FUNCTIONAL DATA ANALYSIS: A NEW METHOD TO INVESTIGATE PACING STRATEGIES IN ELITE CANOE KAYAK SPRINT

Joshua A Goreham[1], Scott C Landry[2], John W Kozey[1], Bruce Smith[3], Michel Ladouceur[1]

Kinesiology, Dalhousie University, Halifax, Canada[1]
Kinesiology, Acadia University, Wolfville, Canada[2]
Mathematics and Statistics, Dalhousie University, Halifax, Canada [3]

The purpose of this study was to investigate pacing strategies used by elite flatwater canoe kayak sprint athletes in 12 Olympic events. Boat velocity data for canoe and kayak A-final races were extracted from the International Canoe Federation's website for five major international competitions during the 2016 and 2017 seasons (n=374 boats; n=87 races). Principal component analysis was used to determine pacing strategies in 200, 500, and 1000 m races. Differences in pacing strategies between medalists and non-medalists were found only in 1000 m races. These findings reflected overall differences in boat velocity but also the timing of changes in boat velocity throughout the 1000 m race. This research shows that certain pacing strategies are associated with more success in canoe kayak sprint during long duration races.

KEYWORDS: velocity, principal component analysis, inertial measurement units, flatwater canoe kayak sprint.

INTRODUCTION: A well-planned pacing strategy is an important component of optimal sporting performance. Pacing strategies have been shown to depend on the length and duration of a race (Abbiss & Laursen, 2008). Few researchers have investigated on-water pacing strategies in canoe kayak sprint, and those who have explored differences in split times or stroke rates only (Borges, Bullock, & Coutts, 2013; McDonnell, Hume, & Nolte, 2013b). Since the sport of canoe kayak sprint includes events of various distances (200, 500, and 1000 m) and durations (approximately 30 seconds to 4 minutes) investigation over these distances is warranted. There has been an increase in pace-related data collection during major international canoe kayak sprint events, in addition to traditional split times, since the International Canoe Federation's approval of adding an inertial measurement unit (IMU) to boats during each race. Functional Data Analysis (FDA) can be used to highlight characteristics, study patterns and variations in data over time, and represent data in ways that aid further analysis (Ramsay & Silverman, 2005). Results from FDA may help improve performance by investigating changes in boat velocity over a race compared to simply observing split times. The aim of the current study was to use FDA techniques to analyse publicly available boat velocity data to better understand the pacing strategies being used by elite flatwater canoe kayak sprint athletes at major international competitions. It was hypothesized that competitors in the 200, 500, and 1000 m race distances would follow all-out, positive, and even pacing strategies, respectively. It was also hypothesized there would be differences in pacing strategies between medalists (first to third place) and non-medalists (final three competitors per race; bottom 3) in all events, and that these differences would be visible during specific time points within the race (i.e. acceleration phase, mid-race, endspurt, etc.).

METHODS: Inertial measurement units (IMU; ST Innovation, Geneva, Switzerland) contained a global positioning system and an accelerometer, and were used to collect boat velocity data from all flatwater canoe kayak sprint events at the 2016 Olympic Games (Rio de Janeiro, Brazil), 2016 World Cup 1 (Duisburg, Germany), 2016 World Cup 2 (Racice, Czech Republic), 2016 World Cup 3 (Montermor-o-Velho, Portugal), and the 2017 World Championships (Racice, Czech Republic). One IMU was placed on the deck of the stern end of each individual boat. All boat velocity data from A-Finals for events being raced at the

2020 Olympic Games (Tokyo, Japan) were downloaded from the International Canoe Federation's website (www.canoeicf.com) and then analysed (n=374 boats; n=87 races). Data from each event were organized by race distance (200, 500, and 1000 m). All single athlete and crew boat (male and female) data for a given race distance were combined for analysis.

Due to differences in boat velocity for each event boat velocities were normalized. Each boat velocity value for each 10-metre split was normalized to the averaged overall race velocity. This process was completed for all races to calculate normalized boat velocity. The normalized boat velocity data for each race were then organized to split the top 3 competitors and the bottom 3 competitors for each race. All data for boats who did not finish in the top 3 or bottom 3 rankings of a race were then removed from the analysis. Principal component analysis (PCA, Matlab®) was then used to determine differences in the normalized boat velocity waveforms (normalised velocity as a function of distance), between the top 3 and the bottom 3 competitors per race of a given distance. A more detailed description of each individual step of PCA can be found in Landry (2007). An unpaired t-test and Cohen's *d* effect sizes were calculated to detect statistical significance between the top 3 and bottom 3 competitors using the calculated Principal Components (PCs). Only PCs that cumulatively explained at minimum 90% of the variance were retained for statistical analysis (Brandon et al., 2013). An alpha value of $p < 0.05$ was used to detect statistical significance. Single component reconstruction was applied to all statistically significant PCs to interpret the waveforms (Brandon et al., 2013).

RESULTS: Averaged boat velocity increased to a peak value and then declined at various rates for all race distances (Figure 1). In 200 m events, boat velocity waveforms followed an all-out pacing strategy, as there were no subsequent increases in boat velocity following the initial acceleration phase at approximately the 60-70 m mark (Figure 1, Panel A). As race distance increased athletes adopted either a positive (500 m) or an even (1000 m) pacing strategy. In 500 m events, boat velocity did not increase again after its peak value; however, there was a slight change in the negative boat velocity slope at approximately the 250 m mark (Figure 1, Panel B). In comparison, boat velocity fluctuated multiple times during 1000 m events. The average top 3 competitor in a 1000 m event increased their boat velocity at approximately the 500 m mark, and again at the 750-800 m mark, whereas the average bottom 3 competitor could not replicate the increase in boat velocity later in the race (Figure 1, Panel C).

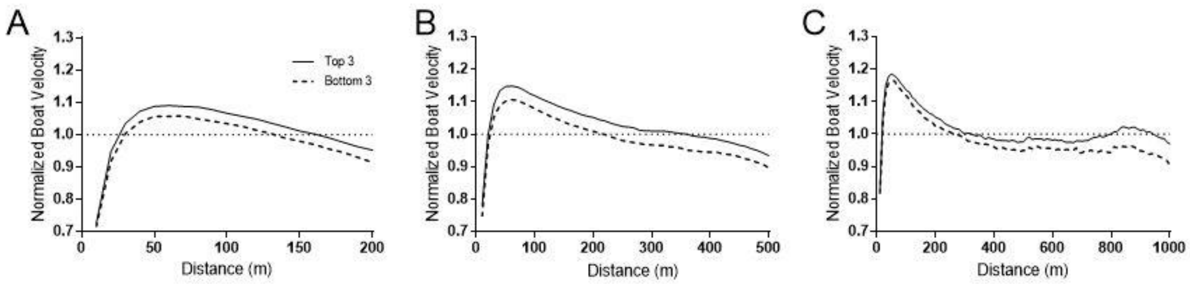

Figure 1: Average normalized boat velocity for the top 3 (solid line) and bottom 3 (dashed line) competitors for (A) 200 m, (B) 500 m, and (C) 1000 m races. Dotted line, race average.

The percent of the variance explained for the normalized boat velocity PC1 was greatest for 200 m race events (97.2%) and decreased as race distance increased (Table 1). The percent of variance explained for PC1 was greater than 90% for both the 200 m and 500 m race distances. Four PCs were needed to explain 90% of the variance for the 1000 m race distance. Normalized boat velocity PC1 (magnitude feature) and PC2 (difference feature, Figure 2) scores were significantly different between the top 3 and bottom 3 competitors for the 1000 m races (Table 1).

Table 1: Statistical and feature description information for PCs explaining variance and normalized boat velocity waveforms.

Race Distance (m)	Number of Races	PC	Percent Variation Explained	Top 3 vs. Bottom 3 Difference?	PC Feature Description
200	11	1	97.2	ns, *p* = 0.33, ES = 0.25	Magnitude
500	16	1	90.8	ns, *p* = 0.66, ES = 0.09	Magnitude
1000	15	1	77.4	s, *p* = 0.39, ES = -0.44	Magnitude
1000	15	2	8.7	s, *p* < 0.01, ES = -0.75	Difference
1000	15	3	2.6	ns, *p* = 0.83, ES = 0.04	Phase Shift
1000	15	4	1.4	ns, *p* = 0.67, ES = 0.09	Unspecified

ns, non-significant; s, significant; *p*, *p*-value; ES, Cohen's *d* effect size.

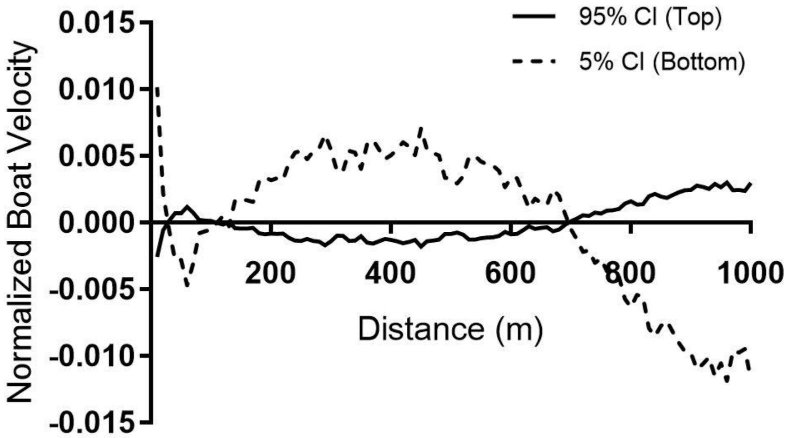

Figure 2. Confidence intervals (CI; 95% versus 5%) for normalized boat velocity PC2 (difference feature) for 1000 m races.

DISCUSSION: Using FDA to investigate boat velocity changes during canoe kayak sprint races of varying distances showed that different pacing strategies are used by elite canoe kayak sprint athletes for the different race distances, and that some pacing strategies are more successful than others. For 200 and 500 m race distances there did not seem to be enough time to follow a specific pacing strategy, aside from reaching a maximum boat velocity and trying to maintain that velocity for the remainder of the race. Boat velocity per stroke is equal to boat displacement divided by stroke time; therefore, it is possible that athletes paced themselves by altering their boat displacement or stroke time (i.e. the inverse of stroke rate), and that would have gone undetected by our methods (McDonnell, Hume, & Nolte, 2013a). PCA showed more than 90% of the variance was explained by PC1 in boat velocity waveforms for the 200 m and 500 m races. This finding means most of the top 3 athletes' pacing strategies in short distance races can be explained by a large boat velocity magnitude alone. This could be because pacing may play a smaller role in shorter race distances due to the energy systems being used (i.e. majority anaerobic energy system compared to aerobic energy system).

An interesting finding from this research was the difference in the normalized boat velocity scores for 1000 m events (Figure 1, Panel C). PCA showed that the top 3 athletes tend to race close to an average boat velocity for the entire race distance until approximately the 700 m mark, where they then increase their velocity to finish the race (Figure 2). Previous research has defined this phase as the endspurt phase (Tucker & Noakes, 2009). The PC2 waveforms (Figure 2) show that the bottom 3 competitors increase their boat velocity in the middle portion of the race (i.e. between the 100 m and 700 m mark), as their velocity is greater than the overall race average velocity during that phase. As bottom 3 competitors reach the 700 m mark of the race their boat velocity decreases rapidly. This pacing strategy does not seem to be effective as they could not finish the race with a strong endspurt phase. PC1 was also significantly different between the top 3 and bottom 3 1000 m competitors, meaning the top 3 competitors had an overall greater boat velocity magnitude than the bottom 3 during the race. These data show medalists and non-medalists followed different pacing strategies for 1000 m races, which could be due to the amount of time an athlete has to alter their work rate (Tucker & Noakes, 2009), and due to the combination of energy systems being used during long duration maximal exercise (Gastin, 2001).

CONCLUSION: Results regarding differences in pacing strategies between competitors at different time points within a race were mixed. Differences in normalized boat velocity PC1 and PC2 scores for athletes competing in long duration events (1000 m) were significantly different between top 3 and bottom 3 competitors; however, this was not found in short duration events (200 and 500 m). In accordance with the results from previous literature and our hypotheses, we found pacing strategies in short-duration events follow an all-out approach, whereas moderate- and long-duration events follow a positive and even-paced approach, respectively (Abbiss & Laursen, 2008). These findings can be used by coaches and practitioners to ensure their athletes are following appropriate pacing strategies. Athletes should ensure they are able to increase their boat velocity during the final 300 m of a 1000 m race to be successful in elite canoe kayak sprint. Athletes are more likely to do this if they do not increase their boat velocity above average during the middle portion of a 1000 m race. Although these findings are intended for canoe kayak sprint, FDA can be used to investigate pacing strategies in other sports using the same methods as shown here.

REFERENCES:

Abbiss, C. R., & Laursen, P. B. (2008). Describing and Understanding Pacing Strategies during Athletic Competition. *Sports Medicine*, *38*(3), 239–252.

Borges, T. O., Bullock, N., & Coutts, A. J. (2013). Pacing characteristics of international Sprint Kayak athletes. *International Journal of Performance Analysis in Sport*, *13*(2), 353–364.

Brandon, S. C. E., Graham, R. B., Almosnino, S., Sadler, E. M., Stevenson, J. M., & Deluzio, K. J. (2013). Interpreting principal components in biomechanics: Representative extremes and single component reconstruction. *Journal of Electromyography and Kinesiology*, *23*(6), 1304–1310.

Gastin, P. B. (2001). Energy System Interaction and Relative Contirbution During Maximal Exercise. *Sports Medicine*, *31*(10), 725–741.

Landry, S. C., McKean, K. A., Hubley-Kozey, C. L., Stanish, W. D., & Deluzio, K. J. (2007). Knee biomechanics of moderate OA patients measured during gait at a self-selected and fast walking speed. *Journal of Biomechanics*, *40*(8), 1754–1761.

McDonnell, L. K., Hume, P. A., & Nolte, V. (2013a). A deterministic model based on evidence for the associations between kinematic variables and sprint kayak performance. *Sports Biomechanics*, *12*(3), 205–220.

McDonnell, L. K., Hume, P. A., & Nolte, V. (2013b). Place time consistency and stroke rates required for success in K1 200-m sprint kayaking elite competition. *International Journal of Performance Analysis in Sport*, *13*(1), 38–50.

Ramsay, J., & Silverman, B. (2005). *Functional Data Analysis* (Second). New York: Springer.

Tucker, R., & Noakes, T. D. (2009). The physiological regulation of pacing strategy during exercise: a critical review. *British Journal of Sports Medicine*, *43*(6), 1–9.

Acknowledgement: This work was supported by Mitacs and Own the Podium through the Mitacs-Accelerate program, and the Nova Scotia Graduate Scholarship.