Defining Racing Profiles in World Championship 2000m Rowing

An Exploration through K-Means Clustering

Ryan Sheehan[1], Dani Chu[1], Dr. Ming-Chang Tsai[2,3] & Dr. Michael Jack Davis[1]

1. Simon Fraser University 2. Canadian Sport Institute Pacific 3. University of Toronto

Contact Information:

Ryan Sheehan
ryan_sheehan@sfu.ca

Dani Chu
dani_chu@sfu.ca
danichusfu.github.io

Dr. Ming-Chang Tsai
mtsai@csipacific.ca

Dr. Michael Jack Davis
jack_davis@sfu.ca

Abstract

The pacing strategy adopted by an athlete is one of the major determinants of a successful performance during competition. Various pacing profiles are reported in the literature and its potential to precede a winning performance depends on the mode of sport. However in 2000m Rowing, the definition of these pacing profiles has been subjective and there is a need to be more objective with the definition. Our objective is to objectively identify pacing profiles used in World Championship 2000 meter rowing races. To do this the average speed and stroke rate (SR) for each 50 meter split for each boat in every race of the Rowing World Championships from 2014-2017 was downloaded from www.worldrowing.com. There were 3496 boats (n = 2112 for men, n = 1384 for women) from 657 races used in the analysis. Pacing profiles were determined using average boat speeds at each 50 meter split. Boat speeds were first standardized with respect to the boat and race (row wise), then standardized with respect to the split (column wise). A principal components analysis was used to extract linearly uncorrelated components in order to reduce the dimensions of the dataset while maintaining maximal information for each dimension. A k-means clustering was used on the first 7 principal components to partition each boats pacing profile and 4 clusters were identified. Additionally, a soft clustering method for longitudinal data was used on the raw speed splits to approach the clustering from a different perspective. Finally, the clusters were described using boat and race descriptors to draw conclusions about who, when and why each pacing profile was observed. Chi-Squared Tests of Independence with Bonferroni corrections were used to test whether variables such as boat size, gender, round, or rank are associated with pacing profiles. 4 pacing strategies (Allout, Consistent, J-Shaped, and Reverse J-Shaped) were identified from the clustering process. Stroke rate, boat size, round (Heat vs Finals) and rank were all found to affect pacing profiles. Whereas, gender, and weight class were not found to affect a boats pacing profile. This novel approach of using clustering is able to objectively define 4 strategies used in 2000m rowing competitions.

Introduction

- Our objective is to objectively identify pacing profiles used in World Championship 2000 meter rowing races.
- With objectively defined pacing profiles we can identify in which situation boats exhibit different pacing profiles.
- The literature of pacing profiles in 2000 meter rowing does not agree on which factors affect a boat's pacing profiles or what those pacing profiles are.
- The literature of pacing profiles in sports contains the identification of All Out, Even Pacing (Consistent), J-Shaped and Reverse J-Shaped strategies among others. [1]

Data Collection

The average speed and stroke rate (SR) for each 50 meter split from each boat in every race of the Rowing World Championships from 2014-2017 was downloaded from www.worldrowing.com along with characteristics about the race including competition class and the round of the race. There were 3496 boats (n = 2112 for men, n = 1384 for women) from 657 races used in the analysis. Races had various categories for sizes of boats and weight classes.

Data Pre Processing

To determine pacing profiles raw speeds at each split are often compared to the mean speed of a boat throughout the race []. So we define $x_{i,j}$, as the speed at split i for boat j and normalize to get

$$y_{i,j} = \frac{x_{i,j} - \bar{x}_j}{\sigma_j}$$

Next, we would like to perform Principle Components Analysis (PCA) to reduce the dimensions of our data. We know that a boat's split times at split time 16 and 17 should capture roughly the same information and we would like to have less dimensions when we cluster our data. Before, performing PCA we normalize column wise to get

$$z_{i,j} = \frac{y_{i,j} - \bar{y}_i}{\sigma_i}$$

Finally, we perform PCA on our data set to reduce the dimensionality while still explaining the majority of the variance. The first 7 principal components were selected to use for the K-Means Clustering.

K-Means Clustering

We used a K-means clustering algorithm to cluster the boats into 4 different clusters based on the 7 principal component loadings computed above. Cluster sizes of 3 and 5 were also attempted but provided worse real world interpretations and no significant change in average silhouette size.

Results

Please note all tests for effects were done with Chi-Squared tests with a Bonferroni Correction for the number of tests we were running.

Identifying Pacing Profiles

Similar to the findings in pacing profiles in other sports we find the All-Out (Conservative or Exhaustive), Consistent, J-Shaped, Reverse J-Shaped pacing profiles in 2000 meter rowing. Interestingly, we clustered on the normalized data but find that the Consistent pacing profile has significantly higher average mean speed then the rest of the strategies.

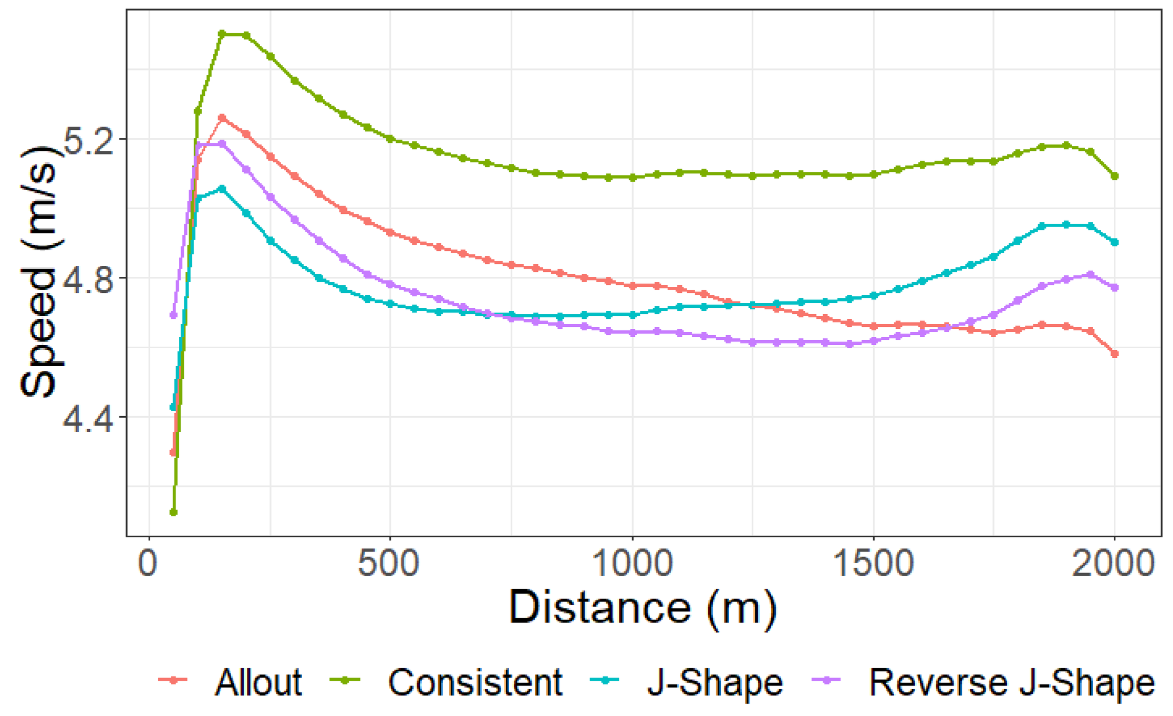

Figure 1: Mean Raw Speeds (non-normalized) at every 50 meter for each cluster

Size of Boats

However, this does not mean that the Consistent strategy is for some reason more efficient. Larger boats have larger mean speeds and so our hypothesis was that the size of boats had an effect on a boat's pacing profile.

We clustered on a normalized boat speed for the race so it is interesting that we pick out size as a factor that affects pacing profile. This is because the strategy of a boat is different when rowing in a group or individually. In larger groups (4s and 8s) there is a bigger need to coordinate rows between teammates and more difficult to change the pace of the boat. This assertion was verified through a chi-squared test.

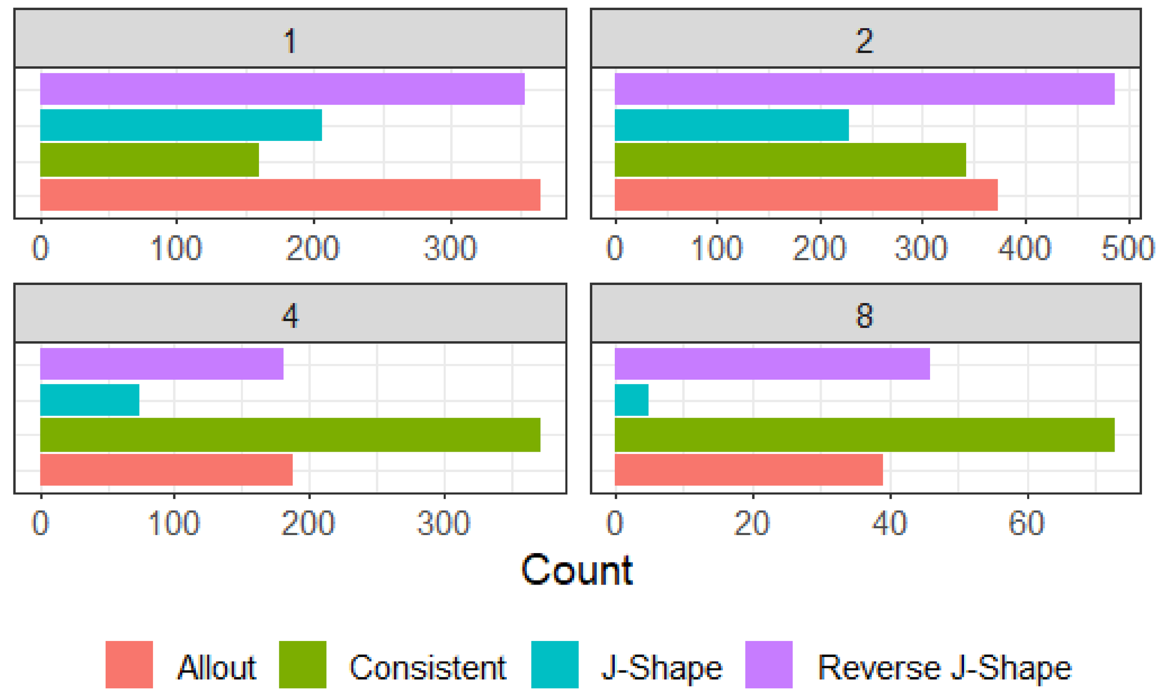

Figure 2: Pacing Profiles for different sized boats in World Championship races.

Heats vs Finals

Next, we looked at whether the race was in a Qualifying Heat or the Final placing race for the competition. The format for the World Championships has a Final (Final A) for placing the winner of the tournament (and top 6 placements) and other Finals races (Final B, Final C...) to place the remaining boats. Any race which is a boat's last race is considered a Final. Depending on the size category we see different trends in changes in pacing profiles from Heats and Finals.

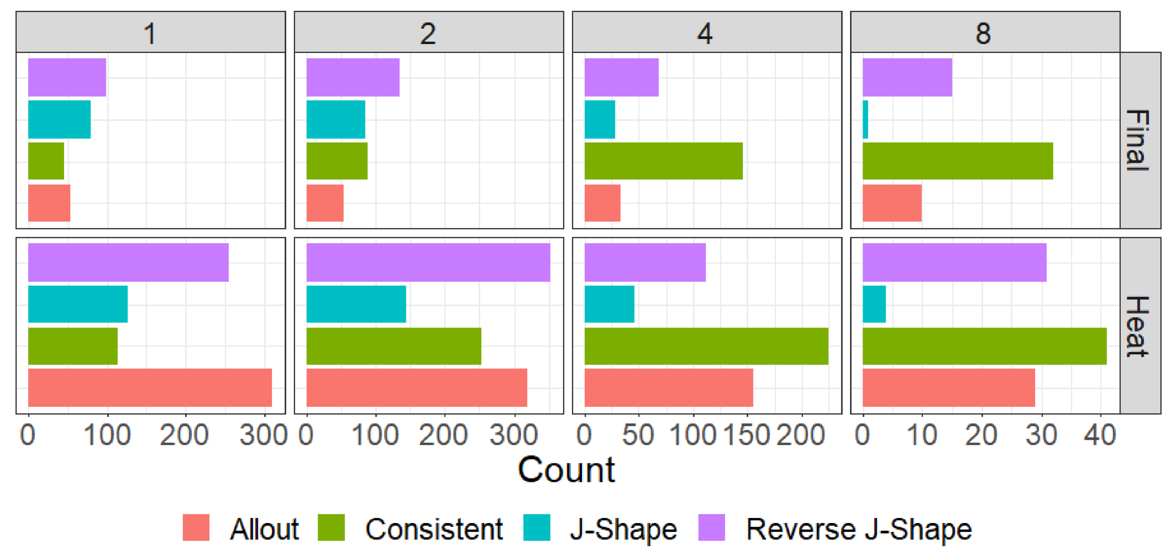

Figure 3: Pacing Profiles by round type accounting for different sized boats in World Championship races.

The All Out (Exhaustive or Conservative) is used more often in Heats compared to Finals across all boat sizes. Specifically for 1 person boats we can also see that the proportion of Reverse J-Shape and J-Shape pacing profiles we observe in Finals is larger compared to Heats. This leads us to a propose a couple of new hypotheses about these 1 person boats.

1. The pacing profile of a boat is affected by their performance during the race
2. Athletes confident that they will qualify for the next round will conserve their energy at the end of the race and exhibit an All Out pacing profile

Race Rankings

Finally, following our hypotheses from the round type test we looked to see if the pacing profile and ranking were related when accounting for round type and boat size. We found that for 1 person boats in Heats that qualified for the next round (typically top 3 finish) had a larger proportions of All Out strategies used. Additionally, in the finals the most common pacing profile we saw was a J-Shape profile in the top 3 and a Reverse J-Shape in the bottom 3. For larger boats we consistently saw the Consistent profile as the most popular across finishing placements and no significant change in distribution.

Gender and Weight Class

We found that after accounting for boat size and round that there was no significant effect of gender or weight class on exhibited pacing profile. Before accounting for these factors there looked like there was an effect as the distribution of the number of boats of different sizes competing in Men's and Women's rowing was quite different.

Conclusions

- We found 4 distinct pacing profiles found commonly in other sports: All Out, Conservative, J Shape and Reverse J Shape
- Size, Round Type and Race Ranking have an effect on the exhibited pacing profile
- Gender and Weight Class did not have a significant effect on pacing profiles

Forthcoming Research

- Clustering using Longitudinal Clustering
- Progression of a boat's pacing profiles throughout a tournament through transition matrices
- Determine optimal pacing profiles to minimize time

References

[1] Abbiss C. and Laursen P. Describing and understanding pacing strategies during athletic competition. 38:239–52, 03 2008.

[2] S. W. Garland. An analysis of the pacing strategy adopted by elite competitors in 2000 m rowing. *British Journal of Sports Medicine*, 39(1):39–42, 2005.

[3] M. D. Kennedy and G. J. Bell. Development of race profiles for the performance of a simulated 2000-m rowing race. *Can J Appl Physiol*, 28(4):536–546, Aug 2003.

[4] Melges T. Muehlbauer, T. Pacing patterns in competitive rowing adopted in different race categories. *The Journal of Strength Conditioning Research*, 25, 2011.

[5] Schindler C. Muehlbauer, T. and Widmer A. Pacing pattern and performance during the 2008 olympic rowing regatta. *European Journal of Sport Science*, 10(5):291–296, 2010.

Acknowledgements

We would like to thank Dr. Dave Clarke for organizing the partnership with CSIP, Chuck Rai for his incredible help with scraping data and Lucas Wu and Kevin Floyd for their consultations on statistics and rowing strategies.