Seat Racing

Christian Lindig

lindig@gmail.com

Draft – May 17, 2020

Seat racing is a procedure used by rowing coaches to find the strongest athletes from a squad for a crew boat. While the general fitness of an athlete can be observed from land training, his or her ability to "move the boat" within a crew is more difficult to quantify. Seat racing is used to bring objectivity and transparency to the selection process. Several procedures exist; a popular one is the Purcer Matrix by Mike Purcer [1]. It works along the following principles:

1. Eight rowers are split repeatedly into two crews of four and race against each other in two boats under controlled conditions for a fixed distance (like 1000m). After each race rowers between boats are swapped according to a fixed plan. The process continues over a total of 6 races.

2. For each race the finishing time for each boat is recorded. The final ranking of the athletes is obtained by ranking them according to the total time each of them raced.

Race	1		2		3	
Seat 1	2	3	4	1	2	1
Seat 2	B	C	A	D	B	A
Seat 3	1	4	2	3	3	4
Seat 4	A	D	C	B	C	D
Time (s)	204.98	204.91	202.49	207.40	202.27	207.62
Race	4		5		6	
Seat 1	1	3	4	3	2	4
Seat 2	C	A	D	C	D	B
Seat 3	2	4	2	1	3	1
Seat 4	D	B	B	A	A	C
Time (s)	206.48	203.41	204.85	205.04	204.93	204.96

Table 1: Six races in coxed fours over 1000m with time in seconds.

Rank	Athlete	Total Time s	Avrg. Time s	Power W
1	2	1226.00	204.33	293.04
2	C	1226.15	204.36	292.93
3	B	1227.87	204.64	291.70
4	3	1227.96	204.66	291.64
5	4	1228.24	204.71	291.44
6	A	1228.47	204.75	291.27
7	D	1236.19	206.03	285.85
8	1	1236.48	206.08	285.65

Table 2: Ranking based on total time. Athlete B raced in 6 races which took in total $1227.87s$. This implies an average time of $204.64s$, an average speed of $v = 1000/t = 4.8766ms^{-1}$ and a power of $P = 2.5 \times v^3 = 291.70W$.

Consider the example in Table 1 of 8 athletes (named 1 to 4 and A to D) who race in two boats 6 races over 1000m each and swap places according to the Purcer Matrix. The resulting ranking is in Table 2 with athlete 2 the fastest before athletes C and B. The data is taken (with minimal modifications discussed below) from an example provided by Mike Purcer [2]. The matrix respects some additional constraints. Most importantly, throughout the races, a rower only uses an even- or odd-numbered seat, which means he or she is always rowing on the same side in a sweep boat. This reflects most rowers preference of rowing on a specific side.

Usually this would be the end of it: an open and transparent process produced a ranking for coach and athletes to use. Athletes might notice that the margins are quite small and not be entirely satisfied despite the openness of the process.

There are good reasons to suspect that the ranking in Table 2 does not reflect the true contribution of each athlete and specifically that the best rower in the squad is actually athlete C by quite some margin. The remainder of this paper discusses what fuels this suspicion and how to find a more plausible ranking.

1 Estimating an Athlete's Power

Power output and boat speed are linked. On a Concept2 rowing machine used for land training the connection is

$$P = 2.8 \times v^3$$

with v being the speed in ms^{-1} and P the power in Watt. In the boat, speed, and the constant (2.8) also depends on the boat class: the same power output per athlete moves a bigger boat faster.

We can use this connection to estimate the power output per athlete based on the results in Table 2. The premise is that power output is a good indicator for an athlete's ability to move the boat. Research by Kleshnev [3] suggests that the factor in a coxed four is closer to 2.5 and we will use that. The exact value is not important because we are mostly interested in the relative power of rowers. Table 2 shows in the last column the estimated power output for each athlete based on their average time t over 1000m, using the formula below:

$$P = 2.5 \times v^3 = 2.5 \times (1000/t)^3$$

Knowing each athlete's power output, we also know how much power was in the boat in each race by taking the sum of the power produced by the athletes in the boat. And this in turn gives us an expected speed (and time) the boat should have taken based on the power in the boat. Table 3 summarises this calculation. For a boat with power P we expect it to go over 1000m in time t:

$$\begin{aligned} v &= (P/(4 \times 2.5))^{1/3} \\ t &= 1000/v \end{aligned}$$

Race	Crew	Power W	Time s		
			Measured	Expected	Diff.
1	12AB	1161.66	204.98	204.95	0.03
2	24AC	1168.68	202.49	204.54	-2.05
3	23BC	1169.30	202.27	204.50	-2.23
4	12CD	1157.47	206.48	205.19	1.29
5	24BD	1162.02	204.85	204.93	-0.08
6	23AD	1161.80	204.93	204.94	-0.01
1	34CD	1161.85	204.91	204.94	-0.03
2	13BD	1154.83	207.40	205.35	2.05
3	14AD	1154.21	207.62	205.39	2.23
4	34AB	1166.05	203.41	204.69	-1.28
5	13AC	1161.49	205.04	204.96	0.08
6	14BC	1161.71	204.96	204.94	0.02

Table 3: Measured time versus expected time based on athlete and crew power. Crew $12CD$ had a combined power of $P = 1157.47W$. Assuming $v = (P/10)^{1/3}$, this leads to an expected race time of $205.19s$ for 1000m, which is $1.29s$ faster than what we observed.

We can see that the difference between the time we can expect based on the power of the crew and the time that we actually observed can be in the order of 2 seconds. This is a lot when the average time between athletes is in the order of tenth of seconds. This suggests that our belief about the power of each athlete might not be correct – and consequently their ranking as well.

The difference between the actual performance of crews and the one to be expected based on their power can be made smaller by assigning a different power to each athlete than the one we learned from an athlete's average so far. Finding these new assignments is an optimisation problem and we present below how to find them. Table 4 shows a new power assignment, resulting in a new ranking, and Table 5 the resulting expectations for each race. You will notice that the ranking changed: previously the top-ranked athletes were $2, C, B, 3$ and now are $C, B, A, 2$.

Rank	Athlete	Total Time s	Avrg. Time s	Power W
1	C	1226.15	204.36	333.44
2	B	1227.87	204.64	318.84
3	A	1228.47	204.75	316.53
4	2	1226.00	204.33	284.43
5	D	1236.19	206.03	284.28
6	3	1227.96	204.66	275.89
7	4	1228.24	204.71	272.85
8	1	1236.48	206.08	238.78

Table 4: Final ranking based on estimated power. Athlete C is the strongest but his or her contribution was not evident in total racing time. It would have taken more races for it to become apparent.

The justification for the power assignment in Table 4 is the now smaller difference between observed and expected race times in Table 5. The power assigned to athletes is still clustered and this leaves some doubt that the ranking is indeed correct. Power difference of less than 10 Watt are probably not meaningful and this would tie athletes B/A, $2/D$, and $3/4$. More racing would be required to learn their true power.

2 Discussion

The original ranking in Table 2 is based on total race time (or average race time, which is equivalent) per athlete. The problem of this method is that the order it produces only converges slowly: after the first race, all athletes in the same boat have the same average (and total) time. With each more

Race	Crew	Power W	Time s		
			Measured	Expected	Diff.
1	12AB	1158.58	204.98	205.13	-0.15
2	24AC	1207.26	202.49	202.33	0.16
3	23BC	1212.60	202.27	202.04	0.23
4	12CD	1140.93	206.48	206.18	0.30
5	24BD	1160.41	204.85	205.02	-0.17
6	23AD	1161.13	204.93	204.98	-0.05
1	34CD	1166.46	204.91	204.66	0.25
2	13BD	1117.79	207.40	207.59	-0.19
3	14AD	1112.45	207.62	207.93	-0.31
4	34AB	1184.11	203.41	203.64	-0.23
5	13AC	1164.64	205.04	204.77	0.27
6	14BC	1163.92	204.96	204.81	0.15

Table 5: Measured time versus expected time based on athlete and crew power. Athlete power is assigned such that it minimises the difference between measured and expected race time.

race, more information is added and the average race time for each athletes drifts towards his or her long-time average. The slow convergence hurts the best athletes the most because their average is most different from the global average and it takes the most races to converge to it. So it is not a coincidence that athlete C was not at the top originally.

The problem is less pronounced when seat racing is done for fewer athletes while still using 6 races – for example ranking 4 athletes racing in pairs. Simply racing more often is usually not an option. Athletes are getting tired and conditions change, which can make comparing races more difficult.

The method based on power gains its strength from these ingredients:

1. Power is considered a measurement for a rower's and a crew's ability to move a boat.

2. Power, speed, and time are linked via $P = c \times v^3$, which provides the underlying statistical model. It is based on knowledge that the original method did not leverage.

3. The model enables to predict race times based on an assignment of power to rowers. This assignment is modified such that the difference between observed and predicted race time is minimised.

As demonstrated by an example, statistical modelling offers a more plausible and hence fairer ranking based on the same amount of information than the original method. We have to suspect that seat racing in the past often not selected the best athletes.

3 Finding the Power Assignment

The power assignment presented in Table 4 obviously improves the difference between observed and expected race times. To explain how it was found we rely on matrix notation.

Rowers are enumerated $1, \ldots, 8$ and races are enumerated $1, \ldots, 12$. Previously a race had two boats but here we are using a race to mean a timed event for a crew. The relationship between rowers and races is captured by a 8×12 matrix C with

$$c_{ij} = \begin{cases} 1, & \text{rower } i \text{ is in crew for race } j \\ 0, & \text{otherwise} \end{cases}$$

The power assigned to rowers is captured by a vector R with r_i denoting the power (in Watt) assigned to rower i. The times observed for races are captured by a vector T with t_j the times in seconds for race j. We now have:

1. $P = R \times C$ is a vector p_j denoting the assumed power of the crew in race j.
2. V with $v_j = (p_j/10)^{1/3}$ is the expected speed of the crew in race j.
3. T' with $t'_j = 1000/v_j$ is the expected time for race j.

The error between observed and expected time is sum of the squared differences.

$$e = \sum_{j=1}^{12} (t_j - t'_j)^2$$

An iterative optimisation algorithm is used to minimize e by assigning values to R. R needs initial values and starting with power values known from land training would work but also using a reasonable guess like 250 Watt would also work. The data in this paper was computed using the *optim* method in R [4].

4 Accounting for Wind and other Differences

It would be unusual to expect that all 6 races can be carried out in the exact same conditions. Wind and stream conditions change and athletes are getting tired as racing progresses. However, the race matrix makes sure that all athletes are at least experiencing the same conditions in each pairwise race. Since we are most interested in the ranking and relative performance of athletes and less in absolute numbers I am proposing the following method to normalise race times prior to processing them:

1. Compute the average race $\bar{t}$ time over all 12 captured times.

2. With two boats racing against each other and results t_1 and t_2, normalise them to t'_1 and t'_2

$$
\begin{aligned}
k &= 2 \times \bar{t}/(t_1 + t_2) \quad (1)\\
t'_1 &= k \times t_1 \quad (2)\\
t'_2 &= k \times t_2 \quad (3)
\end{aligned}
$$

such that $(t_1 + t_2)/2 = \bar{t}$ – the average of the race time is equal to the global average.

This normalisation was applied to the data from Mike Purcer [2] as detailed in Table 6.

	Time s	
Race	Raw	Normalised
1	211.46	204.91
1	211.54	204.98
2	199.46	202.49
2	204.30	207.40
3	200.97	202.27
3	206.29	207.62
4	199.47	203.41
4	202.49	206.48
5	205.98	205.04
5	205.79	204.85
6	205.81	204.96
6	205.78	204.93

Table 6: Normalisation of raw time data such that the average race time of each race equals the global race time average. This maintains relative performance but accounts for changing conditions like wind, stream, or the racing distance.

References

[1] Mike Purcer. Seat Racing Fours. In *NOTES ON ROWING (September 23, 2010)*. https://purcerverance.ca/files

[2] Mike Purcer. Seat Racing Analysis. Excel Spreadsheet. Retrieved May 16. 2020. https://purcerverance.ca/files

[3] Valery Kleshnev. Volume 7 No 76 Rowing Biomechanics Newsletter July 2007. www.biorow.com

[4] R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/