Modelling how should a swimmer perform the 1500m race based on Sun Yang's 1500m World Record

Andro Asatashvili

2019

1 Rationale

I chose this topic because I've done competitive swimming since I was 8 years old. It has always been fascinating and enjoyable, but I always struggled on completing the 1500 metres freestyle race. This is due to the long duration of the trial, as well as the several conditions surrounding it, such as managing fatigue effectively or adapting stroke speed depending on how much distance the swimmer has swam. However, there are several opinions regarding how the 1500m race should be performed, and personally, I've tried some of them (such as saving energy for the last 200m or manage a consistent pace for the duration of the race) with a non- substantial difference. As well, arm span is said to be the determining factor in a swimmer's efficiency. Hence, I am interested in using mathematics to create a model that can determine what is the ideal method for swimming the 1500m race and a function that can determine how many strokes should be performed during the race. I'll use my own measurements and Olympic 1500m medallist Sun Yang's, to compare how ideally both Yang and myself should swim the 1500m race. Moreover, we need to highlight the fact that the model won't be exact, because several variables won't be included (such as surface area of palm, drag, age and experience). The model will aim to be accurate to the extent of how the race should be done in accordance to arm length, as a result, giving a guideline on how a swimmer should swim the respective race.

2 Aim

The aim of this investigation is to construct two models, which reflect how the 1500m swimming race should be performed globally and how many strokes should be done per 50m in order to create an ideal pace based on a swimmer's anthropometry. This will hopefully enable swimmers to compare their meet results with the model, in order to identify performance weaknesses and improve.

3 Introduction

The investigation is divided in two parts. In order to create the first model, the investigation will first focus on constructing a graphical representation of Sun Yang's 1500m race in the 2012 Olympic games, using differential calculus. I'll first analyse my 2018 1500m race to find out how I swam the race, and then compare it with Yang's to see my weaknesses and areas of improvement in regards with structuring the race. Differential calculus is useful in the construction in the model as it provides information like maximum, minimum and inflexion points, which are necessary in order to know when should a swimmer 'stop' swimming at a certain rate and adapt. As Yang's time is the current world record as for 2019, we can assume that the model will represent the most efficient way ever of swimming the 1500m race. Moreover, its important to mention that both sections of the investigation are adapted to a 50m swimming pool. The models won't work on 25m pools, due to more laps and underwater time in every turn. It's relatively easier to swim in a 25m pool, rather than a 50m one.

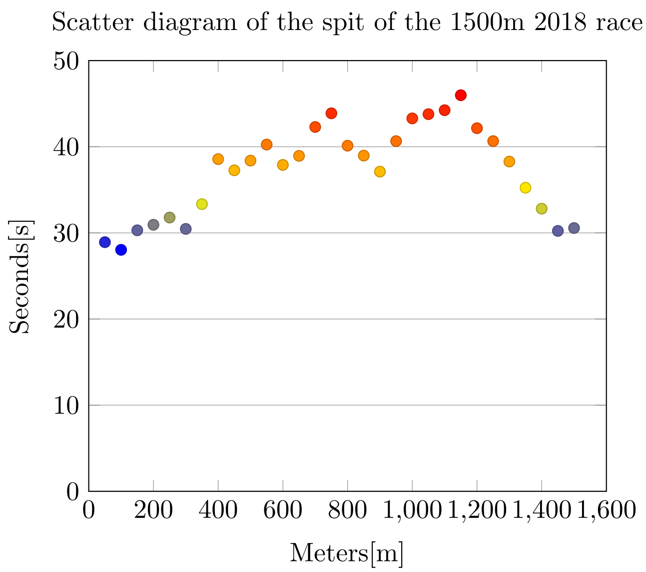
4 Model of my own 1500m race

I consider valuable to first construct a model that shows my 1500m race. Even though I'm an average swimmer in the race, my time of the 1500m is not that bad. My best time ever, which was done in 2018, lacked strategy. This means that I swam the trial in an 'ugly' fashion and without preparation, unlike the majority of long- distance swimmers. However, I still managed to be inside the first 10 swimmers in the 18+ category that year. Here are the values:

Meters (x)	**Split (y)**	**Time**
50	28.92	0:28.92
100	28.03	0:56.95
150	30.30	01:27.25
200	30.94	01:58.19
250	31.78	02:29.97
300	30.46	03:0.43
350	33.34	03:33.77
400	38.56	04:12.33
450	37.26	04:49.59
500	38.39	05:27.98
550	40.25	06:08.23
600	37.89	06:46.12
650	38.94	07:25.06
700	42.29	08:07.35
750	43.88	08:51.23
800	40.12	09:31.35
850	38.97	10:10.32
900	37.11	10:47.43
950	40.65	11:28.08
1000	43.29	12:11.37
1050	43.78	12:55.15
1100	44.24	13:39.39
1150	45.98	14:25.37
1200	42.15	15:07.52
1250	40.65	15:48.17
1300	38.28	16:26.45
1350	35.24	17:01.69
1400	32.81	17:34.50
1450	30.23	18:04.73
1500	30.56	18:35.29

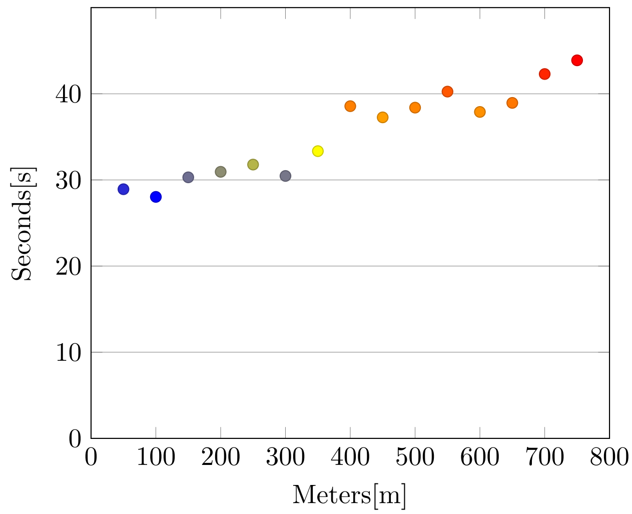
Table 1: My 1500m split time

I'll now graph the split times with the meters as the independent variable and the split times as the dependent variable:



After seeing the resulting plotting of my values, I saw that my race evidently lacked structure. For example, approximately from meters 800 to 1100, the race was a mess. There's no visible attempt into controlling the split times in said interval. However, although I knew that my swimming was not planned during the race, it is still interesting to see how there's a tendency to increase split time from the first 50m to the middle of the race (approximately at the 750m). Moreover, my 'closing time' starts at the 1150m. In swimming, the 'closing time' is where one starts to swim as fast as possible in order to completely burn out and have a better race time. However, in long races such as the 800m and the 1500m, the closing time starts several laps before the end. This value is highly dependent on stamina. Although I swam this 1500m race unconsciously (without being aware of following a structure), I'm pleased to see that my closing time was incredibly steady and even followed a strong tendency to decrease since the 1150m. Therefore, I'm interested to see what mathematics says about this. I'll construct a regression line model to analyse the race in the most important parts of the 1500m according to my coach: the first increase of time and the final decrease of time. In other words, from split 50m, to the 750m, and from split 1150m to 1500m.

Scatter diagram of the spit of the 1500m 2018 race $x : 50 < x \leq 750$



To see how the correlation of the split increases in time in regards with the distance of the trial, we'll use a regression line model. This will tell us if there was a 'natural increase' in this part of the swim, which is expected to be linear and steady. With this I mean that the line will serve as a reference to tell if the initial part of the 1500m (in my case) was well executed: the resulting regression line has the aim of showing at what rate was I increasing my split time in regards to distance. In my opinion, the dispersion was determined by a natural need I felt to preserve stamina.

Regression line $x : 50 < x \leq 750$

We know that:

$$y = mx + b$$

$$y - \bar{y} = \frac{Sxy}{Sx^2}(x - \bar{x})$$

Where:

$$\bar{x} = \frac{\Sigma x}{n}$$

$$\bar{y} = \frac{\Sigma y}{n}$$

$$Sx^2 = \sqrt{\frac{\Sigma(x - \bar{x}^2}{n}}$$

$$Sxy = \Sigma\frac{(x-\bar{x})(y-\bar{y})}{n}$$

Hence:

$$\bar{x} = \frac{\Sigma x}{n} = \frac{6000}{15} = 400\ meters$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{531.23}{15} = 35.41\ seconds$$

$$Sx^2 = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{\Sigma(x-400)^2}{15}} = \sqrt{\frac{7.35*10^{12}}{15}} = \sqrt{4.9*10^{11}} = 700000$$

$$Sxy = \Sigma\frac{(x-\bar{x})(y-\bar{y})}{n} = \Sigma\frac{(x-400)(y-35.41)}{15} = \frac{229853}{15} = 15323.50$$

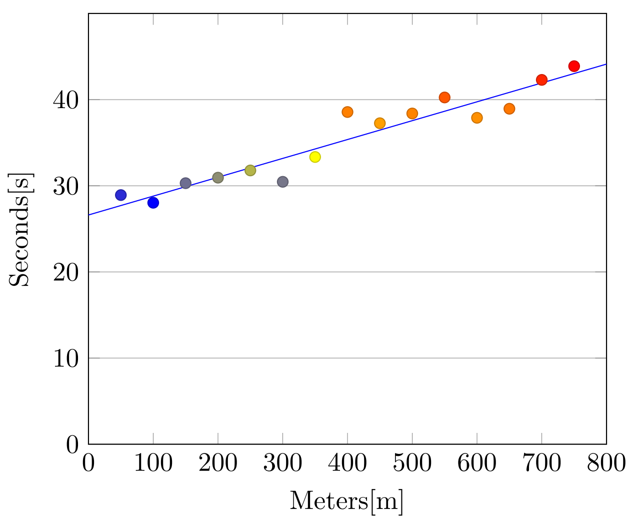
$$y - 35.41 = \frac{15323.5}{700000}(x-400)$$

$$y - 35.41 = 0.022(x-400)$$

$$y - 35.41 = 0.022x - 8.80$$

$$y = 0.022x - 26.61$$

Regression line of the spit of the 1500m 2018 race $x : 50 < x \leq 750$



According to the regression line, for every meter that is swam in the first 750m, the time of the split increases 0.022 seconds. As well, the correlation appears to be strong positive. With the Pearson correlation coefficient, we can find out exactly how strong the correlation between meters and increasing split time is.

Pearson coefficient correlation

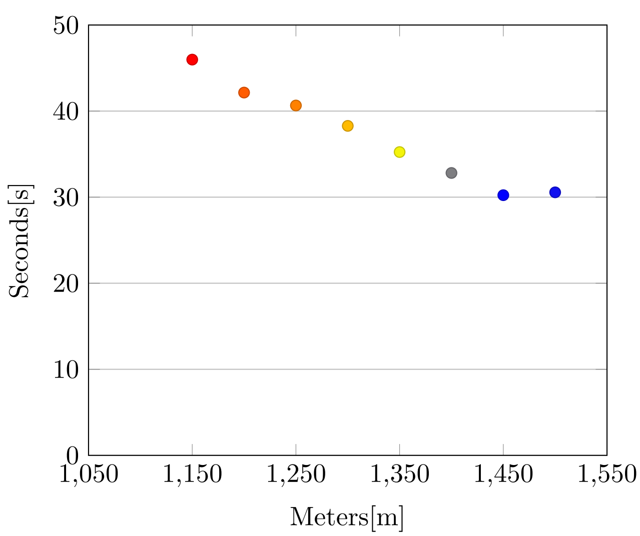
$$\mathbf{r} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$

Following the previous calculations for the regression line, the TI-NSPIRE CX gives the value of $\mathbf{r}$.

$$\mathbf{r} = 0.96$$

This value means that it's a very strong positive correlation. For the value to be such, means that for the first 750m, my race was actually very good. There was a steady increase in time, allowing me to maintain a pace that it's not only good for the race, but also in a psychological aspect. To have such a strong beginning can only be beneficial to my confidence in the 1500m. This first representation of the start led me to believe that work is needed elsewhere, perhaps at the middle of the race. I'll now repeat the calculation process for $\{x : 1150 < x \leq 1500\}$.

Scatter diagram of the spit of the 1500m 2018 race $\{x : 1150 < x \leq 1500\}$

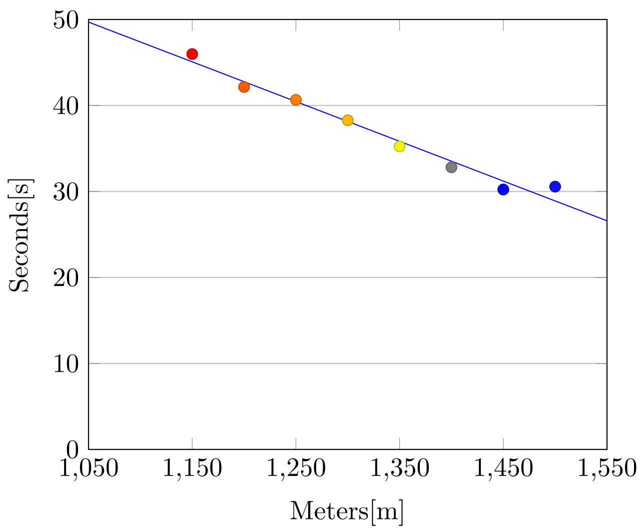


Regression line $\{x : 1150 < x \leq 1500\}$

$$y - \bar{y} = \frac{Sxy}{Sx^2}(x - \bar{x})$$

$$y = -0.046x + 98.22$$

Scatter diagram of the spit of the 1500m 2018 race $\{x : 1150 < x \leq 1500\}$



For every meter that is swam, the time of the split decreases 0.046 seconds. The correlation, again, appears to be strong in the final meters of the race.

Pearson coefficient correlation

$$\mathbf{r} = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{\Sigma(x - \bar{x})^2}\sqrt{\Sigma(y - \bar{y})^2}}$$

$$\mathbf{r} = -0.99$$

In comparison to the first 750 m of the race, the closing 350 meters were even better in performance. The Pearson correlation coefficient (**r**) is even closer to ± 1 than the start of the race, which indicates me that I closed the 1500m race better than I started it. The value of **r** signals that the race was optimal because the stronger it is, it demonstrates how I had control over the splits. According to my coach, to have control over the race means that a swimmer can adjust easier to a race, depending on fatigue and stamina reserves. This is important, because this highlights the capacity I have to exert more pressure in the closing meters.This means that in order to improve the race, my problem doesn't necessarily rely in working on my stamina and endurance, but structure of the race.This model tells me that its imperative for me to have a strategy when I swim.

For example, this problem is evident in the middle of the race. There's a lot of variability in split times, unlike the first 750m and last 350m (that have structure as we've seen). We can use the standard deviation to mathematically see how much variability was in my 1500m race as a whole.

Standard Deviation

$$\sigma = \sqrt{\frac{\Sigma(y - \bar{y}}{n}}$$

$$\sigma = \sqrt{\frac{\Sigma(y - 35.41)}{30}}$$

$$\sigma = 5.36$$

As we can see, the standard deviation (σ) is far away from the smallest possible number (0). This means that there was variability in the split times, which although is normal and to be expected, hinders my progress in this race.

Therefore, by analysing my 1500m race with mathematics, I can get some conclusions and ideas in what should I work in for my next 1500m race. As shown by the regression lines, the problem does not rely on my capacity to open and close the race. I do not require to work on these areas, as I proved to be able to create and steadily increase a pace. I was also able to decrease steadily the pace even when fatigued at the end of the race. Hence, I believe the problem relies in other areas: first, the middle part of the race does not follow a structure like the beginning/closing meters. I was unable to generate a regression line, which makes me question if there's a strategy effective enough to mirror. Moreover, as evidenced by the standard deviation (σ), there's a lot of variability in my split times. The structure then also needs to be constructed around this fact: I cannot afford to have variability in my split times due to loss of time. Energy is wasted and the closing meters become unnecessarily difficult. Having a lower standard deviation is extremely important.

A model based on the 1500m Olympic champion is extremely useful. The aim of the next section is to create a model of Sun Yang's 1500m race. Hopefully, Sun Yang's 1500m will have a visible structure that I can mirror and compare my race with. As well, the mean of Yang's splits needs to become a variable to be taken into consideration for the model. The mean of Yang's split will be useful in the construction of both the model of the raced and the construction of the Sine model.

5 Model of Sun Yang's 1500m World Record

Meters (x)	**Split (y)**	**Time**
50	27.09	27.09
100	28.71	28.71
150	29.46	01:25.3
200	29.05	01:54.3
250	29.35	02:23.7
300	28.97	02:52.6
350	29.53	03:22.2
400	29.34	03:51.5
450	29.23	04:20.7
500	28.89	04:49.6
550	29.26	05:18.9
600	29.27	05:48.2
650	29.25	06:17.4
700	29.34	06:46.7
750	29.41	07:16.2
800	29.3	07:45.5
850	29.49	08:14.9
900	29.38	08:44.3
950	29.46	09:13.8
1000	29.32	09:43.1
1050	29.42	10:12.5
1100	29.21	10:41.7
1150	29.54	11:11.3
1200	29.37	11:40.6
1250	29.17	12:09.8
1300	29.19	12:39.0
1350	29.39	13:08.4
1400	29.14	13:37.5
1450	27.81	14:05.3
1500	25.68	14:31.02

Table 2: Sun Yang's World Record Split Times

Using You Tube's "Sun Yang Smashes Men's 1500m Freestyle World Record", we'll see what are Sun Yang's 1500m race stats, that if we want them to be mirrored, we'll need to make a reasonable comparison of these with our own. The mean and the standard deviation will show us both what is the average swimming split for a 50m split during the duration of the race, and the standard deviation how much variation there is from the mean (which gives us an idea for the distribution of the data set).

5.1 Mean(Split)

$$\bar{x} = \frac{\Sigma x}{n} = \frac{871.02}{30} = 29.03 \; seconds$$

This means that the average 50m split of Yang's 1500m covers 1 meter approximately each 0.58 seconds ($\frac{\bar{x}}{50}$). On the other hand, I cover 1 meter every 0.71 seconds. Unlike a 50m race (which in César Cielo's 2009 world record ideally covers 1 meter in approximately 0.42 seconds), a 1500 race should logically be performed at slower rate.

Even though these are different swimmers, we can assume that the average 50m split of a 1500m race should be performed approximately 38.85% slower than a 50m race ($\frac{29.03*100}{20.91} - 100$). This surprised me, because it's much faster than I thought. I've always attempted to swim the 1500m much slower: this shows me that I have to be faster throughout the race.

5.2 Standard Deviation (Split)

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x-\mu)^2}{n}} = \sqrt{\frac{19.095}{30}} = \sqrt{0.64} = 0.80$$

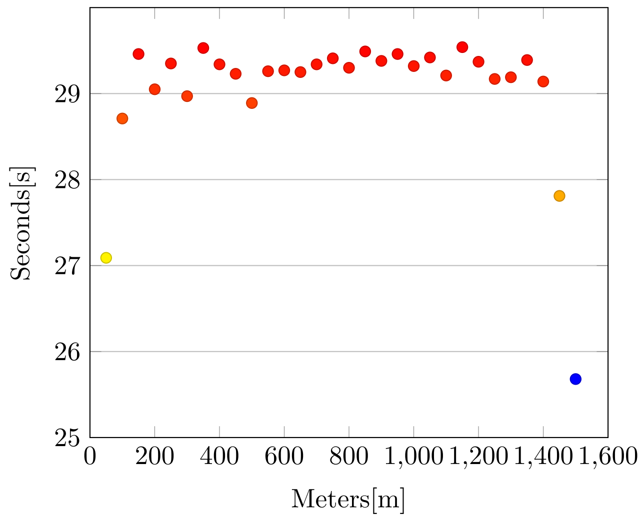
We can interpret that there is a low dispersion of the data set because of σ. In comparison, the standard deviation of my race was 5.36. Yang's standard deviation is vastly inferior to mine. This means that throughout the duration of the race, Yang's pace ($\bar{x}$) was kept extremely similar, which indicates us that it's necessary to keep the 50m Splits relatively the same.

Hence, statistically speaking, we can conclude that a 1500m race should keep:

- A slower time of approximately 38.85% in relation with our personal 50m race.
- The stroke rate regular throughput the race, in order to mantain a similar split time of the whole race.

I'll now graph Yang's data:

Scatter diagram of the spit of Sun Yang's 1500m 2012 race



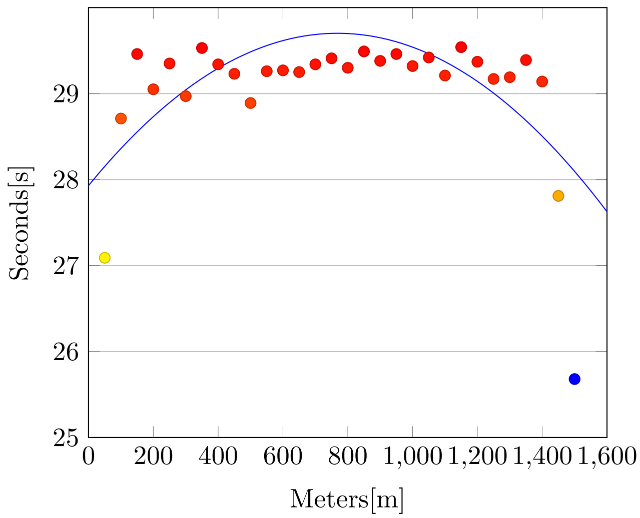
Yang's scatter diagram is very different from my own. It appears that the only thing in common is the closing time strategy, although it appears he starts decreasing his time at the very last 200m. As well, the whole race appears to be static: the opening time ends approximately at the 150m. Instead, Yang appears to have the same split time for all the race. We then need to construct a regression function that can model Yang's 1500 race in order to visualize what should the strategy for swimming be. The function will help both visually and mathematically because a swimmer could emulate the function's behaviour into his own trial. It appears that there are no linear increases/decreases in time, which means that Yang used a method different to what I did in my 1500m. We'll use a quadratic regression model, as we can see by eye that it can fit the dispersion of the points.

$$y = ax^2 + bx + c$$

For finding the coefficients, we'll utilize the TI-Nspire CX calculator. The values are automatically given when selecting the regression quadratic equation for the data of Yang's split:

$$y = -0.000003x^2 + 0.00461x + 27.93$$

Scatter diagram of the spit of Sun Yang's 1500m 2012 race



The quadratic regression implies that emulating a negative parabola is the strategy for performing the 1500m. This would mean that until the maximum point of the parabola is reached, we should increase our time and then decrease it at a steady rate. Thus, we need to find the maximum point for the model in order to estimate at what meter should we start decreasing our split time.

We know that in order to derivate:

$$f(x) = x^n \longrightarrow f^{'}(x) = nx^{n-1}$$

Hence:

$$f(x) = -0.000003x^2 + 0.00461x + 27.93$$

$$f^{'}(x) = -0.000006x + 0.00461$$

$$0 = -0.000006x + 0.00461$$

$$-0.00461 = -0.000006x$$

$$x = \frac{-0.00461}{-0.000006}$$

$$x = 768.333$$

$$\therefore$$

$$f(768.333) = -0.000003(768.333)^2 + 0.00461(768.333) + 27.93$$

$$f(768.333) = 29.70$$

$$\therefore$$

$$\text{Maximum Point} = (768.333, 29.70)$$

Approximately at the 800m (split 16), we should be aiming to start decreasing our split times. Specifically, for Sun Yang, the expected split time of 29.70 seconds would equal the highest possible split allowed in order to maintain the quadratic model. This is 0.7 seconds above $\bar{x}$.

Even though the quadratic model proves to be valuable, from swimming experience I can say that it's extremely difficult to maintain a decrease in split times in that way, more so if it's at the 800m. As seen in my 1500m race, after the first 750m, my race lost the correlation I had in the start. If I was to steadily decrease my time since the 800m all the way to the last meters of the race, I would require incredible stamina. I don't think that it would be smart, considering that that in my 1500m race, the middle part of the race was a mess. I would consider this model, at least for my swimming level, too hard and unrealistic. This leads me to believe that a second model is needed. A quartic one seems fit for the given dispersion of the data, because by eye I can see a slight curvature in the 500m. However, as the calculation gets increasingly longer when **n** increases, we'll use an online LaGrange interpolating polynomial calculator in order to obtain the values of the coefficients.[1] The online calculator gives us the following function for Yang's splits:

$$f(x) = -1.84109 * 10^{-11}x^4 + 5.41975 * 10^{-8}x^3 - 0.000054x^2 + 0.020938x + 26.6905$$

We'll now need the coordinates of the inflexion points visible in the graph to know at what meter should we aim to increase/decrease split time:

$$f(x) = -1.84109 * 10^{-11}x^4 + 5.41975 * 10^{-8}x^3 - 0.000054x^2 + 0.020938x + 26.6905$$

$$f^{'}(x) = -7.36436 * 10^{-11}x^3 + 1.62593 * 10^{-7}x^2 - 0.000108x0.020983$$

$$0 = -7.36436 * 10^{-11}x^3 + 1.62593 * 10^{-7}x^2 - 0.000108x + 0.020983$$

[1] https://www.dcode.fr/lagrange-interpolating-polynomial

The TI-NSPIRE CX gives the following roots for this polynomial; for simplicity, in $f(x), x_{1-3} \in \mathbb{N}$

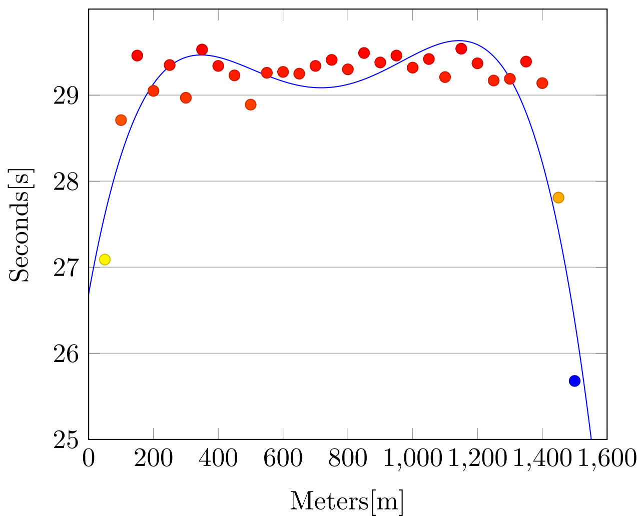
$$x_1 = 347.35 \longrightarrow f_2(x_1) = 29.5$$
$$x_2 = 717.94 \longrightarrow f_2(x_2) = 29.1$$
$$x_3 = 1142.54 \longrightarrow f_2(x_3) = 29.6$$

The second derivative of $f(x)$ allows us to determine which of x_{1-3} are minimum or maximum points:

$$f^{''}(x_1) = -2.20931 * 10^{-10} x_1^2 + 3.25186 * 10^{-7} x_1 - 0.000108$$
$$f^{''}(x_2) = -2.20931 * 10^{-10} x_2^2 + 3.25186 * 10^{-7} x_2 - 0.000108$$
$$f^{''}(x_3) = -2.20931 * 10^{-10} x_3^2 + 3.25186 * 10^{-7} x_3 - 0.000108$$

$$f^{''}(x_1) = -0.000022 \ \therefore \ Maximum \longrightarrow (347, 29.50)$$
$$f^{''}(x_2) = 0.000012 \ \therefore \ Minimum \longrightarrow (718, 29.10)$$
$$f^{''}(x_3) = -0.000025 \ \therefore \ Maximum \longrightarrow (1143, 29.60)$$

Quartic regression line of Sun Yang's 1500m 2012 race



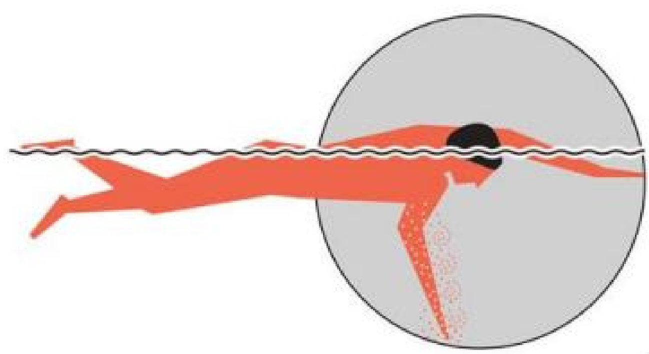
The quartic model shows us how was Sun Yang's split management throughout the 1500m race more accurately than in the quadratic one, because it is more complex. This is impressive, as we can see how Yang's strategy is to start the first 350m relatively fast. As well, when $350 \leq x \leq 1150$, there's a visible increase in split time, probably due to restraining fatigue from the 'closing' meters and allowing Yang to decrease split time. Unlike the quadratic model, the quartic one highlights how there's a tendency for slightly decreasing split time in the middle part of the race. This means that rather than generally keeping the same pace at all splits and decreasing from the 800m, we <u>must</u> have variation in our split times in 4 stages. Clearly, this is an excellent strategy to be applied in the 1500m, as there's chance of conserving stamina as much as possible for the closing meters. However, this idea of variation in 4 stages is very theoretical and difficult: the variation would occur at a hundredth of a second. It's too small to note a change.

6 Modelling Sun Yang's 1500m technique with a Sine function

Now that we have a function for Sun Yang's splits in the 1500m race, its necessary to construct a function that can model the swimming component as well. As we've seen, the 'secret' for swimming the 1500m race relies in split times. In total, there are 30 split times in the race. Opposite to what I did in my race, Yang's splits are ridiculously similar: we need to know what was his pace (based on the number of strokes performed every 50m). Yang is able to swim the 1500m in such a fashion because of his training and primarily his anthropometry (physical characteristics specific to an individual). Therefore, it can t be applied straight away for all swimmers (myself for example). In swimming, it is said that a constant number of strokes per 50m is essential for maintaining a split. Because of the nature of a constant stroke in swimming, I determined that a Sine function would be appropriate to show the technique. We'll be working with Yang's anthropometrics; the input values of the Sine model will be variables that can be adapted to all swimmers. Therefore, the model will be constructed after Yang's technique, and then adapted to my individual anthropometrics to test the accuracy of the model.

As we saw in the first segment of the investigation, in theory, all 50m splits should be kept relatively the same. Following this conclusion, the next values will be taken directly from Yang's world record video in 2012, assuming that they're the same for all the 1500m.[2] Using the following values, we can construct an initial Sine function following the next template:

$$y = aSin(b(x) + c) + d$$



The independent variables are the following: Yang's arm is the determining factor in the construction of the graph, because depending the length of the arm, there's a higher pulling capacity in water. Yang's wingspan (**ws**) is of 2.03m, meaning that an arm's reach is of approximately 1.015m. Now, because

[2]https://www.youtube.com/watch?v=T5FlDy3YmDQ

each positive part of the Sine function should represent only 1 arm's movement, 1 revolution should be equal to 1.90, as it's the time in which Sun Yang's stroke cycle is completed (**sc**). Moreover, each 50m, Yang is underwater for approximately 1.97 seconds (**ut**) at the start of the lap. This means that for each 50m there is a horizontal shift.

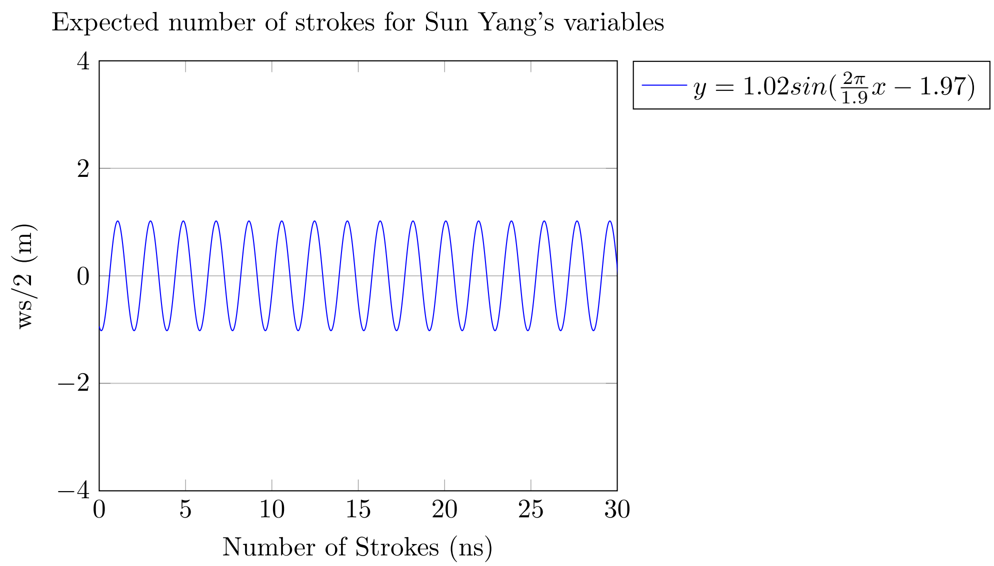
The dependent variables are: the optimal time (**ot**). As seen in the statistical evaluation we did of Yang's splits, the mean of our splits should be 38.85% seconds more than our 50m race time. This value is important because it defines the ideal domain in which the number of strokes per 50m (**ns**) is. For me , this 'ideal mean of the splits' would be 36.10 seconds ($\frac{26seconds*138.85}{100}$). This is surprisingly higher than my actual mean of the 1500m race (35.41 seconds).

Independent Variables

- **ws** - Wingspan
- **sc** - Stroke Cycle
- **ut** - Underwater Time (at the start of each 50m)

Dependent Variables

- **ns** - Number of strokes per 50m
- **ot** - Stroke rate



Each crest represents 1 stroke.

For $0 < x \leq 29$ (seconds), the Sine function predicts that for Yang's arm length, stroke cycle time and underwater time, there will be an approximate of

28 strokes per 50m (**ns**). In the model, one revolution equals to 1 stroke cycle (or 2 strokes). This is accurate, because Yang does exactly 28 strokes in several splits of his race. The Swimming stroke function would then be:

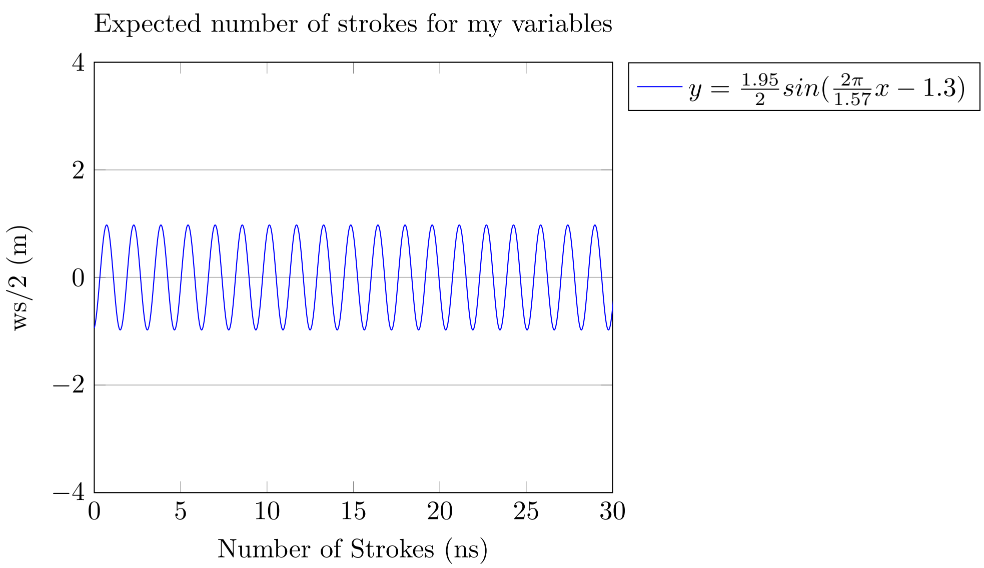
$$ns = \frac{ws}{2} sin(\frac{2\pi}{sc}x - ut);\ \{x : 0 < x \leq ot\}$$

We can then create an estimate number of strokes during a 50m split in a 1500m race for any swimmer that has the above-mentioned personal values. This would serve as a constant guideline of how the swimming pace should be maintained. Nevertheless, this model can variate in several splits depending on small changes in the variables. For example, in my experience, the value of **ut**is the most prone to variate excessively. This is bad in swimming, but fatigue plays a huge role on oxygen demand and supply. This means that the value of ns is ideal, and swimmers should aim to have fewer strokes than **ns**. In other words, **ns** is the maximum limit for strokes in a 50m split of 1500m race. Now I'll compare my average competition variables with Yang's.

Values	Sun Yang (approximate)	Myself
ws	2.03 meters	1.95 meters
sc	1.90 seconds	1.57 seconds
ut	1.97 seconds	1.30 seconds
ot $(\frac{split*138.85}{100})$	29.03 seconds	36.10 seconds

Table 3: Comparison of Yang's variables with mine

The following graph has my personal values in the **ns** equation:



ns indicates that I should perform 45 strokes per 50m to maintain the pace. This is a number that I repeated in some splits of my race. Nevertheless, I found that I had fewer strokes on average (in the majority of splits between 35-38 strokes). Therefore, 45 strokes are my limit for a 50m split. As well, there's an evident difference of swimming styles based on anthropometry. For Yang to perform 17 strokes less than mine not only permits him to have a more efficient stroke (as there's less energy involved), but also, it means that his stroke is significantly stronger than mine. To further highlight how different is Yang's 1500m race comparted to mine, we can integrate both functions to see how much area is covered (**ac**) respectively. With this, we can find how much area does Yang cover with his stroke: comparing this to my own values will represent the difference of power behind our swimming techniques. I believe that this is important, as it further emphasises what should I work in my 1500m race. Perhaps having a stronger stroke (to try to equal my **ac** with Yang's) allows me to reduce my time in the race.

6.1 Area covered by stroke per 50 m

We know that in order to integrate: $\int sin(ax+b) = -\frac{1}{a}cos(ax+b)+c$

Hence:

$$ac = \int \frac{ws}{2} sin(\frac{2\pi}{sc}x - ut)dx$$

$$ac = \frac{ws}{2} \int sin(\frac{2\pi}{sc}x - ut)dx$$

$$ac = \frac{ws}{2} \times \int sin(u) \times (\frac{1}{\frac{2\pi}{sc}})du$$

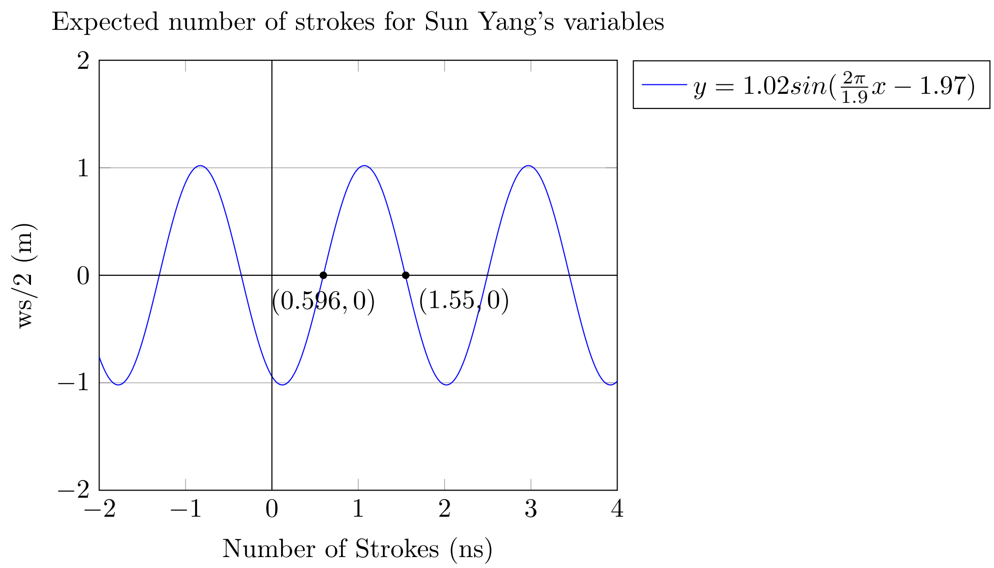
$$ac = \frac{ws}{2} \times (\frac{1}{\frac{2\pi}{sc}}) \times \int sin(u)du$$

$$ac = \frac{ws}{2} \times (\frac{1}{\frac{2\pi}{sc}}) \times -cos(u)$$

$$ac = \frac{ws}{2} \times (\frac{1}{\frac{2\pi}{sc}}) \times -cos(\frac{2\pi}{sc}x - ut)$$

$$ac = -[\frac{ws}{2} \times (\frac{1}{\frac{2\pi}{sc}})] \times cos(\frac{2\pi}{sc}x - ut) + c$$

Where the limits of integration are the zeroes of the Sine function. Therefore, **ac** needs to be multiplied by ns in order to know how much area is covered by strokes in 50m. I'll use the TI-Nspire CX to obtain the values for the first two zeroes of the Sine function and initiate the integration process:



$$ac = \int_{0.596}^{1.55} f(x)dx$$

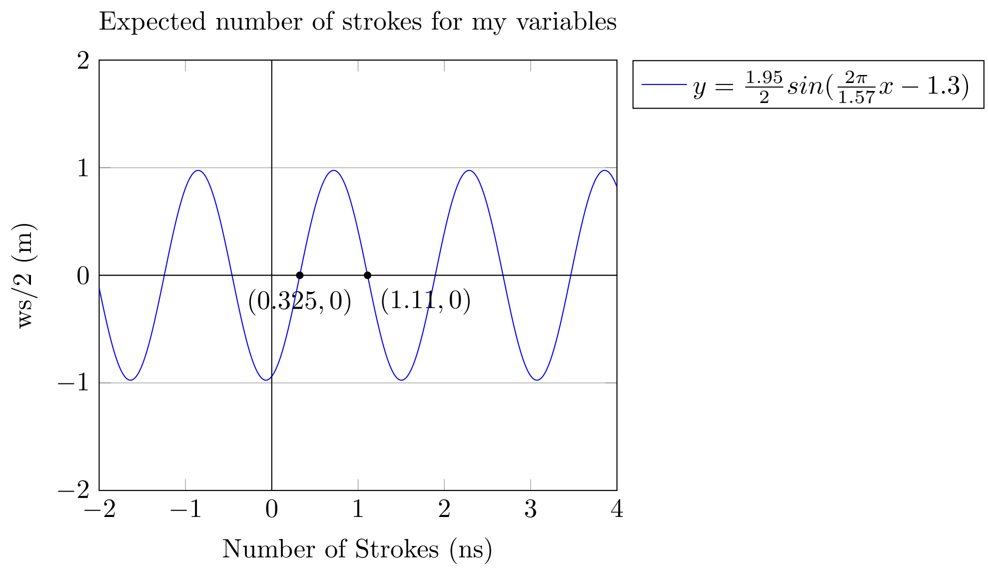
$$ac = \int_{0.596}^{1.55} \frac{2.03}{2} sin(\frac{2\pi}{1.90}x - 1.97)dx$$

$$ac = \{-[\frac{2.03}{2} \times (\frac{1}{\frac{2\pi}{1.90}})] \times cos[\frac{2\pi}{1.90}(1.55) - 1.97]\} - \{-[\frac{2.03}{2} \times (\frac{1}{\frac{2\pi}{1.90}}] \times cos[\frac{2\pi}{1.90}(0.596) - 1.97]\}$$

$$ac = \{0.306465\} - \{-0.30693\}$$

$$ac = 0.61 \times (28)$$

$$ac = 17.18\ meters$$



$$ac = \int_{0.325}^{1.11} f(x)dx$$

$$ac = \int_{0.325}^{1.11} \frac{1.95}{2} sin(\frac{2\pi}{1.57}x - 1.3)dx$$

$$ac = \{-[\frac{1.95}{2} \times (\frac{1}{\frac{2\pi}{1.57}})] \times cos[\frac{2\pi}{1.57}(1.11) - 1.3]\} - \{-[\frac{1.95}{2} \times (\frac{1}{\frac{2\pi}{1.57}}] \times cos[\frac{2\pi}{1.57}(0.325) - 1.3]\}$$

$$ac = \{0.24326\} - \{-0.243626\}$$

$$ac = 0.49 \times (45)$$

$$ac = 21.90\ meters$$

Although I initially thought that the difference of area covered would be higher, it is still impressive to compare how strong is my 50m split in comparison to Yang. This would mean that (ideally) Yang actually swims 515.14 metres and I 657 metres in all the race. This is outstanding. However, I think it's more valuable to keep the Sine model only to a 50m level instead of extending it to 1500m, because it's more relatable and simpler for swimmers to take into consideration in their own 1500m. It's useless to know whether we actually swam 1/3 of the race. As well, the model only takes into consideration the swimming component from the arms. Yang's arms covered 515.14 metres; not all of the swimming technique relies on the arms.

Moreover, the integration of the Sine function is valuable because it can show swimmers how strong their stroke is: the higher the area covered, the less powerful stroke. This is the case because a more powerful stroke allows the swimmer to 'glide' more meters, instead of doing more strokes. Nevertheless, the Sine function that models ns is the most valuable of all, because it gives a predicted number of strokes to be performed per 50m based on individual characteristics. This is extremely useful in training, as both swimmer and coach alike can monitor the maximum number of strokes per 50m, in order for the swimmer to improve based on this result. The Sine model is highly ideal, because it does not take into consideration drag or the kicking movement, which can reduce the number of strokes needed in 50m.

7 Conclusion

I'm very pleased with the investigation because the aim has been successfully met. We were able to create a model than represents Sun Yang's 1500m swimming technique and another model that can give a foreseeable value of how much strokes should a swimmer perform in 50m. Both of these are applicable for swimmers to an excellent degree: Imitating a 1500m based on Yang's model will surely be helpful in managing fatigue and having a structure. The Sine model signals the highest possible number of strokes per 50m if we want to maintain a pace, based on our anthropometry. Therefore, I consider the investigation a success.

However, although the model is a very good guideline, it's not exact. Yang's 1500m strategy changes throughout his race are very small. In fact, if a swimmer tried to imitate Yang's model without understanding that the changes in pace occurred at a hundredth of a second, it can be detrimental. The changes could be exaggerated and the split times would become too inconsistent and regular. On the other hand, the Sine model can be easily changed. A slight increase in underwater time can reduce the number of strokes. The Sine model takes into consideration only the arms as the main component.

I would've liked to introduce other variables such as kicking and drag, because the models would convey more information. For example, it would have been interesting to see what mathematics would say at what level of the pool should the underwater time of every 50m should be performed. There's more drag at a closer level to the surface, rather than deeper in the pool. However, my current knowledge of mathematics was not enough to generate a function that could include variables simultaneously of that fashion. As well, it would've been interesting to develop the quartic regression line, instead of using the Lagrange interpolating polynomial calculator. These limits of my investigation were not detrimental for the models, but the models would've been much more substantial and relevant.

8 Works Cited

- Meena E.. (2014). A Tale of Two Milers: Sun Yang Ryan Cochrane's LCM 1500 Freestyle. February 2019, eScout. URL: https://www.escout.io/postslist/2016/1/31/a-tale-of-two-milers-sun-yang- ryan-cochranes-lcm-1500-freestyle
- Sun Yang Smashes Men's 1500m Freestyle World Record. YouTube video. URL: https://www.youtube.com/watch?v=T5FlDy3YmDQ
- Distance per Stroke vs. Swim Stroke Rate: How to Find Your Swimming Cadence. February 2019, velopress. URL: https://www.velopress.com/distance-per-stroke-vs-swim-stroke-rate-how-to- find-your-swimming-cadence/