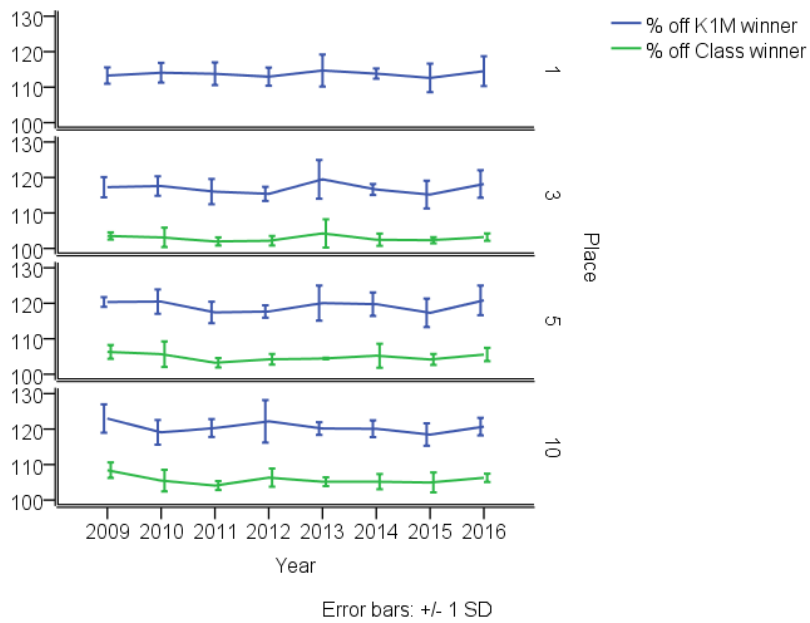
Figure 4.9   Mean K1W percentage times off the K1M winner and K1W winner by place & year (10th place calculated from semi-final winners)

90

In the C1W class (Figure 4.10) the relatively random pattern of variability seen in the other classes was replicated although some improvement in percentages off both K1M and class winner was evident.
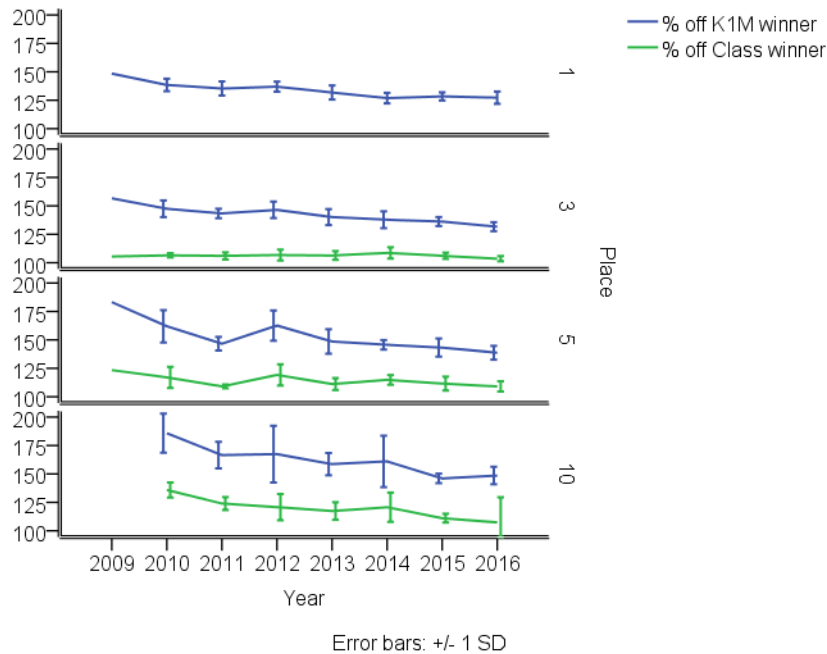
Figure 4.10   Mean C1W percentage times off the K1M winner and C1W winner by place & year (10th place calculated from semi-final winners)

4.3.3.2 Olympic cycles

The lack of obvious pattern in yearly percentages led to the consideration of dividing the data into Olympic cycles, an obvious choice given that canoe slalom is an Olympic sport. This also had the effect of increasing the sample size (to 22 for most classes in 2009-12 and 27 in 2013-16).

Percentiles (25th, 50th and 75th) were calculated for race percentages off K1M and class winning times for the 2009-12 Olympic cycle. This meant that first place percentages could be calculated off K1M for C1M, C2M, K1W, C1W. Third, fifth and tenth place percentages could be calculated off K1M and class winner for K1M, C1M, C2M, K1W,

C1W. This equated to 31 separate analyses where race percentages for the 2013-16 Olympic cycle were classified according to the 25th, 50th and 75th percentile boundaries obtained from the 2009-12 Olympic cycle i.e. the proportion of percentages that fell into 1) the best percentage to the 25th percentile, 2) the 25th to 50th percentile, 3) the 50th to 75th percentile and 4) 75th percentile to worst. Chi-square tests assessed the degree of independence between the two Olympic cycles although there were some instances where some expected counts were less than 5 (indicated by a Key stating the number of cells, Figures 4.12, 4.13, 4.14 and Appendices 4.1 - 4.15, 4.17 - 4.27) suggesting some caution in relation to the veracity of those results.

Three patterns in the data were found, small difference between 2009-12 and 2013-16 Olympic cycles (n = 2); slightly larger differences which exhibited no clear pattern between the two Olympic cycles for the different races (n = 23) and an improvement in the 2013-16 data in comparison to 2009-12 (n = 6). These were based on the effect sizes (0.1 small, 0.3 medium and 0.5 large; Cohen, 1992) which were colour coded in Table 4.2.

Table 4.2 Effect sizes for differences in frequency of percentages within percentile
categories between the 2009-12 and 2013-16 Olympic cycles

Class	Place	Percentage off K1M winner	Percentage off class winner
K1M	3rd	0.1	
	5th	0.2	
	10th	0.3	
C1M	1st	0.3	
	3rd	0.3	0.2
	5th	0.3	0.2
	10th	0.3	0.3
C2M	1st	0.2	
	3rd	0.3	0.2
	5th	0.2	0.3
	10th	0.4	0.3
K1W	1st	0.1	
	3rd	0.2	0.3
	5th	0.2	0.3
	10th	0.2	0.5
C1W	1st	0.7	
	3rd	0.6	0.2
	5th	0.5	0.3
	10th	0.7	0.7
	Mean	0.4	0.3
	SD	0.2	0.1

There were small differences in the proportion of 3rd place K1M percentages off the K1M winner in the four percentile categories between the 2009-12 and 2013-16 Olympic cycles (Chi-square = 0.24, df = 3, p = 0.97, Cramer's V = 0.07; Figure 4.11). See Appendix 4.16 for the other category where small differences were found.
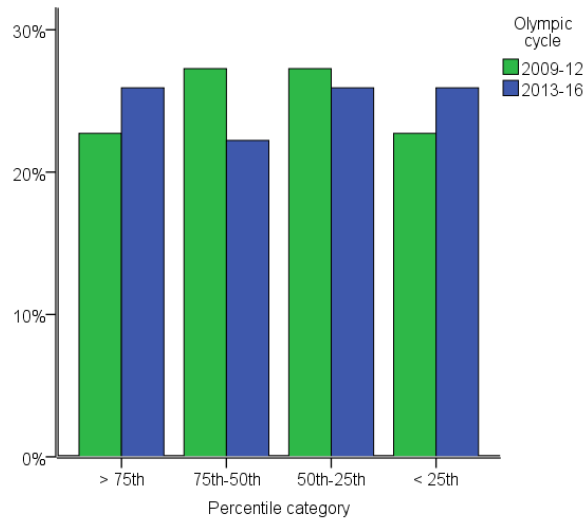
Figure 4.11 Proportion of K1M's 3rd place race percentages off the K1M winner in four percentile categories during the 2009-12 and 2013-16 Olympic cycles

There were more 3rd place C2M percentages off the K1M winner in the slowest percentile category in the 2013-16 Olympic cycle compared to the 2009-12 cycle (Chi-square = 3.35, df = 3, p = 0.34, Cramer's V = 0.28; Figure 4.12).
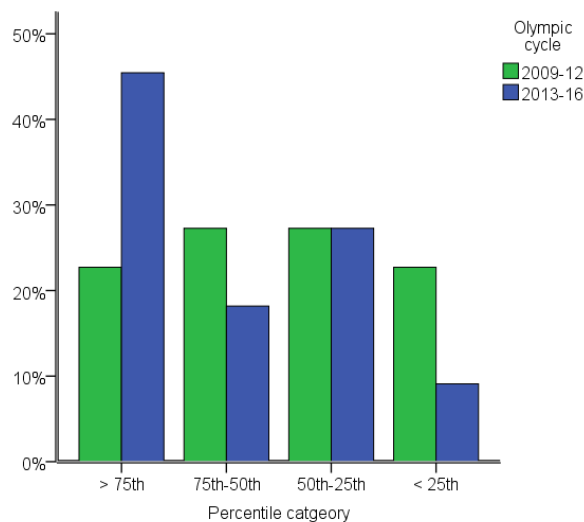
Key: 2 cells had expected counts of less than 5

Figure 4.12 Proportion of C2M's 3rd place race percentages off the K1M winner in four percentile categories during the 2009-12 and 2013-16 Olympic cycles

There were slightly more 10th place K1W percentages off the class winner in the fastest (<25th) and 50th to 75th percentile categories in the 2013-16 Olympic cycle compared to the 2009-12 cycle (Chi-square = 11.30, df = 3, p = 0.10, Cramer's V = 0.48; Figure 4.13). See Appendices 4.1 - 4.15, 4.17 – 4.21, 4.23, 4.25 for the other categories where medium differences were found.
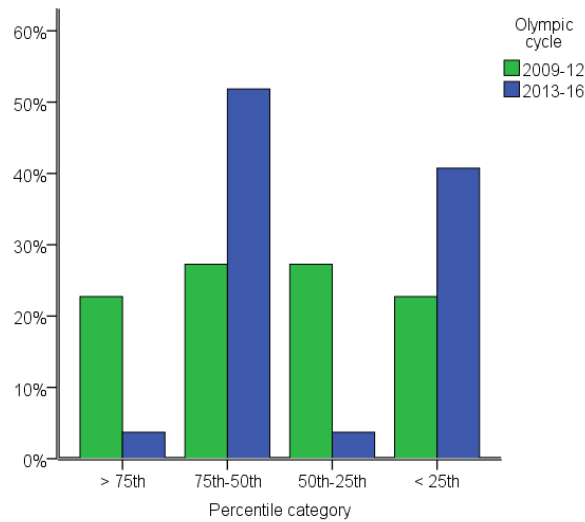
Key: 4 cells had expected counts of less than 5

Figure 4.13 Proportion of K1W's 10th place race percentages off the class winner in four percentile categories during the 2009-12 and 2013-16 Olympic cycles

There were more 1st place C1W percentages off the K1M winner in the fastest (<25th) percentile category in the 2013-16 Olympic cycle compared to the 2009-12 cycle (Chi-square = 20.62, df = 3, p < 0.001, Cramer's V = 0.70; Figure 4.14). See Appendices 4.22, 4.24, 4.26 & 4.27 for the other categories where large differences were found.
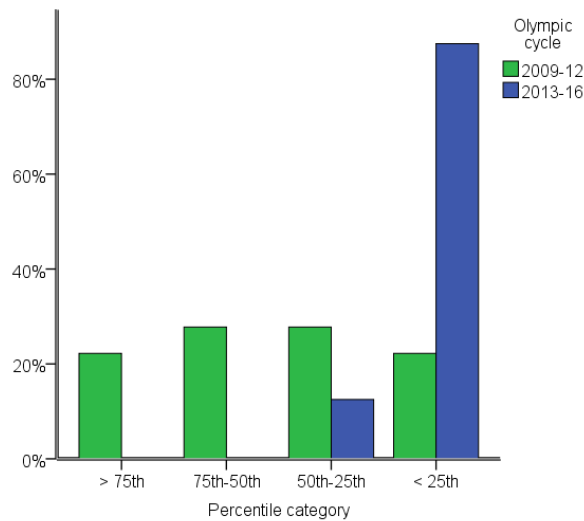
Key: 6 cells had expected counts of less than 5

Figure 4.14 Proportion of C1W's 1st place race percentages off the K1M winner in four percentile categories during the 2009-12 and 2013-16 Olympic cycles

4.3.4 Predicting the probability of achieving 3rd, 5th and making the final in canoe slalom races using percentages off K1M and class winning times (2009 – 2016)

4.3.4.1 Percentage chance

The probability of being placed 3rd, 5th or 10th (making the final) in the 2009-12 Olympic cycle was calculated using the percentage times off the K1M (Table 4.3) and the class winning times (Table 4.4).

96

Table 4.3   The probability of making a place using percentages off the K1M winner
            (2009-12 Olympic cycle)

Class	Placing	Percentage chance				
		95	75	50	25	5
K1M	3rd	100.3	101.3	102.0	103.3	105.1
	5th	101.2	102.3	103.6	104.8	106.6
	10th	101.9	102.7	103.3	105.5	108.6
C1M	1st	100.7	103.4	105.0	105.9	107.9
	3rd	105.5	106.5	107.4	108.5	110.1
	5th	106.0	108.0	109.0	111.2	114.0
	10th	107.8	109.5	110.3	111.6	115.6
C2M	1st	107.7	109.8	113.8	115.9	117.8
	3rd	109.3	112.8	115.6	117.5	120.0
	5th	110.2	116.1	117.7	119.7	124.0
	10th	115.2	117.8	120.9	122.6	129.3
K1W	1st	108.4	112.0	113.2	115.0	119.0
	3rd	111.1	114.7	116.2	118.4	121.8
	5th	113.6	116.8	119.4	120.5	124.6
	10th	115.4	118.6	120.1	123.7	132.3
C1W	1st	126.5	133.1	137.1	143.8	
	3rd	135.7	140.8	146.3	153.3	
	5th	139.7	146.7	161.1	177.2	
	10th	149.7	165.4	195.6	252.1	

Table 4.4 The probability of making a place using percentages off the class winner (2009-12 Olympic cycle)

		Percentage chance				
Class	Placing	95	75	50	25	5
K1M	3rd	100.3	101.3	102.0	103.3	105.1
	5th	101.2	102.3	103.6	104.8	106.6
	10th	101.9	102.7	103.3	105.5	108.6
C1M	3rd	100.6	101.7	102.2	103.6	107.5
	5th	101.2	102.6	104.6	105.7	109.6
	10th	102.2	103.7	105.5	106.2	109.0
C2M	3rd	100.4	101.1	101.7	103.0	104.4
	5th	101.2	102.3	103.7	106.0	108.8
	10th	102.3	104.5	105.3	107.4	114.1
K1W	3rd	100.2	101.3	102.2	103.7	106.9
	5th	101.6	102.8	104.1	106.3	110.5
	10th	102.2	104.2	105.3	108.2	111.1
C1W	3rd	101.2	104.4	105.6	108.6	
	5th	106.5	109.5	111.6	123.4	
	10th	112.7	124.8	139.7	177.8	

To investigate which percentage was the best to use signal detection theory was applied to the probability estimates for making 3rd place (medalling).

4.3.4.2 Using signal detection theory to determine the most appropriate 2009-12 Olympic cycle probability to use

The probabilities associated with making a place (Tables 4.3 and 4.4) ranged from a very tough percentage target (95% certain that achieving this low percentage target or better the athlete will achieve the predicted place) which is likely to result in a lot of athletes who don't make the percentage target but do make the place i.e. misses. Conversely, using a very lenient percentage target (high percentage target but low (5%) chance of success) is likely to result in lots of athletes achieving the predicted percentage but not making the

place i.e. false alarms. Signal detection theory suggests the best probability to use minimises both misses i.e. setting the percentage target very low resulting in lots of medals being gained with higher percentages than the target; and false alarms (percentage target set too low so lots of athletes achieve the threshold but don't make the place predicted).

The 2009-12 Olympic cycle percentages (off K1M and class winner) from Tables 4.3 and 4.4 were used to categorise the 2013-16 Olympic cycle percentages into:

- Hits - the 2013-16 race percentage made a medal and matched or bettered the 2009-12 criterion percentage

- Misses - the 2013-16 race percentage made a medal but was higher than the 2009-12 criterion percentage

- False alarms - the 2013-16 race percentage did not achieve a medal but matched or bettered the 2009-12 criterion percentage

- Correct rejection - the 2013-16 race percentage did not achieve a medal and was higher than the 2009-12 criterion percentage

To assess the best percentage to use (95%, 75%, 50%, 25% or 5%) receiver operating characteristic (ROC) curves (false alarms plotted against hits – green dots which are ideally located in the top left corner of the chart) and detection error trade-off (DET) graphs (the mirror image of the ROC curve where false alarms are plotted against misses – red dots ideally located in the bottom left corner of the chart) were produced. For example, the 50% probability of medalling percentage target i.e. 102.0% (Table 4.4) resulted in relatively few misses (6%) and false alarms (6%; red dot closest to origin in Figure 4.15). This probability level also resulted in a relatively high hit rate (23%; green dot in Figure 4.15.

Figure 4.15 The probability of K1M making a medal in the 2013-16 Olympic cycle using percentages off the 2009-12 Olympic cycle class winner

The sum of the hits and misses always equates to the percentage of athletes who medalled in the 2013-16 Olympic cycle. Hence a hit rate of 23% meant that 79% of all medals in the 2013-16 Olympic cycle were correctly predicted using this percentage. The other races were analysed in the same way with the percentages off the class winner always producing higher hit rates and less false alarms than off the K1M winner (Table 4.5).

Table 4.5 Best probability value, obtained from the 2009-12 Olympic cycle, for accurately predicting a medal in the 2013-16 Olympic cycle

Class		Best probability	Did not medal False alarms	Medals Hit	Miss	Hit rate
K1M	Off K1M wnr*	50%	6%	23%	6%	79%
C1M	Off K1M wnr	25%	7%	21%	9%	70%
	Off class wnr*	50%	3%	23%	6%	79%
C2M	Off K1M wnr	50%	5%	14%	16%	47%
	Off class wnr*	50%	2%	23%	7%	77%
K1W	Off K1M wnr	75%	5%	14%	16%	47%
	Off class wnr*	50%	2%	20%	10%	67%
C1W	Off K1M wnr	95%	9%	22%	7%	75%
	Off class wnr*	50%	4%	23%	7%	77%

Key: * indicates superior prediction (higher hit rate and lower false alarms) between using K1M or class winning times

There were two instances where two red dots were equidistant to the origin. On the basis that false alarms had been deemed more serious errors than misses the probability with the lowest false alarm rate was selected. For example, the 75% chance of medalling probability (114.7%, Table 4.3) was selected over the 50% chance (116.2%) as the false alarms reduced from 11% to 5% (Figure 4.16). ROC/DET graphs for all other classes are presented in Appendix 4.28 – 4.35.
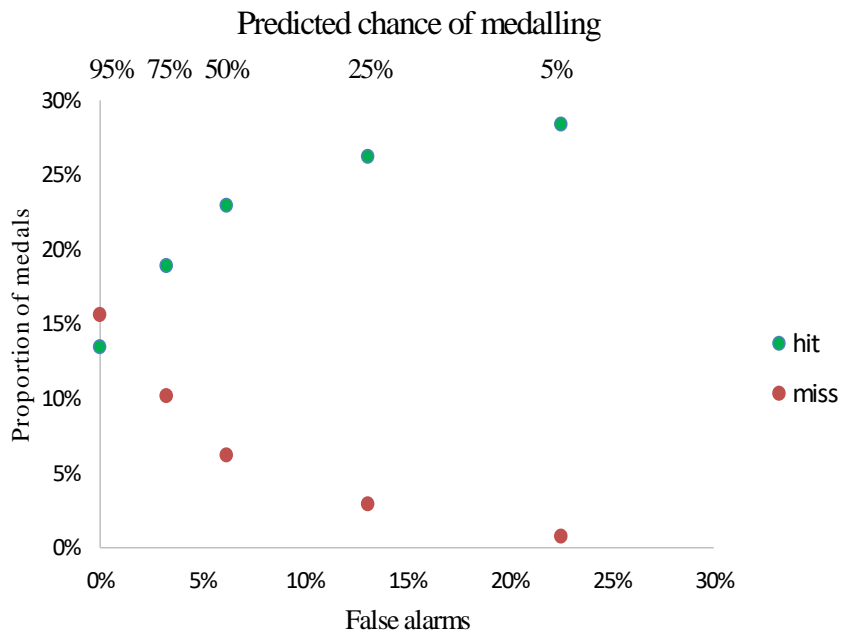
Predicted chance of medalling
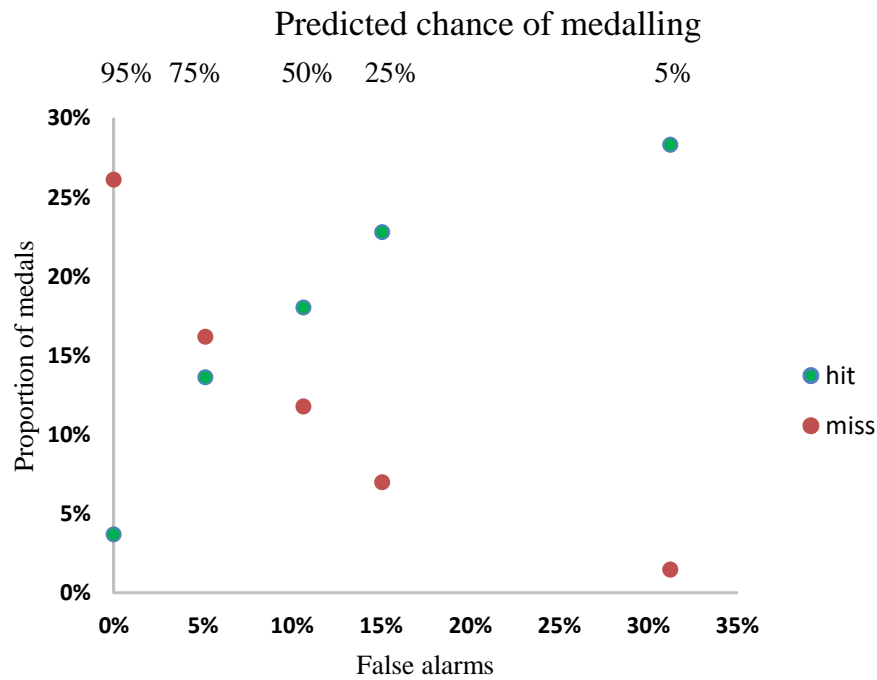
Figure 4.16 The probability of K1W making a medal in the 2013-16 Olympic cycle using percentages off the 2009-12 Olympic K1M winner

4.3.4.3 The stability of probability percentiles in four-year cycles

Having assessed the veracity of using probability percentiles from the 2009-12 Olympic cycle to predict medal chances in the 2013-16 Olympic cycle a final question was posed in relation to whether using 2009-12 data determined percentages was the most appropriate throughout the 2013-16 Olympic cycle. An alternative approach considered was to use the previous 4 years' worth of data to calculate probability percentiles i.e. in the 2014 season the analyst could use 2010-2013 data, in 2015 the 2011-14 data and so on.

Medal (3rd place) percentiles for race percentages for K1M off K1M winners calculated in 4-year cycles suggested that the top 5% of performances had moved closer to theK1M winner whereas the lowest 5% had moved further away (Figure.4.17). Each class presented a slightly different story (Appendix 4.36 – 4.43).
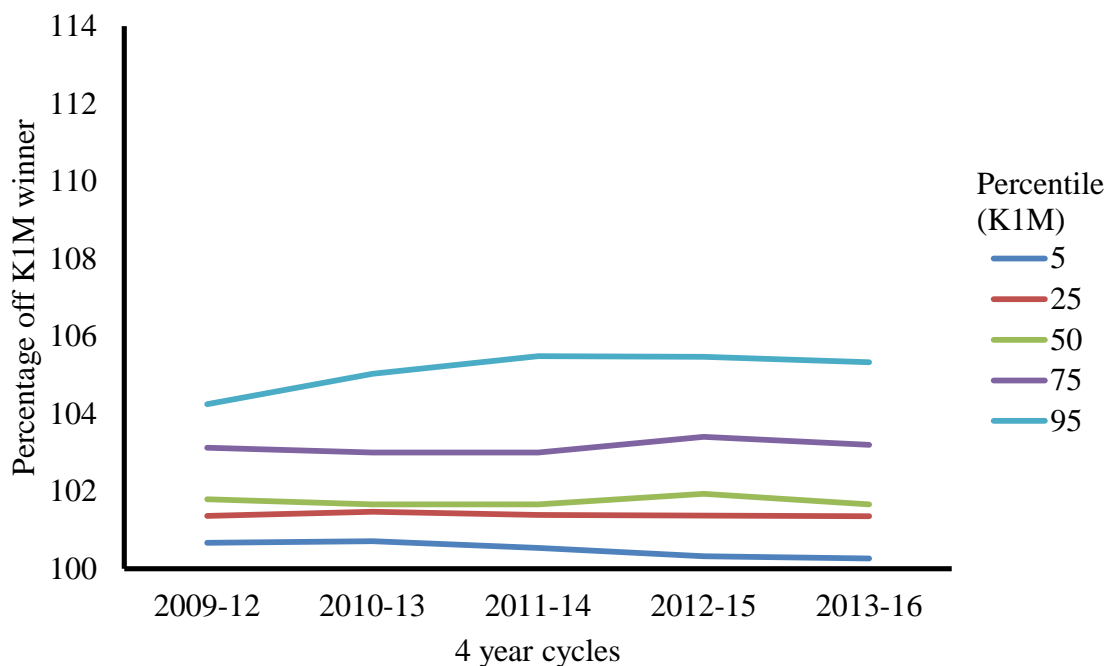
Figure 4.17 Relative 3rd place performance off K1M winner in 4 year cycles for K1M

4.4 Discussion

4.4.1 Characteristics of race performances in canoe slalom

ICF rules stated that a canoe slalom course should be navigable in around 95 seconds by K1M. Race times for K1M in International races between 2009 and 2016 varied between 70s and 108s due to courses being shortened due to prevailing weather conditions or course designs that meant this time was impossible. The introduction of C1W in 2009 introduced a new class where performance levels were not comparable to the other classes due to less experienced athletes competing although their performances have been catching up over time. The variability in course length and difficulty, both within and between classes, promoted the use of race percentages to determine performance levels.

4.4.2 Canoe slalom race percentages

Percentages were produced for each class off the K1M and class winners to determine their appropriateness as a determinant of performance level during a race. Using the K1M winning time as the 100% benchmark, all other race times, from all classes, were converted to percentages off that time. These percentages revealed that the K1M finals produced relatively small differences in performance within and between races in relation to the other classes where the variability increased as the average performance level decreased. This variation was to some extent expected and was in agreement with Nibali et al (2011) who concluded that variability of performance in canoe slalom is greater than that of comparable sports. This variation did not elucidate whether performances were improving in time, classes were improving at different rates, different courses affected different classes differently or whether canoe slalom was sufficiently competitive to produce reasonably consistent times. Nibali (2011) presumed this variability arose from the technical demands of the sport.

The year to year distribution of race percentages for 1st (excluding K1M as this was always 100%), 3rd, 5th place in finals and 10th place in semi-finals (the last place to qualify for the final) suggested that all classes, excluding C1W, were relatively consistent with both within and between year and class variability not clearly demonstrating any between class differential in performance change and no within class change in differentiation between performance levels. However, if the K1M winning times were improving over time i.e. performance on the same course would be better in 2016 compared to 2009, but impossible to measure since course conditions change every race, then athlete improvements over time were relatively consistent across the classes with the exception of C1W. This new class did present improving race percentages in relation to K1M suggesting that this class was improving faster than any other class, the other classes were

not changing in relation to K1M. The C1W 10th place percentages off the class winner suggested that the gap in performance between 1st and 10th was closing over time which is consistent with an increasing number of athletes competing at the elite level.

One problem encountered when analysing the race percentages was the lack of data, there were only 49 International level K1M races (excluding the Olympics which excludes athletes due to nationality restrictions) over 8 years. The number of races is controlled by the ICF which sanctioned an increase in World cup events to 5 per year in 2014. The lack of data meant that statistical methods were limited and hence the decision to consider performances over four-year Olympic cycles, which coincided with development and coaching plans for the National canoe slalom team. Future research could increase the number of competitions by using lower standard races which are on the canoe slalom race calendar. The ICF produce a ranking system which weights the quality of the competition and therefore could extend the analyses to assist in monitoring an athlete's progress over time.

Chi square analyses compared the proportion of race percentages in percentile (25th, 50th and 75th) categories between the 2009-12 and 2013-16 Olympic cycles. The percentiles were calculated from the 2009-12 Olympic cycle data meaning that approximately 25% of the 2009-12 observations were in each of the four percentile categories. Whilst the Chi square test was able to assess the extent to which race percentages had shifted between the percentile categories, the low number of races meant that the assumption of minimum expected cell counts of 5 was violated on 29 occasions out of 31. Clearly a large amount of caution should be applied to these findings from a statistical standpoint. The data tended to back up what was originally thought from the previous analyses although there was more evidence of overall class improvement for C1W, new evidence that K1W's 10th place was closing the gap on K1W's 1st place, and a

consistent suggestion that performances were tending to get closer to both the race and K1M winning times in most events. Each of these shifts in race percentages between 2009-12 and 2013-16 Olympic cycles would be expected if competition was getting closer due to more athletes competing and training in the different classes. Similarly, if advantages due to equipment were being eradicated the expectation would be that race times, and hence race percentages, would get closer.

It was clear, however that race percentages, used in isolation to determine performance levels, such as 'what it takes to make a final or medal', was subject to error due to the variability in race performances, due to unknown factors, but likely to include within athlete variability, course variability and between class variability. For example, when using percentages off K1M, the course design and the number of turns at a particular race, could be advantageous or disadvantageous to K1M versus the other classes or provide no advantage to any class.

One of the presumptions related to using the K1M or class winner was that these times represented something close to the best possible run time possible although this may not always be the case. The assumption of the winning time being nearly perfect is not too farfetched in that the top 10 athletes in the world are competing together and to win it is likely that a very good performance is needed. In this scenario when comparing between races and classes the race percentages would be deemed to be a fair reflection of performance as they are being compared to the best time. However, if a race was won by an inferior performance e.g. if the race conditions changed during the final resulting in an unfair advantage for one athlete who made some errors but still won the race, then the race percentages would be better than if the winner had performed at the expected level. The extent to which this factor influenced the results is unknown but future studies could consider using the virtual best (VB) times (sum of the best split times) for both K1M and

class. If this is considered, then questions should be raised as to how many split sections should be used as the more splits would equate to faster VB times but would also introduce more timing measurement errors. The VB method would be very time consuming if undertaken in event using manual methods as opposed to automatic timing gates. At this point in time a more cost-effective solution was to examine the probability (chance) associated with achieving designated race percentages.

4.4.3 Probability of medalling in canoe slalom

During a race there is time between performances for the athlete to compare their run against other competitors. The coaches and/or analyst will examine the video's looking for time gains and losses to identify a realistic desired outcome and plan for their next run which could be a final. Probability charts off K1M (Table 4.3) and class winning times (Table 4.4) could be examined to determine how close a run was to a theoretical medal or winning performance using the percentage time off the current race winner. The coach can decide a risk versus reward strategy for an athlete using the percentage chance of medalling based on historical data (2019-12 data in this study). For example, depending on the percentage gain thought necessary the coach and athlete can decide to try a faster manoeuvre at a particular point in the race. The 2019-12 probability charts (K1M and class winner) were tested for their predictive ability using the 2013-16 Olympic cycle race results. Signal detection theory was adopted to determine the number of misses i.e. when the prediction stated that an achieved race percentage 'would not medal' but in actuality a medal was achieved, and the number of false alarms where the prediction 'will medal' was erroneous and the race percentage 'did not medal'. Using this terminology, to present a target race percentage that has a high probability of medalling i.e. a low percentage near 100%, there is an increased risk that athletes will have higher race percentages but still medal i.e. a high chance of a miss. Increasing the race percentage target i.e. lowering the

probability of medalling will increase the chance that an athlete will achieve the target but not medal i.e. a false alarm. Hence, increasing the probability of a target race percentage being a hit also increases the risk of a miss but decreases the chance of a false alarm. This approach was used to find an appropriate balance between the risk of misses and false alarms (they are inversely proportional) for the 2019-12 probability charts (K1M and class winner) using the 2013-16 Olympic cycle race results. The results unanimously supported the use of race percentages off the class winner, rather than the K1M, with a 50% level of probability of medalling resulting in a low rate of false alarms (maximum 6%) and a high hit rate (over 70% of medals correctly identified). Targets such as these are commonly used although the basis for the race percentages has not been well documented.

Whilst the creation of race percentage targets was based on an Olympic cycle's data, further analysis could determine whether this was the most appropriate. Indeed, new race percentage targets were presented, but not tested, suggesting that small changes are likely if targets are set using the previous i.e. rolling, four year's data as opposed to using a fixed four year target over the whole 4 year Olympic cycle. Targets can never be 100% accurate, however, since race performances are both variable and closely contested. The main outcome from this study is the verification that race percentages have some predictive validity and that percentages off class winners are more accurate than off K1M. However, this is limited to a small amount of competitions and the percentages do not take into account the quality of the field in each race. This is an area of future research that could be linked to O'Donoghue & Cullinane's (2011) work on evaluating players and teams by determining how much better or worse a player had done then would have been expected in a match against an opponent of the given world ranking. In canoe slalom terms this could examine start lines, allow for more races to be analysed and the potential for athlete ratings based on a range of performance indicators.

4.5 Application to domestic races and training race simulations

Athletes will compete in domestic races and training race simulations which are often limited in terms of the number of international paddlers at the event. This means that the percentages off K1M or class winner are often not comparable to international events such as world cups. The ability to monitor performance in these environments with world standard benchmarks would enhance athlete monitoring, feedback and development. To some extent this can be achieved by monitoring the race percentages of the athletes in a National team and using their virtual best run times, not strictly necessary but advisable if the athlete didn't put together a great run, to estimate a theoretical international race winning time.

4.6 Conclusions

The variability in race percentages was evident throughout this study and understandable in terms of variation within an athlete's performance, between course and class differences and within class (between athlete) performance differences. These can be considered sources of error when trying to establish performance norms for canoe slalom. However, the use of a probability measure e.g. the race percentage off the class winner achieved in the semi-final has a 50% chance of making a medal in the final, has a reasonable degree of predictive validity. This measure can be used in competition to assist performance plans and gauge the risk versus reward benefits. However it is advised that the percentage targets do need to be used with caution and combined with in event intelligence and coaching knowledge. The recommendation for future work is to develop race percentages further by exploring their application to training, determining the usefulness of virtual best times and to improve the predictive power of targets by determining the optimal time scale on which to collect data. However we are still unable to monitor an athlete's performance accurately

over time using percentages in isolation and need to consider widening the number of competitions analysed. Therefore study 3 will focus on the evaluation of the ICF ranking system which uses quality factor criteria per competition. This has the potential to assist in developing individual profiles and tracking of performance to determine whether an athlete is progressing, plateauing or deteriorating.

Chapter Five: Study Three

A TIME SERIES OF INDIVIDUAL PERFORMANCE IN CANOE SLALOM

5.1 Introduction

Many governments around the world are currently providing significant funding to enhance the chance of success in different sports. Consequently, each sport's governing body needs to provide evidence to support their position, both within and between sports, to maximise their funding revenue. Often this is based on expectations of medalling at major International events such as the World Championships or Olympics. This is not straight forward, similar to estimating stock prices, previous performances are not a guarantee of future performance. In canoe slalom this may be even more problematic as Nibali et al. (2011) showed that the variability of performance in canoe slalom was greater than that of comparable sports. However, with an increased pressure to perform at the highest levels to secure funding a worthwhile goal would be to both monitor performance and estimate future potential utilising some statistical evidence. Study 1 also found that coaches wanted the ability to track the performance of their athletes to determine whether they were performing according to pre-determined targets and whether they were improving.

The ability to predict a future performance is determined by the extent to which future performance is a consequence of past performances (James, 2012). Furthermore, James (2012) outlined that this type of prediction is based on the principle that any performance is a consequence of prior learning, inherent skills, situational factors such as motivation, and in some sports the influence of the opposition. This study will apply this reasoning to canoe slalom racing with the proviso that relatively large variations in

performance (Nibali et al., 2011) and small numbers of events per year will limit the use of inferential statistics. Similarly, inappropriate statistics would provide false or unreliable results. For example, Whipp & Ward (1992) used a linear regression model to predict that women marathon runners would run as fast as men by 1998, subsequently proven inaccurate and queried by James (2012) as to why this model was used. Nevill & Whyte (2005) suggested a flattened S shape logistic curve was best for analysing world record times in m/s, showing initial slow improvement followed by rapid improvement and then back to a slow rate of improvement. This model clearly makes more sense than a simple linear one since there is bound to be some limit to human performance, a linear model predicts that at some point in the future running events will take no time at all! James (2012) suggested that Nevill & Whyte's (2005) model also did not consider the limits of human performance and therefore their predictions would become more erroneous over time. James' (2012) predictions in real tennis were concluded to include inherent error related to the uncertainty of future performance suggesting the use of upper and lower confidence limits for any predictions. In the case of a prediction for an individual's future real tennis performance James suggested that 95% confidence limits were too lenient and consideration for lower levels of confidence were recommended.

5.1.1 Smoothing algorithms

In business, predicting future performance based on previous performances is often achieved using a smoothing algorithm. The simplest form of this predicts the next value in a sequence using the average of all previous performances. This places equal weight on all previous observations and is only appropriate if performances do not vary a lot i.e. they are relatively stable. An extension of this involves only using more recent performances and

ignoring past ones i.e. a moving average. This can involve using the previous 3, 5 or any other number (k) of previous observations on which to calculate the average. This method does not predict peaks and troughs very well however with larger values of k having a greater smoothing effect. Smaller values of k place more weight on the recent observations and the moving average therefore tracks fluctuations in the data better. The next progression in terms of prediction involves weighting more recent performances over past ones i.e. exponential smoothing. Consider the situation where early data shows no improvement whereas at some point in time improvements are obvious. By weighting more recent performances over more historical ones, since older ones have less relevance, achieves better forecasts, particularly when trying to remove random variations, i.e. noise, from the data. However, these types of smoothing algorithm, as well as the moving averages approach perform poorly when there is a trend in the data.

An alternative approach, to counter the limitations of basic smoothing algorithms i.e. they do not cope with upwards or downward trends very well, was devised by Holt (1957) who included an adjustment to account for a trend in the data as a function of time. Holt's method also allows the trend to change with the addition of each new observation. The Holt (1957) method was later extended to the Holt-Winters method, also known as double exponential smoothing (Winters, 1960) to consider seasonal variations that repeat each periodic cycle i.e. a year, quarter or any other time frame. This method was not implemented for the canoe slalom data as repeated seasonal variations were not expected.

Some concern has been raised about outliers in a data set (e.g. Gelper, Fried & Croux, 2010) as they have an undue influence on the prediction equations. In canoe slalom, athletes who receive a 50 second penalty at a race are invariably awarded ICF points that are very large and usually outliers in the historical record for that athlete. However, if an athlete gained better points in the semi-final than the final their semi-final

points would be used as their ranking points for this race (ICF, 2015). If an outlier was present, Gelper et al., (2010) suggested three basic choices could be implemented in a smoothing equation i.e. do not adjust the data (as per ICF rules), remove the outlier from the data or replace it with a value considered more representative of the athlete's typical performance.

In study 1, coaches identified the need to know whether an athlete was on track (according to their performance targets) and whether there was potential to identify trends in performance, such as consistency, loss of form or improvements. The ICF use a world ranking system based on an athlete's 5 best (ICF approved events) results over 2 years. A ranking formula uses a quality factor based on the level of the best five competitors at the competition unless it is a World class event where the factor is always zero, and the competitor's race result (Section 5.2.2.1). The World ranking points is therefore, a smoothed performance measure, being relatively insensitive to individual race points. This study will utilise smoothing techniques to individual ICF race points gained by athletes to identify more sensitive evaluations of current form and progression over time. Data will be presented graphically over time as many of the broad general features of a data series can be seen visually and the human eye can be a very sophisticated data analysis tool (Montgomery et al., 2015). Smoothed race points will be used to create a performance time series for an athlete with a view to supporting decision-making for talent identification and team selection.

5.2 Methodology

5.2.1 Participants

All ICF World ranked K1Men (K1M) canoe slalom athletes competing at International races between 2006 and 2016 were eligible for analysis. Races included, ICF World cup, ICF World championships, ICF ranking races & continental championships (n=180).

5.2.2 Procedure

Ranking points for individual K1M races (2006-2016) were obtained from the National Governing body although they were also accessible through the official ICF website (https://www.canoeicf.com/icf-canoe-slalom-world-ranking). ICF points are calculated at each event with points calculated for each individual phase (heats, semi-final or final) and the lowest of the potential 3 scores awarded to the athlete (Section 5.2.2.1). Ranking points were imported into Microsoft Excel for analysis using smoothing algorithms and funnel analysis which are explained in Sections 5.2.2.2 & 5.2.2.3.

5.2.2.1 ICF Race Points formula

ICF race points are used to calculate World ranking points (since 2009), calculated as the average of an athlete's 5 best race points over two years (before 2013 the best 3 race points were averaged). The formula for ICF race points is:

$$\text{Race Points} = \text{Phase Offset} + ((150 * \text{Race time}) / \text{Race time of Leader}) - 150 + \text{Quality Factor}$$

Key: Phase Offset is 0 for a final, 10 for a semi-final and 20 for the heats
Quality Factor is the average ICF World ranking points of the five best athletes competing at the event

Athletes who received a DNS (did not start) in both heats runs are not considered in the Quality Factor calculation. The major events, ICF World Cups, World Championships and Olympic Games are always given a Qualify Factor of 0.00. The phase offset is used to allow athletes who made the finals to generally rank better than athletes who finished in the semi-final and those better than athletes who finished in the heats.

5.2.2.2 Smoothing algorithms for canoe slalom

In this study the formula for producing a simple exponential smoothing forecast was:

$$F_{t+1} = \alpha Y_t + (1 - \alpha)F_t \qquad \text{Equation 1}$$

Key: F_{t+1} is the forecast at t+1; α is the smoothing constant with a value between 0 and 1; Y_t is the ICF race points at time t; F_t is the forecast at time t (the smoothed value at time t)

The above formula can be rearranged to:

$$F_{t+1} = F_t + \alpha(Y_t - F_t) \qquad \text{Equation 2}$$

Hence the basis of a new prediction (at t+1; Equation 2) was the previous forecast (F_t) plus the difference between the actual ICF race points (Y_t) and the predicted ICF race points (F_t) i.e. the error of the previous prediction, adjusted by α (the smoothing constant). On this basis, the exponential smoothing forecast learned from previous forecasts. If the value of the smoothing constant was near to 1 then recent values in the time series were weighted heavily in comparison to the previous forecast whereas values nearer to 0 weighted the previous forecast more heavily than the most recent value. The error

associated with different values for the smoothing constant were compared using the root mean square error

$$RMSE = \sqrt{\frac{\sum(y_t - \hat{y}_2)^2}{n}}$$

Equation 3

Exponential smoothing can be used on small data sets and were hence applicable for the small number of canoe slalom International events. They were also easy to implement but are limited in that the forecasts tend to lag behind the actual data. An alternative approach, to counter this limitation i.e. they do not cope with upwards or downward trends very well, was devised by Holt (1957) who included an adjustment to account for a trend in the data as a function of time (referred to here as T_t, Equation 4)

$$F_{t+1} = \alpha Y_t + (1 - \alpha)(F_t + T_t)$$

Equation 4

As for the simple exponential smoothing forecast (Equation 1) the value of α (the smoothing constant) could vary between 0 and 1 but the trend function could forecast a continual growth or decline in performance. To illustrate this, consider a K1M athlete who has just joined the senior tour after a successful junior career. The expectation is that his performance would improve over the season and his International Canoe Federation (ICF) points awarded at each event (based on finishing position and quality of the field) would tend to improve (lower points being better). Suppose the athlete was awarded 60 points in a race in July, the expectation from the coaches was that he would be awarded 50 points in the next race, he went on to get 40 points. In terms of the forecast two extreme views are possible. The forecast was good and the race outcome an anomaly which suggests not changing the forecast for the next race i.e. he will improve by 10 points and hence gain 30 points next race. Alternatively, if there was minimal faith in the forecasting the actual

117

performance could be selected and hence the prediction of 40 points next time out ("tomorrow will be the same as today" principle). These two extremes suggest the athlete would achieve between 30 and 40 points although a more pragmatic approach would suggest somewhere in between i.e. use a value between 0 and 1 in Equation 4.

Holt's method also allows the trend T_t to vary with time:

$$(T_t = \beta(F_t - F_{t-1}) + (1 - \beta)T_{t-1} \qquad \text{Equation 5}$$

Key: β is the smoothing constant for the trend, with a value between 0 and 1.

Holt's method for forecasting a series with a linear trend thus allowed the trend to change with the addition of each new observation. However, when an upward or downward shift in the time series was expected both the level and slope were adjusted using different smoothing constants for both. This allowed an adjustment to the rate at which the level and trend were tracked. The level estimate was calculated with:

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \qquad \text{Equation 6}$$

The trend estimate was then calculated with:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \qquad \text{Equation 7}$$

Finally, to calculate a forecast m steps into the future:

$$F_{t+m} = L_t + mT_t \qquad \text{Equation 8}$$

To use these equations both the smoothed values (Equation 6) and the trend estimates (Equation 7) needed to be initialised. A number of ways to do this are possible. L_1 could be set to Y_1 and L_2 to Y_2 to initialise the smoothed values and $T_2 = Y_2 - Y_1$ used to

initialise the trend estimates. Another option, when more data is available, is to start the initialisation process at some point such as at the end of the first or second year and use the average values for level and trend. For this study the first five ICF race points for an athlete were used to calculate the trend (T_t) using the slope function in Excel and the level (L_t) using the intercept function. Five ICF race points was selected as this was the minimum number of races to determine World ranking points. If an outlier was present, usually due to a 50 second penalty for the elite athletes, the data was not adjusted and the smoothing equation applied as normal.

The RMSE (Equation 3) was calculated for each permutation of α (between 0 and 1) for the exponential smoothing equation (executed using a macro to record the multiple keystrokes) with the lowest value used to select the α value. Similarly, the solver tool was used to calculate the best values of α and β to minimise the RMSE for the Holt equation. These values were used for subsequent comparisons between the two smoothing methods.

5.2.2.3 Performance funnels

To track an athlete's progress over time and determine if an athlete was on target to win their first race, exponentially smoothed ICF race points were compared to a performance funnel. This was envisaged in a similar way to Spiegelhalter (2005) who suggested that a 'funnel plot' could be a form of control chart where an observed indicator was plotted against a measure of its precision. The purpose of such a plot being to visually identify differences between performances involving high and low error for example. Similarly, De Smith (2015) suggested that time series analysis has advantages in being able to visually present data to observe averages, peaks and troughs and critical turning points.

The basis for the use of a performance funnel here was that it was hypothesised that athlete's performances would change over time i.e. variability would change, potentially becoming smaller as performances improved. A plot of some measure of performance would therefore resemble a funnel if this was the case. On this basis, a 'winners' funnel was created based on the exponentially smoothed ICF race points for all athletes that first entered International competitions from January 2006 and went on to win a race at some point in time (n = 11). Similarly, a 'winless' funnel was created based on the ICF race points for all athletes that first entered International competitions from January 2006 but didn't go on to win a race up to September 2016 when they had to be ranked in the World's top 40 (n = 16).

The 'winners' & 'winless' funnels were created by synchronising each time series at the month each athlete achieved their first ICF world ranking i.e. their 5th race was set as month zero. Data for each subsequent month were presented using the median and 95% confidence limits (95% CL) for the median (as suggested by James et al., 2008) as the data distributions were typically non-normal, suggesting a non-parametric approach.

5.3 Results

5.3.1 Winning K1M athletes

Nineteen athletes won races between 2006 and 2016, 11 of which debuted after 2006 and were included in the study with the 8 athletes who debuted before 2006 excluded. One of the excluded athletes dominated the top of the podium with 22% of the total wins (n = 50) at the major championships during this time (Figure 5.1).
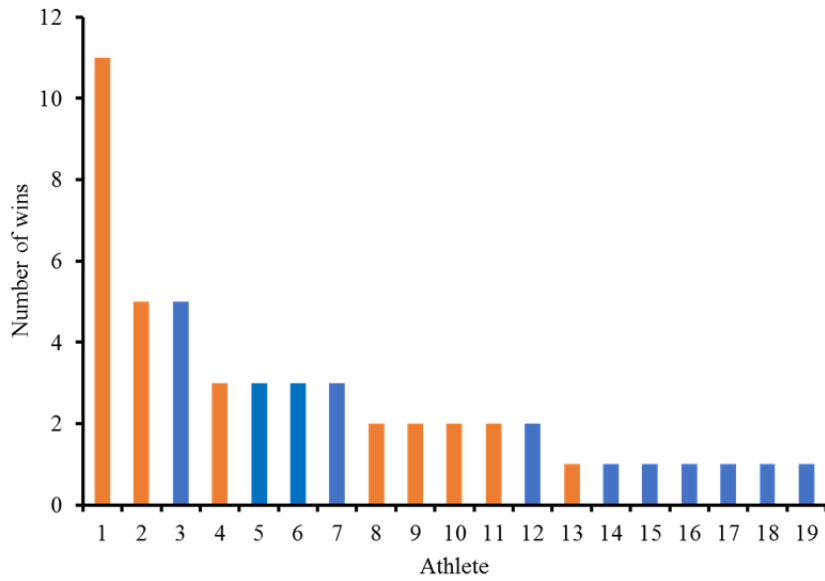
Key: ■ debut before 2006 ■ debut after 2006

Figure 5.1   The number of wins per K1M athlete at the major championships between 2006 and 2016

Of the 11 winning athletes, 55% (n = 6) won their first major championship at the venue for which they competed (home advantage) and 4 of these were won within 3 years of their first ICF world ranking (Figure 5.2).
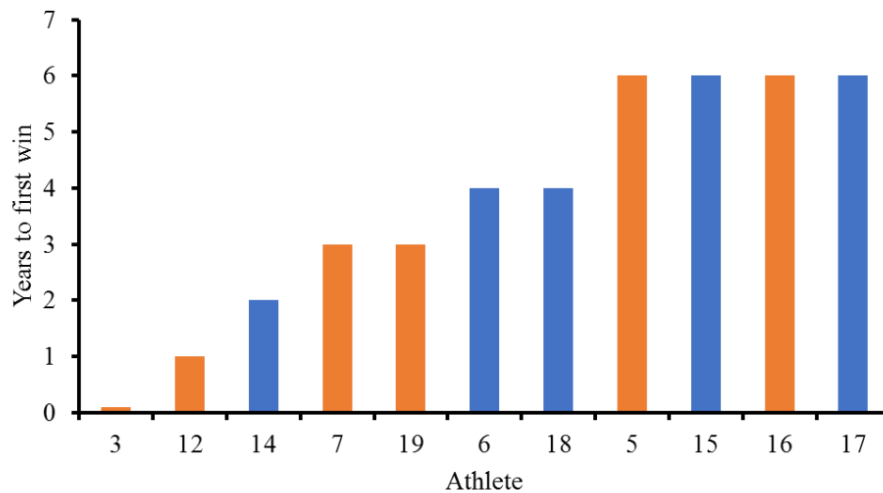
Key: ■ home win ■ away win

Figure 5.2   The number of years to an athlete's (n = 11) first major championship win between 2006 and 2016

5.3.2 Applying smoothing algorithms to canoe slalom

Great race to race variation in ICF ranking points per race were evident (see Figure 5.3 for an example) suggesting some form of smoothing to the time series data to aid interpretation of any trends.
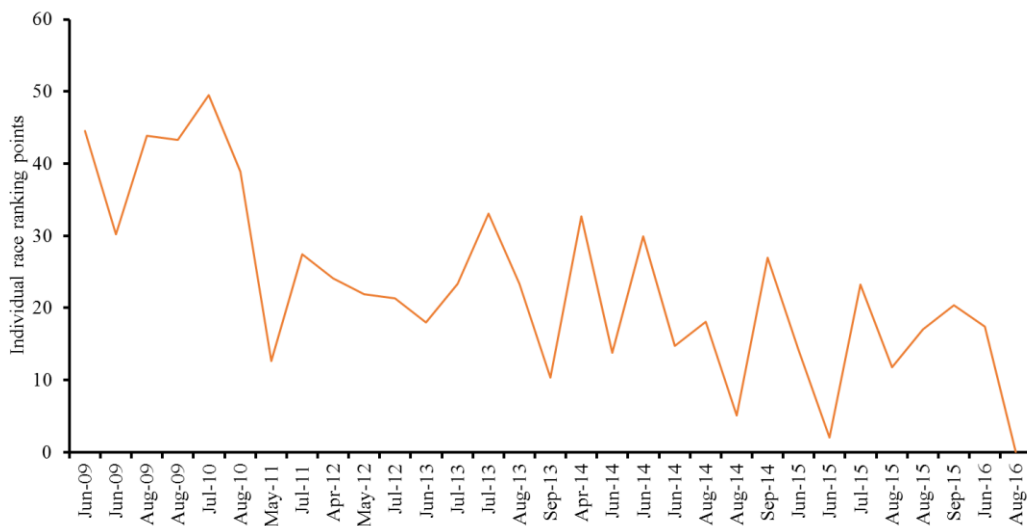
Figure 5.3   Time series of ICF ranking points per race for an athlete (17)

There was typically a reasonably strong relationship ($r = 0.62$, large effect size) between an athlete's current World ranking and their race points (Figure 5.4) although outliers were often present.
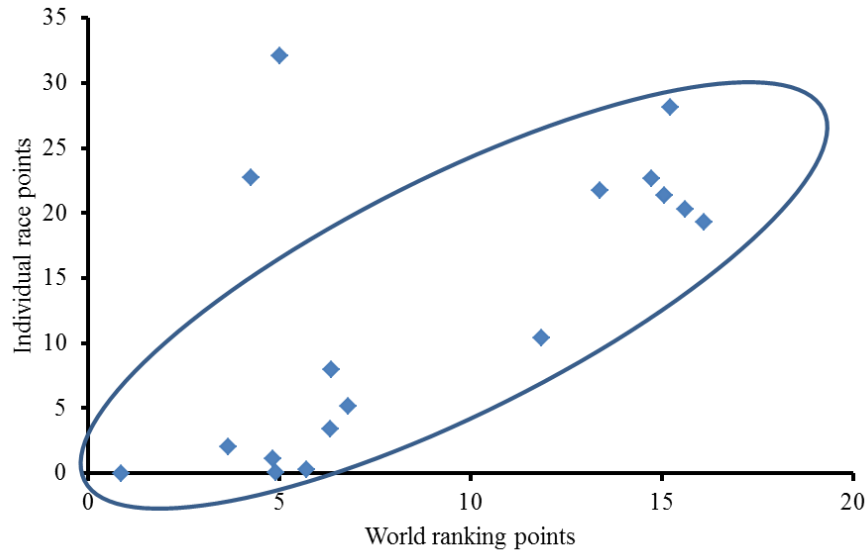
Figure 5.4   Relationship between athletes' world ranking and race points at an
             International ICF competition

Similarly, an athlete's individual race points tended to bear little resemblance to
their ICF World ranking points over time as World ranking points were calculated from
the best five races over two years i.e. over smoothed (Figure 5.5).
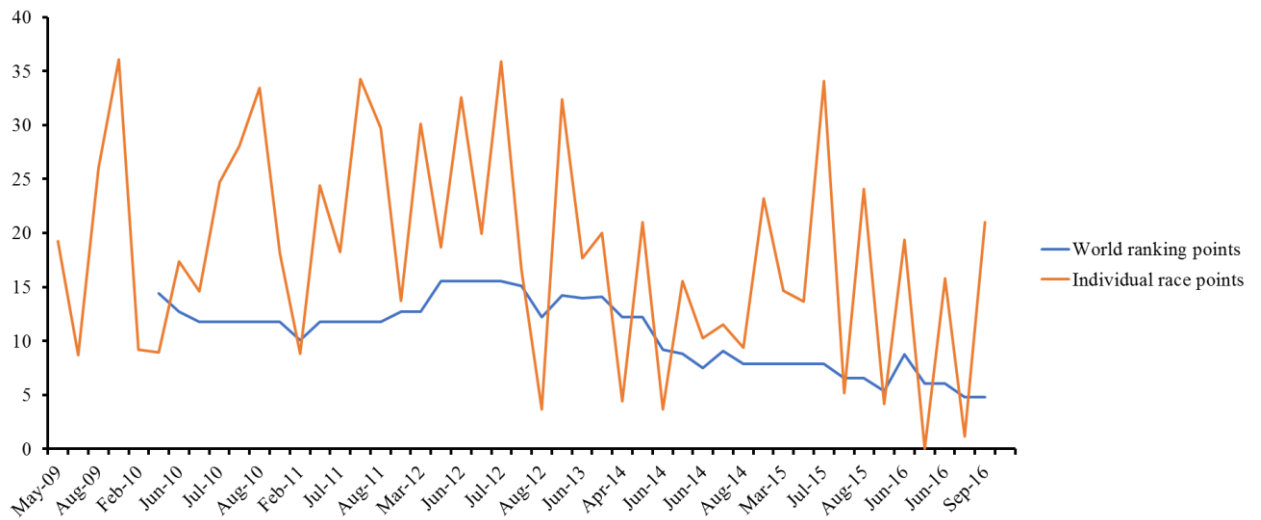
Figure 5.5   ICF World ranking and individual race points for an athlete (15) at ICF
             International competitions

Exponential smoothing and Holt's method were used to produce smoothed race
points using the RMSE to determine the best α and β coefficients to use. The exponential

method tended to be less sensitive to outliers and produce a more representative time series than Holt's (Figures 5.6 - 5.8).

Holt's method tended to over fit any outliers (Figure 5.6).
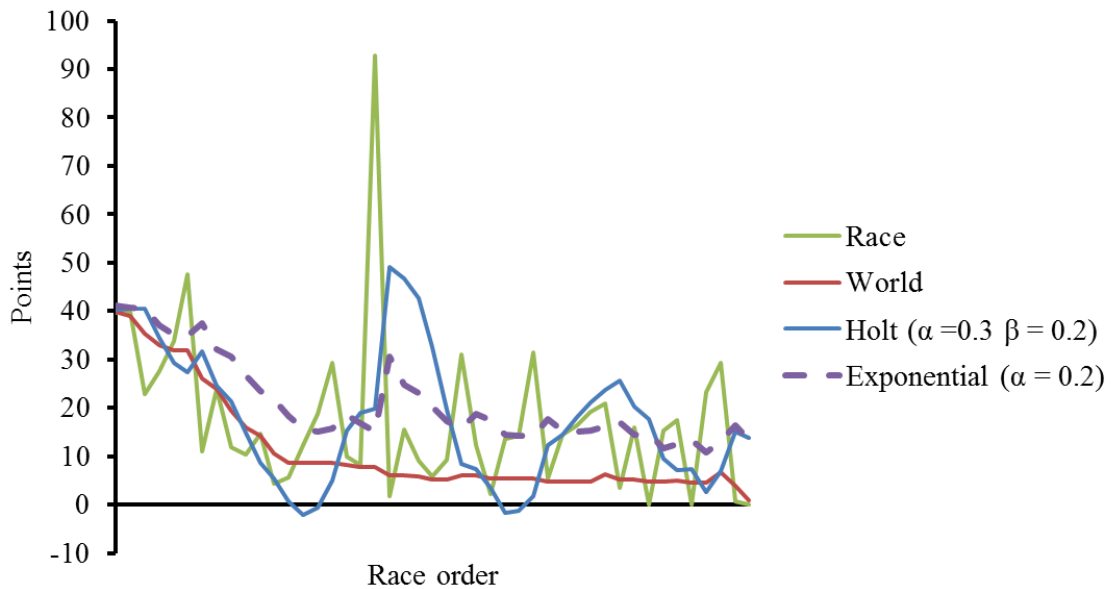
Figure 5.6   Time series of an athlete's (5) race points, smoothed race points and World ranking

Exponential smoothing adequately dealt with a variety of changes in direction (Figure 5.7).
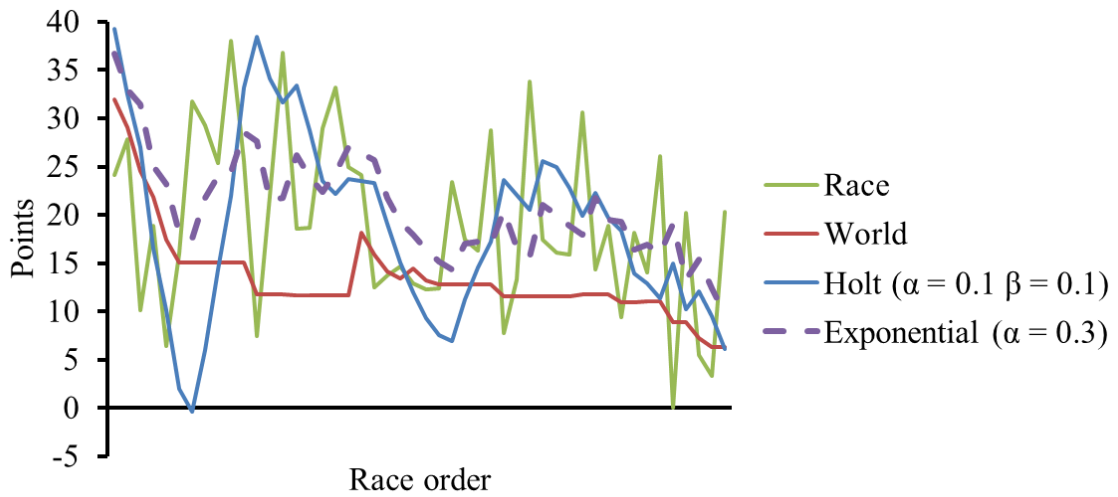
Figure 5.7   Time series of an athlete's (16) race points, smoothed race points and World ranking

For an athlete showing a relatively steady improvement in World ranking the exponential smoothing more accurately portrayed performance fluctuations than either Holt's method or the World ranking points (Figure 5.8). The time series plots for a further 13 athletes are presented in Appendices 5.1 - 5.13.
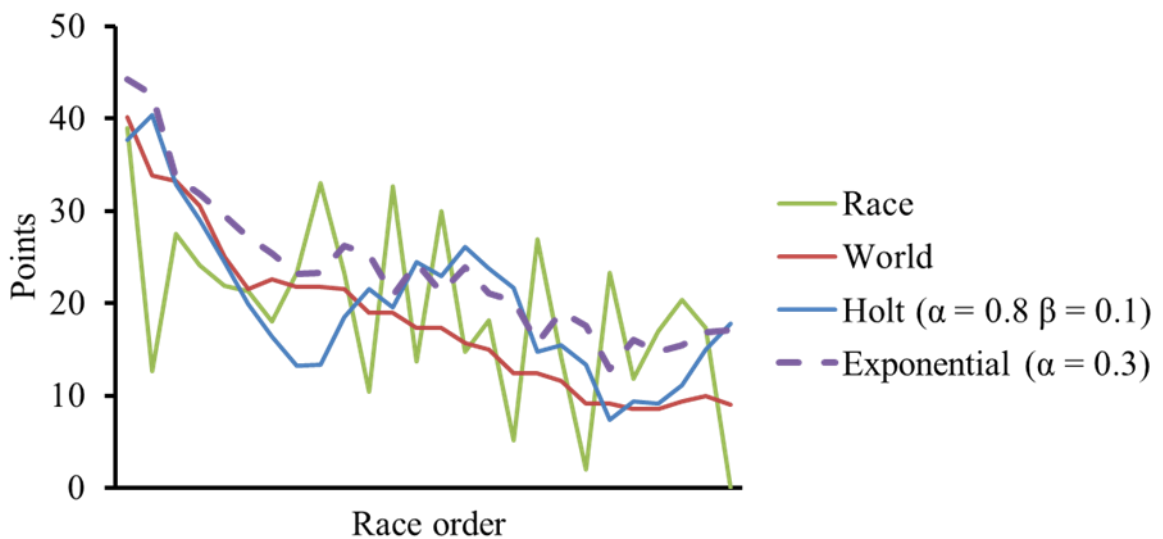
Figure 5.8   Time series of an athlete's (17) race points, smoothed race points and World ranking

5.3.3 Performance funnels for winning & winless athletes

To compare an athlete's performance over time against expected values, funnels were created for athletes who had won a race and for winless athletes. The exponential smoothing algorithm was used to adjust each athlete's race points although different values of α had been selected due to the RMSE values. On reflection, this was thought to add unnecessary complexity and a lack of clarity for future use in the applied setting. On this basis the α values selected by the lowest RMSE were inspected. Any values lower than 0.3 or greater than 0.7 were ignored as being too extreme and the average of all other values (0.4) selected as the constant value for α to be used for all subsequent analyses.

The funnel for athletes who had won a race (n = 11; Figure 5.9) indicted that poor performances became less frequent over time (lower 95% CL decreasing in an exponential manner up to 5½ years) whereas both the median and upper 95%CL remained fairly constant. Increased variability was evident from 5½ years.
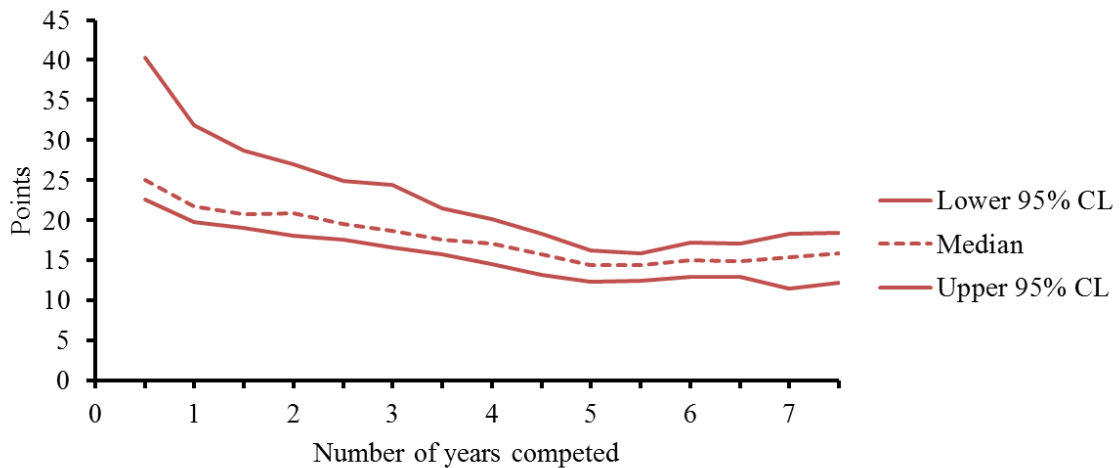
Figure 5.9   An exponentially smoothed performance funnel for winning athletes (n = 11) starting from their first World ranking

The funnel for winless athletes (N = 16; Figure 5.10) indicted relatively linear improvements up to 6 years with some signs of decreasing performance from this point.
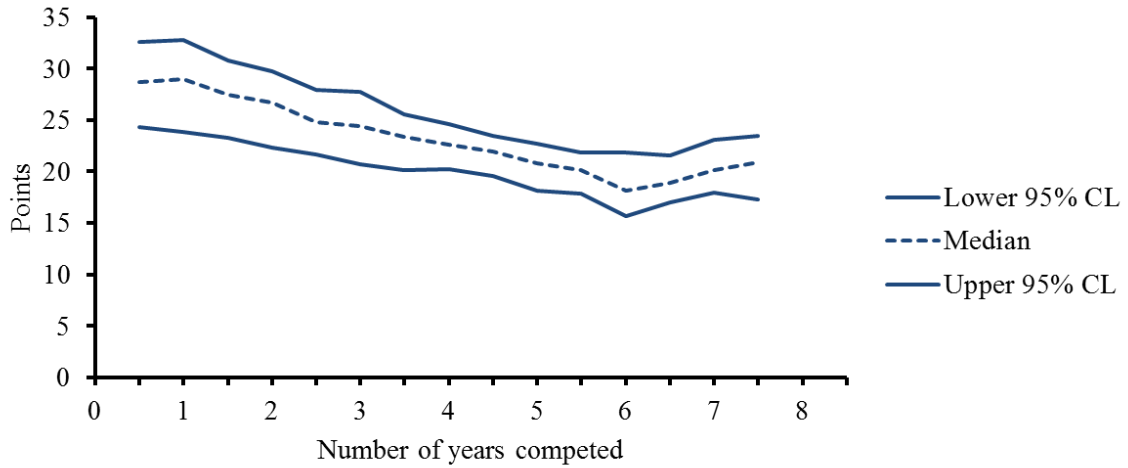
Figure 5.10   An exponentially smoothed performance funnel for winless athletes (n=16) starting from their first World ranking

Both funnels were overlaid to compare winning and winless performances (Figure 5.11). Winless athletes very rarely had exponentially smoothed race points as good as the median for winning athletes with the biggest separation in the two performances occurring after 4 years.
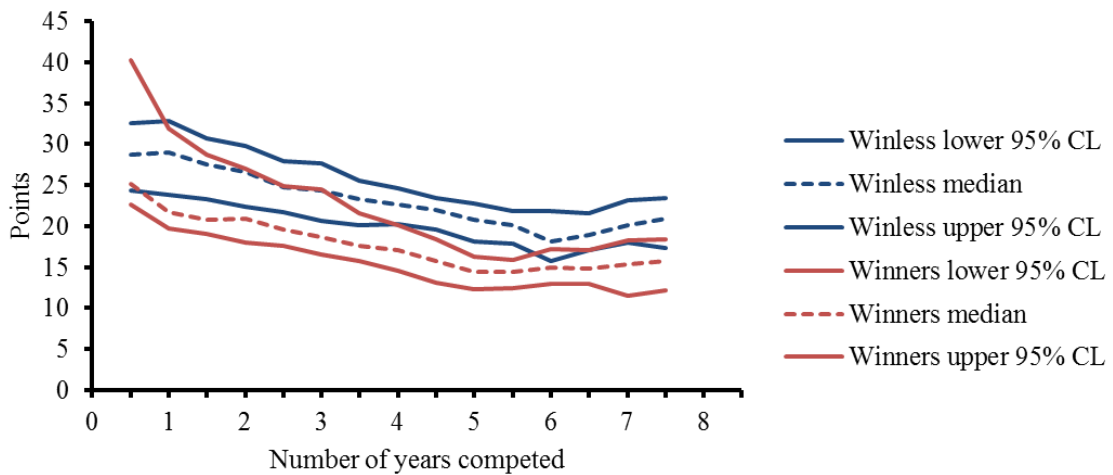
Figure 5.11   Exponentially smoothed performance funnels for winning (n = 11) and winless (n = 16) athletes starting from their first World ranking

5.3.3.1 Time series of a winner's performances

The time series plots of exponentially smoothed race points for all major winners tended to exhibit periods of improvement, times where improvements were countered by poorer performances (plateau) and periods of deteriorating performance.

The exponentially smoothed race points of the world number 1 (as of Sept 2016; Athlete 5) improved consistently during the first two years of competition around which time this score approximated the median of the winning funnel (Figure 5.12). Except for a very poor performance at about 3½ years his performances closely fitted the winning funnel up to his first major championship victory after 6 years.
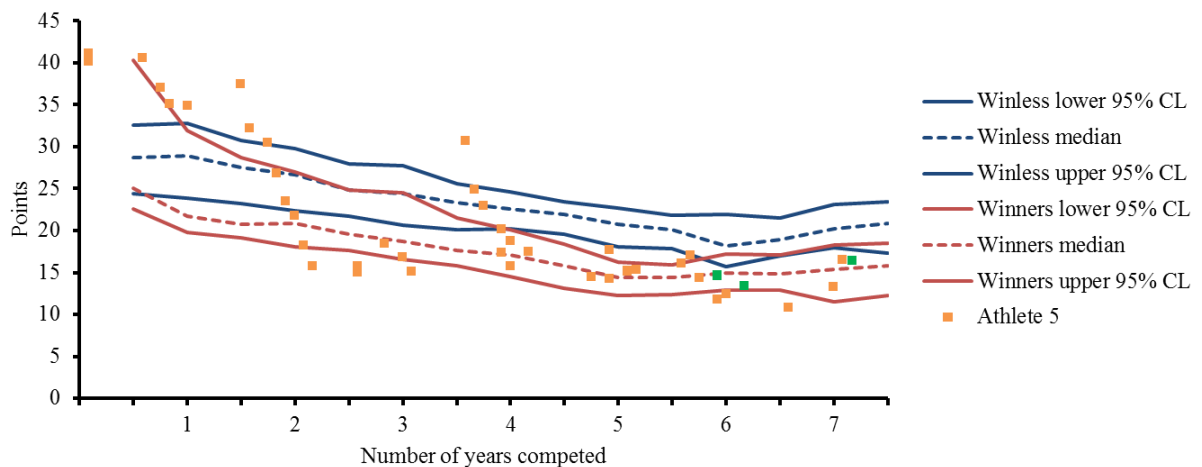
Key ■ major championship win

Figure 5.12   Time series of the World no.1 (Sept 2016) athlete's (5) exponentially smoothed race points to their first major championship win in relation to winning and winless funnels

The 2016 Olympic champion (Athlete 17) improved linearly over the first 4 years after which time his score approximated the median of the winning funnel (Figure 5.13).

Over the next two years, he only raced the summer season and then went on to win his first major championship.
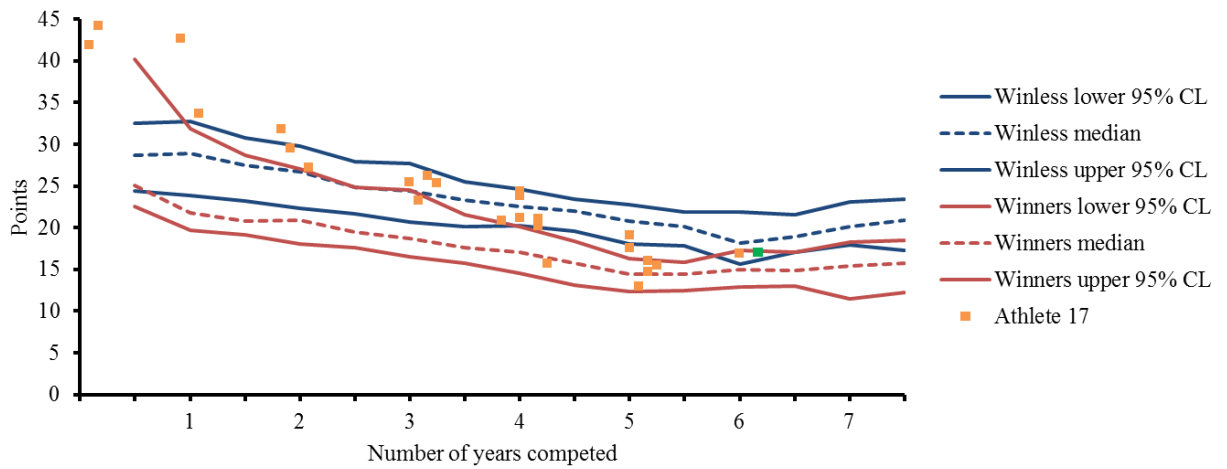
Key ■ = major championship win

Figure 5.13   Time series of the Olympic champion (August 2016) athlete's (17) exponentially smoothed race points to their first major championship win in relation to winning and winless funnels

On his 2nd appearance, and prior to achieving an ICF World ranking, athlete (3) won a major championship, hence it is not visible on the time series chart (Figure 5.14). His smoothed race points tended to improve over the first 4 years, predominately better than the upper 95% CL for the winning funnel, and another major win. A two-year period of relative decreasing performance culminated in two more major wins. Consistently good performance over the next two years also culminated in another major win before a gradual decrement in performance. A further 8 winning athletes are presented in Appendices 5.14-5.21.
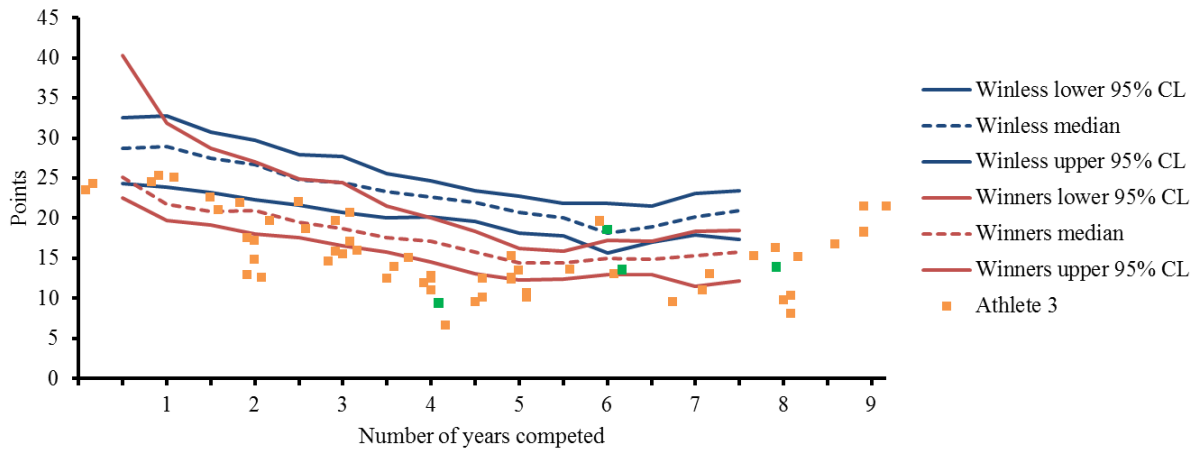
Key ■ = major championship win

Figure 5.14   Time series of an athlete's (3) exponentially smoothed race points to their
first major championship win in relation to winning and winless funnels

5.3.3.2 Time series of likely, possible and unlikely future winners

All athletes were classified against 3 performance criteria based on whether they were
improving, their improvements were countered by poorer performances (plateau) or their
performances were deteriorating. These trends were put into the perspective of the winning
and winless funnels to classify athletes as likely, possible or unlikely to win a major in the
future.

55% of the 11 previous winners and 25% of 16 winless athletes were likely to win
a future race (Figure 5.15) because they were improving towards, or had plateaued at, a
sufficient performance level. Athletes who had plateaued at a sufficient performance level
(31% of winless athletes) were deemed to be possible future winners. 45% of previous
winners and 44% of winless athletes whose performances were deteriorating were deemed
unlikely to win in the future.
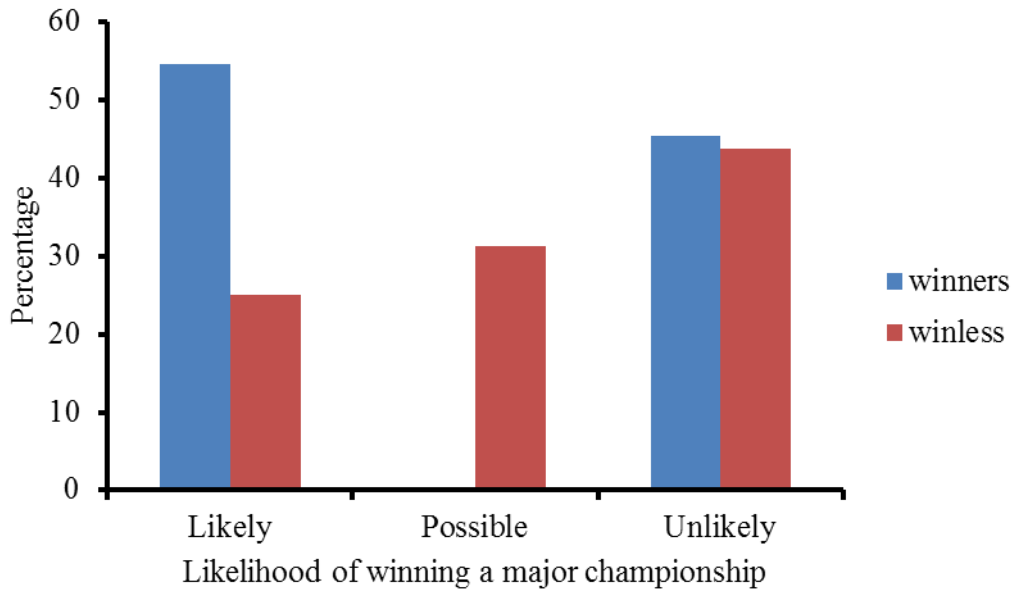
Figure 5.15   The likelihood of an athlete winning a major championship

5.3.3.2.1 Time series of a likely winner

After just under 3 years of competing athlete A demonstrated a relatively linear improvement in his exponentially smoothed race points to a performance level indicative of a likely future major win (Figure 5.16).
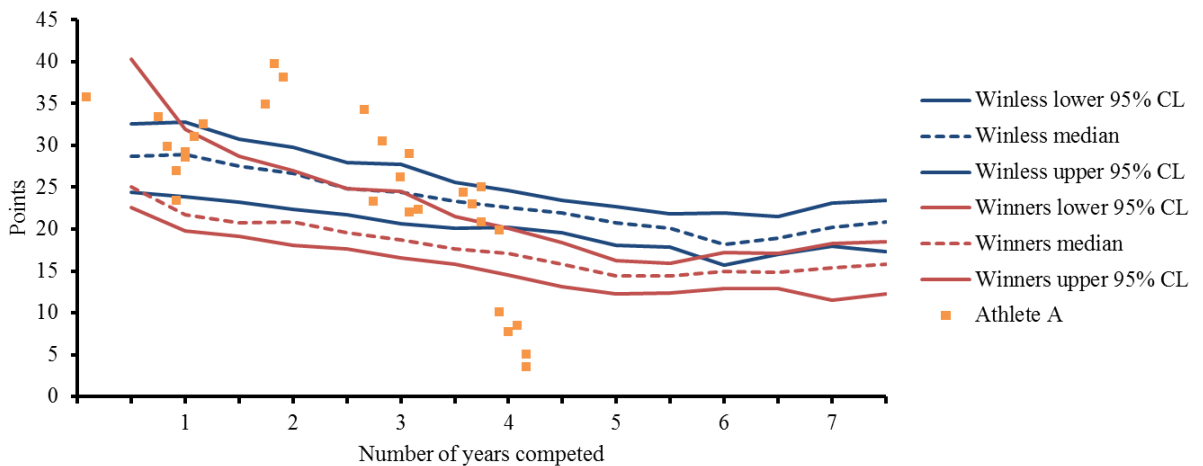
Figure 5.16   Time series of a likely future major championship winner's (Athlete A) exponentially smoothed race points in relation to winning and winless funnels

5.3.3.2.2 Time series of a possible winner

Athlete C has reasonably consistently achieved exponentially smoothed race points at the median or better for the winning funnel throughout his career (Figure 5.17). This consistently high level of performance suggests a future major win is possible.
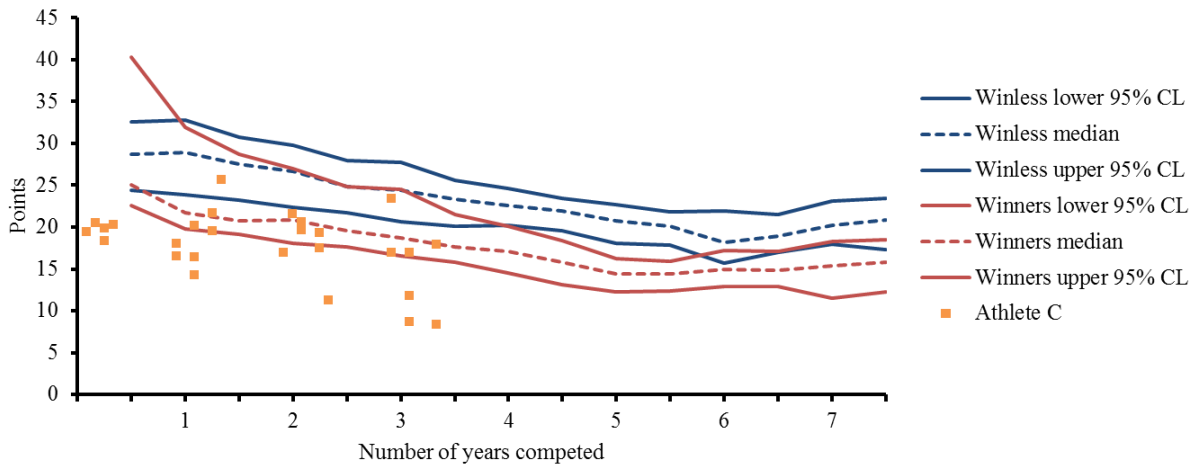
Figure 5.17   Time series of a possible future major winner's (Athlete C) exponentially smoothed race points in relation to winning and winless funnels

5.3.3.2.3 Time series of an unlikely winner

Athlete E gradually improved over the first 4 years to a performance level around the median for the winning funnel (Figure 5.18). A 2-year period of relatively deteriorating performance countered by some improvement was followed by a year of deteriorating performance suggesting a major win in the future is now unlikely. A breakdown for all other athletes likely, possibly and unlikely to win are presented in Appendices 5.22-5.34.
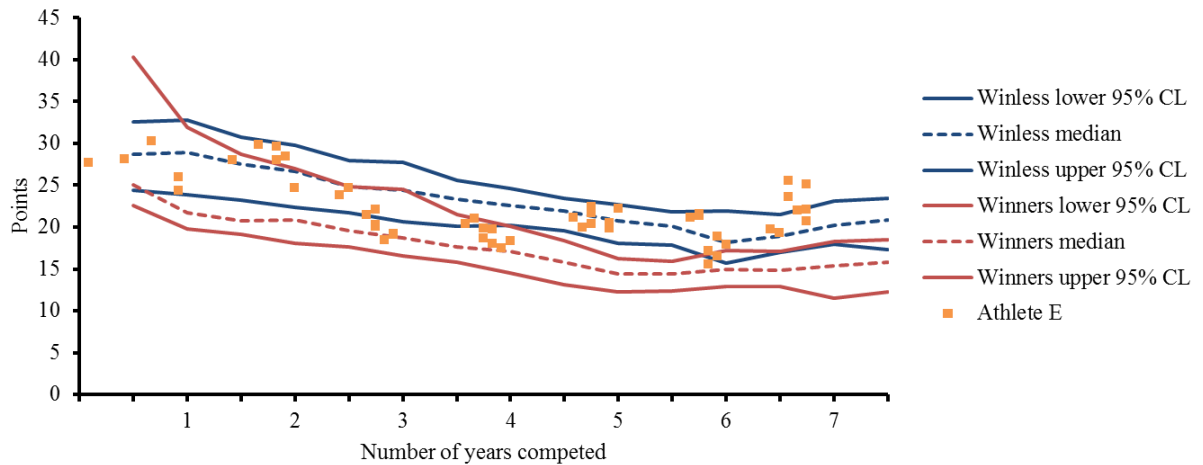
Figure 5.18   Time series of an unlikely future major winner's (Athlete E) exponentially
                smoothed race points in relation to winning and winless funnels

5.4 Discussion

5.4.1 Application of smoothing algorithms

There were clear challenges, due to the high variability in performances (Nibali, et al, 2011), to assessing an athlete over time using raw individual ICF race points. It was therefore understandable that the ICF use a World ranking system based on the best of 5 races over a period of 2 years. This method enables athletes to discard their worst performances and in some instances still maintain a high ICF ranking even when their 'current form' was relatively poor. Similarly, the 2 year window allows an athlete who was not selected for a National team to maintain an ICF ranking, even though they were unable to compete in the major championships (World Cups, Continental / World Championships and Olympics). The ICF World ranking, however, due to the possibility of ignoring poor performances, does not facilitate discerning changes in performance over shorter periods of time and can therefore be described as a highly smoothed performance measure in relation to individual race performances. Indeed, when the relationship between ICF World

ranking and individual race points was assessed it was clear that a positive relationship existed (best athletes in the world tended to finish high in races) but poor individual race performances were possible for top athletes (albeit less often) as well as bottom rated athletes.

The use of smoothing algorithms was explored to produce a performance measure which was more sensitive to current form than World ranking points but would enable time series patterns to be discerned i.e. unlike race points. The Holt and exponential smoothing algorithms were applied using the smallest RMSE value to determine the smoothing constants. When these were plotted alongside the ICF World ranking and individual race points it was evident that both methods could track performance although the Holt method produced larger variations, especially after large peaks or troughs in the data series. Through observations of a number of athletes (n=16), the exponential smoothing algorithm highlighted clear changes in performance level where the trajectory of performance matched the race data better than the Holt method. However the RMSE varied between athletes and sometimes suggested that the algorithm should weight past performances much more highly than recent ones. This caused the smoothed line to be more similar to the World ranking than was hoped for. Indeed, for the purpose of this study the idealised line would have been somewhere between the stability of the World ranking and the variability of the race points. On this basis, a constant value (0.4) for the smoothing constant was selected to reflect the hypothesised ideal and match the values determined by the lowest RMSE. This value gave slightly more weight to the most recent forecast over the most recent race points. The specificity of the fixed smoothing constant value also had benefits for the applied world in that its operationalisation was reliable between analysts and coaches. The exponential smoothing algorithm was thus used to generate performance funnels which could be used to put an athlete's performance into

perspective relative to the time series of previous major winners (winning performance funnel) and the time series of athletes who had never won a major (winless performance funnel). Time series analysis also has advantages in being able to visually present data to observe averages, peaks and troughs and critical turning points (De Smith, 2015).

5.4.2 Winners and winless performance funnels

Performance funnels based on winners and winless athletes' exponentially smoothed race points were created using 95% confidence limits of the median, deemed the most suitable approach for non-normal data distributions (James et al, 2005). Furthermore, presenting the confidence limits, as a form of funnel plot, provided a visual graphical aid for comparison (Spiegelhalter, 2005). The use of funnel plots has been deemed successful in various sectors including the public sector where medical decisions were presented (Rakow, et al., 2015). These authors also suggested that future funnel plots could involve interactive graphics that allow different control limits, or different confidence intervals, to be displayed on request, as determined by a particular decision being required. This interactivity could also be incorporated into future funnel plots for use in the applied world of canoe slalom.

The winners and winless performance funnels tended to suggest that athlete's performances, funnels were based on descriptive statistics of all eligible athletes, incrementally displayed less variability over time from the outset of their International careers up to some point in the future when greater variability reoccurred. These funnels were used as a comparison against an individual's time series of exponentially smoother race points to estimate the likelihood of future success. These funnels were limited to the extent that data for this study was only obtainable from 2006 and there were relatively few

races within the International calendar. Furthermore, athletes that debuted before 2006 were not included in the data analysis, because this data was unobtainable, meaning that the eight athletes who had 28 winning performances during the data analysis period (2006-2016) were not included in the construction of the winner's performance funnels.

The winning and winless funnels overlapped for the initial 4 years of International competition with the degree of overlap diminishing over time. This time period corresponded to the gradual eradication of very poor performances by the winning athletes, hence the improvement seen in the lower 95% CL of the winners funnel. Whilst both funnels depicted the typical improvement in performance seen by internationally ranked canoe slalom athletes (typically over the first 5 to 6 years of competing), only athletes performing around the median level for the winners funnel could be reasonably accurately considered as potential winners. Performances lower than these values may just as likely depict upper levels of performance for athletes who would never go on to win a major race.

The next 18 months to two years (years 4 to 6) typically saw continued improvement in the performance funnels with clear demarcation between the two up until the 6-year mark. During this period athletes could be reasonably accurately (95% certainty) classified as future major winners or not based on their exponentially smoothed race points. Similarly, the final 18 months (years 6 to 7½) of the performance funnels tended to depict slightly decreasing performance levels which were differentiated for winner's and winless athletes.

5.4.2.1 Variability of the time series of a major winner's performances

The individual athletes (n=11) whose performances determined the winners performance funnel displayed individual variation, as would be expected, but a reasonably well defined overall pattern through the 7½ period during which they were assessed. This pattern involved an initial period of improvement (typically between 3 and 6½ years), a period of greater variability where very good performances were countered by some poorer performances (a plateau of up to 4 years) and periods of deteriorating performance (after 6 years on average). There were two outliers in this group which included an athlete who won a major in his first season of competing prior to racing in 5 events and with a World ranking outside 100. His profile therefore didn't include an initial period of improvement as his performance was consistently excellent from the start i.e. defined as in the plateau phase. The other outlier also displayed this trend but didn't win for nearly four years of International competitions. Winning their first major more often occurred during the initial period of improvement, either in the first 3 years of competing (n = 4) or after 6 years (n = 2). However, the athletes (n = 5) who won their first major during their plateau phase typically did so after between 4 and 6 years of competing, excluding the outlier previously mentioned. Just one athlete won a major race after moving into the performance decline phase (5th win of career).

Some of the variability in major race winner's performances can be explained by the fact that individual athletes need to be selected by their National team to compete at the major championships. This means that for the stronger nations World class athletes may not be selected for the major events whereas weaker athletes may get a chance for other nations. Athletes from top nations may have fewer opportunities to gain the best race points due to non-selection and therefore win major races with a seeming lack of experience and ability (Athlete 3 in this study won in his first season). Future research

137

could consider developing performance funnels for different nations (as suggested by Rakow, et al., 2015). Similarly, the positive impact of racing at home was demonstrated in this study with 56% (n = 6) of major winning athletes winning their first major at a home venue. Nibali, et al, (2011) showed that athletes competing at home courses improved their performance time by 0.3 - 0.8%. There may, therefore, be utility in comparing race performance funnels for home and away performances to determine whether athletes are adversely affected (choke) or improve their performance due to home advantage factors.

5.4.2.2 Identifying potential winners

Athlete's time series of their exponentially smoothed race points were examined to predict future performance over time in terms of the potential for winning future major races. The basis for this being whether the most recent part of the time series was improving, had plateaued or the performances were deteriorating. This observation was supplemented by examination against the performance funnels for winners and winless athletes. On this basis athletes were classified as either probable to win a major, possible or unlikely. The extent to which these classifications were accurate was not tested due to the time frame in which the classification was made and the availability of future data on which to test the predictions. This should, however, be done in the future to enable more accurate modelling procedures to be adopted. For example, with more data and more athlete tracking, the extent to which previous performance levels, recent trends in the data and other factors such as age of the athlete determine future winners could be explored. Larger data sets may indicate the appropriateness of when to assess an athlete's performance path as different time points are likely to vary in the ability for prediction. A signal detection approach (see Study 2) could help ascertain this as well as to determine which aspects of a performance profile best reflects future winners. For example, does a consistently high

level of performance (plateau) signal the likelihood of a major win better or worse than a rapidly improving profile.

This study was limited by a relatively small sample meaning that a "surprise" major championship winner (deemed an outlier in this study) may reflect a small but consistently occurring phenomenon. The limited data meant that non-statistical methods were employed for predictions i.e. trends were graphically presented (winners and winless performance funnels) against which individual athlete's time series of exponentially smoothed race points were compared and patterns visible to the human eye discerned (Montgomery et al, 2015). Larger data sets would enable statistical methods such as control charts, logistic regression (win or no win) or cluster analysis (heterogeneity of variables between winning and losing athletes) to aid discerning the important variables that help predict future major winners.

Some athletes exhibited different profiles season to season and year to year suggesting that there may have been a seasonal component to their performance. Future studies may therefore be improved by adopting the Holt-Winters method, also known as double exponential smoothing (Winters, 1960). This was not the case in this study although this may have been a consequence of the small sample size.

The use of performance funnels based on major winners and winless athletes provided some context against which performances could be assessed. However the availability of data (race points were first awarded at races from 2006 with the purpose of creating World rankings) meant that individual athletes were compared against their peers, including themselves, in this study. Future studies will be able to compare athletes against past performances to assess the extent to which previous performances reflect current performances. However, in this type of future performance prediction (e.g. Neville James)

it is clear that changes in performance over time due to maturation i.e. equipment, sports science support etc. render any prediction liable to error. Acknowledgement of this error in terms of a probability estimate for any prediction is therefore preferable. Similarly, the recognition of context behind a performance outcome is important. Hence in canoe slalom a race where a gate had been hit and a two second penalty applied could be accounted for within a prediction. At a practical level coaches develop race strategies with their athletes on the basis that a risky manoeuvre may be required to win a race final but hitting a gate would result in no medal at all. In sports determined by fine margins this is inevitable but modelling performance without consideration of the finer points of race strategy introduces more error into the prediction. Future studies should therefore consider the use of finer grained performance variables such as upstream and downstream moves, penalties and course design factors into models used to determine future performances.

5.5 Conclusions

A method for tracking individual race performance in canoe slalom using exponential smoothing algorithms was successfully applied. This method provided the analyst with a more sensitive measure of performance variation compared to World ranking points, which were based on the best 5 performances over a two year period, or race points which varied too much and did not facilitate the discerning of a trend. Performance funnels were created using median and 95% confidence intervals for previous major winners and winless athletes who had reached the top 40 in the World rankings. These provided a visual representation of performance over time that could be used to compare individual athletes against. Trends (improvement, plateau and decreasing performance) were evident in reasonably consistent ways that suggested these could be used to support talent

identification and athlete selection. However strategic planning based on these performance funnels would not be without risk and therefore until more robust methods are developed the performance funnels should be used as a guide to future performance along with other, perhaps more multi-disciplinary team and individual information.

Chapter Six: General Discussion

6.1 Reflections on the thesis

These studies came about as a consequence of the researcher becoming the first analyst to work with British Canoeing as an employee of the English Institute of sport. At the outset there was a desire from both organisations to develop PA with an applied emphasis from the sport but a more mixed approach from the employer who encouraged and supported the PhD study. Study one reflected these goals with a needs analysis and system development the starting point for this new collaboration. Within the sport the coaches embraced this direction and along with the head of science and the senior elite athletes the needs of the sport were identified. This was a crucial process as PA was little understood within the sport and the potential direction unknown. The timeline of this thesis reflects the development of PA which has since increased the number of support staff, increased the PA support and adopted a multi-disciplinary approach with far more sophisticated equipment routinely used to capture data. These developments can be partly attributed to the success of the pioneering work presented in this thesis. Without buy-in from the coaches and athletes these developments and the subsequent growth of PA in this sport would not have taken place. It was the acknowledgement of the importance of the coach-analyst relationship that shaped the evolution of the PhD with the implicit understanding that all processes had to have worth in the minds of the coaches. Balancing this with the need for academic rigour was always the challenge. Hence this thesis attempted to answer two fundamental questions posed by Groom, Cushion & Nelson (2011), namely 'what of PA' and 'how' to best implement PA within the coaching process.

Study one described the process of developing a reliable system for collecting the performance variables, deemed important by the coaches, in canoe slalom. Coach and

athlete feedback suggested split sections (timed), techniques (moves) and overall race times were important variables and so the method of collection of this data was devised and assessed using reliability tests. Overall race times were published at race events using timing gates but were collected at practice using stop watches. Errors less than 0.2s were deemed acceptable as coaches believed differences less than this value were meaningless. Following extensive practice intra-analyst (the PA) tests found acceptable errors for timing the split sections live and post event (3.8% were over 0.2s). However inter-analyst tests (coaches and PA interns) found 20% of differences were greater than 0.3s. Similarly, the technique analysis for upstream gates found 95.1% (Kappa agreement) for the intra-analyst and 91.6% for the inter-analyst tests. These results supported Hunter et al.'s (2007) contention that it was necessary to use one well-trained analyst to collect this type of data reliably. In practice this was not possible as the one full-time analyst could not collect all of this type of data. Consequently, the analysis of split sections and techniques continued to be developed but within the time frame of the PhD thesis were not sufficiently rigorous to be included. The result of this piece of work supported the development and exploration of using advancements in technology that could overcome this limitation. In 2013 British Canoeing with support from UKSport and EIS invested in a Local Positioning System that enhanced quality of video positions and split timings for coaching analysis. Within British Canoeing investment has been made with timing gates, GPS and local positioning measurement systems (LPM) that can automatically track and time an athlete's performance now available at Lee Valley. Future research studies will no doubt use these systems along with wireless and mobile technology to identify stroke patterns and micro analyse performance.

Given the constraints after Study 1 it was decided to focus on race times for Study 2. The coaches routinely used percentages off the K1M winner, both in training and

competitions, on the basis that between and within course difficulties were not uniform. Some debate existed between the coaches on the merits of using percentages off the K1M (assumed close to the best possible run time) or the class winner and this formed the basis of Study 2. This type of analysis was problematical in that the variability in race times, and hence percentages, were due to course, class and athlete differences. Course and class variation were therefore sources of error when attempting to establish athlete performance levels. Indeed, Nibali et al. (2011) identified variability as a key factor when analysing and interpreting canoe slalom performances. The 2019-12 probability charts (K1M and class winner) were tested for their predictive ability using the 2013-16 Olympic cycle race results. Signal detection theory was adopted to determine the number of misses i.e. when the prediction stated that an achieved race percentage 'would not medal' but in actuality a medal was achieved, and the number of false alarms where the prediction 'will medal' was erroneous and the race percentage 'did not medal'. This approach was used to find an appropriate balance between the risk of misses and false alarms (they are inversely proportional). The results unanimously supported the use of race percentages off the class winner, rather than the K1M, with a 50% level of probability of medalling resulting in a low rate of false alarms (maximum 6%) and a high hit rate (over 70% of medals correctly identified). Signal detection theory can be adapted to other sports performance questions that require using historical data to inform decision making processes. The power of probability over guaranteed outcomes is a realistic method to adopt as sport in many ways is unpredictable. Whilst the creation of race percentage targets was based on an Olympic cycle's data, further analysis could determine whether this was the most appropriate. For example, new race percentage targets were presented, but not fully tested, suggesting that small changes are likely if targets are set using the previous i.e. rolling four year's data, as opposed to using a fixed four-year period such as a 4-year Olympic cycle. Since the

evaluation of race percentages, the coaches have adopted the use of class percentages as a key measure of performance and are exploring the use of virtual best times based on a standardised 'world's best' criteria (e.g. top *x* boats in final and semi-final).

Study 2 was based on a relatively small number of competitions, but this was to some extent inevitable, and no account was taken regarding the quality of the field at each race. Race percentages could be further explored in relation to their application to training and in conjunction with virtual best times (best split times for all athletes combined).

On the basis of the main limitation of Study 2, namely the lack of eligible competitions, and the need to monitor an athlete's performance over time, the ICF ranking system which used a quality factoring algorithm per competition was considered a logical source of data for the Study3. The need to track an athlete over time was driven by UK Sport who required evidence based performance records and targets to determine funding across the Olympic sports. Study 3 therefore attempted to derive a methodology for tracking an athlete over time, with some contextual basis to allow comparisons and predictions to be made. The ICF race points were considered the best performance measure as the quality of the athletes competing at an event were factored into them. This enabled more races to be analysed and overcame some of the difficulties previously encountered in Study 2. At the outset it was recognised that individual athletes tended to perform better than expected, as expected or relatively poorly at races, probably due to the small differences in performance that determined race outcomes. This meant that simply producing a time series of race points gained did not portray trends very easily. The ICF also produced a world ranking based on the best 5 race points gained over a two-year period. This essentially was a smoothed version of race points, although this measure was not sensitive to individual race performances. To rectify this situation, a smoothing algorithm approach was adopted in Study 3 with the aim of providing a sensitive but

145

understandable measure of performance over time. The best technique was found to be an exponential smoothing algorithm with a constant smoothing coefficient of 0.4. To make this time series more understandable a funnel plot using the median and 95% confidence interval of the median s was created using previous major winners and winless athletes who were ranked in the World's top 40. All time series plots were synchronised at the point when an athlete gained their first World ranking i.e. after 5 competitions. Athlete profiles were then examined over time against the winners and winless performance funnels. This novel approach provided indications of an athlete's progress over time where periods of improvement, plateau and decreases in performance were evident to the extent that predictions of future performance could be made. However, these techniques require more refinement and consideration of factors so far not tested e.g. the impact of penalties, choice of techniques, home advantage etc. In study 3 it was concluded that modelling performance without consideration of the finer points of race strategy introduced more error into the prediction. Because of this the coaches and analysts in canoe slalom are now examining finer grained performance variables such as upstream and downstream moves, penalties and course design. Over time as the data increases these factors can be tested in the performance models to determine future successful performances. Also, with more data signal detection techniques could be applied to assess the validity of any predictions. These smoothing methods and funnel plots could also be applied to Study 2's race percentages to compare the suitability in tracking performance over time.

6.2   Conclusion

This research has contributed positively to the sport of canoe slalom and to the discipline of performance analysis. The sport has clearly grown its use and reliance upon performance analysis and has since invested money to develop more in the future. This

thesis has explored canoe slalom at the elite level where results determine funding and funding facilitates development. Fortunately, the period in which this thesis has been undertaken has seen great success for British canoe slalom. Advances in procedures and methodologies have been pioneered in this thesis but the future has great potential for developing these far more due to technological advances and an increase in man hours devoted to PA in this sport. The future of canoe slalom is, however, multi-disciplinary and sports science support will impact areas such as mental toughness and lifestyle, physiological and biomechanical improvements as well as PA and coaching interventions.

Chapter Seven: References

Atkinson, G. & Nevill, A. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med, 26, 217-238.

Arundel, S. & Holmes, L. (2013). Performance analysis of show jumping strategy during British Eventing. In D.M. Peters & P. O'Donoghue (Ed), Performance Analysis of Sport IX (pp183-188). Oxon: Routledge.

Bland, J.M. & Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet, 1, 307-310.

Bland, J.M. &Altman, D.G. (1999). Measuring agreement in method comparison studies. Statistical Methods in Medical Research, 8, 135-160.

Bonetti, D.L., & Hopkins, W.G. (2010). Variation in performance times of elite flat-water canoeists from race to race. International Journal of Sports Physiology and Performance, 5(2), 210-217.

Braghin, F., Cheli, F., Melzi, S. & Sabbioni, E. (2008). Race driver model. Computers and structures, 86, (13-14), 1503-1516.

Bullock, N., Hopkins, W.G., Martin, D.T. & Marino, F.E., (2009). Characteristics of performance in skeleton World Cup races. Journal of Sports Sciences, 27 (4), 367–372.

Butterworth, A, O'Donoghue, P. & Cropley, B. (2013). Performance profiling in sports coaching: a review. International Journal of Performance Analysis in Sport, 13, 572-593.
Carling, C., Reilly, T. & Williams, M. (2009). Performance Assessment for Field Sports. Routledge: Oxon.

Choi, H., O'Donoghue, P.G. & Hughes, M. (2006). A Study of team performance indicators by separated time scale using a real-time analysis technique within English national basketball league. In H. Dancs, M. Hughes & P. O'Donoghue, (Ed) Performance Analysis of Sport 7 (pp. 138-141) Cardiff: CPA Press, UWIC.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.

Cooper, S., Hughes, M., O'Donoghue, P., & Nevill, A.M. (2007). A simple statistical method for assessing the reliability of data entered into sport performance analysis systems. International Journal of Performance Analysis in Sport, 7(1), 87-109.

D'Angelo, R. (2013). Analysis of competition results, methods of analysis, evolution of the analysis tools and limitations and elements that may condition the analysis. 2013 ICF Canoe Slalom Coaches Conference, Ivrea, Italy, 28th November – 1st December.

De Smith, M. J. (2015). STATSREF: Statistical Analysis Handbook - a web-based statistics resource. The Winchelsea Press, Winchelsea, UK.

Gelper, S., Fried, R. & Croux, C. (2010). Robust forecasting with exponential and Holt-Winters smoothing, Journal of Forecasting, 29, 285-300.

Green, C. (2012). Performance analysis of canoe slalom: performance indicators at Cardiff international white water (CIWW). Cardiff Metropolitan University, Degree of Bachelor of Science (honours), Sport and Exercise Science.

Groom, R., Cushion, C. J., & Nelson, L. J. (2011). The delivery of video-based performance analysis by England youth soccer coaches: Towards a grounded theory, Journal of Applied Sport Psychology, 23, 16-32.

Heeger, D.J. & Landy, M.S. (2010). Signal detection theory and procedures. In B. Goldstein (Ed.), Encyclopedia of Perception (pp. 888-892). Thousand Oaks, CA: Sage Publications, Inc.

Holt, C.C. (1957). Forecasting trends and seasonals by exponentially weighted averages, Carnegie Institute of Technology, Pittsburgh ONR memorandum no. 52.

Hopkins, W.G. (2005). Competitive performance of elite track- and-field athletes: variability and smallest worthwhile enhancements. Sportscience, 9, 17–20.

Hughes, M.D. & Franks, I.M. (1994). Dynamic patterns of movement of squash players of different standards in winning and losing rallies. Ergonomics, 37(1), 23-29.

Hughes, M. D., Wells, J. & Matthews, C. (2000). Performance profiles at recreational, county and elite levels of women's squash. Journal of Human Movement Studies, 39, 85-104.

Hughes, M. (2004). Notational analysis – a mathematical perspective. International Journal of Performance Analysis in Sport, 4(2), 97-139.

Hughes, M., Cooper, S. & Nevill, A. (2002). Analysis procedures for non-parametric data from performance analysis. International Journal of Performance Analysis in Sport, 2(1), 6-20.

Hughes, M., Evans, S. & Wells, J. (2001). Establishing normative profiles in performance analysis. International Journal of Performance Analysis of Sport, 1, 4-26.

Hunter, A., Cochrane, J. & Sachlikidis, A. (2007). Canoe slalom - competition analysis reliability. Sports Biomechanics, 6(2), 155-170.

Hunter, A., Cochrane, J. & Sachlikidis, A. (2008). Canoe slalom competition analysis. Sports Biomechanics, 7(1), 24-37

Hunter, A. (2009). Canoe slalom boat trajectory while negotiating an upstream gate. Sports Biomechanics, 8(2), 105-113.

James, N, Mellalieu, S.D. and Hollely, C. (2002). Analysis of strategies in soccer as a function of European and domestic competition. International Journal of Performance Analysis in Sport, 2(1), 85-103.

James, N., Mellalieu, S.D. & Jones, N.M.P. (2005). The development of position-specific performance indicators in professional rugby union. Journal of Sports Sciences, 23(1), 63-72.

James, N., Taylor J. & Stanley, S. (2007). Reliability procedures for categorical data in Performance Analysis. International Journal of Performance Analysis in Sport, 7(1), 1-11.

James, N. (2012). Predicting performance over time using a case study in real tennis. Journal of Human Sport & Exercise, 7(2), 421-433.

Kirby, R. (2009). Development of a real-time performance measurement and feedback system for alpine skiers. Sports Technology, 2(1-2), 43-52.

Kuper, G.H. & Sterken, E. (2007). Modelling the development of world records in running. In J. Albert & R.H. Koning (Ed). Statistical Thinking in Sports (pp. 7-32), Chapman and Hall/CRC.

Lane, D.M. & Osherson, D. (Online Statistics Education: A Multimedia Course of Study (http://onlinestatbook.com/), Rice University.

Lyons, K. (2005). Performance Analysis in Applied Contexts. International Journal of Performance Analysis in Sport. 5(3), 155-162.

Messias, L.H.D, Masselli dos Reis, I.G., Ferrari, H.G. & Manchado-Gobatto, F. (2014) Physiological, psychological and biomechanical parameters applied in canoe slalom training: a review. International Journal of Performance Analysis in Sport. 14, 24-41.

Montgomery, D.C., Jennings, C.L. & Kulahci, M. (2015). Introduction to Time Series Analysis and Forecasting. John Wiley & Sons.

Nevin, J.A. (1969). Signal detection theory and operant behavior - A Review of David M. Green and John A. Swets' Signal Detection Theory and Psychophysics. Journal of the experimental analysis of behavior, 12(3), 475-480.

Nevill, A.M., Lane, A.M., Kilgour, L.J., Bowes, N. & Whyte, G.P. (2001). Stability of psychometric questionnaires. Journal of Sports Sciences, 19, 273 – 278.

Nevill, A., & Whyte, G. (2005). Are there limits to running world records? Medicine & Science in Sports & Exercise, 37(10), 1785-1788.

Nibali, M., Hopkins, W., & Drinkwater, E. (2011). Variability and predictability of elite competitive slalom canoe-kayak performance. European Journal of Sport Science, 11(2), 125-130.

O'Donoghue, P.G., Dubitzky, W., Lopes, P., Berrar, D., Lagan, K., Hassan, D., Bairner, A. & Darby, P. (2003). An evaluation of quantitative and qualitative methods of predicting the 2002 FIFA World Cup. Proceedings of the World Congress of Science and Football V, Lisbon, Portugal, 44-45.

O'Donoghue, P.G. & Longville, J. (2004). Reliability testing and the use of Statistics in performance analysis support: a case study from an international netball tournament, In P.G. O'Donoghue, & M.D. Hughes (Ed.), Performance Analysis of Sport 6 (pp.1-7). Cardiff, CPA Press, UWIC.

O'Donoghue, P.G. & Williams, J. (2004). An evaluation of human and computer-based predictions of the 2003 rugby union world cup. International Journal of Computer Science in Sport, 3(1), 5-22.

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

O'Donoghue, P.G. (2007). Reliability issues in performance analysis. International Journal of Performance Analysis in Sport, 7(1), 35-48.

O'Donoghue, P.G., Mayes, A., Edwards, K. & Garland, J. (2008). Performance norms for British national super league netball. International Journal of Sports Science and Coaching, 3(4), 501-511.

O'Donoghue, P. & Cullinane, A. (2011). A regression-based approach to interpreting sports performance. International Journal of Performance Analysis in Sport, 11, 295-307.

O'Donoghue, P., Ball, D., Eustace, J., McFarlan, B. & Nisotaki, M. (2016). Predictive models of the 2015 Rugby World Cup: accuracy and application. International Journal of Computer Science in Sport, 15(1), 37-58.

Paton, C.D. & Hopkins, W.G. (2006). Variation in performance of elite cyclists from race to race. European Journal of Sport Science, 6(1), 25– 31.

Potter, G. & Hughes, M.D. (2001). Modelling in competitive sports. In M.D. Hughes (Ed), Notational Analysis of Sport III (pp. 58-74). Cardiff: UWIC.

Pyne D, Trewin C.B. & Hopkins W.G. (2004). Progression and variability of competitive performance of Olympic swimmers. Journal of Sports Sciences, 22, 613–620.

Rakow, T., Wright, R., Spiegelhalter, D.J. & Bull, C. (2015). The pros and cons of funnel plots as an aid to risk communication and patient decision making. British Journal of Psychology, 106, 327-348.

Robertson K (1999). Observation, Analysis and Video. The National Coaching Foundation, Leeds.

Robinson, G. & O'Donoghue, P. (2007). A weighted Kappa statistic for reliability testing in performance analysis of sport. International Journal of Performance Analysis in Sport, 7(1), 12-19.

Schmidt, R. A., & Lee, T. A. (2005). Motor control and learning: A behavioural emphasis (4th edn.). Champaign, IL: Human Kinetics.

Smith, T.B. & Hopkins, W.G. (2011). Variability and predictability of finals times of elite rowers. Medicine and Science in Sports and Exercise, 43(11), 2155–2160.

Slot, O. (2017). The Talent Lab: The secret to finding, creating and sustaining success. EBury Press, UK.

Spiegelhalter, D. (2005). Funnel plots for comparing institutional performance. Statistics in Medicine, 24, 1185-1202.
StatSoft, Inc. (2013). Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook/.

Spencer, M., Losenegard, T., Hallen, J. & Hopkins, W.G. (2014). Variability and Predictability of Performance Times of Elite Cross-Country Skiers. International Journal of Sports Physiology and Performance, 9, 5-11.

Treadwell, P. & Lyons, K. (1997). The use of video in notational analysis. In M. Hughes (Ed), Notational Analysis of Sport. Abingdon: Taylor & Francis.

Wells, J., Robertson, C., Hughes, M. & Howe, D. (2004). Performance profiles of elite men's squash doubles match play. In A. Lees, J.-F. Kahn and I. Maynard (Ed), Science and Racket Sports III (pp. 196 – 201). London, E & FN Spon.

Whipp, B.J. & Ward, S.A (1992). Will women soon outrun men? Nature, 25(1), 355:25

Winters, P.R, (1960). Forecasting sales by exponentially weighted moving averages. Management Science, 6, 324-342.

Appendix 3.1 Manipulating data linked to video within Dartfish tagging module
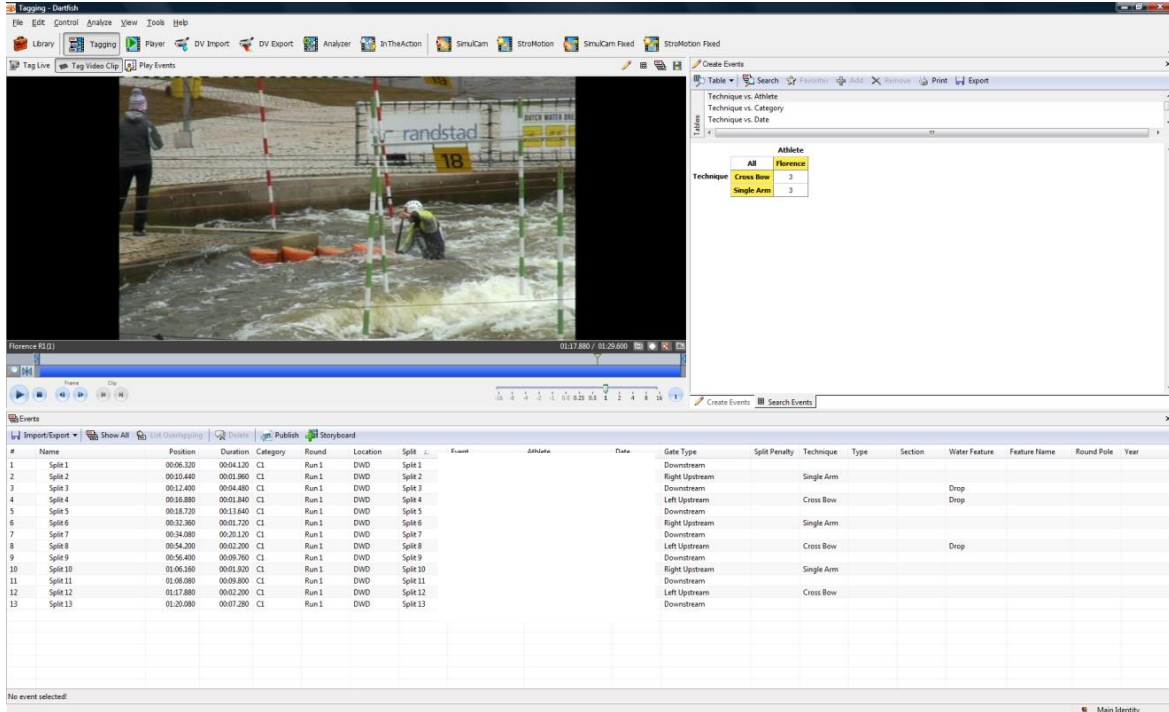
Appendix 3.2 An example of the split times that was exported from Dartfish tagging software to Microsoft Excel for further analysis and visualisation

Athlete	Run	Splits													Overall		
		1	2	3	4	5	6	7	8	9	10	11	12	13	Run time	Pens	Total
Athlete A	Run 1	12.5	2.2	4.2	2.7	17.8	1.8	4.8	2.1	16.3	2.3	13.2	2.5	7.2	89.5		89.5
Athlete A	Run 2	12.3	2.0	2.8	2.9	18.6	1.4	5.1	2.4	17.7	2.4	12.6	2.2	7.1	89.8		89.8
Athlete B	Run 1	12.0	2.1	2.9	2.9	18.7	1.8	5.3	2.2	16.6	2.8	13.1	2.4	7.4	90.3		90.3
Athlete B	Run 2	12.2	2.2	3.1	4.0	17.4	1.6	5.2	2.7	15.6	2.5	12.6	2.8	6.5	88.5		88.5
Athlete C	Run 1	12.4	2.2	2.8	2.3	18.2	1.8	5.3	2.4	15.8	3.1	12.8	2.5	7.0	88.5		88.5
Athlete C	Run 2	13.1	2.0	2.6	2.8	19.5	1.5	5.5	2.2	16.4	2.6	12.8	2.3	7.0	90.4		90.4
Athlete D	Run 1	12.3	1.9	3.4	**2.5**	18.2	1.6	5.0	2.1	16.8	2.4	12.2	1.8	7.2	87.3	2.0	89.3
Athlete D	Run 2	11.9	2.0	3.1	1.8	18.6	1.6	5.4	2.1	16.2	2.3	**11.9**	2.5	6.7	86.0	2.0	88.0
Virtual Best		11.9	1.9	2.6	1.8	17.4	1.4	4.8	2.1	15.6	2.3	11.9	1.8	6.5	81.9		81.9

Key: Green box = fastest split; Bold = penalty

Appendix 4.1 Proportion of race times for K1M 5th place off the K1M winner across Olympic cycles
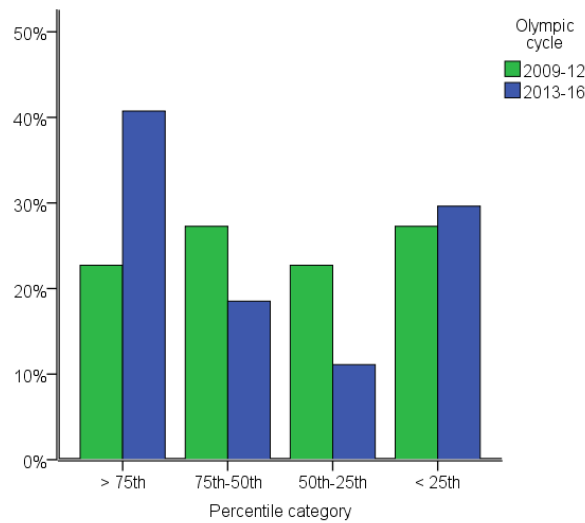
Key: 3 cells had expected counts of less than 5. There were more 5th place K1M percentages in the slowest percentile category and slightly less in the 75th – 50th & 50th – 25th percentile in the 2016 Olympic cycle compared to the 2012 cycle (Chi-square = 2.64, df = 3, p = 0.45, Cramer's V = 0.23).

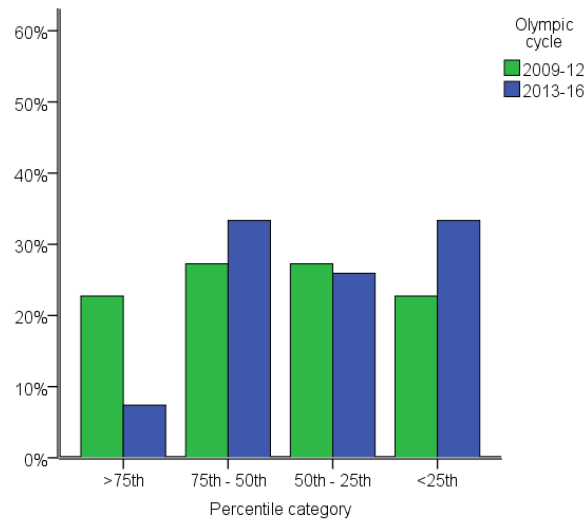Appendix 4.2 Proportion of race times for K1M 10th place off the K1M winner across Olympic cycles
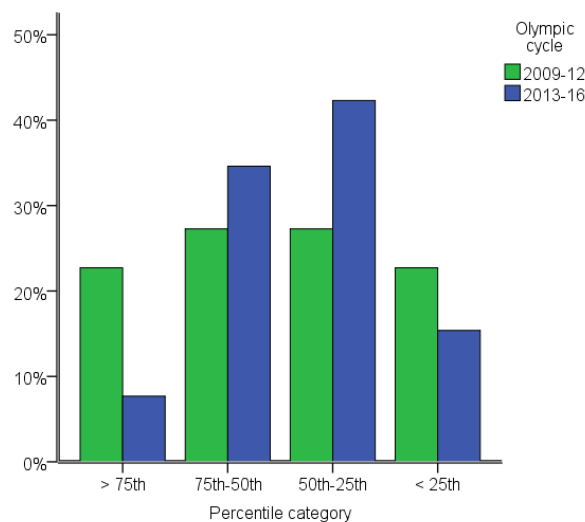
Key: 4 cells had expected counts of less than 5. There were more 10th placed K1M percentages in the 75th & 50th percentile category in the 2016 Olympic cycle compared to the 2012 cycle (Chi-square = 3.12, df = 3, p = 0.37, Cramer's V = 0.26).

154

Appendix 4.3 Proportion of race times for C1M 1st place off the K1M winner across Olympic cycles
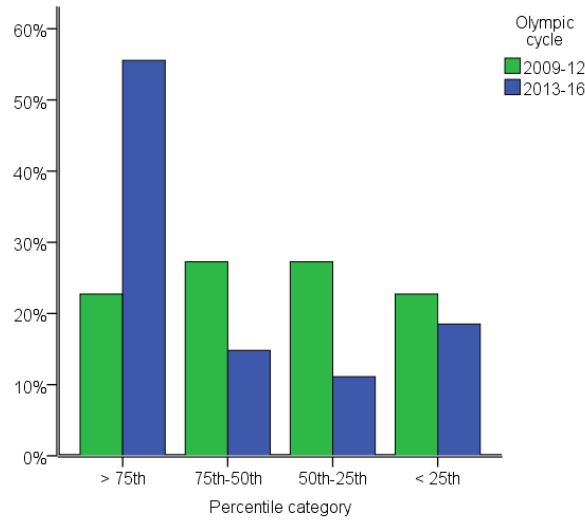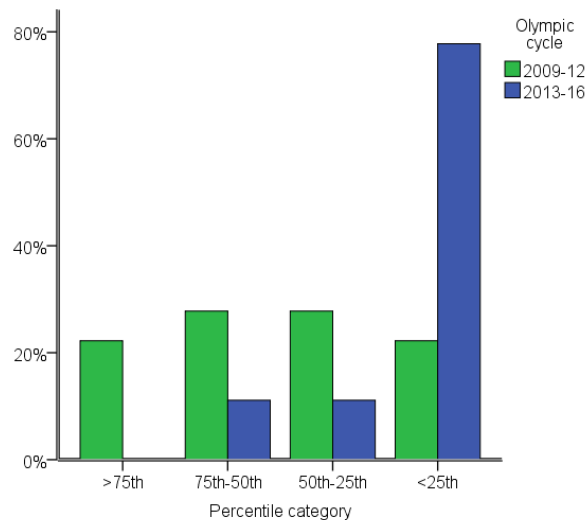
Key: 4 cells had expected counts of less than 5. There were more 1st placed C1M percentages in the slowest (>75th) percentile category in the 2016 Olympic cycle compared to the 2012 cycle (Chi-square = 5.92, df = 3, p = 0.11, Cramer's V = 0.35).

Appendix 4.4 Proportion of race times for C1M 3rd place off the K1M winner across Olympic cycles
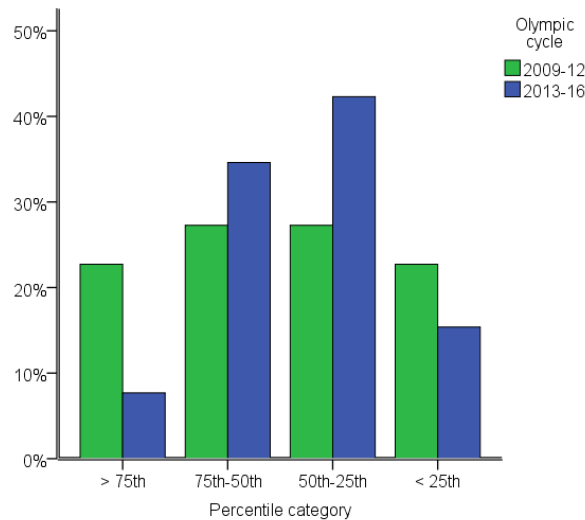
Key: 5 cells had expected counts of less than 5. There were more 3rd placed C1M percentages in the slowest (>75th) percentile category in the 2016 Olympic cycle compared to the 2012 cycle (Chi-square = 5.8, df = 3, p = 0.12, Cramer's V = 0.34.

155

Appendix 4.5 Proportion of race times for C1M 3rd place off the C1M winner across Olympic cycles
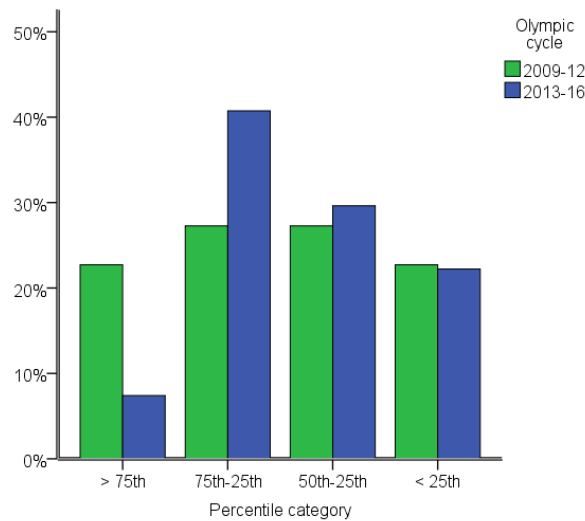
Key: 3 cells had expected counts of less than 5. There were less 3rd placed C1M percentages in the slowest (>75th) percentile category in the 2016 Olympic cycle compared to the 2012 cycle (Chi-square = 2.65, df = 3, p = 0.5, Cramer's V = 0.23).

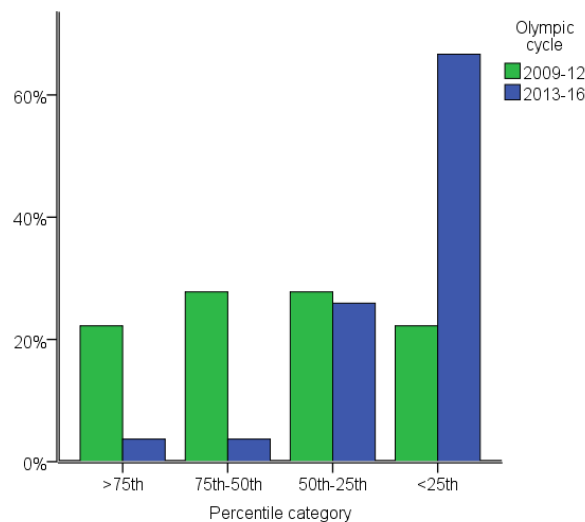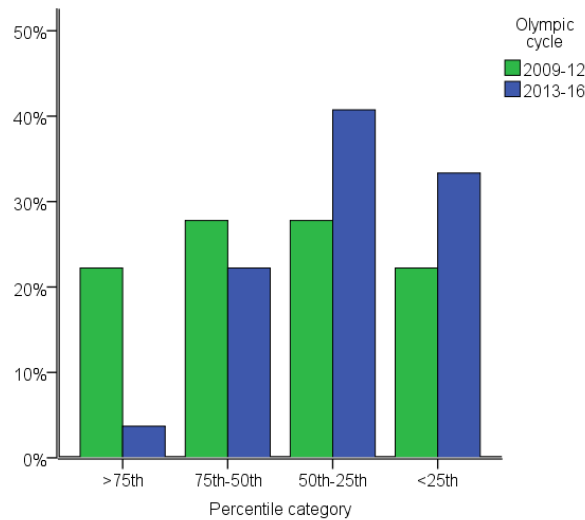Appendix 4.6 Proportion of race times for C1M 5th place off the K1M winner across Olympic cycles.
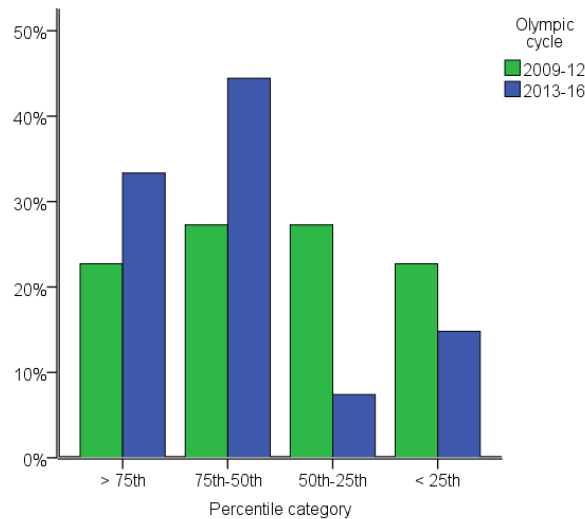
Key: 4 cells had expected counts of less than 5. There were more 5th placed C1M percentages in the slowest percentile category (>75th & 75th) in the 2016 Olympic cycle compared to the 2012 cycle (Chi-square = 4.79, df = 3, p = 0.19, Cramer's V = 0.31).

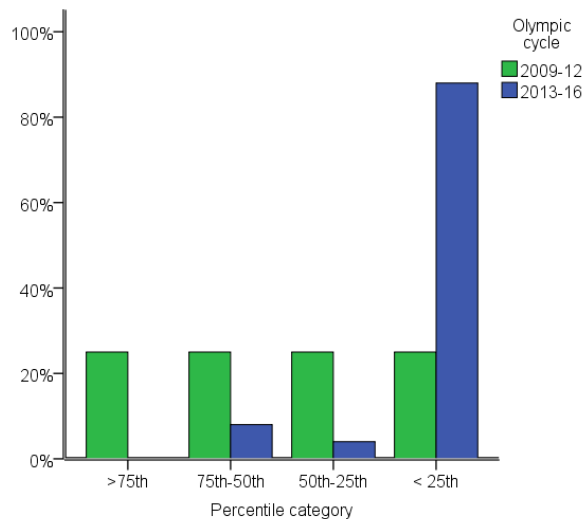Appendix 4.7 Proportion of race times for C1M 5th place off the C1M winner across Olympic cycles.
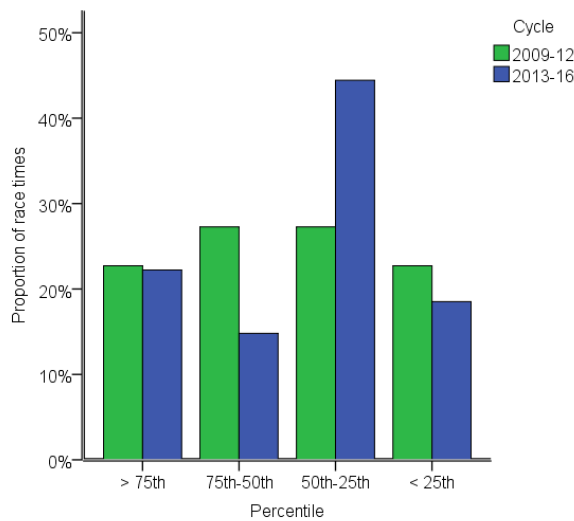
Key: 3 cells had expected counts of less than 5. There were more 5th placed C1M percentages in the 50th percentile category in the 2013-2016 Olympic cycle compared to the 2009-2012 cycle (Chi-square = 2.00, df = 3, p = 0.57, Cramer's V = 0.20).

Appendix 4.8 Proportion of race times for C1M 10th placing semi-final off the K1M semi-final winner across Olympic cycles
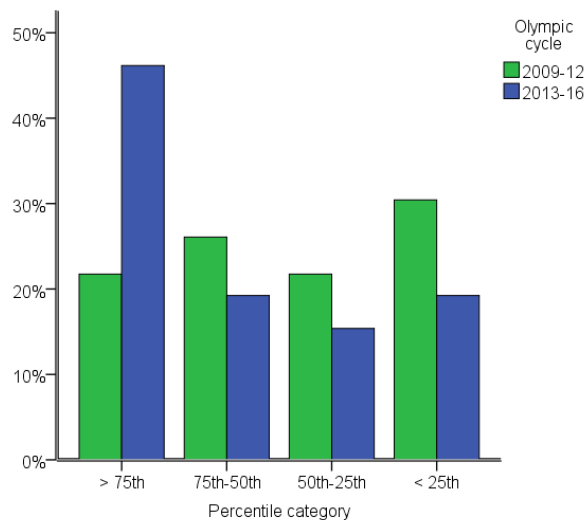
Key: 2 cells had expected counts of less than 5. There were more semi-final 10th placed C1M percentages in the slowest (>75th) percentile category in the 2016 Olympic cycle compared to the 2012 cycle (Chi-square = 3.25, df = 3, p = 0.36, Cramer's V = 0.26.

Appendix 4.9 Proportion of race times for C1M 10th placing semi-final off the C1M semi-
final winner across Olympic cycles
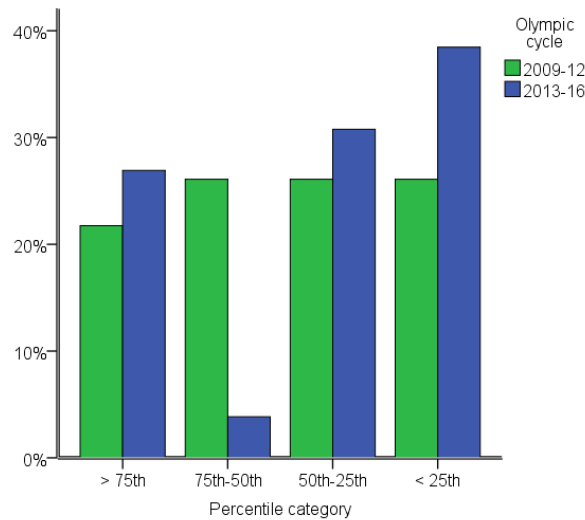
Key: 2 cells had expected counts of less than 5. There were less semi-final 10th placed C1M percentages in the 75th percentile category in the 2016 Olympic cycle compared to the 2012 cycle (Chi-square = 5.03, df = 3, p = 0.17, Cramer's V = 0.32).

Appendix 4.10 Proportion of race times for C2M 1st place off the K1M winner across Olympic cycles
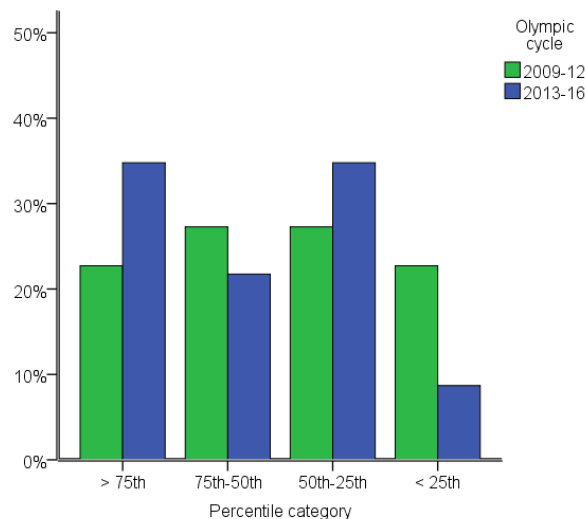
Key: 2 cells had expected counts of less than 5. There were more 1st placed C2M percentages in the slowest (>75th) & 50th percentile category in the 2016 Olympic cycle compared to the 2012 cycle (Chi-square = 2.33, df = 3, p = 0.51, Cramer's V = 0.23).

158

Appendix 4.11 Proportion of race times for C2M 3rd place off the C2M winner across Olympic cycles
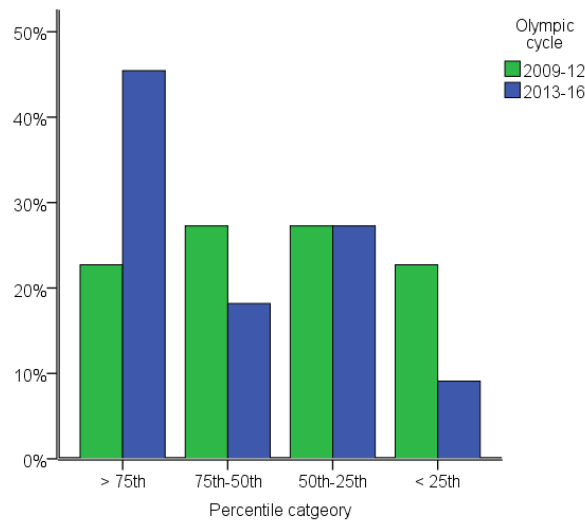
Key: 4 cells had expected counts of less than 5. There were more 3rd placed C2M percentages in the slowest (>75th) percentile category in the 2013-2016 Olympic cycle compared to the 2009-2012 cycle (Chi-square = 1.21, df = 3, p = 0.75, Cramer's V = 0.17).

Appendix 4.12 Proportion of race times for C2M 5th place off the K1M winner across Olympic cycles
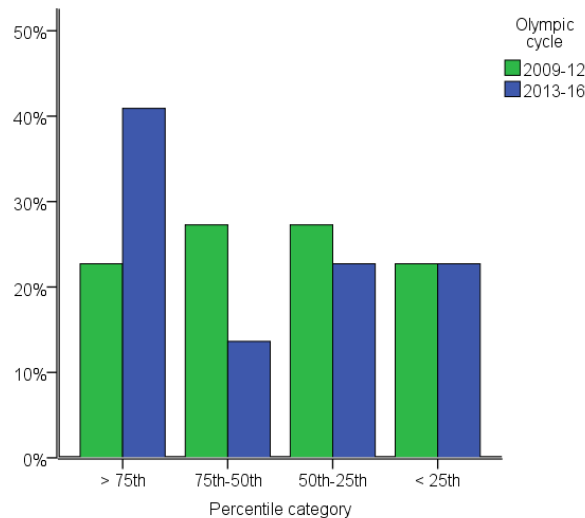
Key: 2 cells had expected counts of less than 5. There were more 5th placed C2M percentages in the slowest (>75th) and less in 75th percentile category in the 2016 Olympic cycle compared to the 2012 cycle (Chi-square = 2.23, df = 3, p = 0.53, Cramer's V = 0.23).

159

Appendix 4.13 Proportion of race times for C2M 5th place off the C2M winner across Olympic cycles
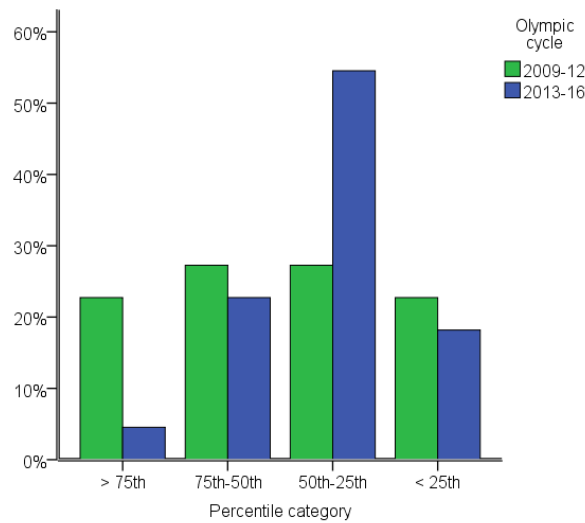
Key: 4 cells had expected counts of less than 5. There were more 5th placed C2M percentages in the 50th percentile category in the 2013-2016 Olympic cycle compared to the 2009-2012 cycle (Chi-square = 4.87, df = 3, p = 0.18, Cramer's V = 0.33).

Appendix 4.14 Proportion of race times for C2M 10th placing semi-final off the K1M semi-final winner across Olympic cycles
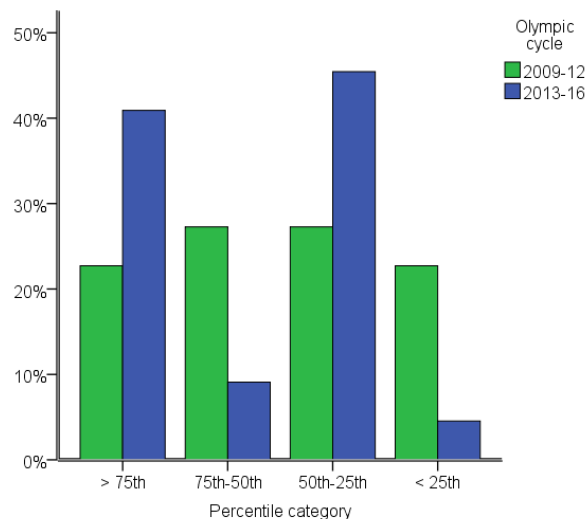
Key: 4 cells had expected counts of less than 5. There were less semi-final 10th placed C2M percentages in the fastest (<25th) and slower (75th) percentile category in the 2013-2016 Olympic cycle compared to the 2009-2012 cycle (Chi-square = 6.81, df = 3, p = 0.08, Cramer's V = 0.39).

Appendix 4.15 Proportion of race times for C2M 10th placing semi-final off the C2M
semi-final winner across Olympic cycles
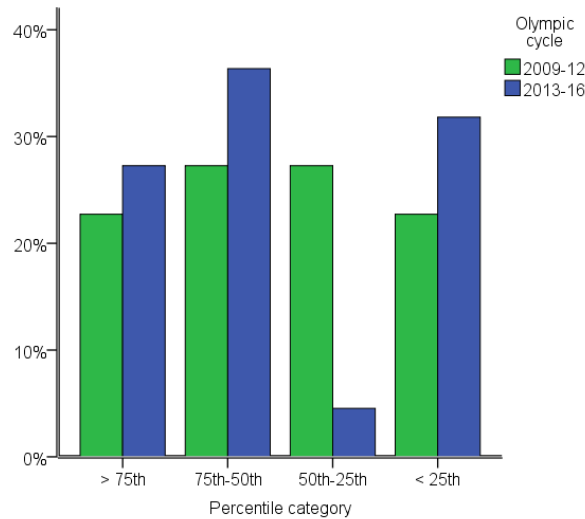
Key: 2 cells had expected counts of less than 5. There were less semi-final 10th placed
      C2M percentages in the 50th percentile category in the 2013-2016 Olympic cycle
      compared to the 2009-2012 cycle (Chi-square = 4.28, df = 3, p = 0.23, Cramer's V
      = 0.31).

Appendix 4.16 Proportion of race times for K1W 1st place off the K1M winner across
Olympic cycles
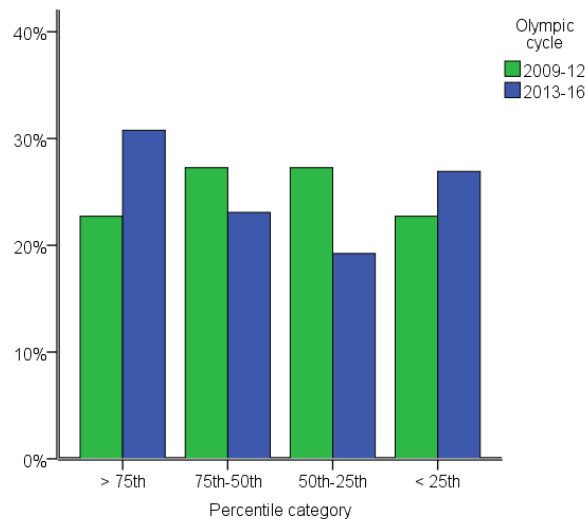
Key: There were small differences in the proportion of 1st place K1W percentages off
      the K1M winner between the 2012 Olympic cycle and 2016 Olympic cycle (Chi-
      square = 0.79, df = 3, p = 0.85, Cramer's V = 0.13).

161

Appendix 4.17 Proportion of race times for K1 women's 3rd place off the K1 men's winner across Olympic cycles
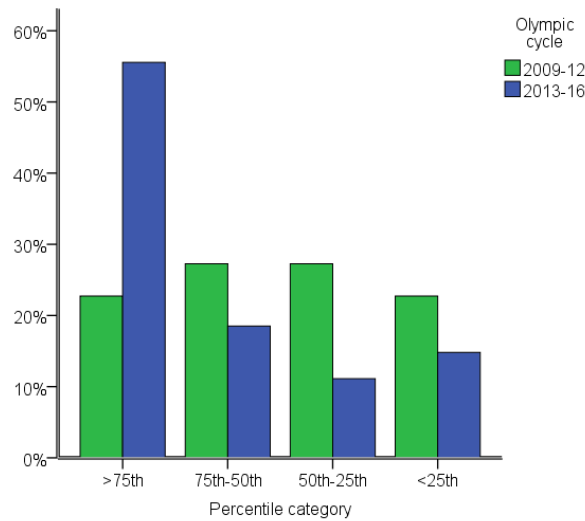
Key: 1 cell had expected counts of less than 5. There were less 3rd placed K1W percentages in the 50th & less in the slowest percentile category in the 2013-2016 Olympic cycle compared to the 2009-2012 cycle (Chi-square = 1.46, df = 3, p = 0.69, Cramer's V = 0.17).

Appendix 4.18 Proportion of race times for K1W 3rd place off the K1W winner across Olympic cycles
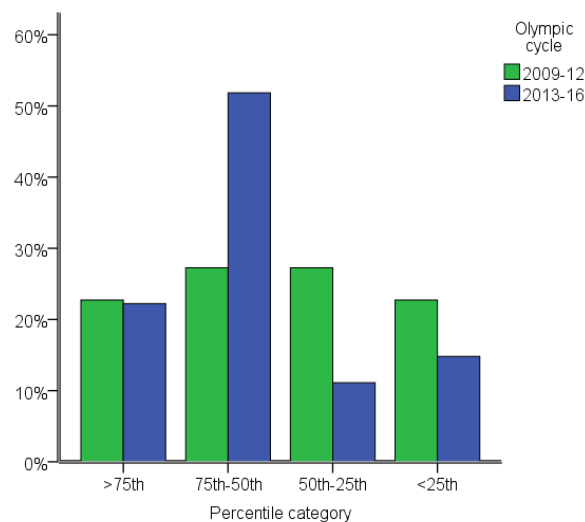
Key: 5 cells had expected counts of less than 5. There were more 3rd placed K1W percentages in the 75th percentile category in the 2013-2016 Olympic cycle compared to the 2009-2012 cycle (Chi-square = 3.93, df = 3, p = 0.27, Cramer's V = 0.28).

Appendix 4.19 Proportion of race times for K1W 5th place off the K1M winner across
Olympic cycles
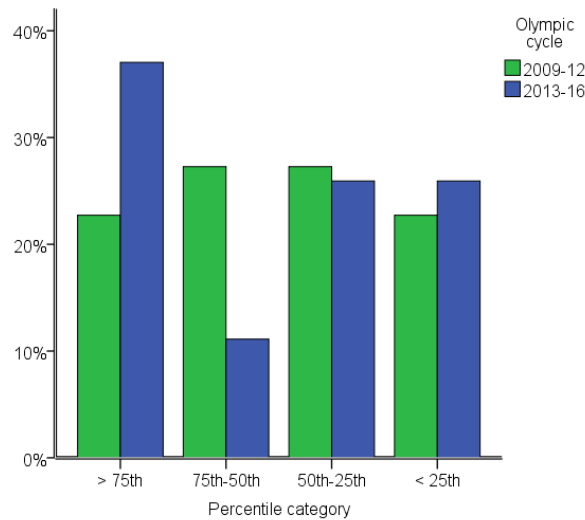
Key: 2 cells had expected counts of less than 5. There were more 5th placed K1W percentages in the slowest (>75th) and less in the 75th percentile category in the 2013-2016 Olympic cycle compared to the 2009-2012 cycle (Chi-square = 2.59, df = 3, p = 0.46, Cramer's V = 0.23).

Appendix 4.20 Proportion of race times for K1W 5th place off the K1W winner across
Olympic cycles
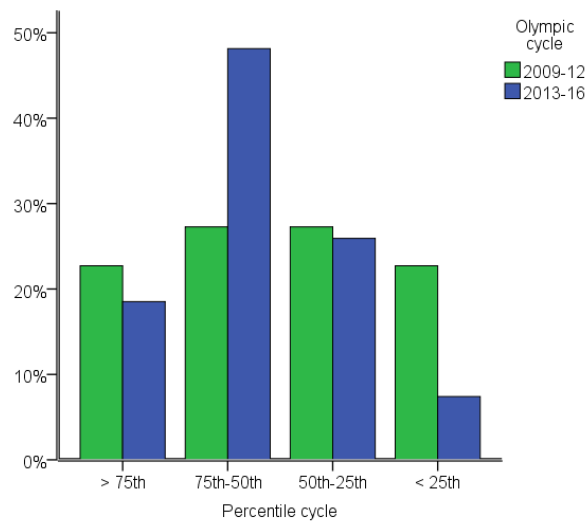
Key: 3 cells had expected counts of less than 5. There were more 5th placed K1W percentages in the 75th percentile category in the 2013-2016 Olympic cycle compared to the 2009-2012 cycle (Chi-square = 3.47, df = 3, p = 0.33, Cramer's V = 0.27).

Appendix 4.21 Proportion of race times for K1W 10th placing semi-final off the K1M
semi-final winner across Olympic cycles

Key: 2 cells had expected counts of less than 5. There were less semi-final 10th placed
K1W in the slowest (>75th) percentile category in the 2013-2016 Olympic cycle
compared to the 2009-2012 cycle (Chi-square = 2.62, df = 3, p = 0.45, Cramer's V
= 0.23).

Appendix 4.22 Proportion of race times for C1W 3rd place off the K1M winner across
Olympic cycles

Key: 6 cells had expected counts of less than 5. There were significantly more 3rd placed
C1W percentages in the fastest (<25th) percentile category in the 2016 Olympic
cycle compared to the 2012 cycle (Chi-square = 15.38, df = 3, p = 0.002, Cramer's
V = 0.59).

Appendix 4.23 Proportion of race times for C1W 3rd place off the C1W winner across
Olympic cycles
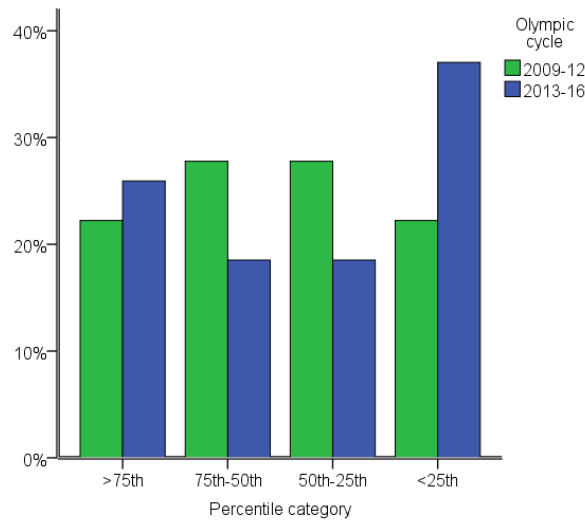
Key: 3 cells had expected counts of less than 5. There were more 3rd placed C1W
       percentages in the fastest (<25th) percentile category in the 2016 Olympic cycle
       compared to the 2012 cycle (Chi-square = 1.66, df = 3, p = 0.65, Cramer's V =
       0.19).

Appendix 4.24 Proportion of race times for C1W 5th place off the K1M winner across
Olympic cycles

Key: 5 cells had expected counts of less than 5. There were more 5th placed C1W
       percentages in the fastest (<25th) percentile category in the 2016 Olympic cycle
       compared to the 2012 cycle (Chi-square = 12.41, df = 3, p = 0.01, Cramer's V =
       0.53).

165

Appendix 4.25 Proportion of race times for C1W 5th place off the C1W winner across Olympic cycles

Key: 3 cells had expected counts of less than 5. There were less 5th placed C1W percentages in the slowest (>75th) percentile category in the 2016 Olympic cycle compared to the 2012 cycle (Chi-square = 4.44, df = 3, p = 0.22, Cramer's V = 0.31).

Appendix 4.26 Proportion of race times for C1W 10th placing semi-final off the K1M semi-final winner across Olympic cycles

Key: 6 cells had expected counts of less than 5. There were significantly more 10th placed C1W percentages in the fastest (<25th) percentile category in the 2016 Olympic cycle compared to the 2012 cycle (Chi-square = 18.51, df = 3, p = 0.00, Cramer's V = 0.66).

166

Appendix 4.27 Proportion of race times for C1W 10th placing semi-final off the C1W semi-final winner across Olympic cycles
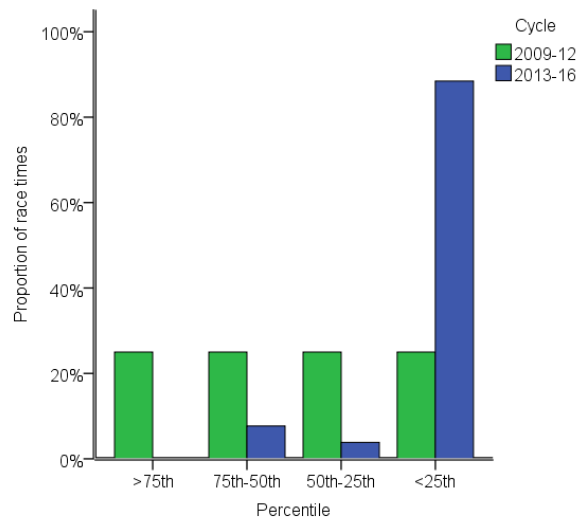
Key: 6 cells had expected counts of less than 5. There were significantly more 10th placed C1W percentages in the fastest (<25th) percentile category in the 2016 Olympic cycle compared to the 2012 cycle (Chi-square = 17.81, df = 3, p = 0.00, Cramer's V = 0.66).

Appendix 4.28 The probability of C1M making a medal in the 2013-16 Olympic cycle using percentages off the 2009-12 Olympic K1M winner
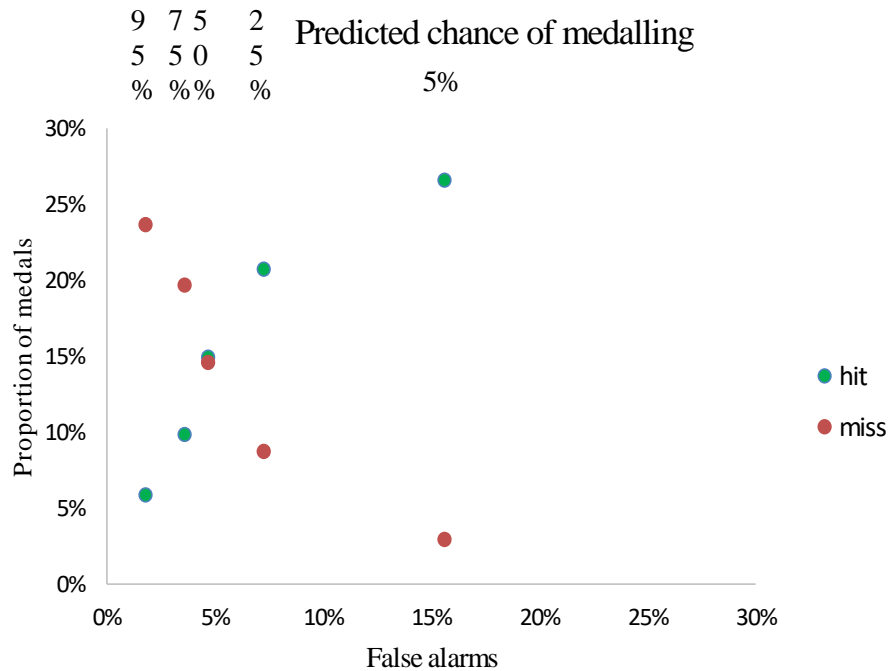
Appendix 4.29 The probability of C1M making a medal in the 2013-16 Olympic cycle
using percentages off the 2009-12 Olympic cycle class winner
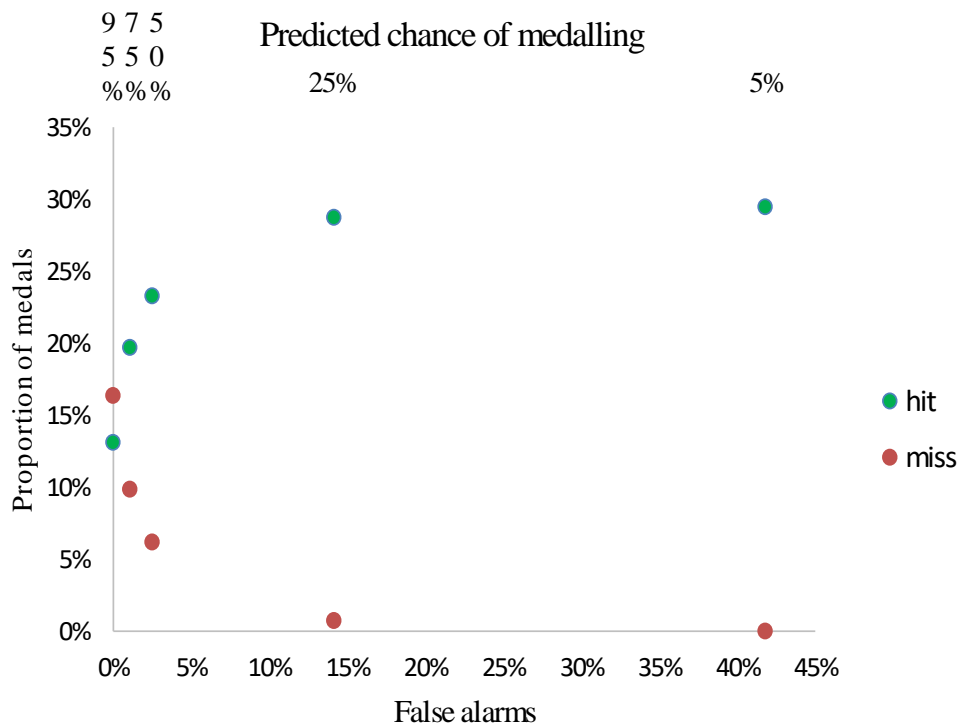
Appendix 4.30 The probability of C2M making a medal in the 2013-16 Olympic cycle
using percentages off the 2009-12 Olympic K1M winner
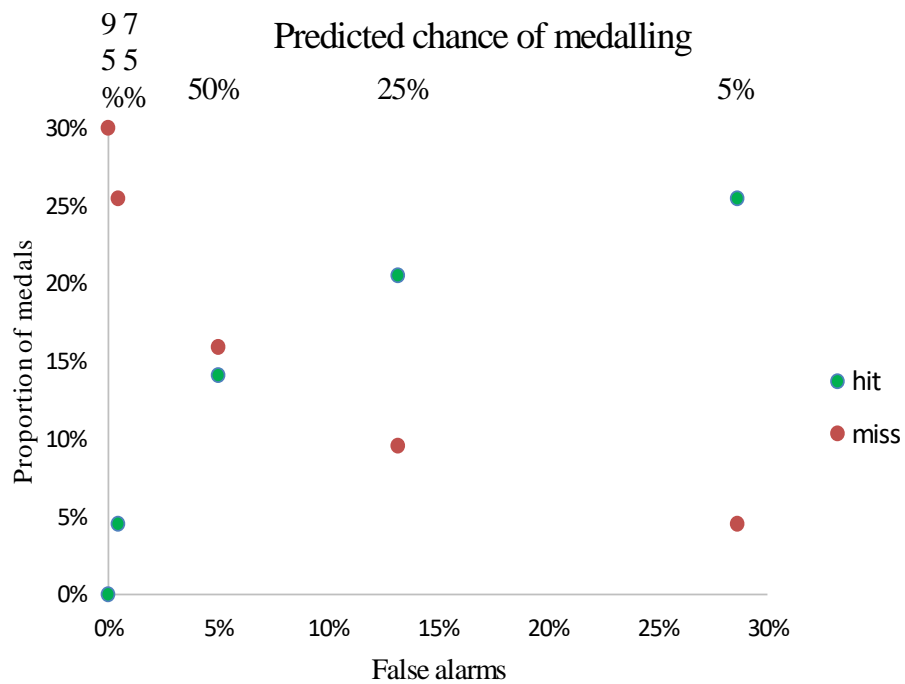
Appendix 4.31 The probability of C2M making a medal in the 2013-16 Olympic cycle using percentages off the 2009-12 Olympic class winner
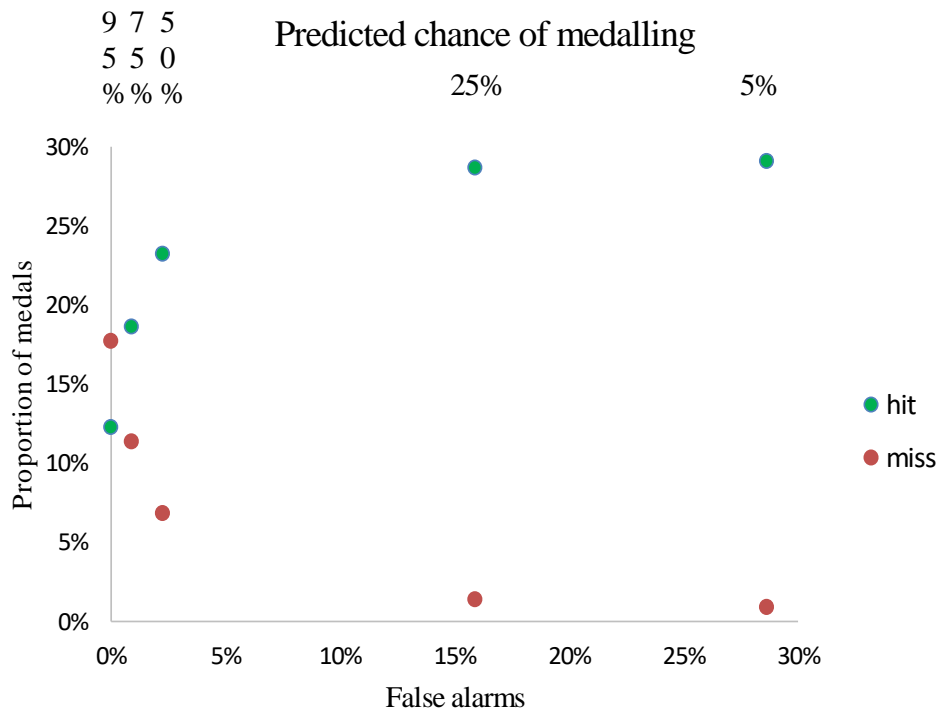
Appendix 4.32 The probability of K1W making a medal in the 2013-16 Olympic cycle using percentages off the 2009-12 Olympic K1M winner
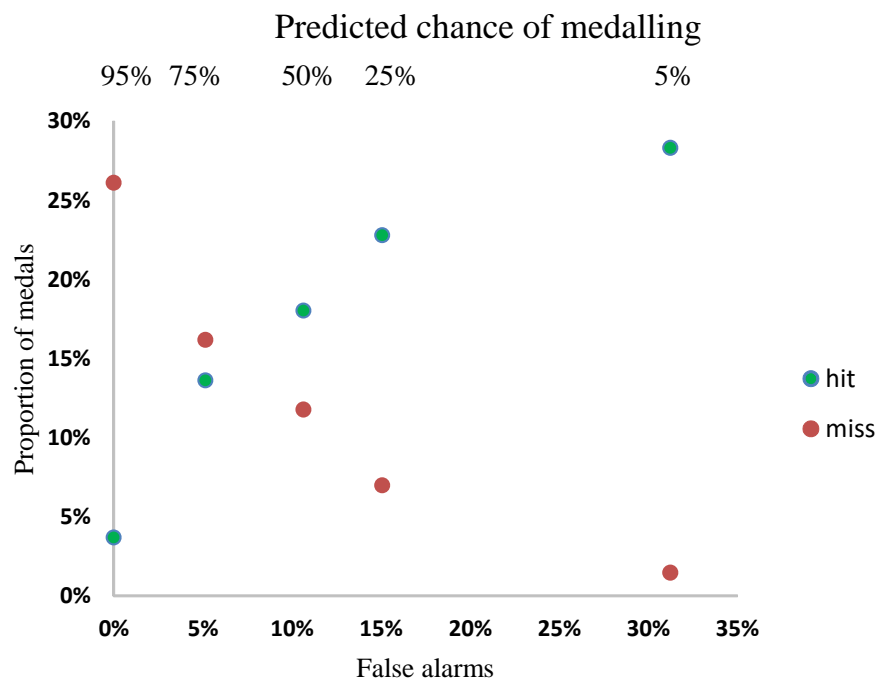
Appendix 4.33 The probability of K1W making a medal in the 2013-16 Olympic cycle
using percentages off the 2009-12 Olympic class winner

Predicted chance of medalling
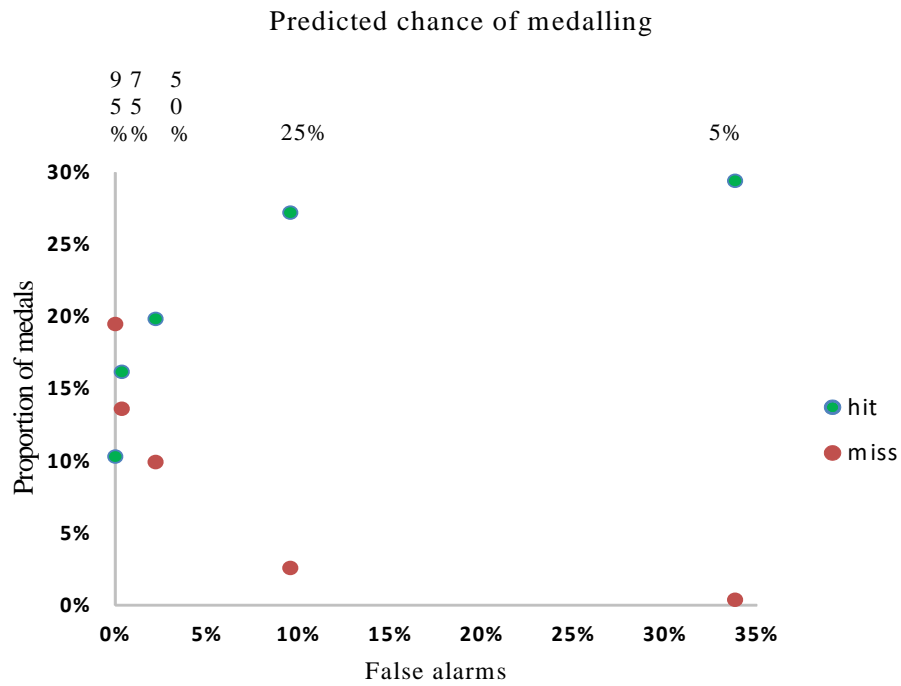
Appendix 4.34 The probability of C1W making a medal in the 2013-16 Olympic cycle
using percentages off the 2009-12 Olympic K1M winner

Predicted chance of medalling
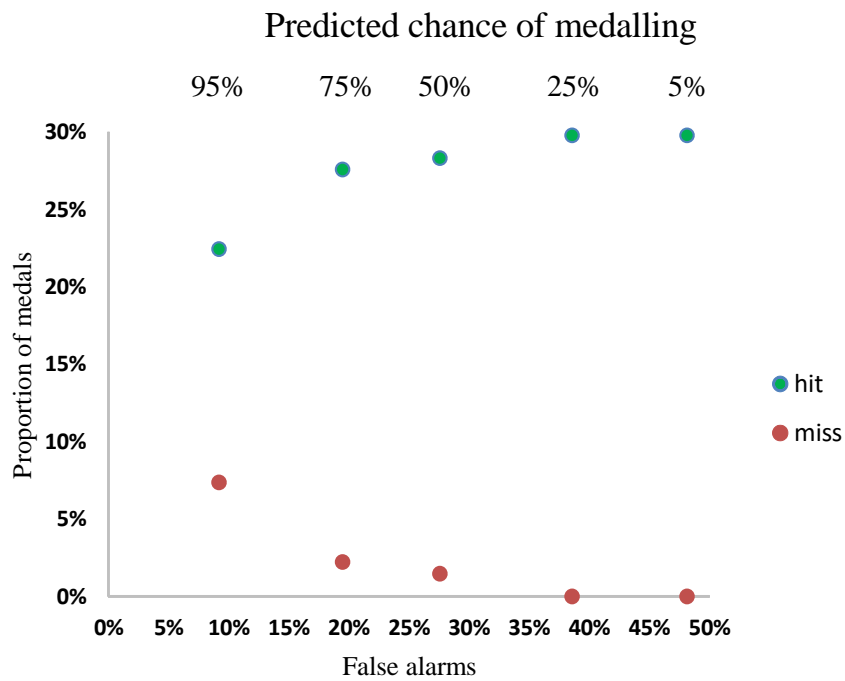
Appendix 4.35 The probability of C1W making a medal in the 2013-16 Olympic cycle using percentages off the 2009-12 Olympic class winner

Predicted chance of medalling
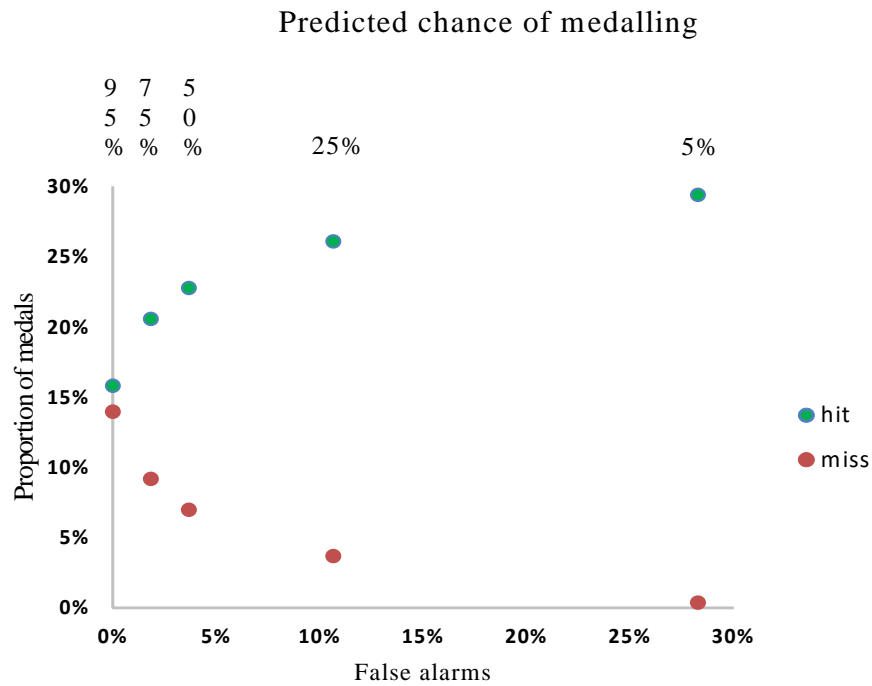
Appendix 4.36 Relative 3rd place performance off K1M winner in 4 year cycles for C1M
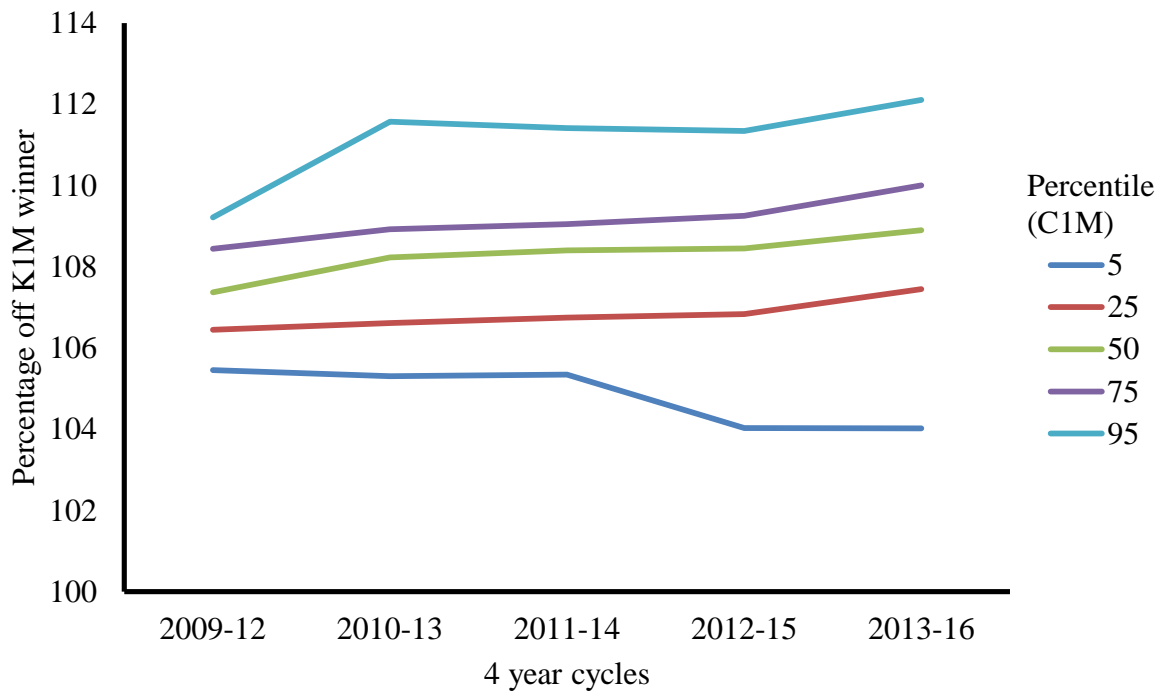
Appendix 4.37 Relative 3rd place performance off K1M winner in 4 year cycles for C2M
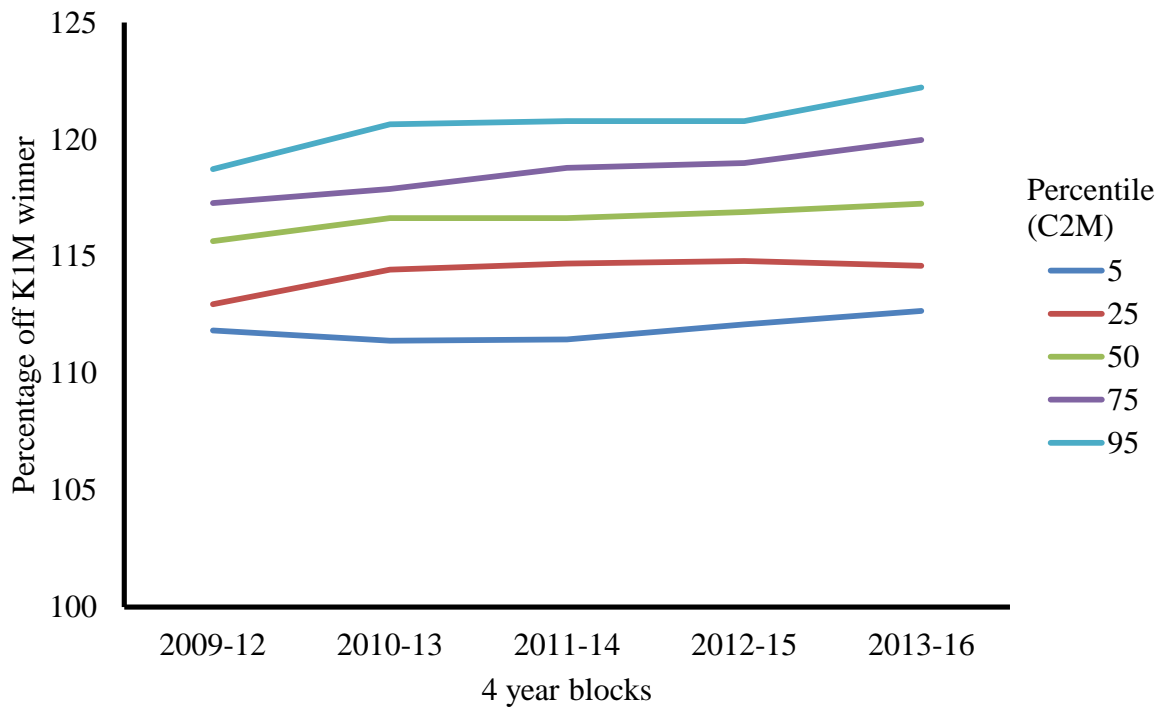
Appendix 4.38 Relative 3rd place performance off K1M winner in 4 year cycles for K1W
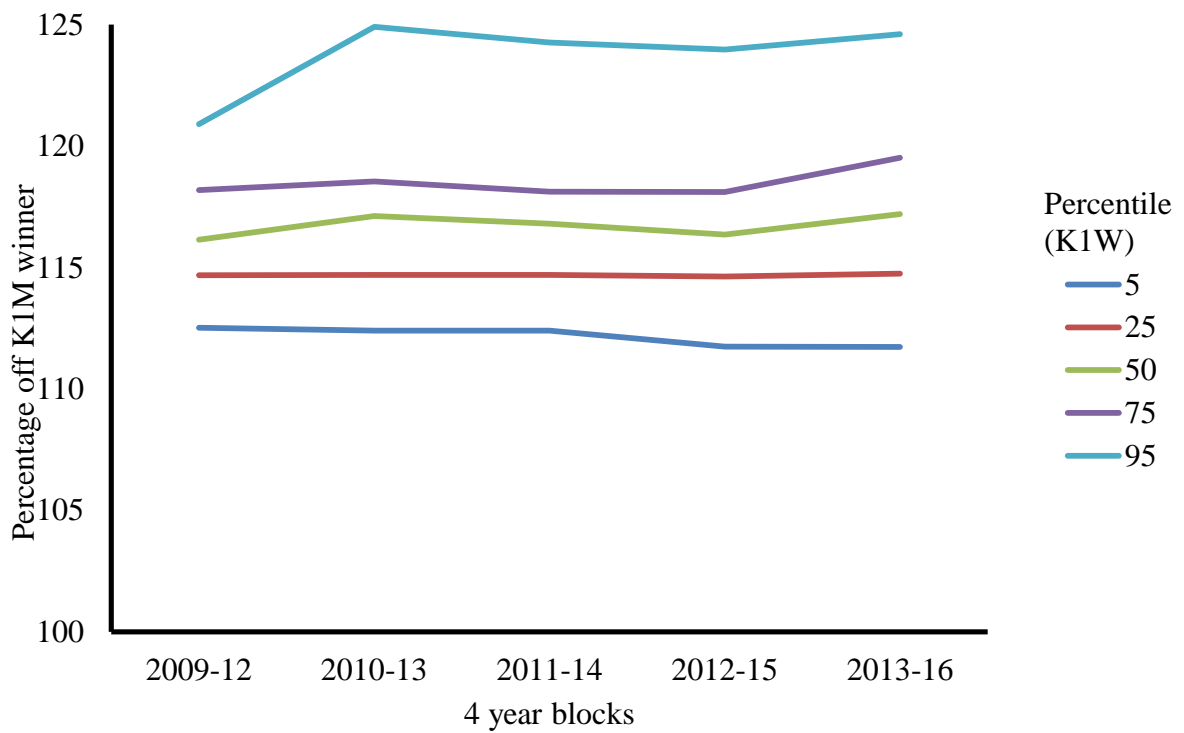
Appendix 4.39 Relative 3ʳᵈ place performance off K1M winner in 4 year cycles for C1W
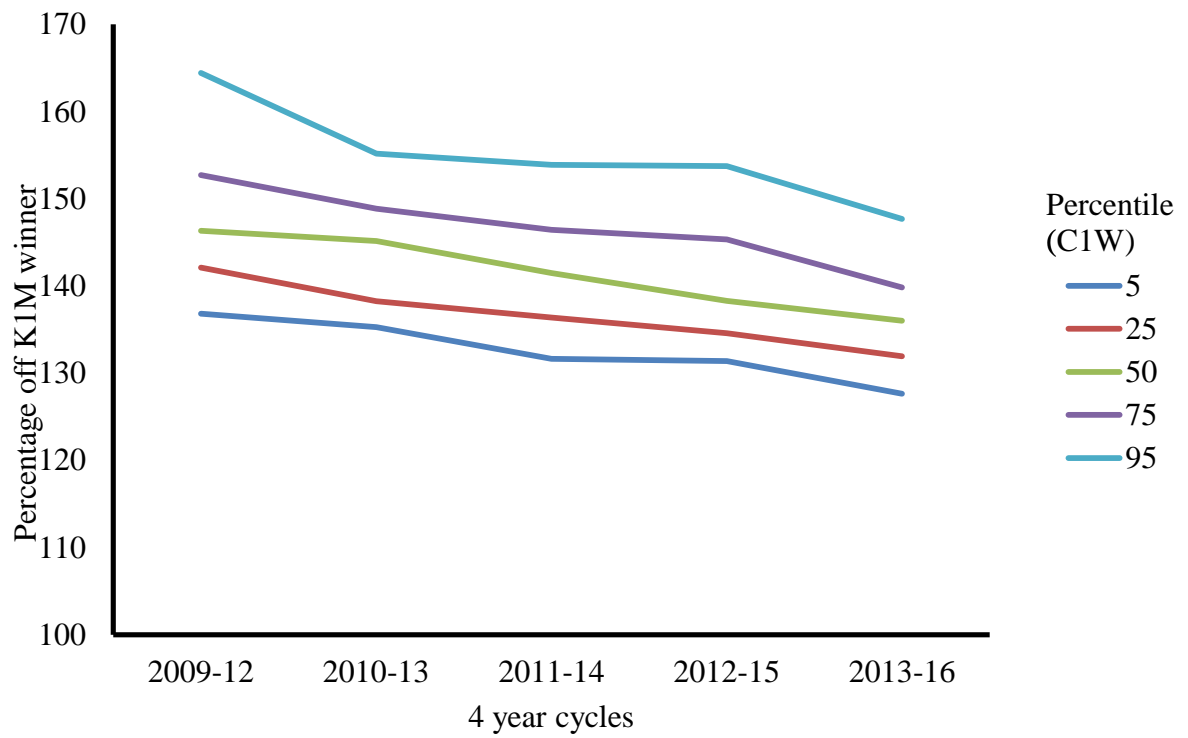
Appendix 4.40 Relative 3ʳᵈ place performance off class winner in 4 year cycles for C1M
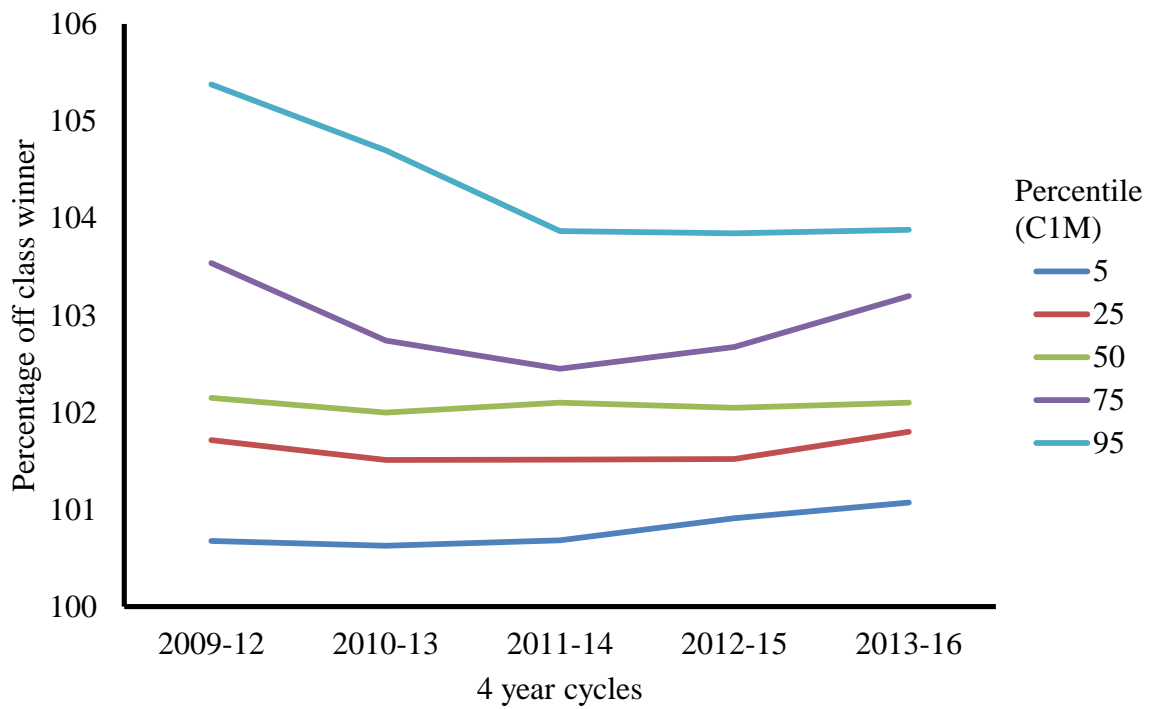
Appendix 4.41 Relative 3rd place performance off class winner in 4 year cycles for C2M
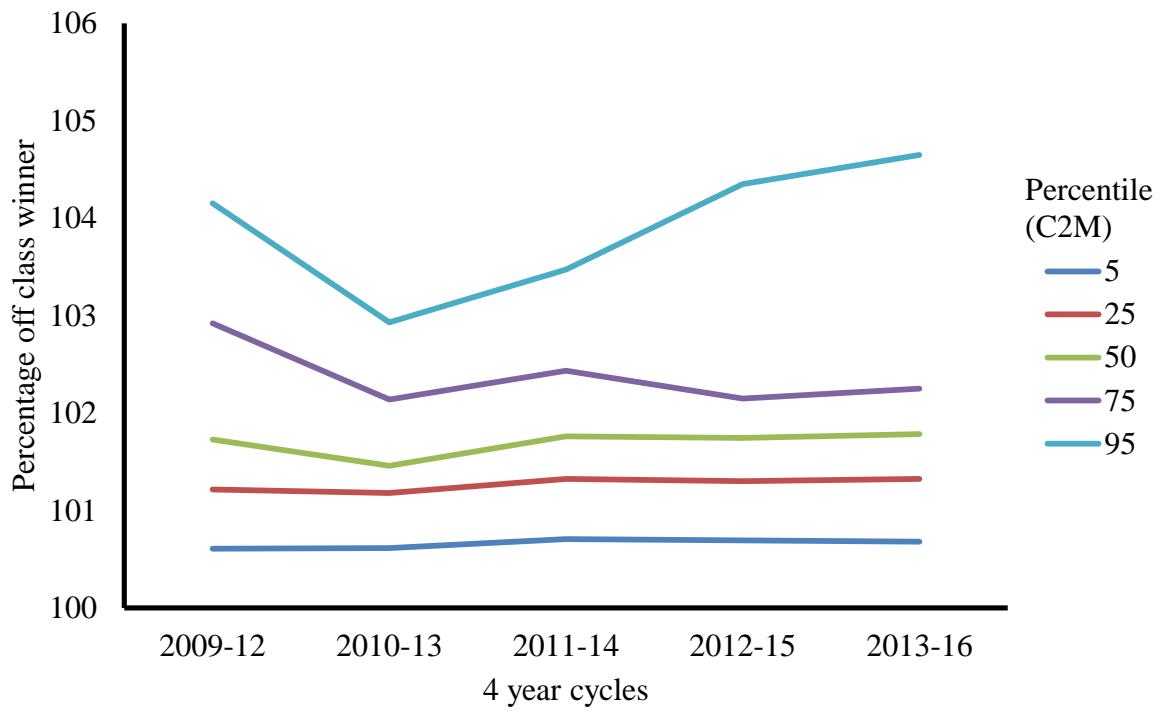
Appendix 4.42 Relative 3rd place performance off class winner in 4 year cycles for K1W
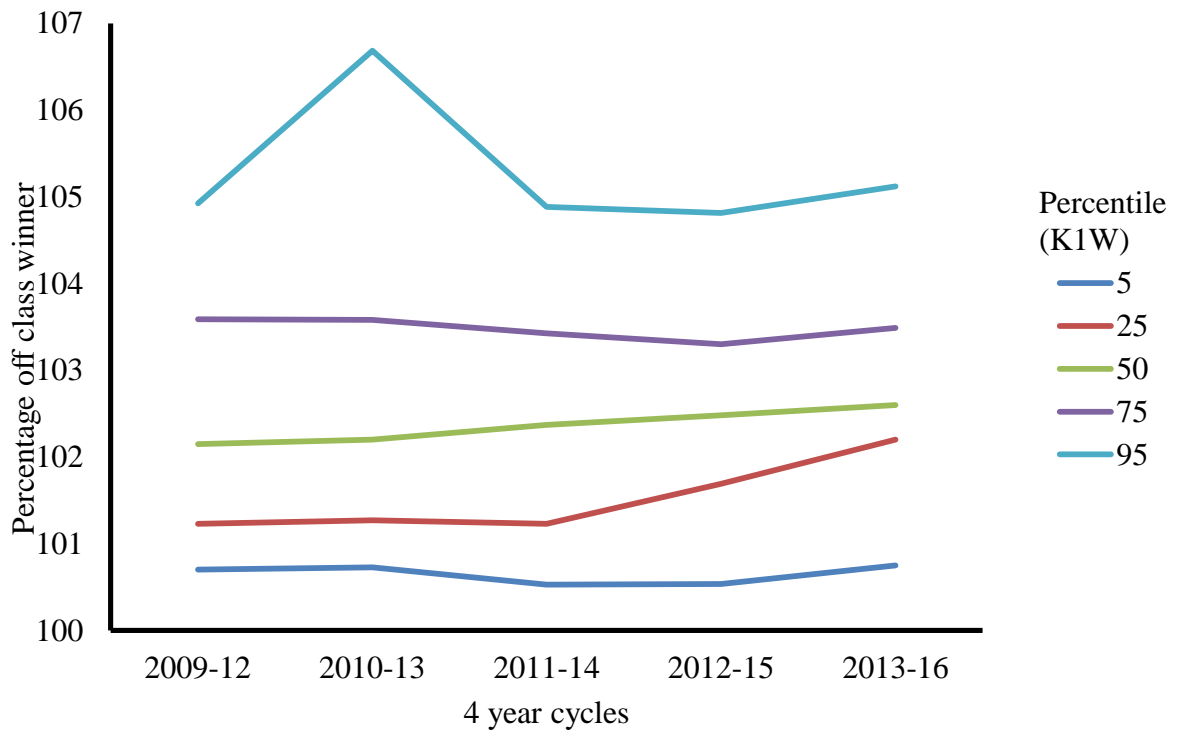
Appendix 4.43 Relative 3rd place performance off class winner in 4 year cycles for C1W
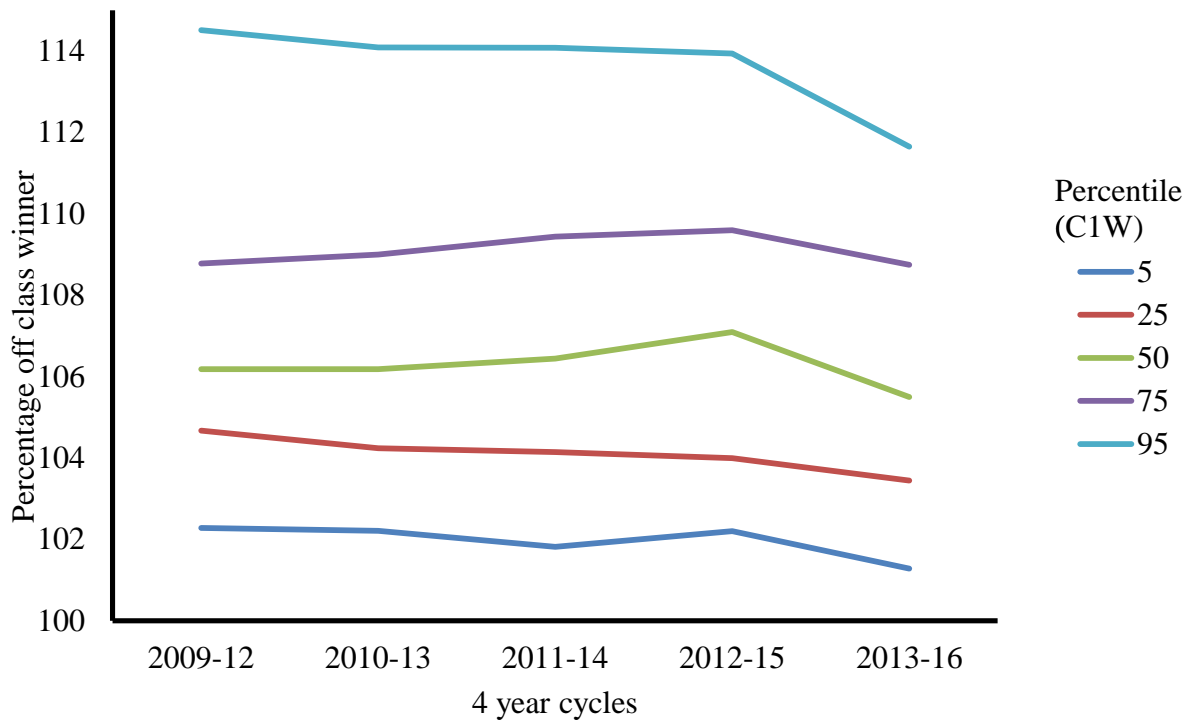
Appendix 5.1 Time series of an athlete's (3) race points, smoothed race points and World ranking
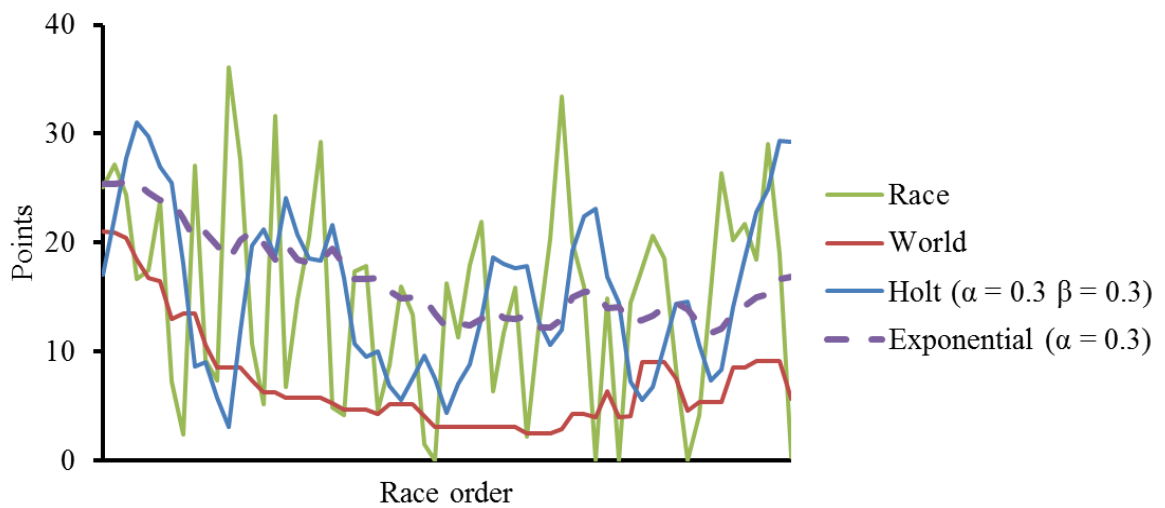
Appendix 5.2 Time series of an athlete's (6) race points, smoothed race points and World ranking
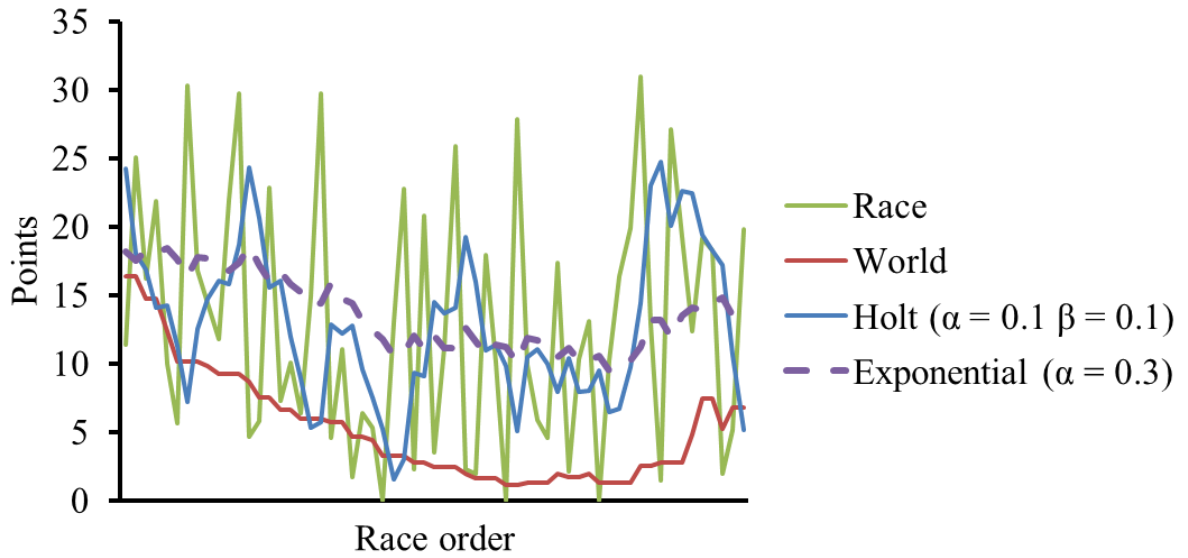
Appendix 5.3 Time series of an athlete's (7) race points, smoothed race points and World ranking
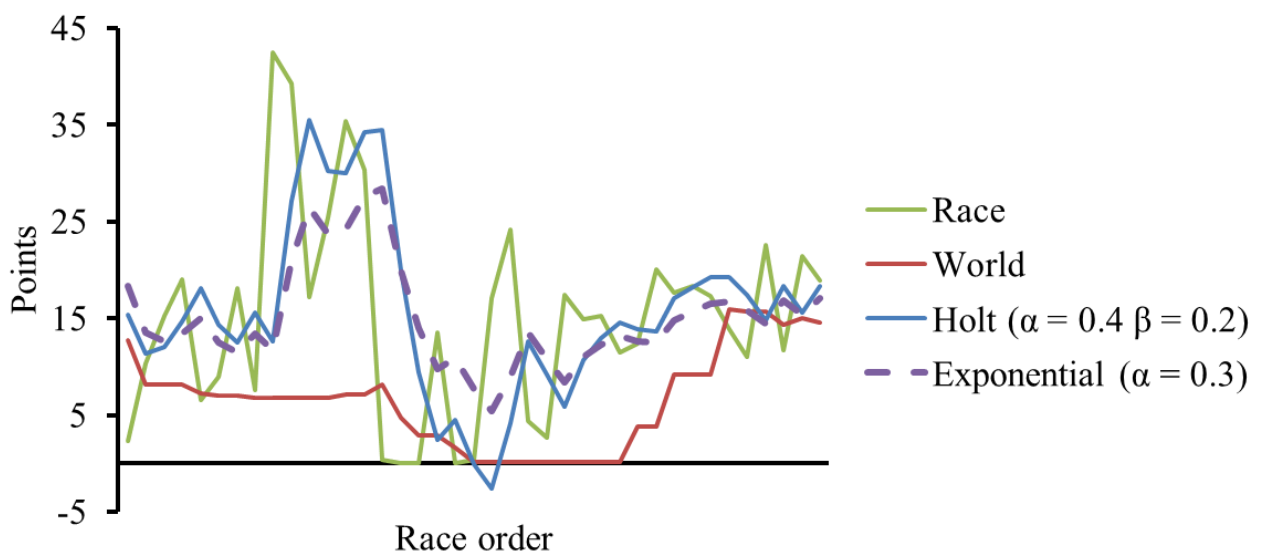
Appendix 5.4 Time series of an athlete's (12) race points, smoothed race points and
World ranking
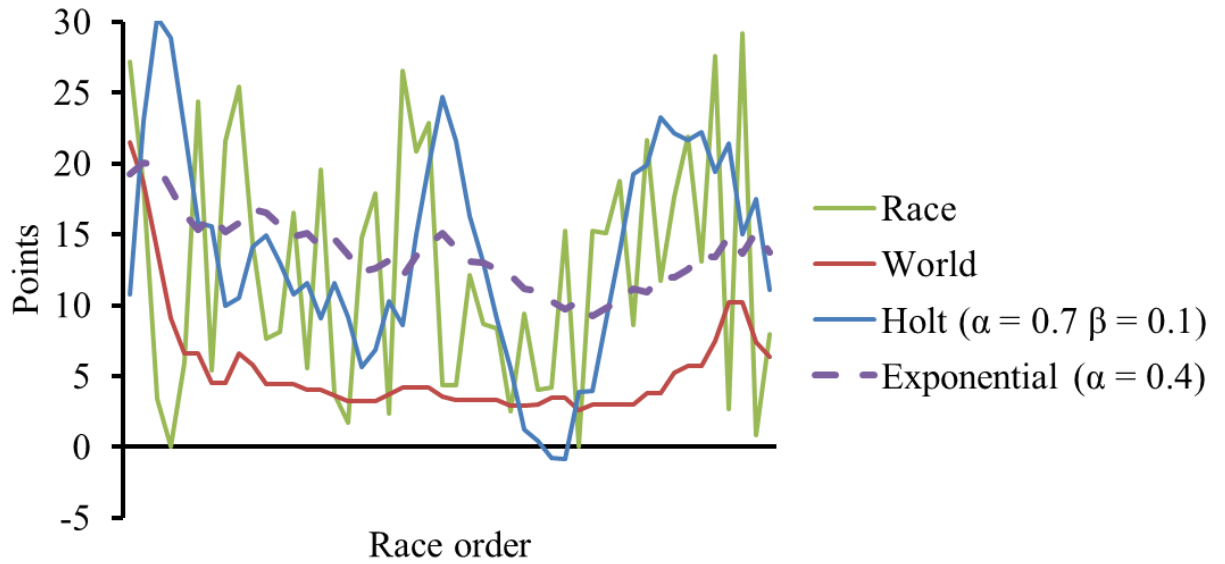
Appendix 5.5 Time series of an athlete's (14) race points, smoothed race points and
World ranking
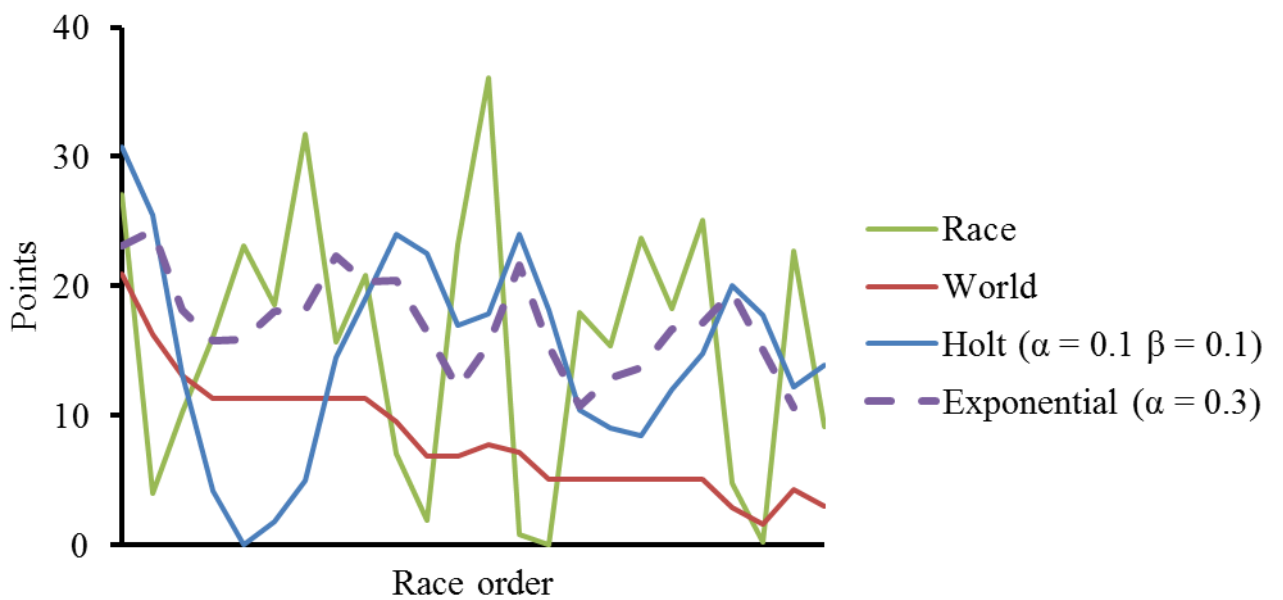
Appendix 5.6 Time series of an athlete's (15) race points, smoothed race points and
World ranking
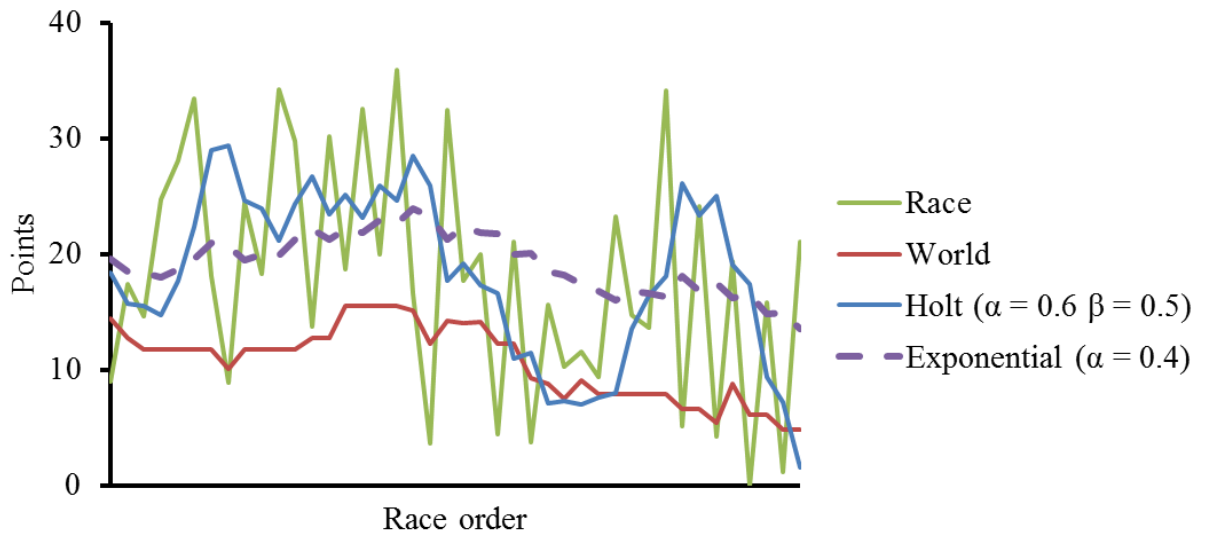
Appendix 5.7 Time series of an athlete's (18) race points, smoothed race points and
World ranking
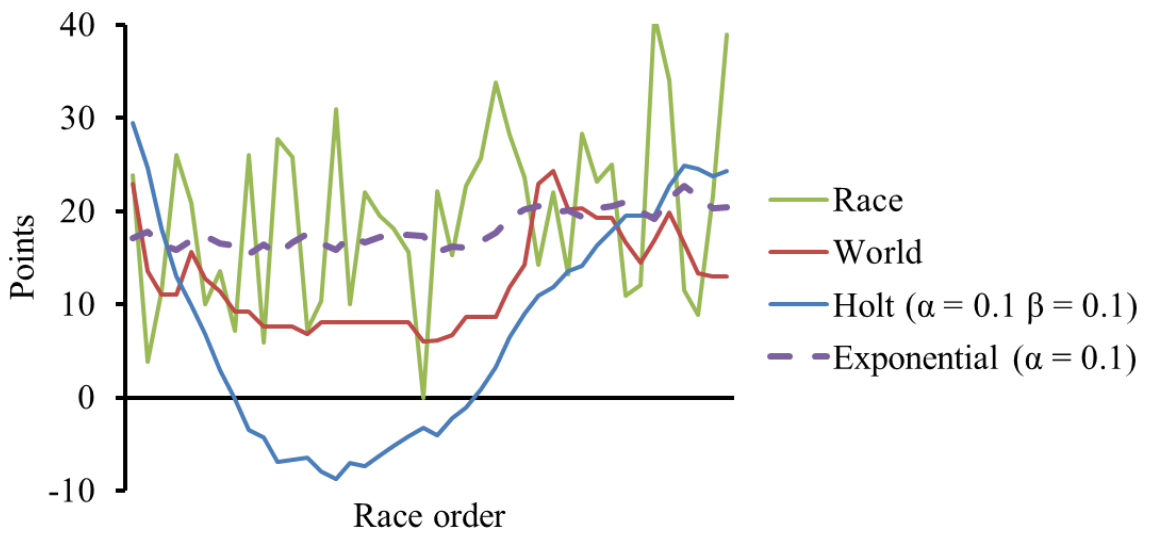
Appendix 5.8 Time series of an athlete's (19) race points, smoothed race points and World ranking
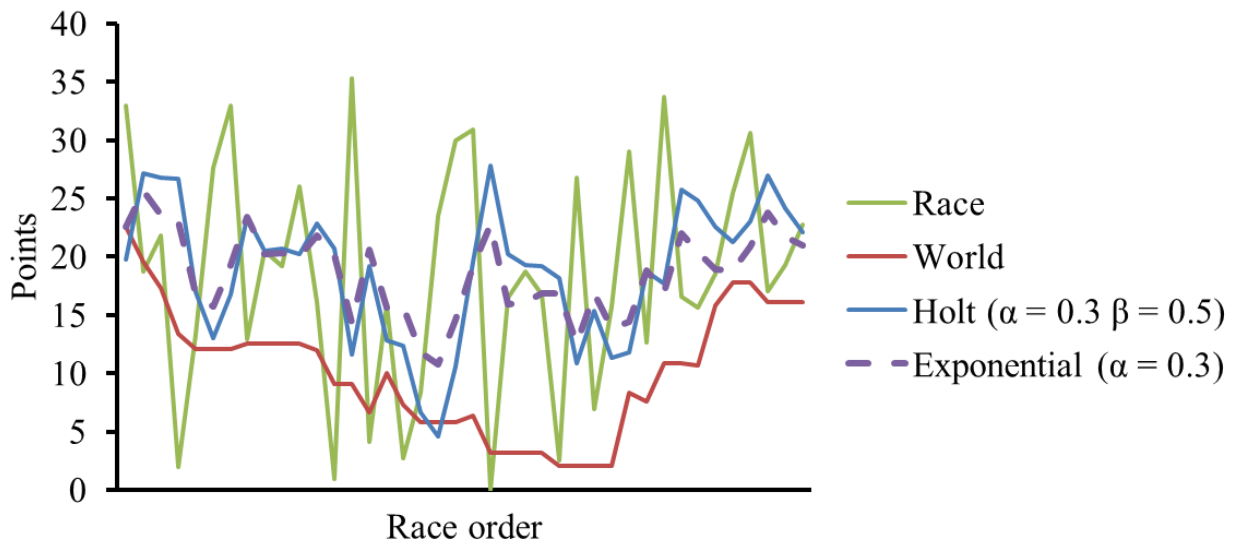
Appendix 5.9 Time series of an athlete's (A) race points, smoothed race points and World ranking
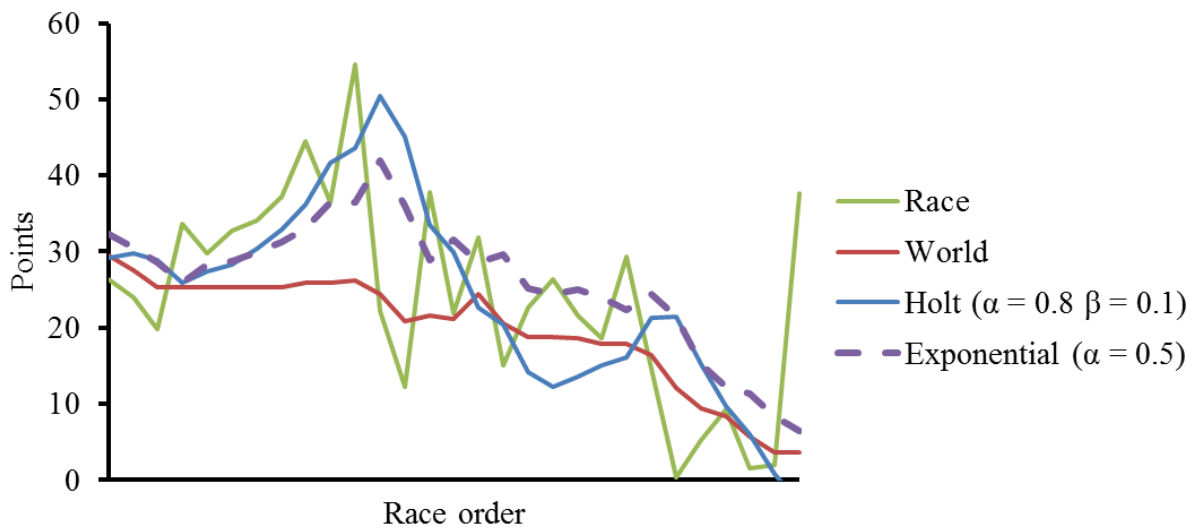
Appendix 5.10   Time series of an athlete's (B) race points, smoothed race points and
                     World ranking
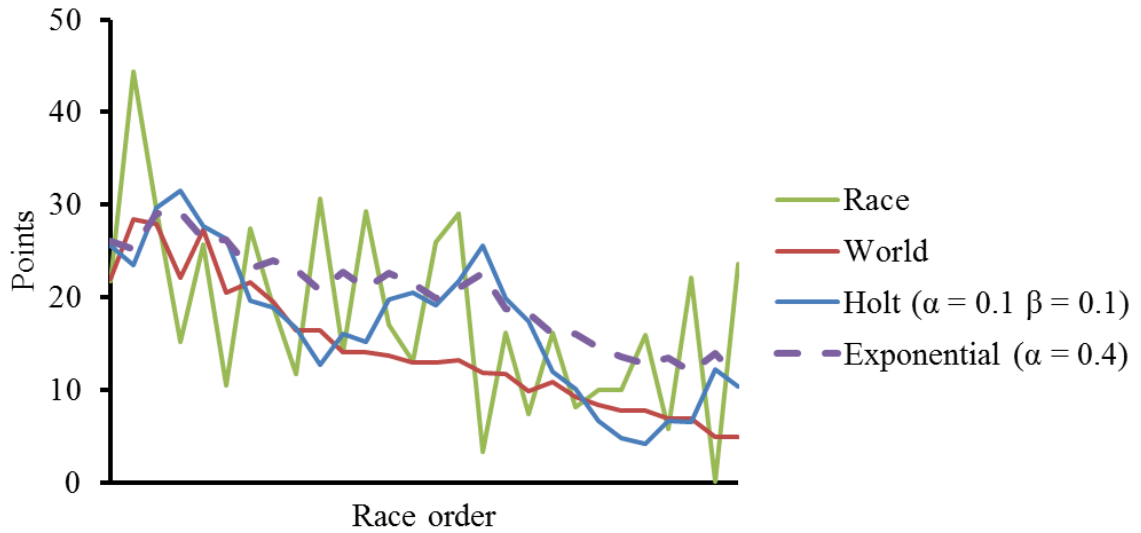
Appendix 5.11   Time series of an athlete's (C) race points, smoothed race points and
                     World ranking
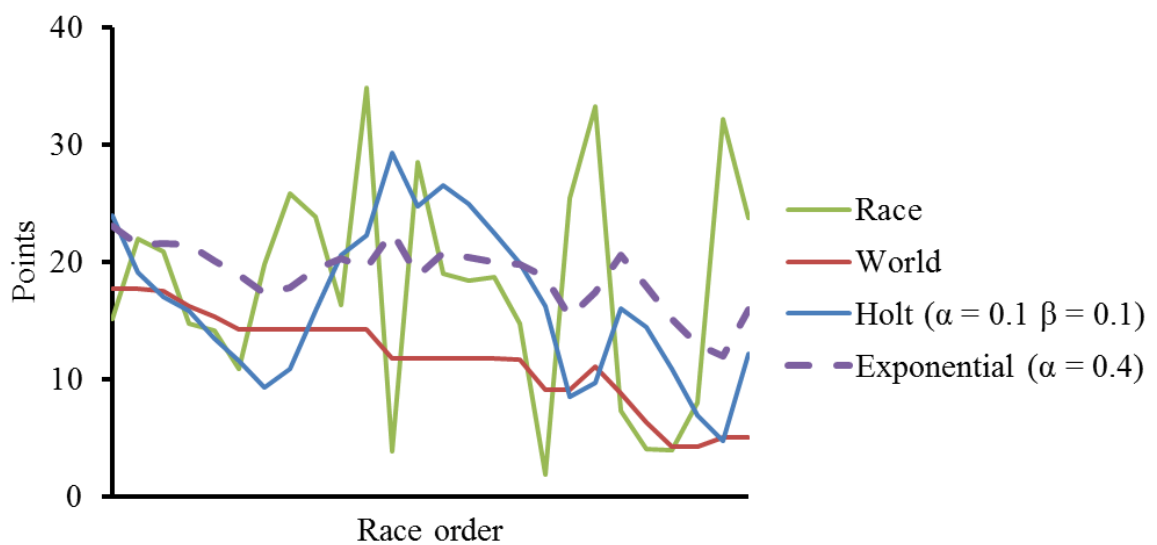
Appendix 5.12   Time series of an athlete's (D) race points, smoothed race points and World ranking
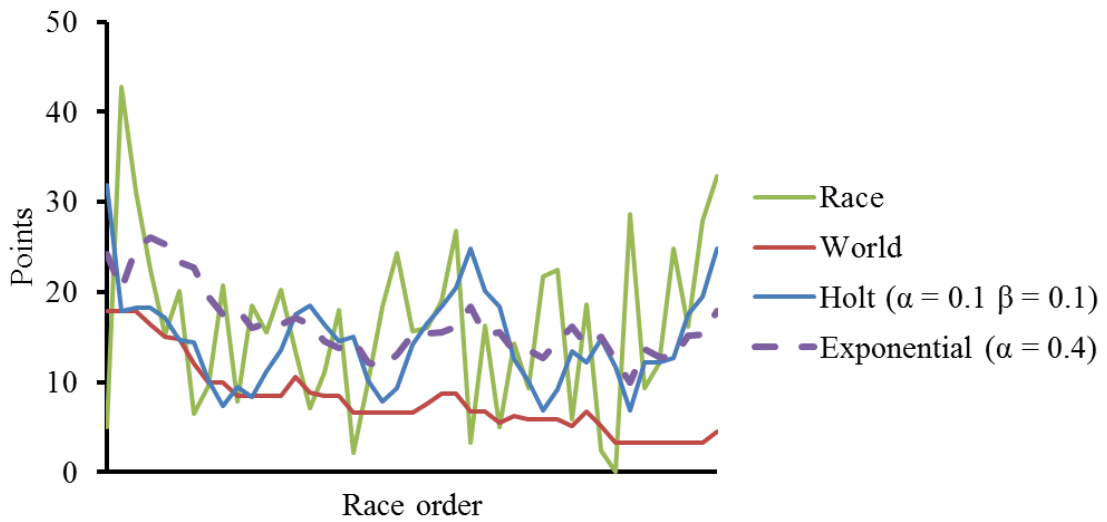
Appendix 5.13   Time series of an athlete's (E) race points, smoothed race points and World ranking
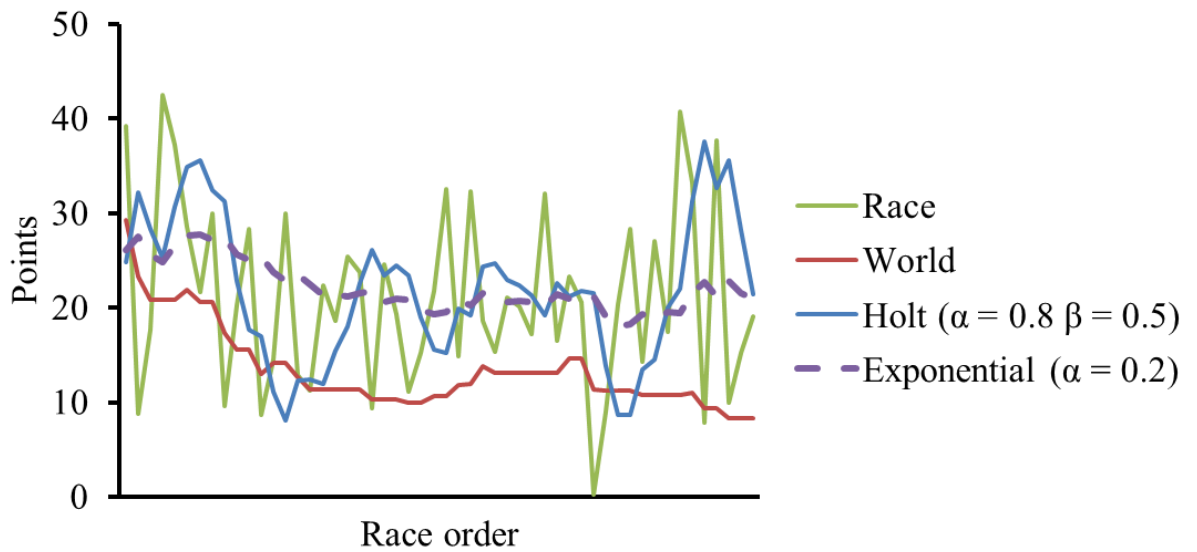
Appendix 5.14   Time series of an athlete's (6) exponentially smoothed race points to their first major championship win in relation to winning and winless funnels
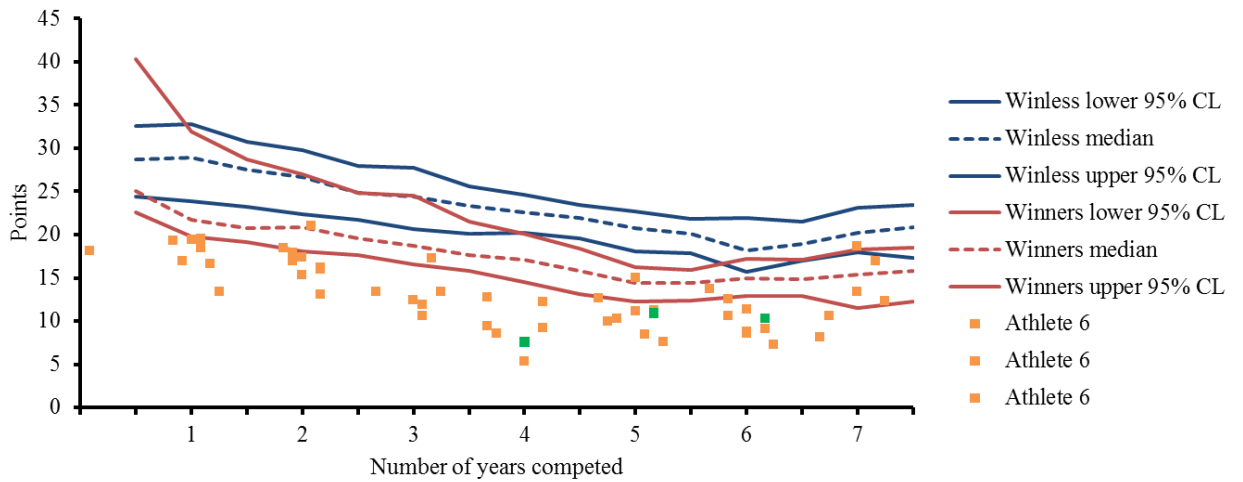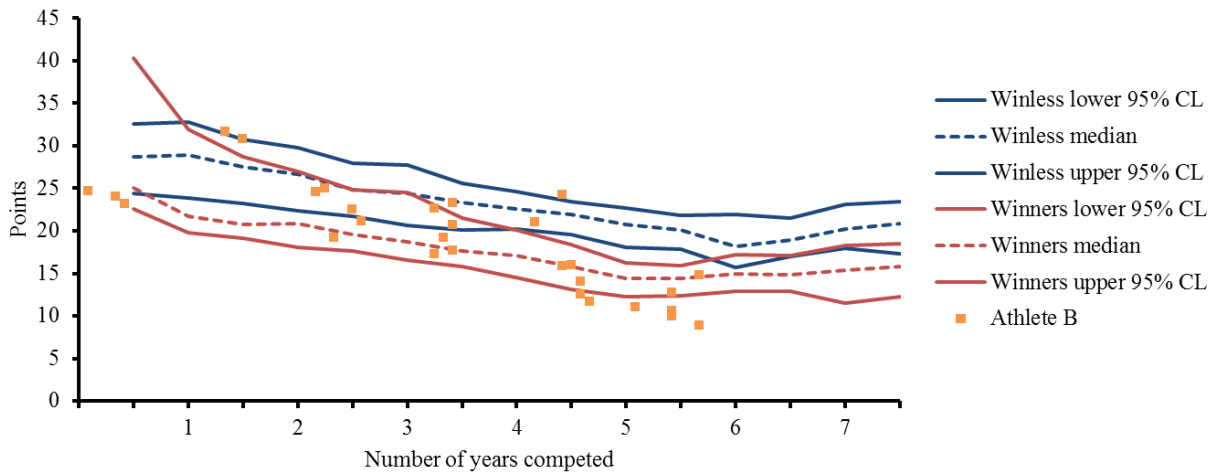

Appendix 5.15   Time series of an athlete's (7) exponentially smoothed race points to their first major championship win in relation to winning and winless funnels
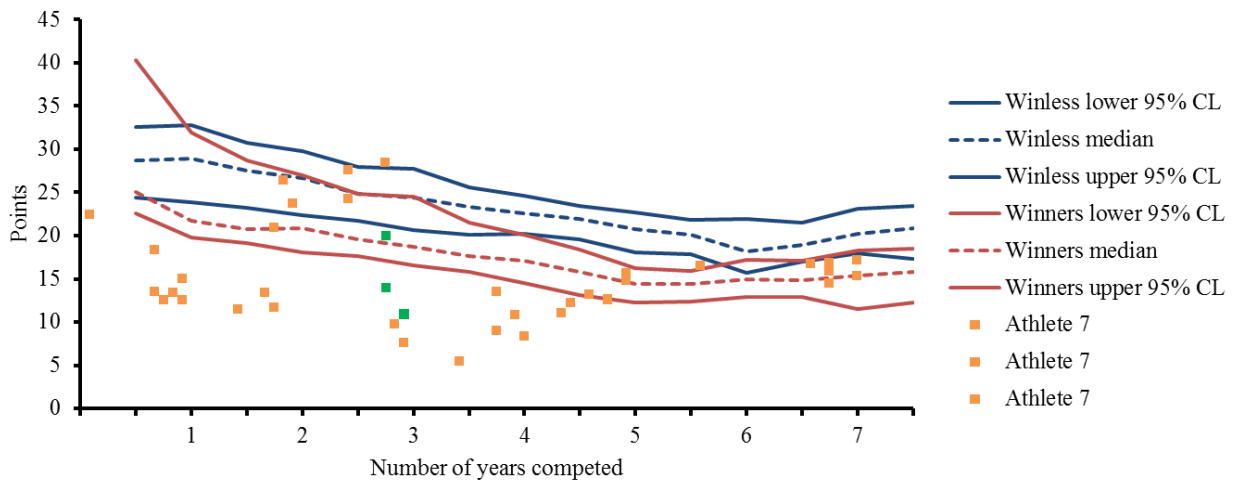
Appendix 5.16   Time series of an athlete's (12) exponentially smoothed race points
                 to their first major championship win in relation to winning and
                 winless funnels
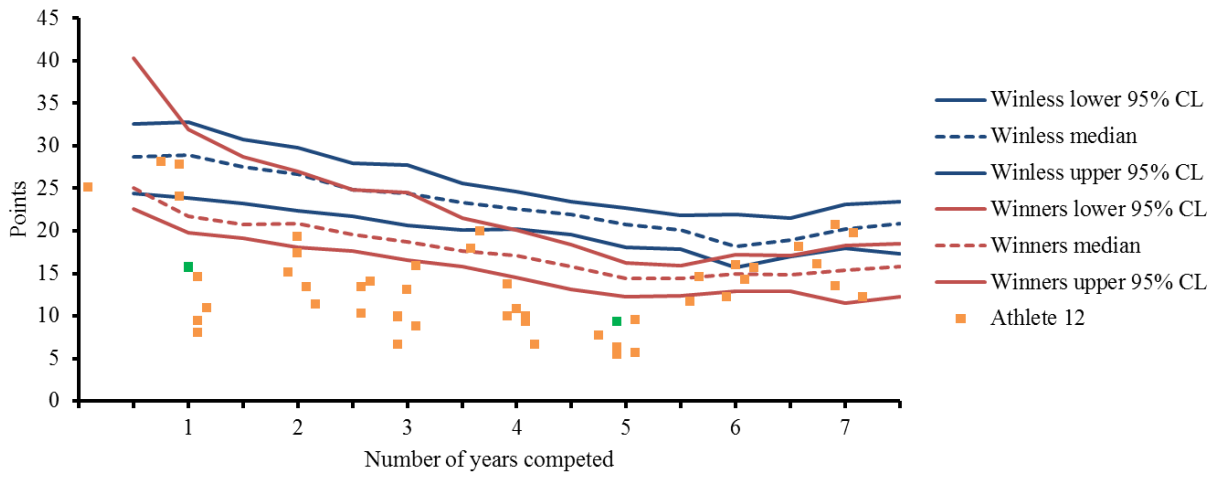
Appendix 5.17   Time series of an athlete's (14) exponentially smoothed race points
                 to their first major championship win in relation to winning and
                 winless funnels
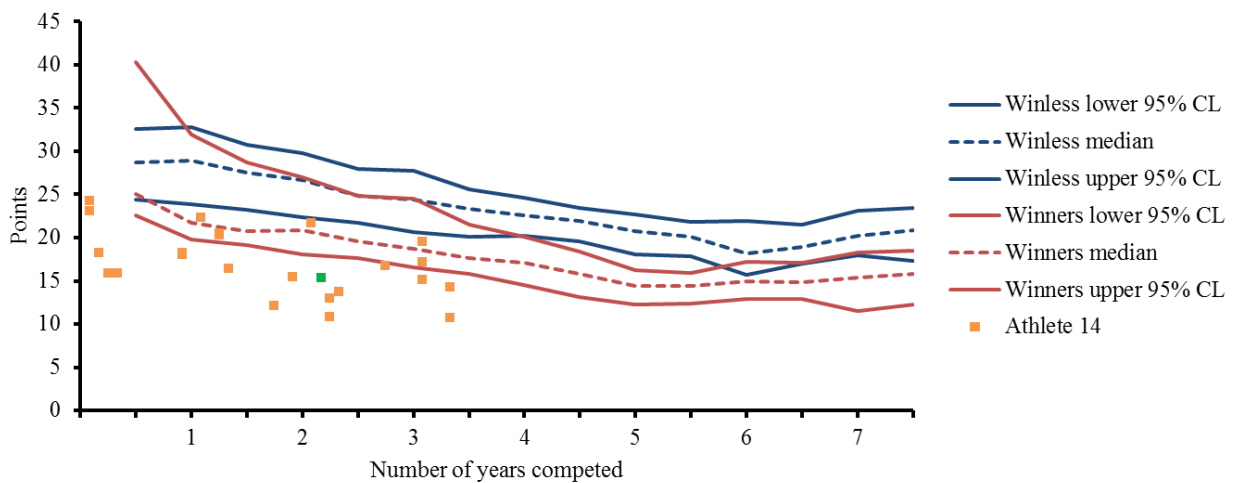
Appendix 5.18   Time series of an athlete's (15) exponentially smoothed race points
to their first major championship win in relation to winning and
winless funnels
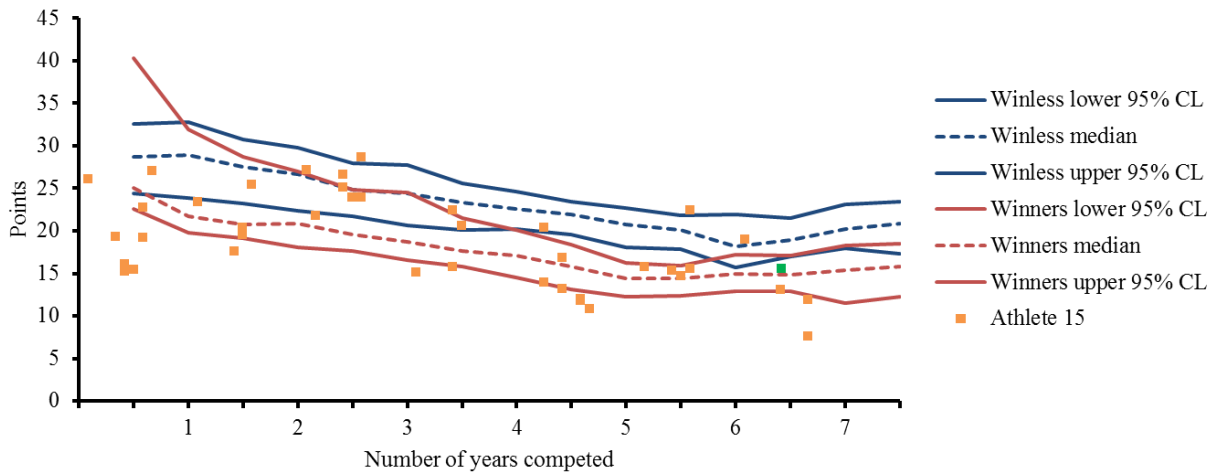
Appendix 5.19   Time series of an athlete's (16) exponentially smoothed race points
to their first major championship win in relation to winning and
winless funnels
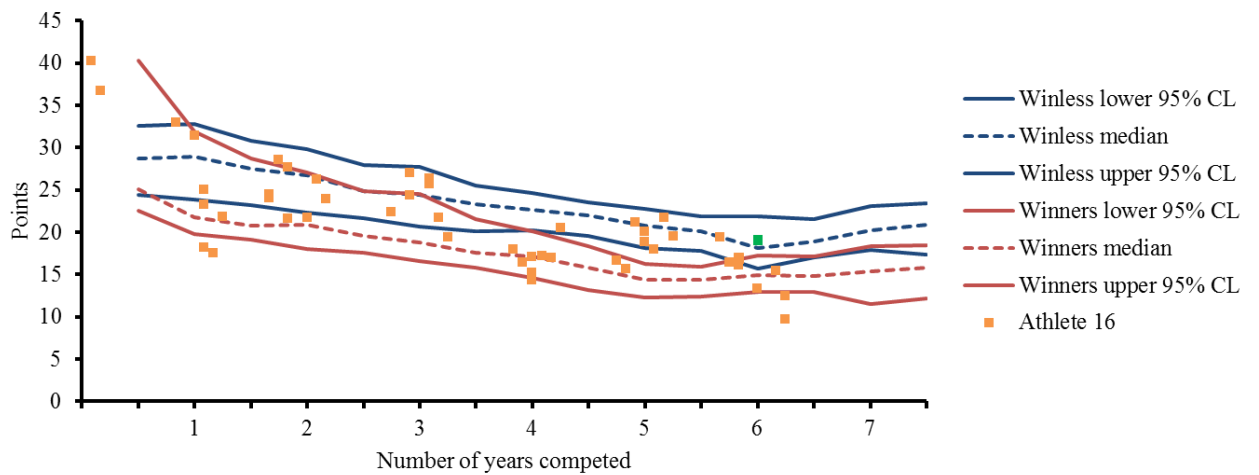
Time series of an athlete's (18) exponentially smoothed race points to their first major championship win in relation to winning and winless funnels
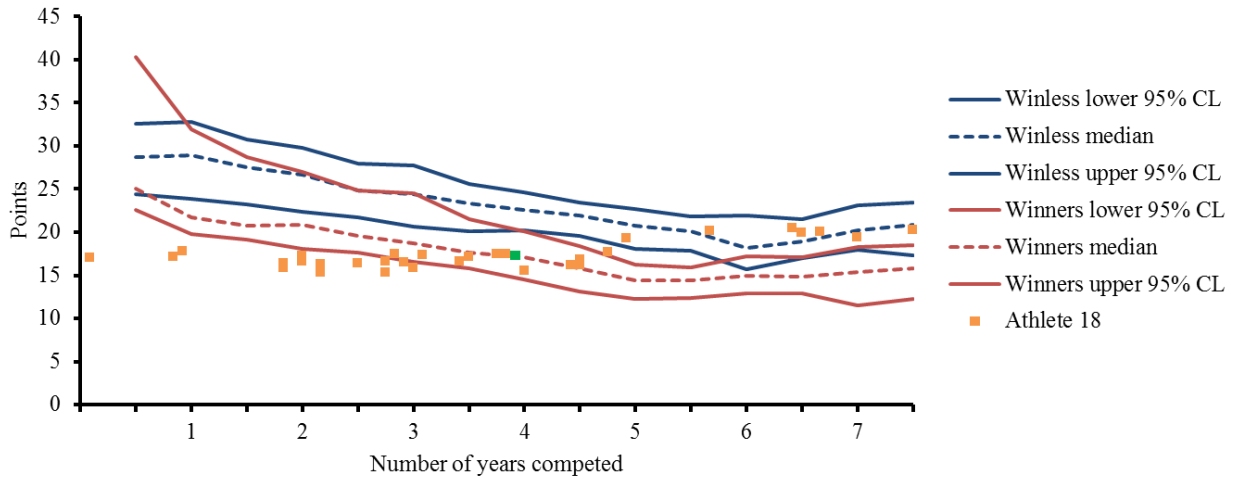
Time series of an athlete's (19) exponentially smoothed race points to their first major championship win in relation to winning and winless funnels
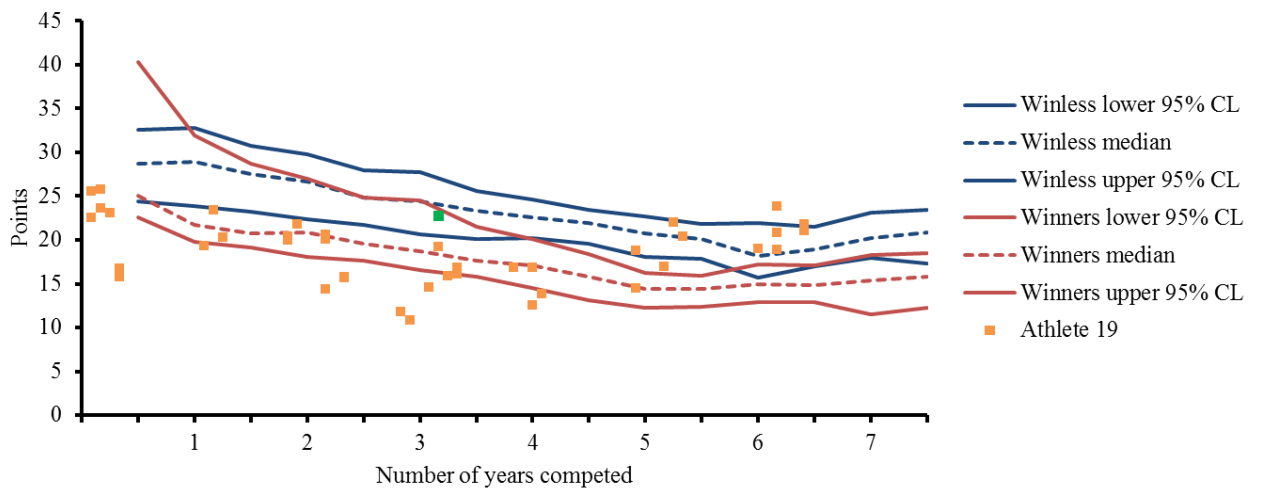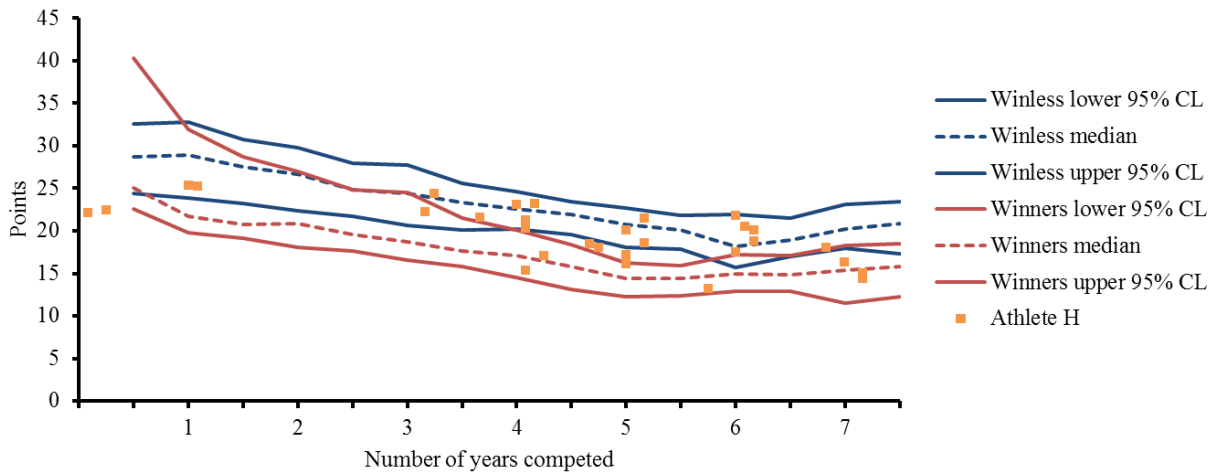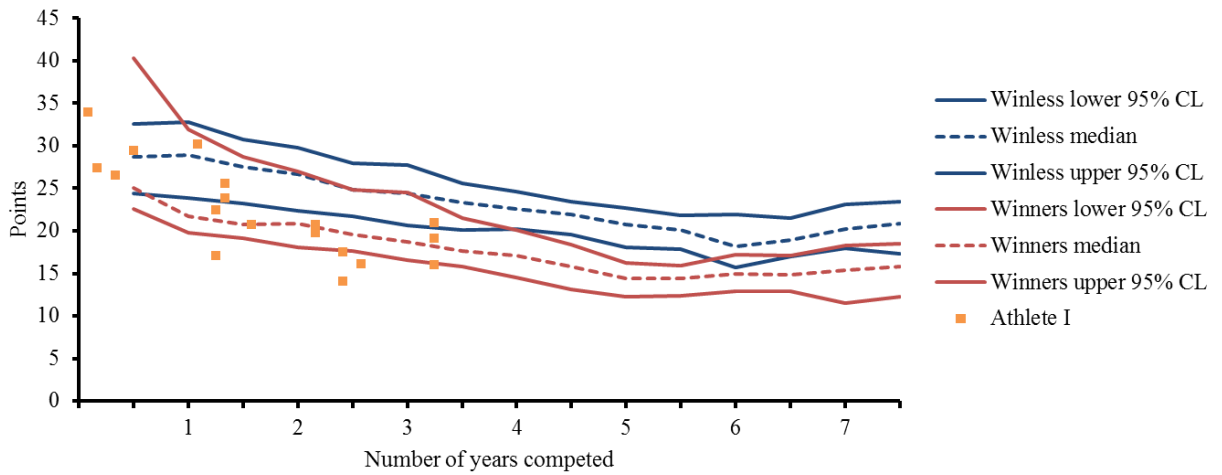
Appendix 5.22   Time series of a likely future major championship winner's (Athlete B) exponentially smoothed race points in relation to winning and winless funnels
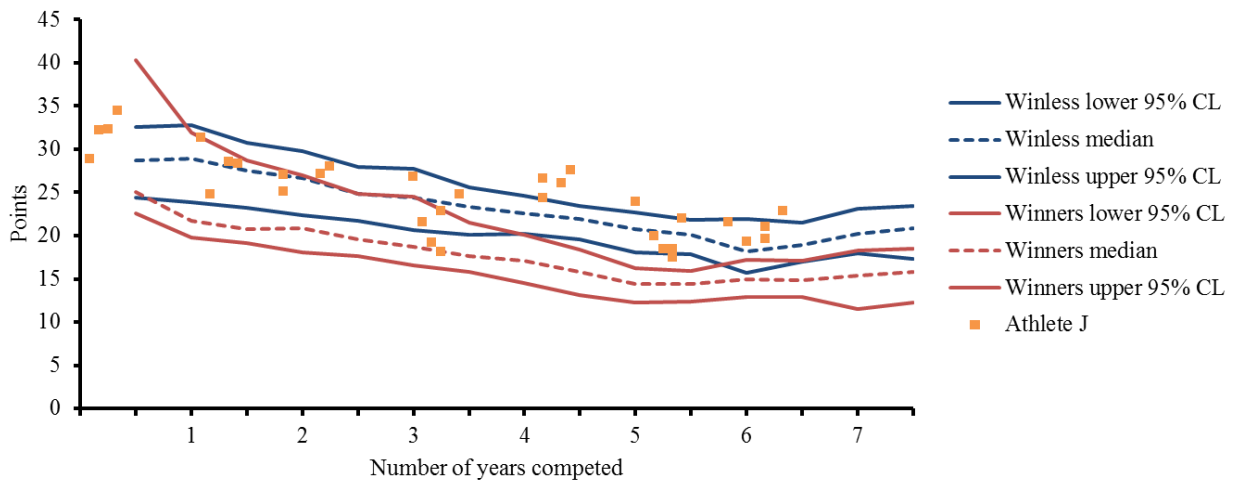

Appendix 5.23   Time series of a possible future major winner's (Athlete D) exponentially smoothed race points in relation to winning and winless funnels
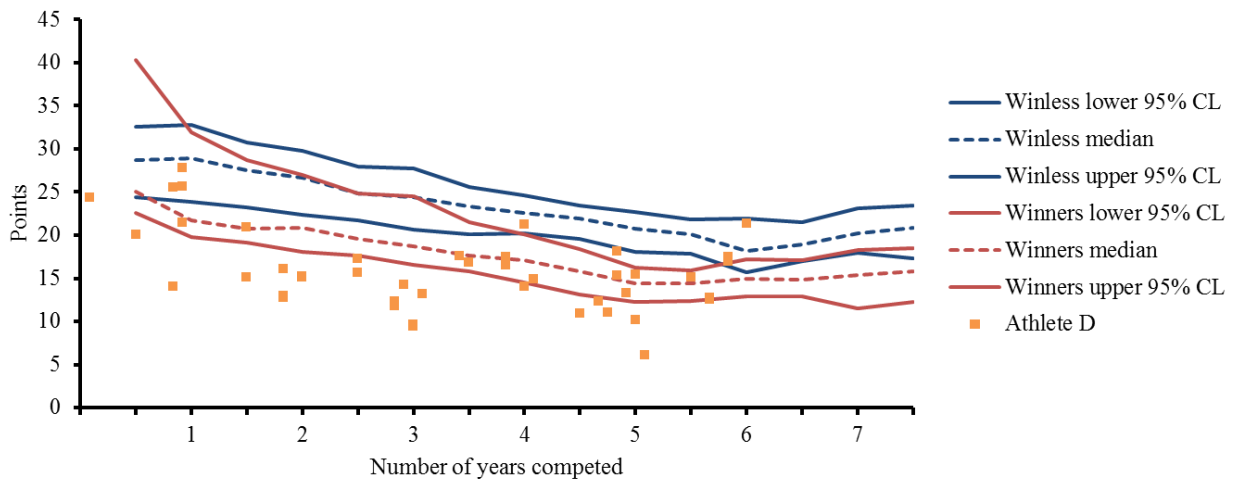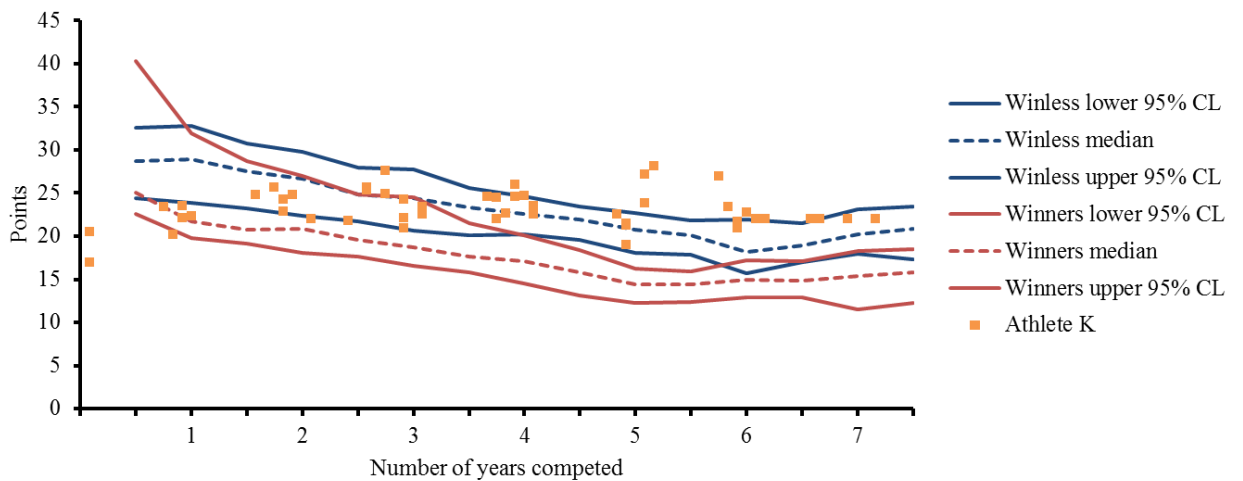
Appendix 5.24   Time series of a possible future major winner's (Athlete F) exponentially smoothed race points in relation to winning and winless funnels
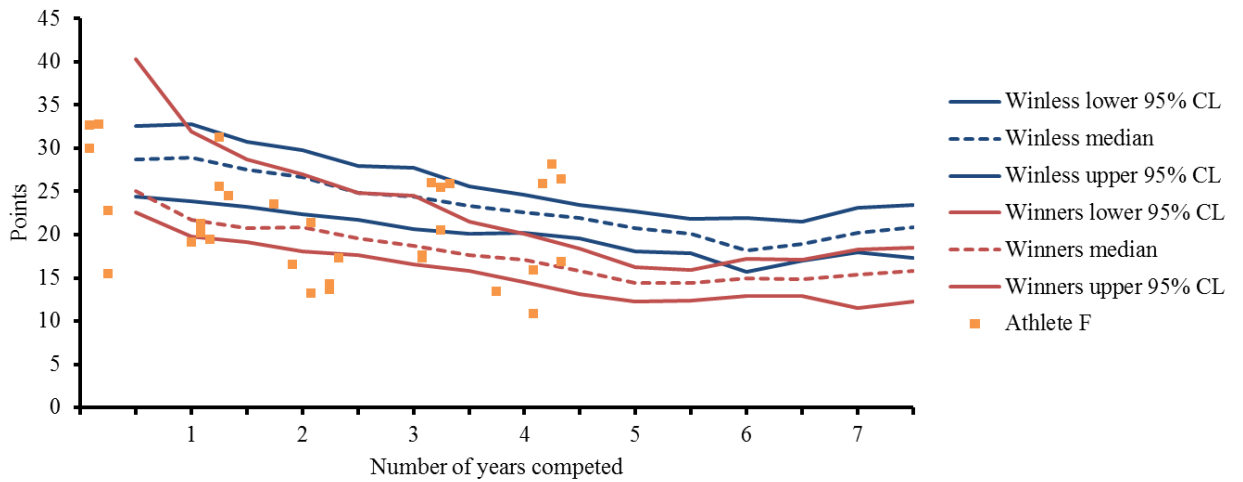
Appendix 5.25   Time series of a possible future major winner's (Athlete F) exponentially smoothed race points in relation to winning and winless funnels
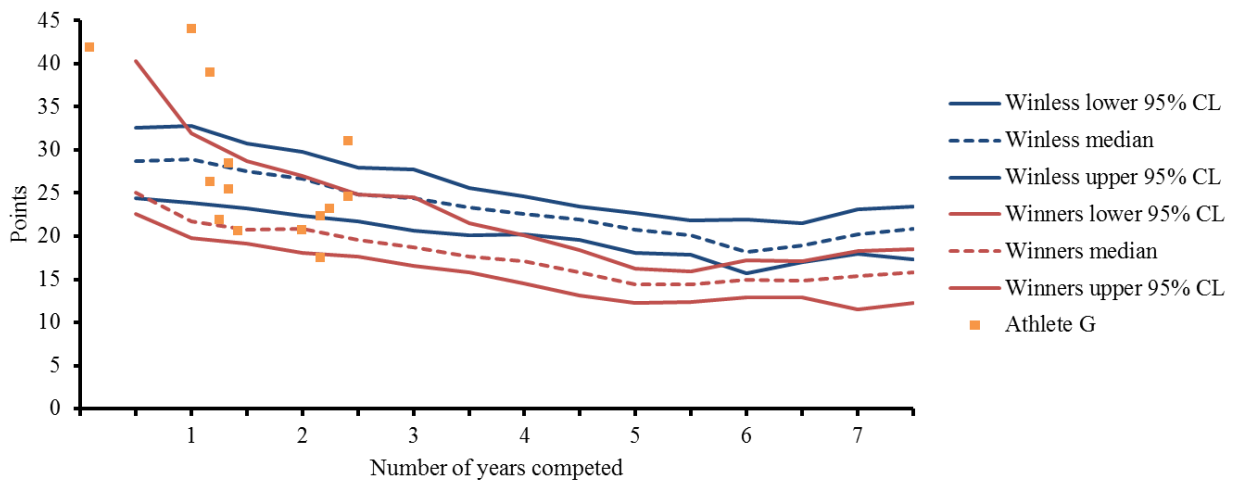
Appendix 5.26   Time series of a possible future major winner's (Athlete H) exponentially smoothed race points in relation to winning and winless funnels

Appendix 5.27   Time series of a likely future major winner's (Athlete F) exponentially smoothed race points in relation to winning and winless funnels
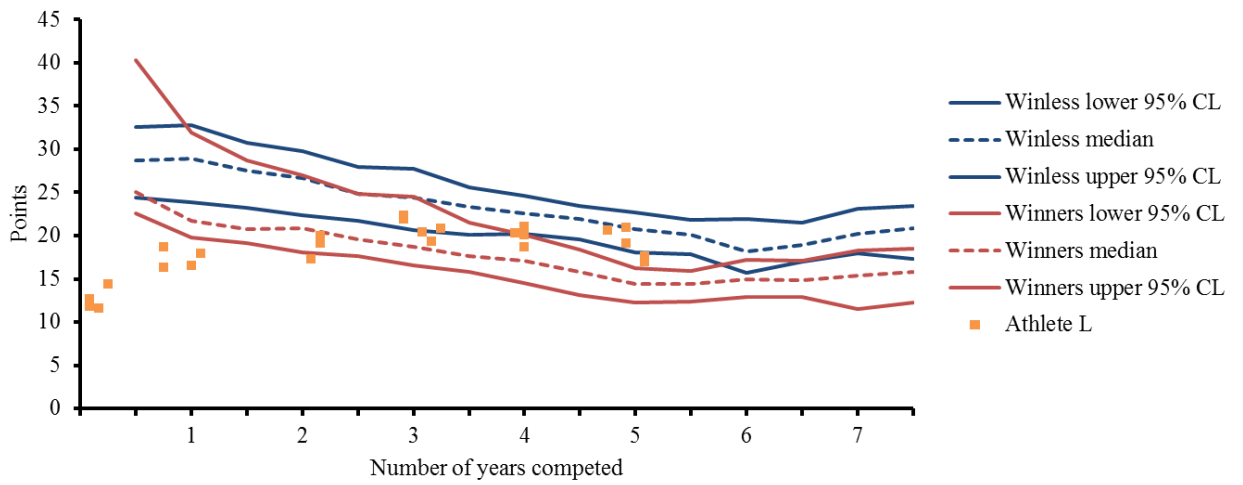
Appendix 5.28   Time series of an unlikely future major winner's (Athlete J) exponentially smoothed race points in relation to winning and winless funnels
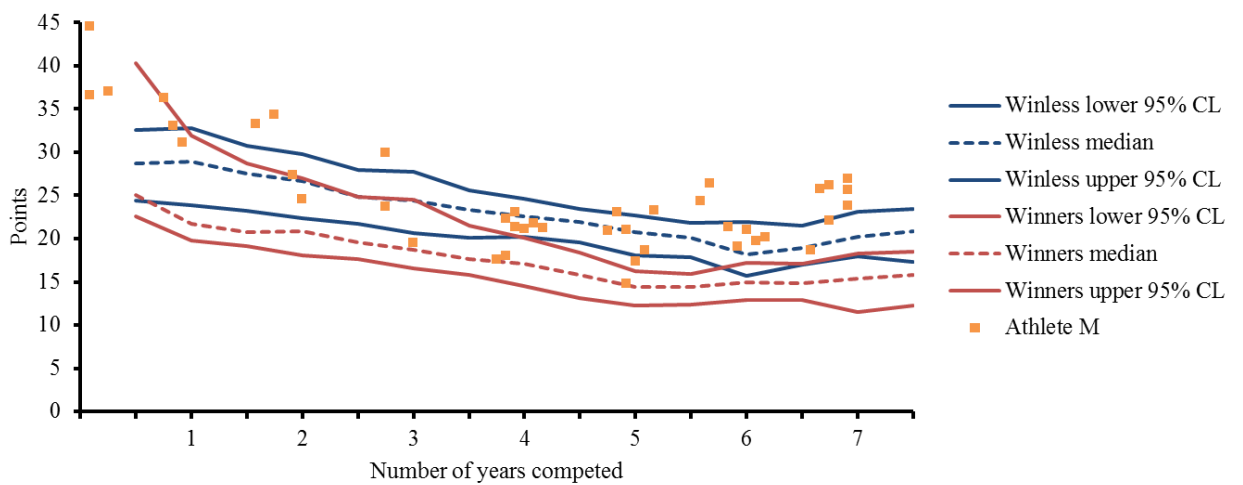

Appendix 5.29   Time series of an unlikely future major winner's (Athlete K) exponentially smoothed race points in relation to winning and winless funnels
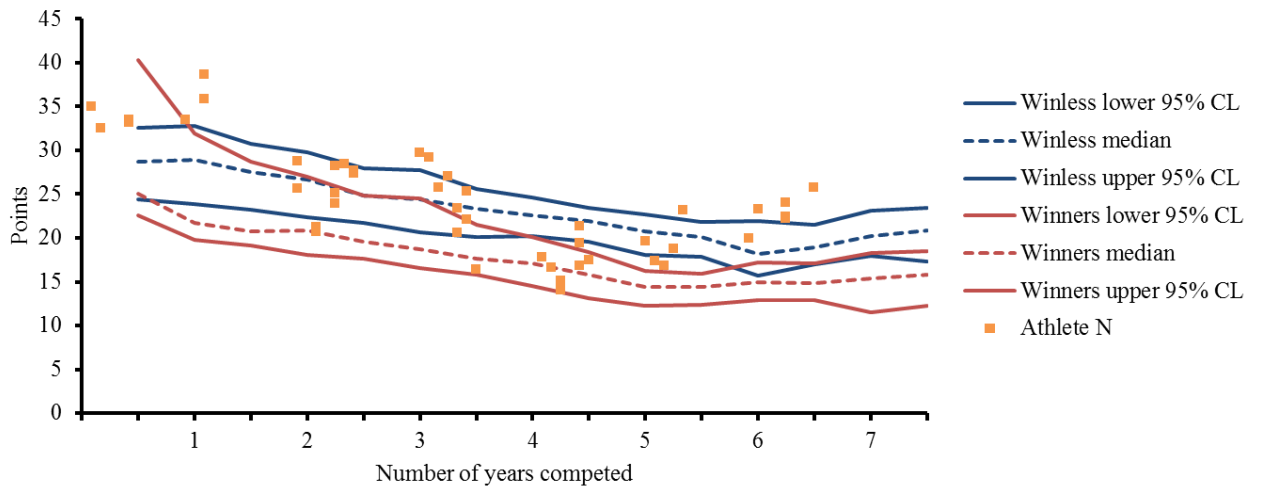
Appendix 5.30   Time series of an unlikely future major winner's (Athlete L) exponentially smoothed race points in relation to winning and winless funnels
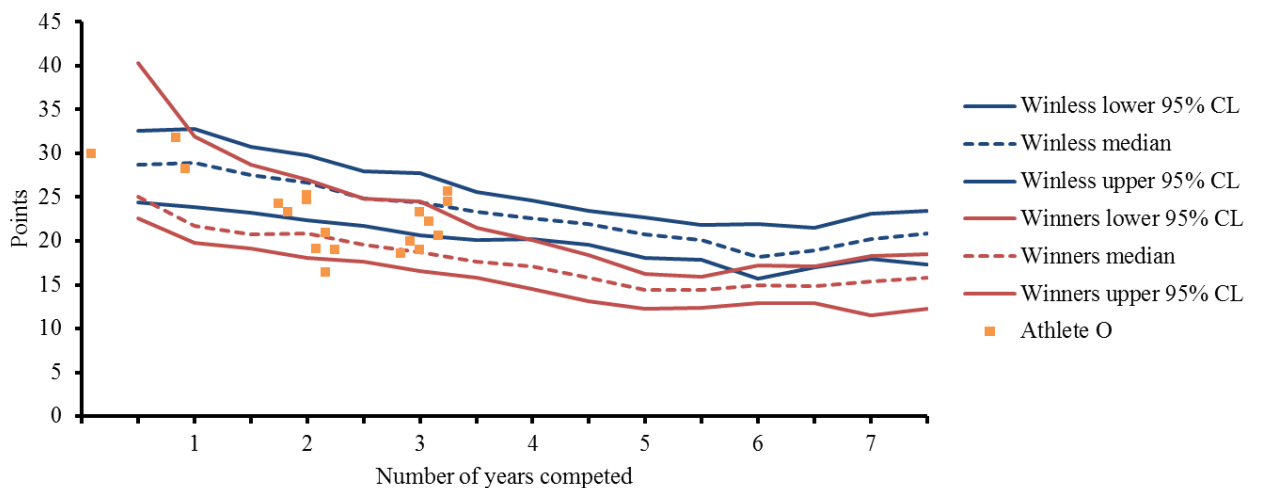

Appendix 5.31   Time series of an unlikely future major winner's (Athlete J) exponentially smoothed race points in relation to winning and winless funnels

Appendix 5.32   Time series of an unlikely future major winner's (Athlete J) exponentially smoothed race points in relation to winning and winless funnels

Appendix 5.33   Time series of a likely future major winner's (Athlete O) exponentially smoothed race points in relation to winning and winless funnels

Appendix 5.34 Time series of an unlikely future major winner's (Athlete P) exponentially smoothed race points in relation to winning and winless funnels