Quantitative Analysis of 200 Meter Track Times

by

Brandon Corts

Submitted to the
Department of Mechanical Engineering
in Partial Fulfillment of the Requirements for the Degree of

Bachelor of Science in Mechanical Engineering

at the
Massachusetts Institute of Technology

June 2017

© 2017 Massachusetts Institute of Technology. All rights reserved.

Signature of Author: ____________________________

Department of Mechanical Engineering
May 12, 2017

Certified by: ____________________________

Anette Hosoi
Professor of Mechanical Engineering
Thesis Supervisor

Accepted by: ____________________________

Rohit Karnik
Associate Professor of Mechanical Engineering
Undergraduate Officer
Quantitative Analysis of 200 Meter Track Times

by

Brandon Corts

Submitted to the Department of Mechanical Engineering
on May 12, 2017 in Partial Fulfillment of the
Requirements for the Degree of
Bachelor of Science in Mechanical Engineering

ABSTRACT

Using male, varsity, division 3 collegiate track and field results from the past decade, critical coaching decisions such as optimizing meet scheduling, targeting efficient training programs to mimic, and identifying potential performance-influencing factors on athletes can be made more easily.

To come to these conclusions, 200-meter race times were normalized using seasonal athlete improvement factors and wind data to identify at which facilities athletes tended to run faster times and what factors make those facilities fast. It can be concluded that the variation in banked track and field facilities makes the banked-to-flat track conversion factor implemented by the NCAA in 2012 is potentially too harsh for athletes to compete on some banked indoor tracks compared to others. The data also has the potential for many other applications such as identifying the highest quality training programs, analyzing conversion factors and facility speed for races other than the 200-meter dash, and applying similar principals to variations in swimming facilities.

Thesis Supervisor: Anette Hosoi
Title: Professor of Mechanical Engineering
Acknowledgements

This thesis would not have been possible without the support of a few key individuals who gave fantastic advice and vast amounts of time to help with many different aspects of the research.

Firstly, Professor Hosoi deserves many thanks for her unending enthusiasm for every aspect of the project. She never once said anything negative in any of the dozens of meetings we had, and had countless fantastic ideas for analysis that were key insights for making the analysis possible.

I would also like to thank Devin Neal for his help in many of the programming aspects of the project, from providing support in developing the data analysis steps and coding in Matlab to developing the data scraper that allowed this project to be possible. He put in many hours to make sure everything ran well and fixed anything that went wrong over the duration of the project.

Finally, I would like to thank the MIT Track and Field Coaching staff for their advice on many aspects of the sport that no one else could have known about. They have been incredible coaches for the past four years and I am happy to have had the honor to run for them and receive help from them on this project.
# Table of Contents

Abstract ................................................. 3
Acknowledgements ........................................ 4
Table of Contents ......................................... 6
List of Figures ............................................. 7
List of Tables .............................................. 7
1. Introduction ........................................... 8
   1.1 F.A.T. and TFRRS .................................... 8
   1.2 Overview of Variation in Racing Conditions .... 8
2. Preparing the Data ...................................... 9
   2.1 Data Acquisition .................................... 9
   2.2 Normalizing the Data ................................ 10
3. Analysis of Data ........................................ 12
   3.1 Variations in Facilities ............................... 12
4. Applications and Future Work ......................... 15
   4.1 Current Applications ................................ 15
   4.2 Further Study ....................................... 17
6. References ............................................... 19
List of Figures

Figure 2-1: Raw Time Vs. Date 11
Figure 2-2: Wind Speed Vs. Raw Time 11
Figure 3-1: Facilities Deviation Averages (Indoor Vs. Outdoor) 12
Figure 3-2: Indoor Vs. Outdoor Histogram 13
Figure 3-3: Track Type Comparison 14
Figure 3-4: Championship Vs. Regular Meet 14
Figure 4-1: 200 Meter Time Vs. Differential Time on Banked Tracks 15
Figure 4-2: Banked Angle Vs. Differential Time 16
Figure 4-3: Track Size Vs. Differential Time 17

List of Tables

Table 2-1: Example Data Set 9
1. Introduction

1.1 F.A.T. And TFRRS

According to the IAAF Official Competition Rule 165, "A Fully Automatic Timing (F.A.T.) and Photo Finish System approved by IAAF should be used at all Competitions. [4]" These systems are required to accurately report athlete finishing times within 1/100th of a second by processing an image every 1/1000th of a second into a single row of pixels and then combining those lines in what is called a linescan. These accurate systems have been used in the Olympics since 1972, and results have been put online publically as early as 2005 by companies such as Direct Athletics Inc. on sites such as the Track and Field Results Reporting System (TFRRS) site. The data available on the TFRRS site currently includes results such as the athlete's name, home school, race location, wind speed, and race time. This system is usually used as the final say in athlete rankings for qualifications to races such as the New England Division 3 Championship (Requires a certain time) and the Division 3 National Championship (Requires a top place nationally) [6].

1.2 Overview of Variation in Racing Conditions

Track and field athletic facilities vary wildly in many different factors. The length of the track gives the radius of curvature of the track. If the track is shorter, it tends to have tighter turns in which the athlete uses more energy to stop himself from being pulled into the outer lanes while running around the turns. Outdoor tracks tend to adhere to fairly strict dimensioning but indoor tracks are inclined to vary more due to a lack of space within a facility. This can lead to unusually long indoor straight-aways (such as what the MIT indoor track has) in order to reduce overall facility area taken up, leading to high radius of curvature and thus potentially slower times. Some indoor tracks are built with a bank angle to help counteract the antagonistic forces of the turns. The International Association of Athletics Federations (IAAF) standard 200-meter regulation track has a bank angle of 10 degrees but some tracks such as Boston University’s are as much as 18.5 degrees [5].

Facilities tend to have two types of materials that are important in the springiness of the track. The first is what the track is built on (wooden platform for banked tracks or the solid ground for outdoor tracks), and the second is the material that the surface of the track is made out of (rubber, etc.) reacts. Quantitatively these can be measured using a durometer or measuring the coefficient of restitution of the surface [2] [3].

An athlete’s lane might also be a potentially significant factor in his or her performance. The further inside of the track a lane is, the tighter the turn the athlete faces and thus the more energy he puts into turning. Conversely, when an athlete is in an inner lane, he can see all the other competitors, which may be an advantage compared to an athlete in lane 8 that can’t see anyone else in the race.

Less quantitatively, the durability of track and field facilities might need to be taken into account. For example, some tracks need to be shoveled out of snow which can damage them by pulling up sheets or can develop bubbles underneath the surface from water. Facilities might also
be updated over time which could cause jumps in the data, such as a slow facility being
renovated to have a banked turn might cause a large disturbance in time differentials. In addition,
athletes may perform better during championships due to mental factors such as excitement or
due to the fact that friends and family might be watching. They also might perform worse due to
stress, etc. In outdoor competitions, weather such as rain or high wind can also impact an
athlete's ability to race.

In summary, there are many potential factors that may or may not cause noise in the data
analysis. There are few elements that can be used to normalize 200-meter race data to better
compare athlete performance at different facilities. The first factor is the natural improvement of
athletes over time. As an athlete's season progresses, he or she will tend to race faster due to
technical and physical improvements from practicing and competing. These variations
significantly affect athlete performance and therefore if comparing differences between facilities,
needs to be accounted for or else tracks that are traditionally raced on later in the season (such as
outdoor tracks) will appear to be much faster than they actually are. Wind can also cause
significant changes in the outcome of race results, especially in the outdoor 200-meter dash
where the direction of the race will usually be adjusted so that the wind is to the athletes' backs
and thus theoretically helps “push” them to faster times.

2. Preparing the Data

2.1 Data Acquisition

All data collected for final analysis was taken from the Track and Field Results Reporting
System (TFRRS) website. The data was scraped by cycling through each year of available 200-
meter data dating back to 2005 and recording each unique athlete (men and women). After this
collection was complete, each athlete’s profile was scraped for all 200 meter races. Each line of
data is representative of a single 200-meter race performance. The line includes the athlete's
name, time, place, location of race, whether the race was indoor or outdoor, the date of the meet,
the athlete’s home team, the length of the race (all cases considered in this paper are 200 meters
but this is applicable for further research) and the recorded wind speed if applicable. Wind speed
and place required the scraper to enter the TFRRS meet data. With full data acquisition, the
application, using primarily the python library beautiful soup, acquired approximately 900 lines
of data per hour with an ultimate data set size of 56,282 lines.

Initial test data was acquired by hand of past and present MIT track and field athlete data
through the same logical method on the TFRRS website. This data was on the order of
approximately 300 total lines. An example of a typical line can be seen in Table 2-1.

<table>
<thead>
<tr>
<th>Name</th>
<th>Time</th>
<th>Place</th>
<th>Location</th>
<th>In/Out</th>
<th>Home</th>
<th>Date</th>
<th>Length</th>
<th>Wind</th>
</tr>
</thead>
<tbody>
<tr>
<td>John_Doe</td>
<td>23.67</td>
<td>6</td>
<td>MIT Henry G. Steinbrenner Track</td>
<td>Outdoor</td>
<td>MIT</td>
<td>04-23-17</td>
<td>200</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Table 2-1: This table shows the type of data scraped from the TFRRS website. This data
includes some potentially redundant fields such as “Length” which will be potentially
useful for future research on different events.
2.2 Normalizing the Data

Several changes had to be made to the data set to make it useable. Each column was cycled through Microsoft Excel to find cases of missing data. The most prominent case was a No Time or NT as a time for the athlete indicating that there was either an issue acquiring a time or more likely, the athlete did not complete the race or false started. From the set of 56000 lines available, approximately 900 were unusable due to this or similar issues. These lines are not entirely useless however, as they can be useful if one were to try to find programs with the most injured athletes or facilities with the least reliable timing services.

Some of the dates within the data were unusable as they were given as ranges because a large number of championship meets are run over the course of several days. It is not possible to know for sure which date within that range an athlete ran his or her race, so the first date in the range was used by cropping the end of any date cell longer than 10 characters then converting the string representations to date values (numbers representing a date) and finally converting those date values to actual dateTime cells. These dateTimees were also split into additional day, month, and year columns for additional processing in Matlab and graphing.

To analyze the data and develop a differential time for each athlete’s performance, a seasonal average had to be constructed for each season for each athlete. To do this, first a list of indexes was created that represented line locations in the data that indicated the start of an athlete’s season. This list is useful for many different applications. Firstly, it was used to cycle through the data and find the average improvement of each athlete each season. This was done by creating a linear fit for every season and averaging the slopes. This ultimately gave a value of \(-0.008 \text{[s]/[day]}\) over all seasons and years. A similar fit can be seen in figure 2-1. This value was then used to scale each race relative to the athlete’s first performance in a given season by again cycling through the list of races and finding each season using the list of seasonal indexes.
Figure 2-1: Every race with a legible date and race time were plotted to show the increase in athlete speed over the course of a season. Note the immediate cutoff and drop in times in late May when only the fastest athletes compete in nationals.

From here, the data needed to be adjusted for wind before any differential times could be created. To do this, all races with wind data were compiled and plotted against their respective date. The resulting linear relationship showed that for each meter/second the wind was going with the athlete, the resulting performance was improved by approximately \(-0.0962 \pm 0.0101\) seconds as seen by the linear fit in figure 2-2. This correlation was confirmed using a Student’s t-Test. The wind adjustment was then added to any race with wind data.

Figure 2-1: Every outdoor race with a recorded wind speed can be plotted to show that athletes tend to run faster times when the wind faster and at their back. The observed linear slope of the data was \(-0.0962\) seconds per meter/second of wind speed. At the maximum
allowed wind speed of 4 meters per second for national championship qualification by the NCAA, an athlete will tend to run 0.385 seconds faster than with no wind.

This gave a new list of times scaled by how far into the season each race was and with available wind data, which could then be averaged again to give the final adjusted seasonal average of the athlete’s season using the same process as described above. From here, each race was compared to its seasonal average, giving a differential value for that single race. If the value was negative, it meant that the race was faster than the average of that season, and if it was positive, that race was slow compared to the seasonal average.

3. Analysis of Data

3.1 Variation in Facilities

Several trends can be picked out of the data that highlight good and bad facilities and their properties. Firstly, we can look at the aggregate deviations of only tracks with over 100 logged 200 meter races to thin out the massive amount of data available. The resulting figure 3-1 gives a good representation of the field of most used tracks by division 3 athletes.

![Figure 3-1](image-url)

**Figure 3-1:** This plot shows every track with over 100 logged 200-meter races. Note that the outdoor tracks in blue tend to be faster than indoor tracks with the few exceptions of oversized and banked tracks such as the Boston University indoor track facility.

To further analyze the data, we will look at a handful of properties of facilities that the MIT track and field team competes at regularly. Firstly, the curvature of the track seems to be a strong indicator of how well an athlete performs. We can see in this figure that as the curvature increases, the speed relative to the athlete’s average decreases. This may be because the athlete uses a significant portion of his or her energy in fighting the turn on the smaller indoor tracks and much less so when on the much wider outdoor tracks.

Conversely, the larger the bank angle of a track, the better the athlete will tend to do relative to his or her seasonal average, with the Boston University indoor track being significantly more banked than any other track observed in this data set. This puts it ahead of all
the indoor tracks observed in terms of time differential, along with the other two banked tracks looked at in this case as seen in the histogram of figure 3-2.

![Histogram of Speed Differential](image)

**Figure 3-2:** Outdoor tracks tend to have a better speed differential compared to indoor tracks with the exception of the few banked and oversized tracks seen in the bins on the far left.

Other track properties considered included simply indoor versus outdoor. Indoor, however, can be split into three categories: 200 meter flat tracks, 300 meter oversized flat tracks, and 200 meter banked tracks. These three's average differential speeds are compared in figure 3-3. It can be seen that clearly indoor flat tracks are the slowest of the grouping on average with banked tracks and oversized tracks tending to be faster than even outdoor tracks. This could be because of the downsides to running outdoors such as wind and other undesirable weather conditions. Looking only at indoor versus outdoor, it can be seen in figure 3-4 that there tends to be indoor tracks that are significantly faster than some outdoor tracks while the average outdoor track is faster than the average indoor track.
It can be seen in this plot that there are significant differences between the different types of tracks. Banked and oversized tracks stand out far more in terms of speed differential than flat, indoor tracks and outdoor tracks.

Championships versus regular meets were compared to see if more important meets had a better differential than regular ones. As seen in figure 3-4 they are weakly correlated but not significant enough to warrant a change in data analysis.

Championships and regular meets tend to have a much closer average speed differential than that seen between the flat, banked, oversized, and outdoor tracks in figure 3-3. Because the difference seems to be negligible, and differentiating championships and regular meets is a difficult process, the status of a meet was not used to normalize meet data.
4. Summary and Conclusion

4.1 Current Applications

This data can be used to compare track and field facilities to isolate what factors make a track particularly fast or slow and use that information to best place athletes for qualification purposes. Using the data, we were able to conclude that, within the context of division 3 track and field, the Boston University indoor track (a differential speed of $-0.409 \pm 0.0259$ seconds) is the fastest track in the country followed closely by the Lynchburg’s outdoor track ($-0.281 \pm 0.766$ seconds) and the Armory indoor track ($-0.214 \pm 0.0318$ seconds).

According to one industry expert, the best way the data can be used is to see how NCAA blanket banked track adjustments compare to actual data. Currently, the penalty for running on a banked track is a 1.0173 times what an athlete’s banked 200-meter time is \([1]\). For example, an athlete running a 21.48 second 200-meter dash on a banked track will have a converted time of 21.87 seconds. Comparing the 3 most popular banked tracks on the east coast (The Boston University Track and Tennis Center, Armory Track and Field Center, and the Reggie Lewis Center) shows that this correction factor is not ideal because of the range in speed differentials for the facilities. It can be seen in figure 4-4 that when comparing the three banked tracks with the NCAA penalty to all tracks (0 differential benefit) and all indoor tracks (0.053 seconds slower than average), only the Boston University track is viable to race on, and only at certain speeds.

![Diagram showing NCAA athlete's benefit from running at different tracks](image)

**Figure 4-1:** This plot shows an NCAA athlete’s benefit from running at a particular track based on what time he or she is running in the 200-meter dash. The banked tracks have a slope of -0.0173 from the NCAA banked-to-flat penalty \([1]\). A track is viable to race on compared to all tracks if it has a benefit above 0, and is viable compared to a random indoor flat track if its benefit is larger than -0.0573 seconds. The fastest indoor flat track found (UW Oshkosh) is added for comparison. It can be seen that unless the athlete is running the 200-meter world record, he will prefer to race at UW Oshkosh over any Banked track.
When comparing to all tracks, it is only viable to race at BU when an athlete is running 23.65 or faster (note this takes into account outdoor and other banked tracks). When comparing to only indoor flat tracks, BU is much more viable and tends to be faster for athletes running 26.70 or faster. There are exceptions to this however, including the University of Wisconsin Oshkosh indoor flat track with an average speed differential of 0.079 seconds faster than the average track, and 0.13 seconds faster than the average indoor flat track. When comparing BU to UW Oshkosh, we can see that because of the NCAA penalty, an athlete would always prefer to race at Oshkosh, as the benefit for running at BU is only larger when the athlete is running below 19 seconds, which would be a world record. Looking at the Reggie Lewis Center and Armory Track and Field Center, the NCAA penalties always outweigh the benefits of racing on a banked track, even at the highest levels of competition. The reason for this may be because all banked tracks are lumped together by the NCAA when calculating the benefit of banked tracks. This means that faster tracks such as BU (which may then also be overrepresented because of their speed) bring the banked-to-flat speed ratio up above that of the slower banked tracks.

A solution to this issue is to make the banked-to-flat penalty more flexible by either setting the conversion on a track to track basis or making it variable based on bank-angle. It can be seen in figure 4-2 that bank angle is correlated with time differential and therefore could be used as an additional factor in calculating a more efficient conversion.

Figure 4-2: The time differential of an indoor track seems to correlate with the bank angle of the track. The trendline of the data has a slope of -0.0234 seconds per degree of bank angle. This data most likely is not linear at scale as once a bank angle gets too high it may be difficult for the athlete to stay upright. This also may not be the same slope for different races than the 200-meter dash.

To further enhance this correction factor, a time adjustment could be applied based on track size to allow athletes to qualify for outdoor events at indoor meets. It can be seen in figure 4-3 that the size of a track correlates with speed, where the larger the track is, the faster it is.
Figure 4-3: The time differential of tracks seems to correlate with the size of the track. The slope of the trendline for this data is -0.0003 seconds per meter.

4.2 Further Study

Additional areas of study for this type of data are extensive. On the surface, other factors that may need research simply to enhance our understanding of the 200 meter include preliminary races versus finals, seeding, lane placement, and the spring constant of both the surface of the track and the base it is built on (including the different surface of a banked track as well) [2] [3].

In-race injury rates can be studied by looking at cases of abnormally slow or no times in races. This could provide information on which races, facilities, or track programs tend to have the highest injury rates and could use reforming or further study.

Wind data could be used to find which outdoor track has the harshest wind conditions. The positioning of a track is critical for a 200-meter race because if the straightaways are parallel to the usual direction of the wind, the 200 will tend to be faster since it can always be run with the wind. However, for a 400-meter or longer race, wind of any direction can be harmful to race times since the athlete won’t be getting back all the energy he used to fight the wind which he will absolutely have to do at some point as he is running in a circle on the track. The orientation of tracks with respect to the wind could be a useful case to study for the engineering of track facilities.

As suggested by industry experts, track and field is a mentally exhaustive and intensive sport. Further research could look into grouping athletes from high-achieving schools to compare to the aggregate group of all athletes to see how they improve over time and overall. Anecdotally, high-achieving students tend to overthink their races which can affect their competitiveness while some athletes are able to ignore these outside factors and focus on competing. This would be interesting to study in conjunction with outdoor weather conditions and championships where athletes tend to be more anxious and less comfortable.
Finding aggregate or program seasonal peaks could help coaches in adjusting their practice conditions to better taper their athletes for championships. The data for a particular program could be checked for trends in athlete performance during a season to find more than just the linear correlation used for seasonal adjustments seen in section 2. If a program tends to peak earlier than other teams, their athletes may qualify well for their targeted championships but may not perform at their peak ability at the actual championship. This could also be used to see which track programs have the best or worst coaching by looking at athlete season-to-season and year-to-year race time progression. Programs with a large percentage of athletes improving more than the average runner over the course of their careers may be training their athletes better than others and could be good points of contact for coaching tips and ideas.

The value of a track’s renovation or introduction can be investigated with this data. By looking at speed differential trends on a particular facility before and after renovation one can develop and approximate value of an improvement in performance for competitors by comparing to the cost of renovating the track.

Similar data could be used in other sports or events as well, from simply different length races to the jumps, throws, or even swimming. Observing other races would allow a coach to better decide which athlete competes at which facility in order to maximize athlete performance when attempting to qualify for events. Similarly, some facilities may simply have qualitative factors such as large crowds that could result in improved performance in the throws or jumps. There are also many factors to be studied in swimming including the depth of the pool, change of depth in the pool, location of lines on the bottom, and location of filters and other required instruments for a pool.
References


