HYPERQUANTIFYING ATHLETES: OPPORTUNITIES AND PROBLEMS IN MODERN COLLEGIATE SPORTS

by

AMELIA HECKMAN

A THESIS

Presented to the Department of Biology
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

May 2022

An Abstract of the Thesis of

Amelia Heckman for the degree of Bachelor of Science
in the Department of Biology to be taken in May 2022

Title: Hyperquantifying Athletes: Opportunities and Challenges in Modern Collegiate Sports

Approved: *Ben McMorran, PhD*
Primary Thesis Advisor

In recent years, collegiate sports have started turning to data analysis to assist in improving performance and training tactics. There are many opportunities in utilizing data-driven – or "hyperquantified" –approaches, such as talent identification, injury reduction, in-game decision making, and increasing profits. Many universities and professional organizations utilize models to predict success. While there are many benefits, there is less emphasis on the broader wellbeing of the athlete—which "success" in this context does not include. This thesis investigated one specific example of creating a predictive model of success at the University of Oregon as well as the issues that arise from using such a model. Three ethical implications that arise from hyperquantifying athletes discussed in this thesis include data reliability, data security, and athlete autonomy. Further research is recommended into how athlete wellbeing can be emphasized at the collegiate sport levels.

Acknowledgements

I would first like to thank my wonderful Primary Advisor, Dr. Ben McMorran for offering invaluable support and guidance through writing this thesis, as well as the rest of my thesis committee Dr. Courtney Cox and Professor Michael Moffitt. This work would never have been possible without Ben McKay and Jordan Troester at the Marcus Mariota Center. I would like to thank both of my parents for the countless hours sitting on the phone with me offering words of encouragement whenever needed. And to my best friend Hannah Small, I never would have gotten through this without you by my side every step of the way from the dorms until sitting at Starbucks submitting our final drafts. Finally, I am beyond thankful for my roommates and everyone else who has believed in me through this whole process.

Table of Contents

List of Figures

Introduction

Mary Cain was a promising young distance runner from Beaverton, Oregon, whose career ended after what she describes as 'four miserable years at the Nike Oregon Project' (Manning, 2021). The data-driven— "hyperquantified"— training approach Cain's coach took heavily emphasized her weight, which took a toll on her physical and mental health. If Cain had been able to voice the impacts of this approach earlier, the four years of her professional career could have been drastically different. Unfortunately, Cain is not alone in her experience. Cain's story pushed several other athletes from high-profile collegiate and professional track and field programs to come forward with similar stories, sharing their experiences in feeling devalued as an athlete by coaches and trainers. Hyperquantifying athletes offers a competitive edge for talent identification, injury reduction, in-game decision making, and increasing profits. But it is easy to lose sight of the athlete in focusing purely on data. This is especially easy at a collegiate level, as young athletes lack the authority and professional experience necessary to voice when a coach or trainer is doing something that is harming them. Collegiate sports are competitive and success-oriented, which can be intimidating to confront.

Sports are one of the largest cultural phenomena in history, with a big emphasis on performance and success. In order to keep up with the demands of data usage in sports, there is a necessity to develop laboratory environments and techniques that can aid in learning about enhancing performance. "The use of analytics in sports has slowly transformed everything from how talent is identified and assessed to how athletes are trained and managed to how games are played on the field" (Cohen et al., 2019). In this era of the hyperquantified athlete, the increasingly urgent question is how to move from data collection to actionable insight, to potential monetization, all the while protecting athletes' rights and long-term health, ensuring fair play and competitiveness, and meeting the financial needs of leagues, players, and owners (Jarvis & Westcott, 2020). But through all this, the athlete should be at the center of every decision and conversation. There are ethical implications that must be addressed if hyperquantification is to come into its own. Enormous potential is shown with using data analytics to maximize performance and success. If left unregulated and focused on too heavily, hyperquantification can lead to adverse effects such as what happened with Mary Cain and work must be done to improve these predictive models and change the emphasis to the well-being of the athlete as a whole.

Increasingly sophisticated techniques are being taken advantage of in sports programs to measure player data. Aspects of sports have been quantified for hundreds of years, but an explosion in utilizing athlete data is transforming how athletes train, compete, and manage their careers. We are living in the age of *hyperquantified* athletes- where virtually every aspect of an athlete's performance can be quantified. This is happening on a large scale in professional sports, but collegiate sports offer a unique playing field for researchers to test new technologies and techniques and look at the repercussions of doing so. This explosion of athlete data is raising many questions of how best to use it, and how to do so ethically.

Data Usage in Sports

Fantasy sports is perhaps the most visible manifestation of statistics in mainstream sports (Shipman, 2009). Additionally, the book *Moneyball*'s released in 2003 provides insights into how data can be a big enabler for performance outcomes. This book detailed how the Oakland A's— a low budget baseball team in Oakland, California— used an analytical, evidence-based approach to assemble a competitive team (Lewis, 2004). While data analytics was not new to baseball, it propelled the world into an obsession with predictive statistics in sports. The *Moneyball* narrative sequentially snuck its way into most aspects of the modern world, including track and field.

Sports analytics is defined as the "examination and modeling of sports performance using scientific techniques" (Morgulev et al., 2020), and gains popularity with each year. It combines a wide domain of specialties from human physiology and kinetics, sports science, big data, data science, data mining, mathematics, and statistical analyses (Passfield & Hopker, 2016). When these data are processed in tabular or graphical construct, it helps to observe trends and find valuable insights that can influence decision-making. Although sports analytics has been around for a long time, it is becoming increasingly more popular recently, and this is in part due to biometric wearable devices. These devices, such as Apple Watch, Fitbit, and Garmin, are making sports analytics more accessible to the general population. These have high potential to prevent injuries, improve performance, and extend athletes careers.

Data collected can come in two forms: positional/tracking and biometric. Positional/tracking data measures location in three dimensions, and can include position, acceleration, lateral motion, speed, jump height, and other measures. Biometric data encompasses more biological information from an individual player. This can include pulse rate, blood glucose levels, oxygen levels, sweat rate, sleep rhythms, and more.

The sports analytics market was valued at $1.05 billion US dollars in 2020, and it is expected to grow to $5.11 billion by 2026 (*Sports Analytics Market- Growth, Trends, COVID-19 Impact, and Forecasts (2022-2027)*, 2021). With this expanding market, there is an increased pressure on athletes, trainers, coaches, and everyone involved to produce the best results possible. At the height of lockdowns due to COVID, teams were forced to scout players virtually, leading them to utilize automated video analysis and positional/tracking data in talent identification. The biggest influence for wanting to use sports analytics is arguably for injury reduction. Teams want to have the ability to predict when conditions may heighten the risk of injury. This benefits both the team and the athlete- teams want more wins and more revenue, and athletes want to extend their careers and earnings potential as much as possible.

Unhealthy Uses of Data Analytics in Track Teams

Eugene, Oregon is 'Track Town USA.' The town hosts many major track and field competitions, and the state of Oregon is home to world-famous running companies *Nike* and *Adidas,* so it does not come as a shock that running is a huge part of the culture in Eugene and in Oregon as a whole. One infamous running group in Oregon was the *Nike Oregon Project*. *Nike* created this group in 2001 with a goal to get Americans to be more competitive in long distance events in a global setting (Hartzell, 2018). This group was home to many famous athletes such as Mary Cain, Jessica Hull, Galen Rupp, and more. The head coach's extreme training methods created a toxic culture where female athletes' bodies were demeaned and spoken about publicly and the women were encouraged and pushed to maintain unhealthy weights. After a four-year ban of head coach Alberto Salazar, the group dissolved in 2019 (Futterman, 2019).

This group is not alone in being outed as abusive towards their athletes and expecting them to maintain unhealthy habits to stay as competitive as possible. An *Oregonian* article published in October of 2021 described six women's experiences with the University of Oregon's track and field program. The article claims that the program uses a data-driven approach to tracking athletic progress with weight and body fat percentages. Athletes participate in a DEXA (dual-energy X-ray absorptiometry) scan once per term to precisely measure bone density and body fat percentage. These six women all left the program saying they felt devalued as individuals and at risk for eating disorders because of the nature of the program. The online newspaper *Tucson* published a similar article in August of 2020 outlining eight women's experiences with the University of Arizona track and field team (Schmidt, 2020). The team required athletes to participate in public weigh-ins as well as track their food and calories consumed. Both practices are shown to cause and perpetuate eating disorders (Arthur-Cameselle et al., 2017). The perceived correlation between athletic success and being lower in weight is deeply ingrained into our society.

Predictive Models of Success

With so much money pouring into athletics, both professional and collegiate sports have long strived to optimize performance of their athletes. Due to this, athletics at the professional and collegiate levels have become laboratories devoted to "preventing injury and enhancing performance" (King & Robeson, 2007, 2013). It can be difficult to use traditional methods of predicting a player's performance due to the multitude of factors interacting in complex ways. Accurate prediction of sports achievement can be used to uncover trends that aid in decision making. Most collegiate sports teams have developed models that can aid in future decisions of individual training plants, game-time decisions, diet choices, and more. These models can be used to track and athlete's progress or success and are referred to here as 'predictive models of success.' This encompasses all models that athletic teams and trainers use to assess their athlete's progress and to create individualizes training plans. To explain why I chose to define these models this way, I will further break down each part of the term. The word *predictive* means "relating to or having the effect of predicting an event or result." Coaches and trainers desire to know where to focus their athletes' training. The word *model* is defined as "an informative representation of an object, person, or system." Models in sports can come in the form of codes (in python, java, R, etc.), graphs, or tabulations. Finally, the word *success* is defined as "the accomplishment of an aim or purpose." For runners, the goal is to become faster, hence the aim or purpose is speed.

Coaches use previous and current data from athletes to predict how their athletes can best achieve success using a model that can take many forms. Individualized training plants can then be created from what this model tells the coaches and trainers. For the purpose of this thesis, I will be focusing on models that apply to runners. These models can encompass many things and can be created by coaches, trainers, or athletes themselves. Predicting an athlete's performance is essential for creating a scientific sports training plan. This is beneficial not only to the athlete, but to the development of the sport as a whole. Athletes and coaches can use sports performance predictions as a basis for creating a training schedule.

Research Question

In this thesis, I aim to investigate both the challenges and opportunities presented when utilizing data analytics and predictive models in collegiate sports. Hyperquantifying athletes is a relatively new field of research and there is little known about the ethical implications of doing so as well as there being a multitude of methods of creating these models. I am focusing on data analytics in collegiate sports at the University of Oregon to see if there are any lessons to be learned from this example of a predictive model. I will break this research into multiple parts: 1) How are predictive models in sports created to produce well-rounded, accurate outcomes, 2) What are the ethical implications of hyperquantifying athletes, and 3) What should collegiate teams do moving forward to maximize an athlete's performance while still protecting the athlete.

Methods

My thesis will explore how data are used in track and field programs. More specifically, I am looking at how predictive models of success are used, and how these models have been the center of several track and field programs in creating individualized training programs. This work will be a mixture of literature review, and first-hand experience working with a predictive model of success on the UO campus.

I will first examine what University of Oregon specifically is doing with data and how the school uses predictive models to drive their training plans. Since the fall of 2021, I have been working with the Marcus Mariota Center for Athletic Performance to create a code to streamline the data analysis processes used for various athletic teams. Working on this code has given me first-hand experience with having access to athlete data and thinking through how to analyze it. Through this experience, many questions have come up addressing how data is accessed, who has ultimate control over the spread of the data, and safety of the data. While these issues do prevail in professional sports, the power dynamic between a student athlete and a coach or university is much different and must be considered. These questions are what I seek to explore in this thesis as well as how we can better protect student athletes.

Since I am dealing directly with athlete information, I have finished the CITI NSF/NIH Responsible Conduct of Research Training for Researchers and the CITI Social-Behavioral-Educational Researchers courses. The Marcus Mariota Center has an outstanding IRB certification to take and analyze these data from their athletes.

The literature review will cover three main things. The first part will look at what predictive models in sports are. This will look at previous work on the topic on what these models emphasize and whether they work. The next part will look at the ethical implications of hyperquantifying athletes such as privacy protection, athlete autonomy, and data reliability. I will then look at what we should do moving forward to best protect student athletes.

My expectation for this research is to start a conversation of how we can best protect student athletes while hyper-focusing on data and shift toward a more holistic regimen that also emphasizes psychological health. My hope is that in the future, with more research in the area, collegiate sports programs will aim to be more aware of their athlete's psychological health on top of their physiological health in creating training programs.

Sports Analytics at the University of Oregon

In an average year, the University of Oregon puts $108 million into sports, namely their football and track teams. This money goes toward paying coaches, transportation for athletes to competitions or games, and to technology to improve the performance of their athletes. Most universities that have access to large sums of money such as this have data-driven processes of analyzing their athletes' progress. Oregon chooses to use their funding for technology to create robust, multivariable models to analyze athletes' data.

The algorithm used by Oregon is modeled after a method created by the Morin lab at the University of Jean Monnet in Saint Etienne, France. This lab created a method to analyze sprint acceleration force-velocity-power profiling that has been validated in several publications. This model was created with the purpose of exploring an athlete's performance as well as gaining insight into creating a beneficial individualized training plan for the athlete based on underlying mechanical variables (J. B. Morin, 2017). The group shares a user-friendly Excel sheet for anyone to use (J. Morin, 2019). Users can input as little as 4 split times (i.e. 5, 10, 15, and 20 meters), the runner's weight and height, and air temperature and pressure at the time of measurement. These can then output maximum power output, initial velocity, DRF, maximum speed, and other useful measurements for analyzing a sprint trial.

This model is being used in the form of a code in the programming language R by researchers at the University of Oregon's Marcus Mariota Center. A common way to obtain data from sprint trials is through using a radar or lidar gun, which makes using the programming language much easier to use and more robust than the excel sheet, as the data from the technology can be uploaded straight into the software. Radar guns use radio waves to provide information on speed versus time. Lidar guns use light waves and can produce information on speed, distance, direction, and time. University of Oregon researchers use both Radar and Lidar data in their data collection. The researchers currently use the Stalker Acceleration Timing System II radar gun, which has a +/- 3% accuracy and has a resolution in the tenths of a second. The Lidar gun is from Light Ware and can record 20 thousand readings per second but can only record up to 100 meters away (*The Stalker Acceleration Testing System*, n.d.). Both methods produce massive data sets, which are difficult to analyze by hand or with an excel sheet. This warrants the use of algorithms in coding software. Uploading sprint trials to the R system can quickly output maximum speed, initial velocity, optimum velocity, DRF, maximum power, and split times.

Despite the apparent trustworthiness of the technology, data resulting from this technology can still be noisy from faulty measurements, signal interruptions, or other factors. This noise can lead to faulty calculations. For example, outliers in velocity data can lead to inaccurate calculations of acceleration. Hence, it is vital to remove these outliers to yield reliable results. The process of doing so by hand is very time consuming, and the researchers who were doing so sought out a new code that no longer required as much tedious work.

There were multiple goals in creating the new code: 'smoothing' the data so that hand-picking the outliers out of data was no longer necessary, de-emphasizing weight, enabling the code to run through a file of trial data and exporting straight into an excel sheet rather than going individually through each trial. In order to keep the process of creating the code as objective as possible, athlete names were anonymized through a separate code. Athletes could be identified through a 5-letter code. Coaches and trainers working directly with the athletes have access to the key that links the 5-letter codes to athlete names.

Data Smoothing

In smoothing data, there is always the risk of removing meaningful outliers and emphasizing analysts' biases. Lidar and radar technology are not 100% accurate. Outliers in this data means there was a signal interruption, or something happened that caused the technology to pick up another signal separate from the athlete it was supposed to be tracking. If the athlete completing a trial run moves off course or moves their arm in the line of data measurement, it is easy for the technology to pick this data up as having a much higher or lower velocity or acceleration as the torso does (which is what the technology aims to measure.

A sprinter's position versus time and velocity versus time curves can be modelled through the acceleration phase of a sprint using a mono-exponential function. A runner's 100-meter sprint can be divided into three main phases (Healy et al., 2022). The acceleration phase is characterized by a rapid increase in velocity followed by a subsequent gradual increase in velocity. The maximum velocity phase is represented as a relatively flat section of the velocity-distance curve, with the velocity staying close to 100% of maximum velocity (V_{max}). It is common for sprinters to reach maximum velocity between 50 and 80 meters into a 100-meter sprint. The final phase is the deceleration phase, where there is an inevitable decrease in velocity due to fatigue. The mono-exponential function is accurate in modelling position-time and velocity-time curves over short sprints as well (<55 meters) (Clark et al., 2019; Furusawa et al., 1927; Samozino et al., 2016). This model can be used to analyze athlete's sprint trials for athletes at UO, but to accurately fit this function, data must be smoothed, and outliers removed to assure only reliable data is used.

Low pass filters can be applied to data to reduce noise effects. There are several methods of filtering data called low pass filtering, which includes moving average, Butterworth, and polynomial filters (Crenna et al., 2021). The two filters used in this code are a moving average filter and a Butterworth zero-phase filter. Moving average filters are very simple filters, often called a boxcar average. This creates a series of averages of different subsets of the full data sets and returns an array the same size as the original. The Butterworth filter is very common in filtering biomechanical signals and utilizes frequency. If a component of the data is varying too rapidly in time, the filter dampens it.

Before applying either low-pass filter, any clear outliers were removed. Plotting the accelerations as a histogram creates a spread of calculated accelerations, which can aid in determining which points to flag and potentially get rid of, as shown in Figure 1.

Histogram of Raw Acceleration Values

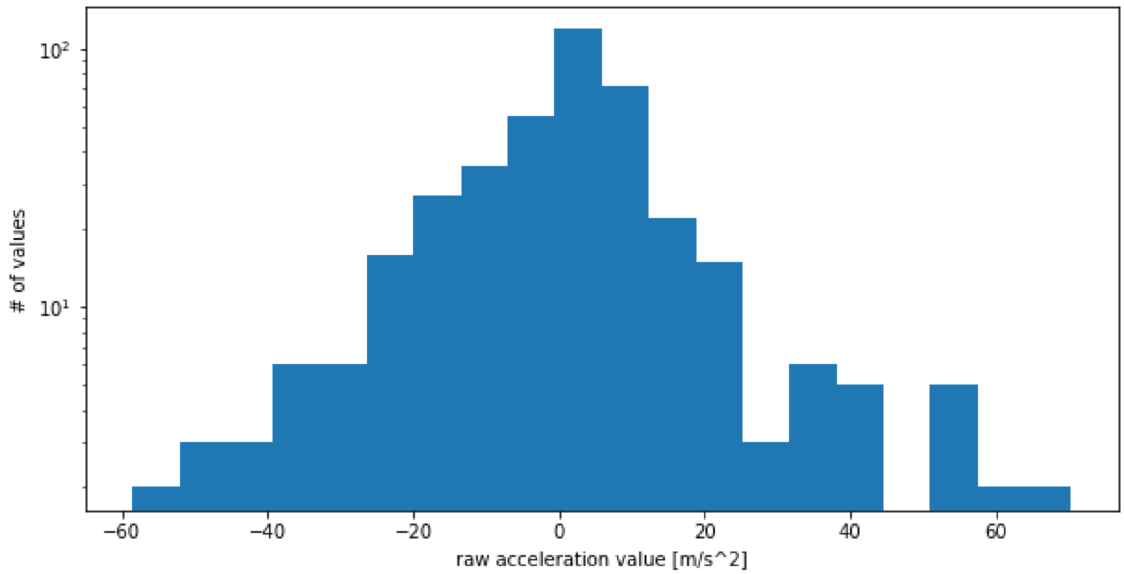


Figure 1. Creating a histogram of raw acceleration values and how often these occur can show which accelerations to flag and potentially remove. The user can input maximum and minimum acceleration values. Any values outside of these parameters are removed from the corresponding velocity data.

The user of the code can manually input maximum and minimum accelerations, and any points with accelerations that are outside of this range are removed. The velocities are then removed and replaced with NANs (not a number) that have too large of an acceleration. Replacing them with NANs rather than completely removing the data point helps to not alter the time variable. The raw velocity versus time data is shown in figure 2. Figure 3 shows velocity versus time data with acceleration outliers removed.

Raw Velocity Data

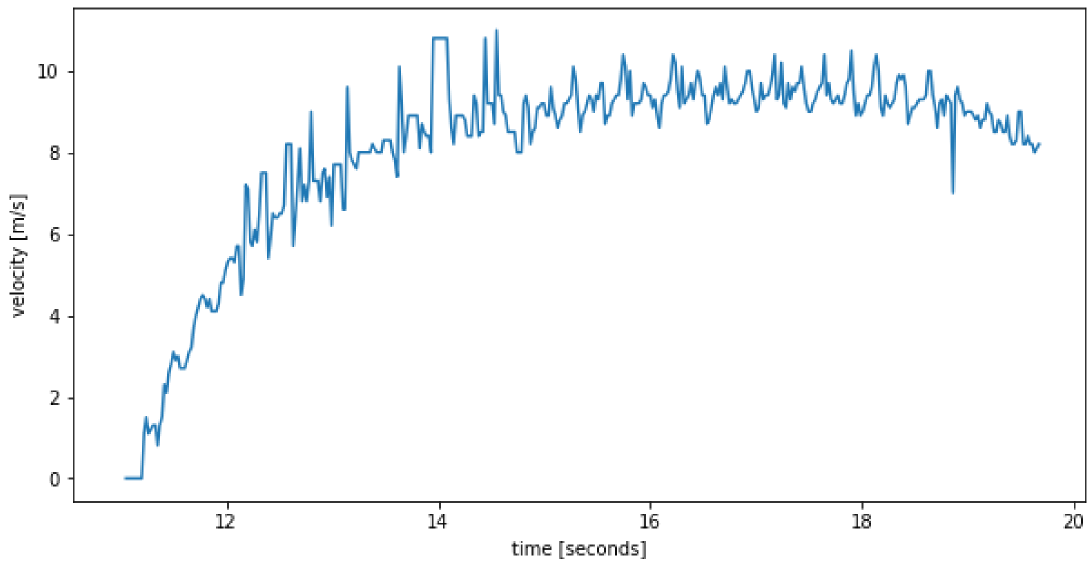


Figure 2. Raw velocity data directly from the LIDAR gun shows many spikes, particularly in the 14-15 second range. Including these data in calculations can lead to erroneous values that are not indicative of the athlete's true capabilities.

Velocity data with outliers removed

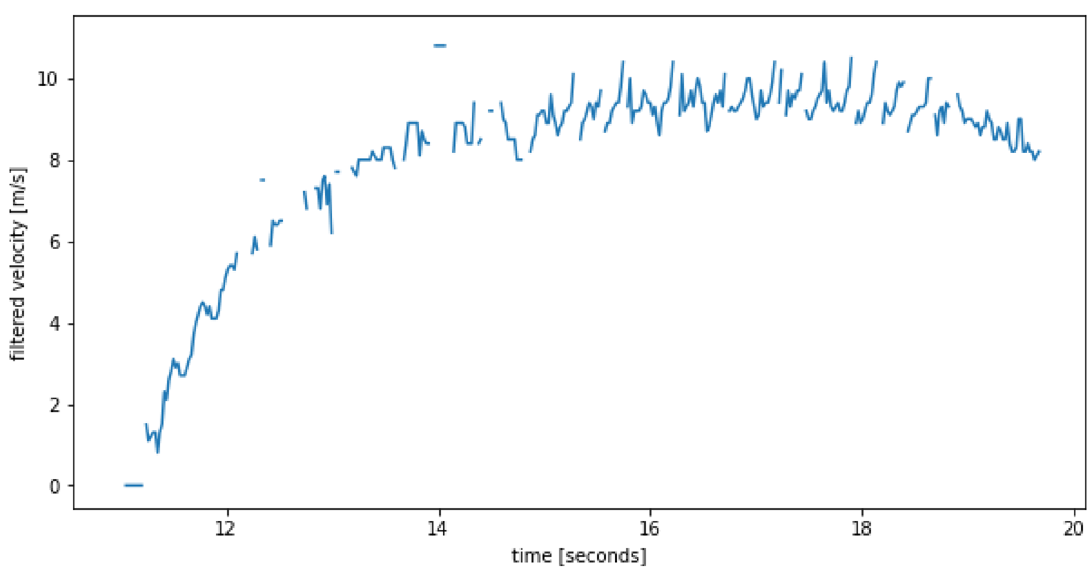


Figure 3. Velocities with too large of accelerations less than -20 meters per second or larger than +30 meters per second have been removed.

After removing the outliers, a moving average filter was applied with a window of 3 values. The moving average trend-line is shown in figure 4 below on top of a scatter plot of the raw velocity data.

Moving average filter applied

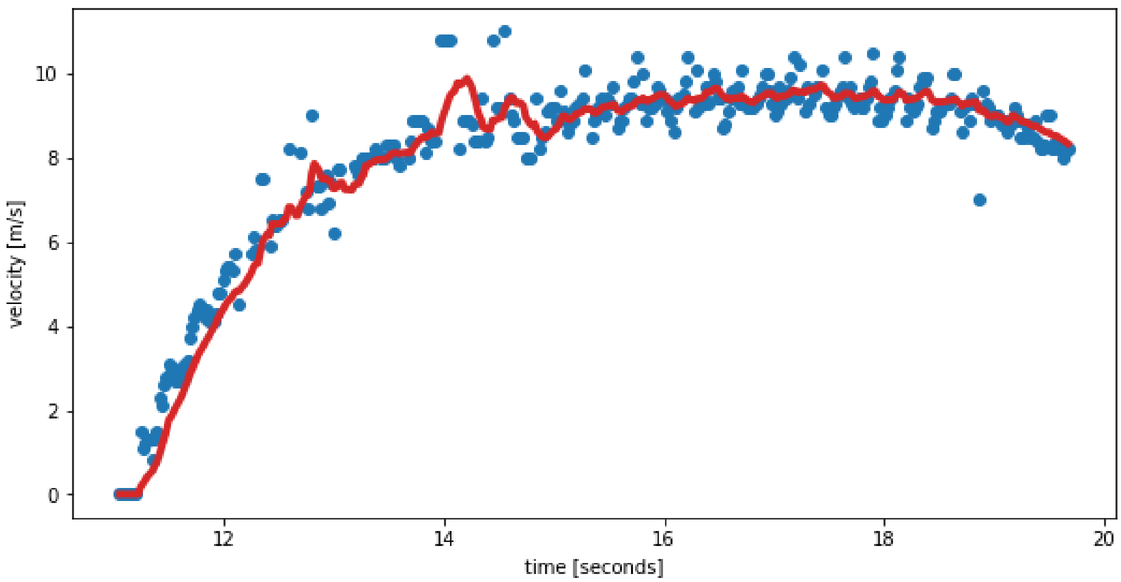


Figure 4. This figure shows the raw velocity data in blue against the smoothed velocity in red after a moving average filter was applied. The smoothed velocity has much less veracity and therefore is more trustworthy for calculating any values.

As the graph above shows, calculating maximum velocity and acceleration from unfiltered data would lead to erroneous calculations. The maximum velocity of the velocity with outliers removed (in blue) is close to 11 meters per second, while calculating the same value for the filtered data shown in red is closer to 10 meters per second. As mentioned above, it is not always advisable to remove outliers since they could be meaningful, but some outliers can lead to faulty calculations based on data that does not physically make sense, which can be seen by looking at the acceleration data.

Acceleration is the change in velocity over time. Based on the raw data for this specific trial, the maximum acceleration would be much higher than expected, over 60 meters per second squared. The smooth data shows much more believable data about the velocity. Since the technology used records velocity, this smoothed velocity data was used to calculation position, which then was used to calculate split times. The user is asked what distances split times are desired in meters or yards, then outputs an array of those split times.

Another low pass filter used is called the Butterworth filter. This was applied to data with outliers removed using the histogram. This filter is among the most used digital filters in motion analysis as well as audio circuits. A 10 Hz Butterworth filter was used to smooth out the data while minimizing loss in sensitivity through reduced frequencies (Poehling, 2018). The Butterworth filter has high performance in three key areas of smoothing biomechanical models: (1) numerical differentiation followed by low pass filtering, (2) polynomial local approximation and direct differentiation, and (3) optimal Fourier filtering (Crenna et al., 2021). The moving average is acceptable in these areas but far less efficient.

Using the filtered velocity, the mono-exponential function shown below can be fit to the data. This function requires three input parameters: maximum horizontal velocity (V_{max}) and the acceleration time constant (τ), which is the ratio of V_{max} to initial horizontal acceleration, and time delay (d) (Furusawa et al., 1927). All parameters can be calculated from the filtered velocity data. The function fitted to the raw data is shown in figure 5.

$$V(t) = V_{max} * (1 - e^{-\frac{(t-d)}{\tau}})$$

Raw Data with Fitted Function

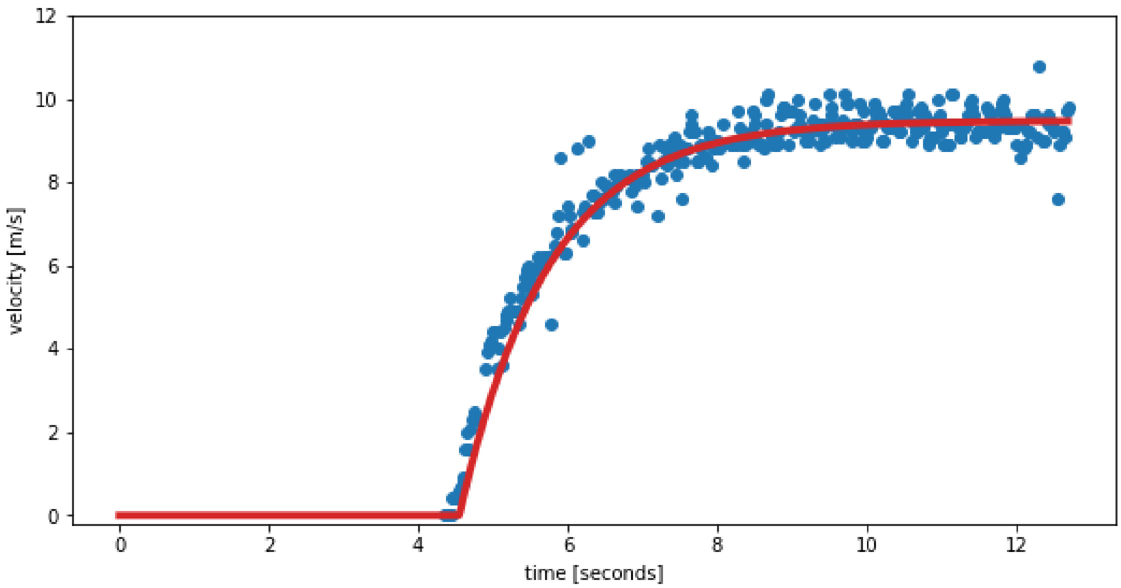


Figure 5. Raw data shown in blue with fitted mono-exponential function created by K. Furusawa shown in red.

One study found that faster sprinters had a significantly higher V_{max} than slower sprinters by 6.0%, demonstrating that V_{max} was among the most important variable to measure in elite 100-meter sprinters. There are mixed results in the variability of the acceleration time constant τ. A study comparing τ for NFL players with a high V_{max} to those with a lower V_{max} found no significant difference (Clark et al., 2019). Other studies found a moderate difference in τ values between the two groups among sprinters (Healy et al., 2022).

The three parameters are unique to each trial and allow for the derivation of multiple outputs. Theoretical maximum force (F0), theoretical maximal velocity (V0), absolute and relative maximum power (Pmax), maximum ratio of force (RFmax), maximum velocity (Vmax), and all sprint distance times (5 to 30 m) show acceptable reliability using this model. Using the average of two or three trials was also shown to be the best method of establishing reliable sprint times and force-velocity-power characteristics between sessions (Edwards et al., 2022).

Using the code

Researchers with the University often test entire teams at a given time. For example, the entire woman's soccer team will go in for sprint trial testing at the same time. This makes it especially time consuming to manually analyze each sprint trial. The code is set up in a way that it can loop through a series of RDA files in a folder that are named with the 5-digit code created for individual athletes. After removing clear outliers, smoothing the data, and fitting the mono-exponential function, several outputs are calculated and exported to an excel file. The outputs are maximum velocity (maxV), position in which maximum velocity is achieved (maxV position), time at which maximum velocity is achieved (maxV time), initial force production (F0_N.kg), the slope of the force-velocity time curve (FVslope), initial velocity (V0_m.s), maximum power output (PowerMax_W.kg), the maximum ratio of horizontal force to total force (RFmax), desired split times entered by the user either in yards or meters, the start time of the trial, and the acceleration time constant tau. Breaking down the sprint into these values can help trainers and coaches know where to focus training efforts, such as if the runner is reaching their maximum speed earlier or later than ideal.

Since this code uses different methods of smoothing and cleaning the data than the original methods, the data outputted is slightly different. This goes to show that different methods of analyzing a data set can yield varying results. But in all, the data shows the same trends, and the differences are not statistically significant.

There are numerous calculations and measurements that can be utilized in predictive models such as this code, and thus there are many additions that can still be made. One of these possibilities is to add error bars around split times that would allow researchers to easily visualize how accurate the calculated split times are. These error bars would represent a built-in uncertainty of taking data with radar or lidar technology, smoothing and cleaning the data, and making the calculations. They could also help account for the differences seen when analyzing the data with differing methods. Another addition that could be made is to alter the code to calculate running cadence—sometimes referred to as turnover rate or revolutions per minute (RPM). Cadence is a valuable metric to look at in running to attempt to maximize running efficiency. A periodicity in data can be observed in a plot of kinetic energy versus time. This can be verified using a fast Fourier transform (FFT), which provides frequency information about the signal. Utilizing this can give insight to researchers into frequency of strides and how it changes over the course of the sprint. While the current state of the algorithm fulfills the goals set out for it, there is always room to add more variables it is calculating and make it more accurate. The beauty of a university working to create their own algorithm to streamline sports analytics is how dynamic the process can be. These algorithms and models should change as more research is done and more information is gathered on sports analytics. The techniques and technologies being used should be up-to-date and continuously being improved.

Opportunities for Improvements

There are many improvements to be made to the University of Oregon's usage of sports analytics. University of Oregon's facilities and the technologies used are very advanced. Researchers at the Marcus Mariota Center prides themselves on using state-of-the-art technologies to aid in sports science and analytics. While they are extremely advanced, working directly with the researchers here has given me an inside look at the interworking's of data analysis. The process in which the University's facilities go about utilizing and protecting data must be revised.

A major downfall of the University of Oregon's process of using athlete's data is a disconnect in communication between athletes and those who are processing and analyzing their data. In my own experience with being given all these data, I was given the bare minimum instructions of how to deal with these data. After meeting the researchers at the Marcus Mariota Center, hundreds of athlete's trial run data were put into a dropbox folder to be used to create the code. This was necessary to make sure the code worked with the file type and could accurately analyze these data. However, in doing so, there were several files put into this dropbox that were clearly not meant to be there. I ended up gaining access to old diet plans for certain athletes. This information was not necessary to my research and a huge invasion of privacy. Improvements to the protection of these files are needed to assure that mistakes such as these no longer occur.

I had no experience interacting directly with the athletes, so we determined it was inappropriate to be working with data that directly says the athlete's name. The first thing we did when we received the data was run it through a separate algorithm to anonymize the data. Researchers who work directly with the athletes had access to a key stating the 5-digit code corresponding with each athlete. This also assured that bias did not play a role in analyzing these data.

A huge area for opportunity and expansion could be to assure connections between the athletes and all who have access to athlete information. This would help to assure that any external consultants that have access to the data stay true to what the focus of all sports analytics should be—the athlete.

Ethical Implications

Sports analytics is growing exponentially and revolutionizing the sports industry, but the demands that come from hyper-fixating on the data can be too much on the athletes. An important potential downside is the ethical issues of focusing only on the athletic performance rather than the bigger picture. The element that is not including in “hyperquantification” is the athletes’ emotional wellbeing. Athletes such as Mary Cain have been coming forward recently to speak about how pressure from coaches and trainers can lead to athletes turning to unhealthy behaviors and relationships with food and their sport. There are ethical implications that must be considered when hyperquantifying athletes.

Big data entered the world of sports hundreds of years ago but became popular though baseball in the mid-20th century. Big data is characterized through 3 V’s: volume, variety, and velocity. These immense data sets can uncover hidden patterns and trends, but it can only paint the big picture. But big data is “neither neutral & objective, nor necessarily valid and reliable” (Spaaij & Thiel, 2017)—essentially, while there is a lot of potential with big data, it is difficult to get objective, reliable results due to a multitude of factors. Seeing successes in focusing on pure data such as the Oakland A’s pushes athletic organizations to want to have every decision be backed by data. In a paper on the applications and risks of big data in sports analytics, Vermeulen et al. poses the question “Where is the line between data working for the athlete and data working against the athlete?” (Vermeulen, 2018). Data analytics is a fantastic supplement to traditional training, but it should not become the sole driver of performance. The current University of Oregon track coach said in an interview with

Oregon Live that “track is nothing but numbers” (Goe, 2021), which discounts the entire psychological aspect of sports. Sports psychologist Dr. Doug Gardner says that his work with athletes is guided by one simple principle: “For every physical and fundamental act in sport, there is an equally important and equally-related mental component that must be addressed.” (Gardner, n.d.). No sport, including track, is just a numbers game. There is an immense mental demand that come with being an athlete that is often overlooked because much of the athlete’s training and attention is put on the physical demands of the sport.

Three specific ethical implications that I will detail include data reliability, data security, and athlete autonomy. These factors may halt progress of research into athlete performance and are important to consider when creating a predictive model that aims to drive success in a sport.

Data Reliability

There are two areas of focus when considering data reliability— whether the data is accurate and represents what it claims to, and whether the data being shown matters to athletic performance. When subjecting athletes to biometric tests or trial runs, the technology being used must be valid and accurate, as performance decisions are based on them. Researchers must also decide whether the data they are taking are indicative of success. It is unethical to run tests or use technology on athletes if the data being collected does not correlate with any performance improvement.

Are the data accurate and does it correctly represent what it claims to represent? Data veracity is how accurate or truthful a data set may be. Incorrect readings or faulty technology can lead to over- or under- determination of performance capabilities and have the possibility of leading to harmful decisions. Athletes may push themselves too far physically or assume fatigue falsely due to a performance detriment (Karkazis & Fishman, 2017). The veracity is influenced by the noise created by faulty technology or signal interruptions. Not all data can be completely trustworthy, and there is always the chance of noisy data. Algorithms that are designed to clean up this noise should be objective, but still are subject to bias due to the assumptions of the developer. Data analytics in sports using extremely large data sets is still relatively new, so there is an undersupply of historical, validated data to develop a valid algorithm while there is an overload of data that requires interpretation. There are many ways to clean up noise in a data set, and no two are the same. Two algorithms that set out to clean a data set may present different outcomes simply because of a difference in methods of the cleaning. For algorithms to become reliable, true representatives of the actual data generated by the runner, this data must somewhere be validated. This is shown when comparing the previous code used to test athletes at the University of Oregon and the current one that uses a slightly different approach in a separate algorithm. While the two codes use the same mono-exponential function to fit to the data, there are different approaches to smoothing the data that yield slightly different results. There is a certain degree of uncertainty that comes along with any data collection or analysis that must be considered when applying this knowledge to creating training plans.

One study tested multiple commonly used fitness tracking devices on their accuracy. When measuring distance, the error rate ranged from 3.72 to 11.17%. For step count, the error rate ranged from 1.05 to 27.28% (Guo et al., 2013). These errors can easily lead to inaccurate calculations of stride frequency (cadence), running speed, and pace—all of which are data points that runners are commonly advised on by coached and trainers These inaccurate readings can lead to an athlete falsely believing that they are either over-or under- performing. Unnecessary changes will thus be made to training or race tactics.

With the increase in popularity and accessibility of wearable biometric devices, it is easier than ever to track any daily activities of athletes. Apple released the fitness-oriented Apple Watch in 2015, with an estimated 30.7 million ordered worldwide in 2019 alone. The newest version of the watch can keep data of minutes of standing, exercise time, steps, sleep, blood oxygen levels, breath rate, heart rate, location, calories burned, and more. There are countless other technologies in existence that can do the same and more. With biometric wearables in addition to laboratory techniques of taking data measurements, there are countless data points to be measured and analyzed. The question is, how does a researcher decide what matters to performance and success?

A common data point highlighted by athletics is an athlete's body mass index (BMI). There are cheap instruments that can calculate BMI in a matter of seconds used in schools as early as high school. But BMI is an outdated, indirect way of defining weight. The only two points it takes into consideration is height and weight. This disregards important details about age, sex, bone density and structure, and fat distribution (Rothman, 2008). Despite these downfalls of BMI, it persists in many aspects of society. There is a correlation between weight/BMI and performance. An athlete must be in shape and healthy to perform at such a high level. But this relation is not causation. Lower BMIs do not cause improved performance, just how successful athletes do not necessarily have low BMIs. Muscle weighs much more than body fat does, thus people with more muscle and who are more athletic would be placed in higher BMI categories. Running periodic BMI tests (or similar tests such as DEXA scans) subjects athletes to useless tests and puts too much of an emphasis on a part of their body that does not lead to improved sprint times or performance enhancements.

Another point to consider is whether the measurements and the models being created are actionable. Meaning, after taking a measurement and analyzing the data, it can be used to prescribe additional training regimens that will improve performance. Focusing on BMI is not 'actionable' in this sense. If coaches know, for example, that having a fast reaction time improves sprint performance, they can work to identify those who can benefit from a specific training that can improve their reaction time.

These lessons on using BMI can be applied to all other data points used in analyzing performance. Research should be done to assure that aspects of an athlete's performance that are being tracked and advised on are impacting their performance and can help them to succeed. One danger of hyperquantification is the desire to quantify every aspect of an athlete's life. With biometric wearables such as the Apple Watch or FitBit, researchers now have the ability to track sleep, diet, heart rate, blood oxygen levels, and more. While these are undoubtedly interesting things to look at, we must ask ourselves whether quantifying these aspects of an athletes lives actually can improve their performance.

Data Security

When an athlete joins a team-whether that be collegiate or professional- that team has access to some portion of the health information of that athlete. Professional teams have nearly unrestricted access to health information. This open access creates expectations for signed players to submit to medical exams and surveillance. Trainers often use biometrics to make athletes healthier and maximize their careers. But in taking these data, who is granted access to players' data, and for what purposes? Who grants such access? Do athletes have any control over the dissemination of their personal data, especially to third parties? Are there protections in place? These are all questions that ran through my mind when I was first given access to these data to create the code.

There is no clear answer to these questions. The Health Insurance Portability and Accountability Act (HIPAA) was enacted in 1996 with the aim to improve the portability and continuity of health insurance coverage. HIPAA is also responsible for much confusion and controversy in collegiate sports settings, specifically one portion of it called the privacy rule. This rule of established the first set of national standards for the protection of personal health information. This rule specifies that all covered entities must follow five steps to ensure the privacy of patients' health information (Dolan, 2003). These steps include: notifying patients of their rights and how their information will be used, adopting & implementing privacy procedures, training employees on privacy procedures, designating an individual to be responsible for ensuring privacy procedures are followed, and ensuring that patient records containing individual identifiable health information are secure. The challenge that faces all sports organizations is determining whether HIPAA applies to them, and therefore what protocols need to be established to perform duties adequately while complying with federal regulations. Any 'covered entity' is expected to adhere to the policies of the privacy rule. But a college, university, or high school is not automatically a covered entity simply because there is an athletic trainer on staff (Bell et al., 2004).

The interaction of HIPAA and the Family Educational Rights and Privacy Act (FERPA) also adds to the confusion. FERPA applies to schools that receive federal funding and is intended to allow parents access to information about their children without safeguarding information from release to other parties. As FERPA takes precedence over HIPAA (Pitz, 2003), information about college students can be released without consent to school officials. There are many exceptions and rules that can allow the free movement of athlete data and information. Any sport entity that is covered under HIPAA needs to review its existing practices, policies, and procedures to assure their athletes are protected as much as needed. HIPAA and FERPA, however, only apply to health information. Much of the data being collected may not relate to an athlete's health and therefore will not be protected under these acts.

Information about players' injuries is communicated with physicians, athletic trainers, coaches, school administrators, and even the media. A huge part of modern sports is the media, and a big concern following the passing of these acts is how these procedures would be impacted. For athletes, health and injury information is considered criteria for employment. Therefore, an injured athlete would not be able to legally withhold this information from a team to whom he or she is contractually obliged. Yet, the athlete still is entitled to some degree of control over the dissemination of their information and data. There must be protections in place to assure this information does not fall into the wrong hands or is misused. Some universities have started requiring athletes to sign mandatory authorization forms to participate in athletics. These authorization forms should have specific explanations of the information that will be shared and should include both an expiration date and a statement assuring that the athlete will not be denied treatment for any injuries if they refuse to sign (Hill, 2003). There should also be consent forms for each new media entity that is brought in.

All collegiate sport organizations should periodically review existing policies, practices, and procedures to assure they are adhering to the guidelines of HIPAA. Clear consequences must also be communicated for inappropriate release of information.

Athlete Autonomy

Another pressing ethical implication of hyperquantifying athletes—and perhaps the most difficult to address—is the risk of compromising athlete autonomy. Robert Johnson, the UO's head track and field coach since 2012, says that "track is nothing but numbers… A good mathematician probably could be a good track coach." (Goe, 2021). But the drive to compete and success must be separate from these numbers. Johnson is a very powerful name in the running community, and it is difficult for any athlete who runs for him to speak up against his use of periodic DEXA scans to create individualized training plans. Autonomy is the capacity to make an informed, uncoerced decision. Sports analytics remains a valuable supplement for athletes to improve and to succeed, but it should not become the driver of performance. If biometric data is relied on too heavily for performance efforts or training tactics, athletes risk losing their autonomy and intuition (Vermeulen, 2018). Putting too much emphasis on the numbers can diminish the enjoyment of the sport, which is contradictory to the desired effect of sport participation in the first place for many athletes.

As mentioned previously, an athlete's weight or BMI is an element that many predictive models focus on, which can add to the list of demands that come with being an athlete that add pressure to want to lose weight. Women in particular are often required to wear physically revealing uniforms. The main instance of such that may come to mind are dance or cheer uniforms, but women runners are also required to wear what is called ‘runderwear’ during races, which is no larger than a swimsuit. This can push athletes to want to look a specific way because they are constantly in small, tight, revealing clothing. Comments from coaches or trainers can perpetuate these thoughts. This issue impacts all athletes of all genders. For runners specifically, there is a general consensus that being lower in weight will correlate to running faster, yet this is not necessarily a causal connection. Easy measurements to base training on are an athlete’s weight or BMI. But putting too much emphasis on an athlete’s weight can be harmful to the athlete’s physical and mental health. It also has the possibility of pushing athletes to develop eating disorders.

Disordered eating among athletes can range from subclinical behaviors to diagnosable clinical eating disorders. Disordered eating is defined as an abnormal eating behavior that can potentially become dangerous (Zucker, n.d.). Not every person who shows disordered eating habits will develop a clinical eating disorder. But if left untreated, it can lead to an eating disorder. One study followed a group of 496 adolescent girls in the general population for 8 years until they were 20 and found that 5.2% of the participants met criteria for clinical anorexia, bulimia, or binge eating disorder (Stice et al., 2009). This can be compared to the 27-32% of female athletes that can be classified as having a subclinical or symptomatic eating disorder (Reel et al., 2013). Another study found that 25.5% of collegiate athletes specifically exhibit *clinical* eating disorder symptoms (Greenleaf et al., 2009). Revealing team uniforms is a unique pressure that exists within sports environments that can contribute to the development of disordered eating habits. In addition, coaches or teammates can push the narrative that there will be performance advantages gained at a certain weight or body size and shape. This can come in the form of comments from peers, public weigh-ins, analyzing body fat percentage, and more. Eating disorders are common among the entire population, but athletes are subject to far different pressures that can perpetuate and encourage these behaviors.

The University of Oregon track and field and cross country team's usage of periodic DEXA scans has pushed current and former athletes to develop eating disorders and other destructive behaviors(Goe, 2021).

A main reason for the push to use data analytics in sport is prolonging athlete's careers through higher performance, but a big factor in this is making sure the athlete is still enjoying the sport. The athlete must have a drive to perform and compete separate from the analytics. Data analytics is an extremely valuable tool in improving athletes' performances, but athletes risk losing their autonomy and intuition when biometric data becomes the sole governor of performance efforts and training plans. Reducing an athlete to just numbers diminishes the enjoyment of the sport and may work contradictory to the desired effect of extending the athletes' career.

Moving Forward

The emerging craze with hyperquantification has the potential to become much more complex in the near future. Data analytics shows immense possibilities for athletes, coaches, trainers, and sports that were not possible with traditional methods. Through data science, raw data can be organized and transformed into useful mathematical and statistical models, which in turn can be used for innovative approaches to training. If hyperquantification is to come into its own, ethical implications must be considered. While analytics should be used as a tool to make informed decisions about training it should not become the ultimate decision maker.

Three ethical implications detailed in this thesis were data reliability, data security, and athlete autonomy. Collegiate sport researchers, trainers, and coaches must consider all ethical implications of implementing data analysis into their organizations to assure the athlete is the center of all decisions being made. Organizations should assure that all parties dealing with data are knowledgeable about the responsible use of the technologies used. Athletes should be made aware it is in their best interest to share their data and allow for its application in their training. To do so, the program should make sure that the athlete is involved in decisions being made about their training and their data.

The algorithm created for the University of Oregon Sports Performance Center shows how complex data analysis now is. It also is becoming much more accessible through the use of biometric wearables. Trainers and coaches now can easily have access to a plethora of metrics simply through wearing a watch or running an algorithm to analyze a trial run. While these models are already complex and sophisticated algorithms, there is always other levels or variables that can be added. This could include adding a holistic view of the athlete and accounting for emotional wellbeing.

As mentioned earlier in this thesis, more communication between researchers and athletes should be encouraged. This would aid in allowing athletes to better understand what is happening to their data and information one it is taken. Recently, the NCAA (National Collegiate Athletic Association) changed a policy allowing student athletes to profit from their own image. This essentially gives athletes the ability to create their own 'brand' much like professional athletes are able to do. This is hopefully just the start of a much larger movement to give athletes the rights to and control over their own data.

Collegiate sports are in the unique position of working with adult athletes who have limited experience in the professional world, on top their bodies and personalities not being completely matured. Structured professional programs focused on performance and winning is new and can be intimidating. All involved in the program should make it their top priority to assure the athlete is prepared emotionally for the pressure of performance and the strain being on a highly competitive team will take on them. There is immense value in building a program designed to focus on the whole athlete. Sports are not just physical but are psychological too. Both aspects must be catered to in attempts to hyperquantify the athletes.

Hyperquantification brings the power of Big Data to solve the problem of how to best improve athlete performance. However, it is only focused on the objective physical outcomes of the athlete and neglects to consider the emotional wellbeing of the athlete. Additions to the hyperquantification approach that add value to the emotional and mental performance aspects would not only better protect the athlete but also can find elements that better translate into success in performance. The athlete should be the center of every decision and conversation. Trust must be built and maintained among all involved because, without it, potential gains may never be fully realized.

Appendix

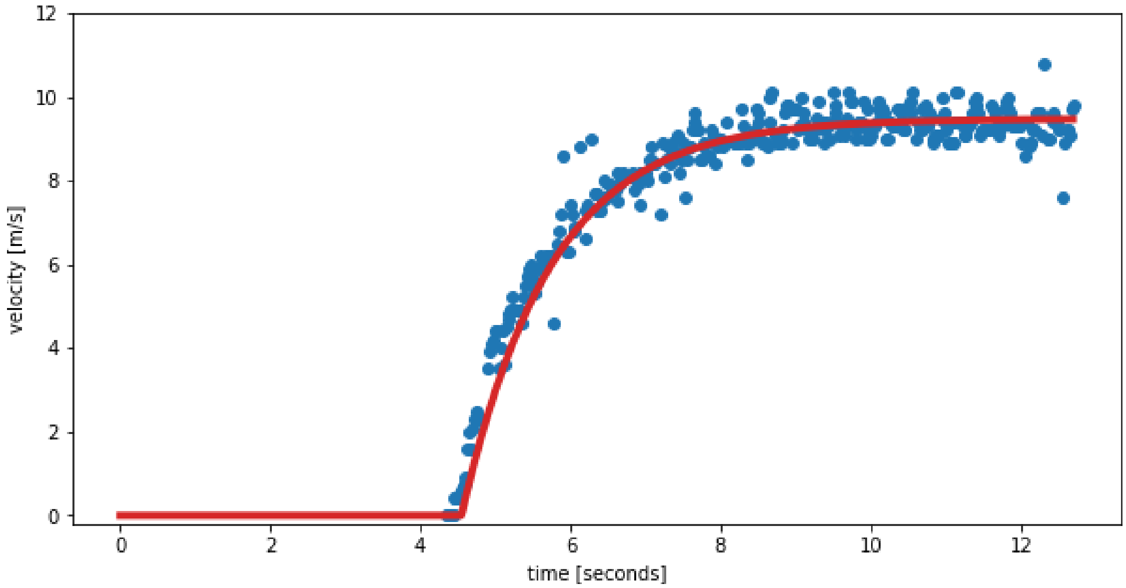
Python Code

```
import os
import numpy as np
import numpy.ma as ma
import bisect
from scipy import fftpack
from scipy import signal
import matplotlib.pyplot as plt  # To visualize
import pandas as pd  # To read data
from sklearn.linear_model import LinearRegression
from scipy.optimize import curve_fit
from xlsxwriter.utility import xl_rowcol_to_cell
import nbconvert

sl = pd.read_excel('sprintLookups.xlsx')
df = pd.DataFrame(sl)

dirname = os.getcwd()          # giving directory name

ext = ('.rda')    # giving file extension

Output = pd.DataFrame(columns=['student','trial#','maxV','maxV position','maxV
time','split_5m','split_10m','split_20m','split_30m','Start Time','Tau'])

# iterating over all files
for filename in os.listdir(dirname):
   if filename.endswith(ext):
      filepath = os.path.join(dirname, filename)
      tf = filename.split("_")
      playerid = tf[0]
      trialnum = tf[1].split(".")[0]
      df[sl['playerName'] == playerid].index[0]
      idcode = sl['idCode'][df[sl['playerName'] == playerid].index[0]]
      print("processing",filename)

      raddata = pd.read_csv(filename)  # load data
      vel_raw = raddata.iloc[:, 0].values.reshape(-1)  # values converts it into a
numpy array
      t_tossout = 0.5 #### specify interval of time to zero out at the beginning
(always seems to be initial spikes)
```

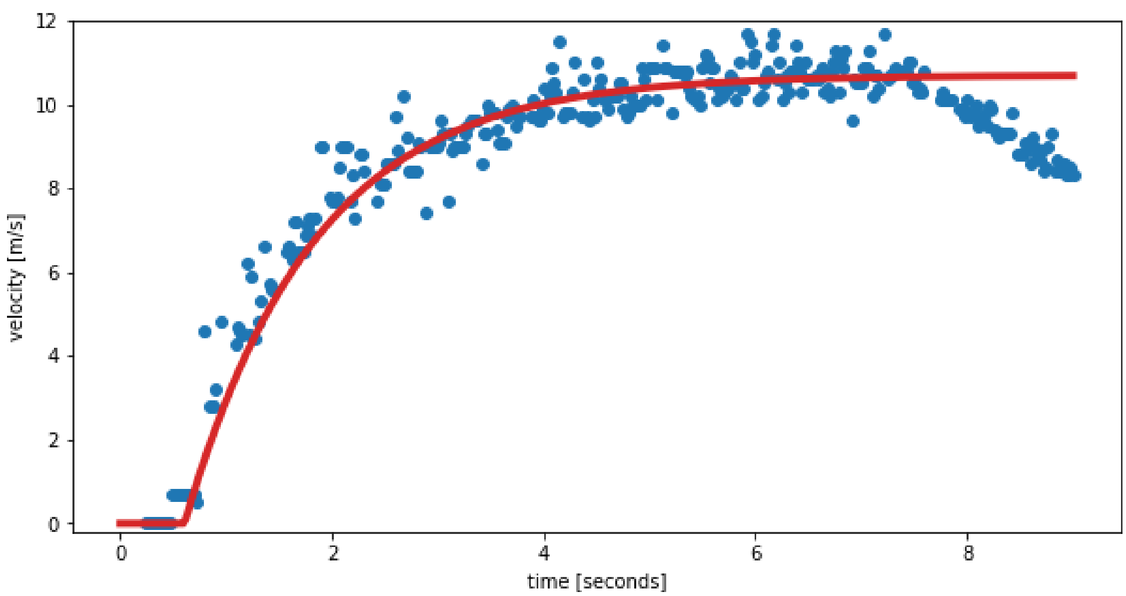
```
    vel_raw[:int(round(t_tossout*fs,0))] = 0 # sets an equivalent number of data
points to zero at the beginning
    # by the way, that use of the colon in the line above is called a "slice operation"
- search for that to figure out what it does
    t = np.linspace(0,(vel_raw.size-1)*dt,vel_raw.size) # time series to plot against

    mass = sl['bodyMass'][df[sl['playerName'] == playerid].index[0]]

#       # to automatically identify the start, calculate the energy but first shift the
vel data by 1 pixel to cancel out outliers
#       KE = 0.5*mass*vel_raw*shift(vel_raw, 1)
#       # Find the sprint start time using a kinetic energy threshold (since KE is
smoother) and trim
#       KE_start_threshold = 30 #### used to determine the start of sprint
#       rise_time_to_threshold = 0.15 # estimated time to reach threshold (just from
eyeballing a few runs)
#       start_index = np.argmax(KE > KE_start_threshold)-
int(round(rise_time_to_threshold*fs,0)) ## shift back slightly earlier to capture the
true start of the sprint
#       start_time = start_index*dt

    pars, cov = curve_fit(f=initial_v_fit, xdata=t, ydata=vel_raw)
    start_time = pars[0]
    start_index = int(round(start_time*fs,0)) # calculates the number of pixels
corresponding to start_time

    vel_trim = vel_raw.copy()
    vel_trim[:start_index] = np.NAN
    KE[:start_index] = 0

    accel_raw = np.gradient(vel_trim, dt) # calculate the raw acceleration

    t_smooth = 0.5 #### set the time interval over which to average various
signals
    w = int(round(t_smooth*fs,0)) # calculates the number of pixels
corresponding to t_smooth
    accel_smooth = moving_average(accel_raw,w)
    accel_smooth[:w] = np.nan # smoothing messes up the beginning, so set this
back to zero

    # Remove outlier velocities at times with too-large an acceleration, TO DO:
remove both rise and fall values around outliers
    accel_min = -20
    accel_max = 30
    mm = (accel_raw < accel_min) | (accel_raw > accel_max)
```

```
    mm2 = shift(mm,-1) ## shifts the mask slightly earlier to capture the true start
of the sprint
    accel_filtered = ma.array(accel_raw, mask = mm)

    vel_OR = ma.array(vel_trim, mask = mm) # "outliers removed"

    # to further smooth data, apply a low-pass filter and then compute a moving
average
    #following R. A. Poehling, "Monitoring explosive performances in relation to
training load accumulation in adolescent female soccer players," University of
British Columbia (2018).
    f_nyq = 0.5 * fs #Nyquist frequency
    f_hi_cut = 10 #cut down on frequencies over this many Hz
    b, a = signal.butter(6, f_hi_cut/f_nyq, btype='low', analog=False) #applies a
Butterworth low-pass filter
    vel_filtered = signal.lfilter(b, a, vel_OR) #this doesn't work well with NaNs

    t_smooth = 0.3
    w=int(round(t_smooth*fs,0))
    vel_smooth = moving_average(vel_OR,w) # used vel_filtered instead of vel_OR
    vel_smooth[:w] = np.NAN

    t_smooth = 0.3
    w=int(round(t_smooth*fs,0))
    KE_smooth = moving_average(0.5*mass*vel_smooth**2,w)

    #KE_smooth = 0.5*mass*vel_smooth**2

    power_smooth = np.gradient(KE_smooth, dt) # calculate the power
    pos = np.cumsum(vel_smooth)*dt
    n_max = np.argmax(vel_smooth)
    max_vel = vel_smooth[n_max]
    max_v_time = n_max*dt - start_time
    max_v_pos = pos[n_max]
    print("Maximum velocity of" , round(max_vel,2) , "m/s achieved at" ,
round(max_v_pos,1) , "meters," , round(max_v_time,2) , "s after start.")

    split_times = (find_split_times(pos,[5,10,20,30],dt)) #TO DO: add error bars
    print(split_times)

    valid_range = ~(np.isnan(t) | np.isnan(vel_smooth))
    valid_range[n_max:] = False
    pars, cov = curve_fit(f=model_v, xdata=t[valid_range],
ydata=vel_smooth[valid_range], p0=[max_vel, start_time,1.5],
bounds=[[0,0,0],[np.inf,np.inf,np.inf]])
```

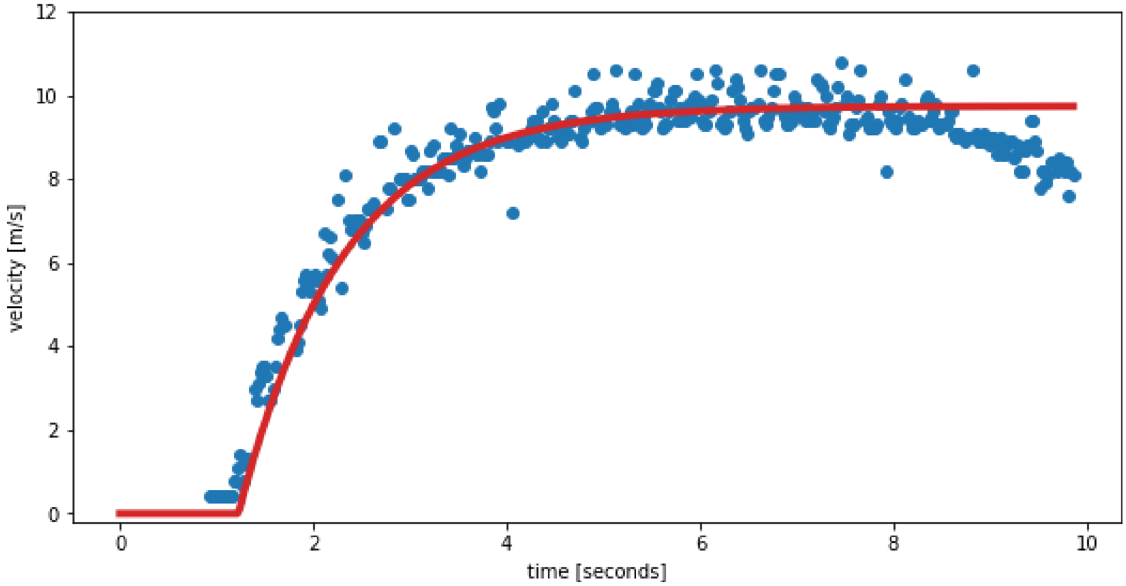
```
print(pars)

      best_fit_v = model_v(t,*pars)

      fig = plt.figure()
      ax1 = fig.add_subplot(111)
      ax1.scatter(t,vel_OR)
      #ax1.plot(t,vv)
      ax1.plot(t,best_fit_v,linewidth=4,color='tab:red')
      plt.xlabel('time [seconds]')
      plt.ylabel('velocity [m/s]')
      #plt.xlim(left = start_time - 0.5)
      #plt.xlim((start_time - .1,start_time + 1))
      plt.ylim((-.2,12))
      #plt.xlim((0,15))
      plt.show()

      Output = Output.append(
         {'student': [playerid], 'trial#':[trialnum], 'maxV':[pars[0]], 'maxV
position':[max_v_pos], 'maxV time':[max_v_time], 'split_5m':[split_times[0]],
'split_10m':[split_times[1]], 'split_20m':[split_times[2]],
'split_30m':[split_times[3]], 'Start Time':[pars[1]], 'Tau':[pars[2]]}, ignore_index =
True)

   else:
      continue

processing 211112-ZHXMN_1.rda
Maximum velocity of 9.58 m/s achieved at 46.0 meters, 6.23 s after start.
[5.862410643581034, 6.558158720867775, 7.769197567726959,
8.878253201978632]
[9.47654202 4.54734775 1.19844342]
```



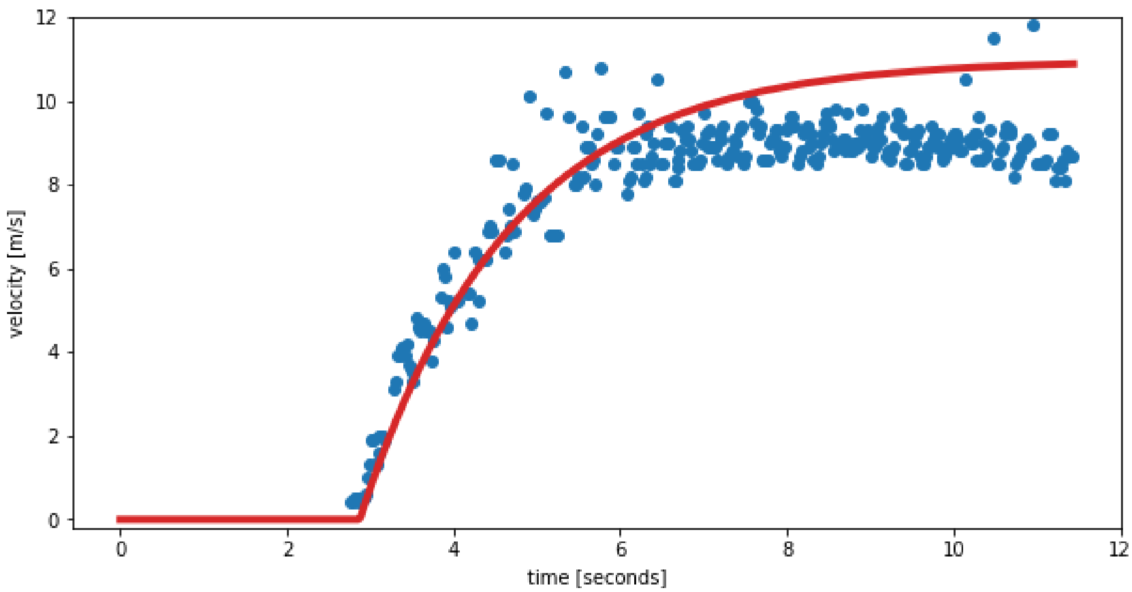
processing 211112-LNOIX_1.rda
Maximum velocity of 11.23 m/s achieved at 63.1 meters, 7.5 s after start.
[1.8422684225350892, 2.469820735283941, 3.594730568558211, 4.6023701907237715]
[10.70154435 0.60860214 1.23097145]



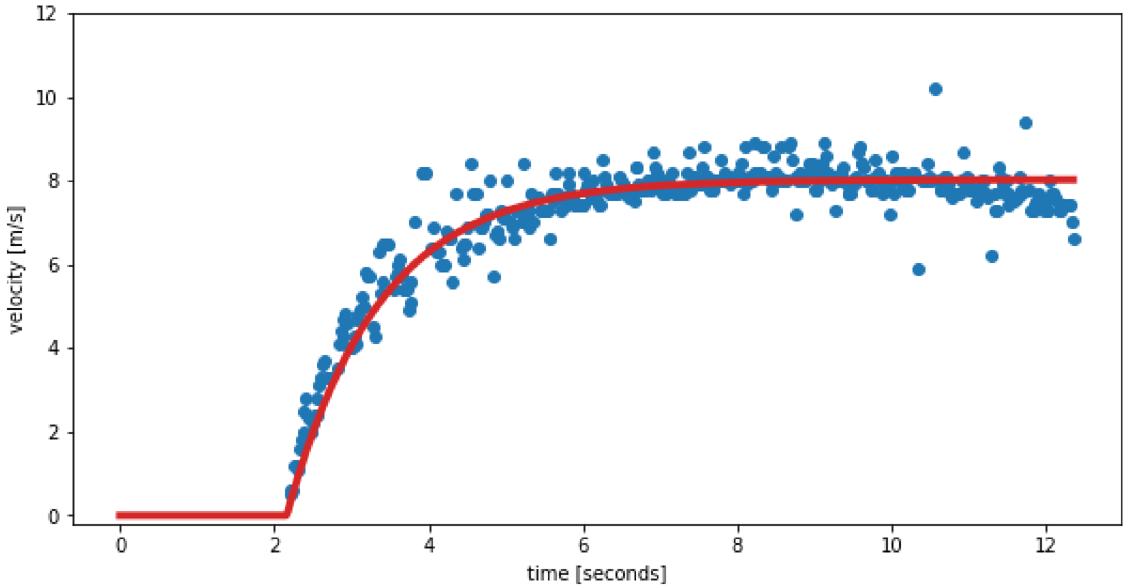
processing 211112-MBXYE_2.rda
Maximum velocity of 9.88 m/s achieved at 50.8 meters, 6.57 s after start.
[2.4511274645857366, 3.111725153503772, 4.273884738819166, 5.345080387830189]
[9.74078894 1.22471256 1.080065]



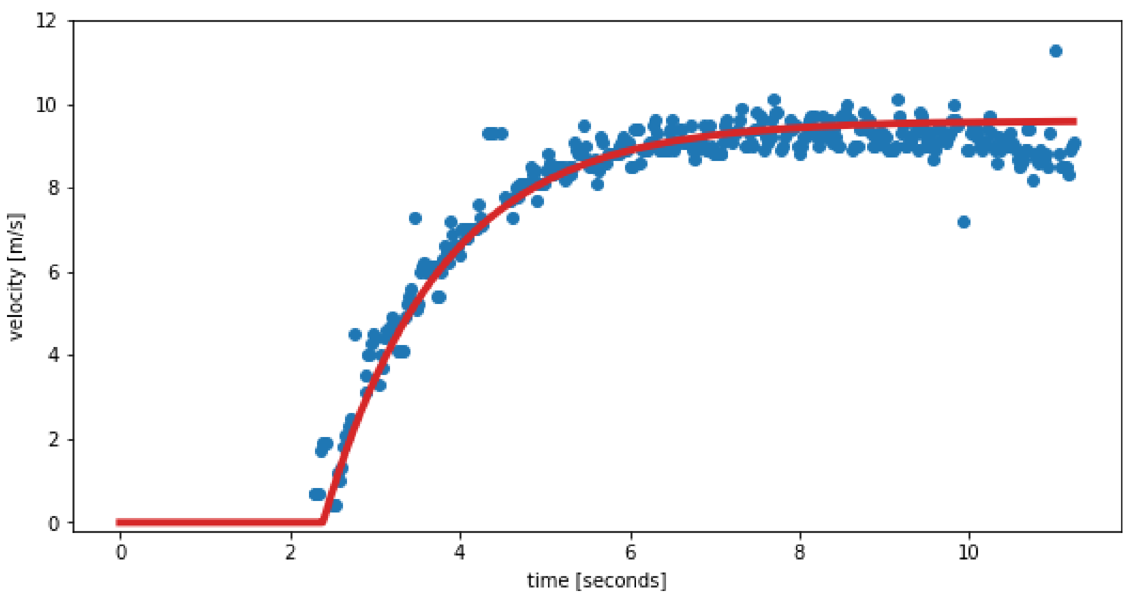
processing 211112-EWPAR_2.rda
Maximum velocity of 9.44 m/s achieved at 18.3 meters, 3.25 s after start.
[4.310395329949692, 5.020247452833868, 6.203319682375238, 7.3337126615180095]
[10.97955594 2.874695 1.79754371]



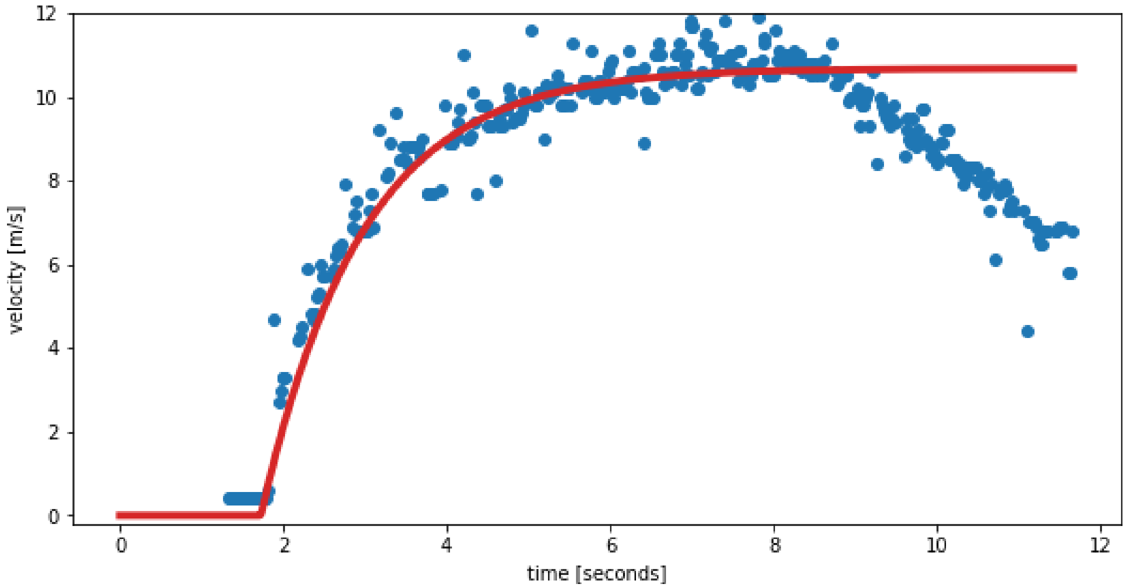
processing 211112-RNAEW_1.rda
Maximum velocity of 8.31 m/s achieved at 40.5 meters, 6.18 s after start.
[3.610130429250651, 4.395422551522551, 5.797637470305021, 7.0826441318834865]
[8.01889907 2.15994117 1.19728387]



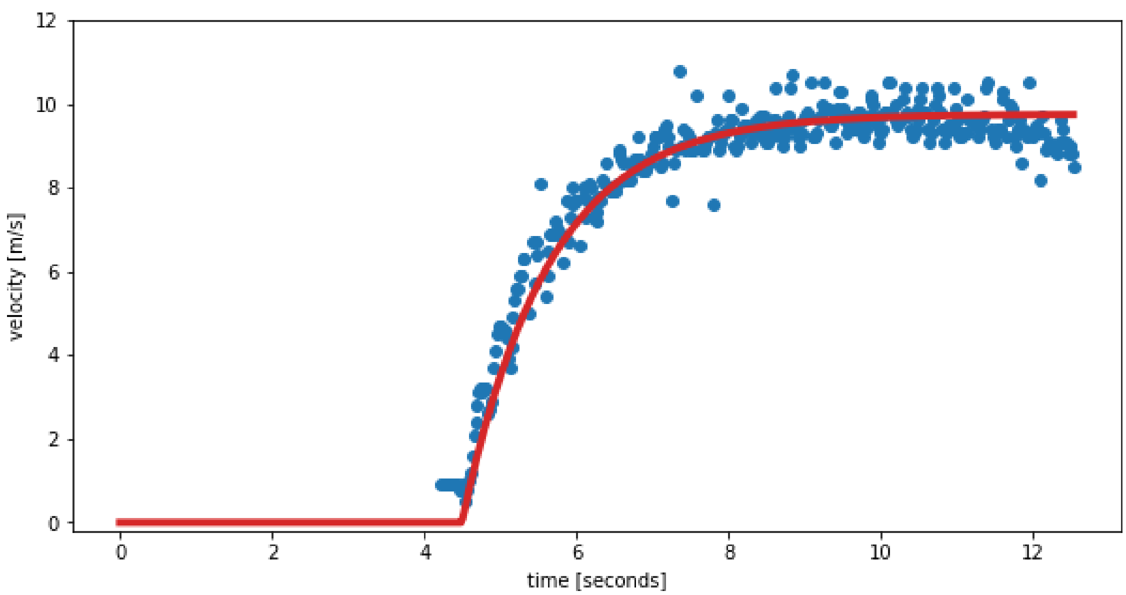
processing 211112-ZHXMN_2.rda
Maximum velocity of 9.5 m/s achieved at 38.8 meters, 5.48 s after start.
[3.784981934263862, 4.506266528899426, 5.702669508680955, 6.815387258973464]
[9.59983008 2.38814585 1.3880434]



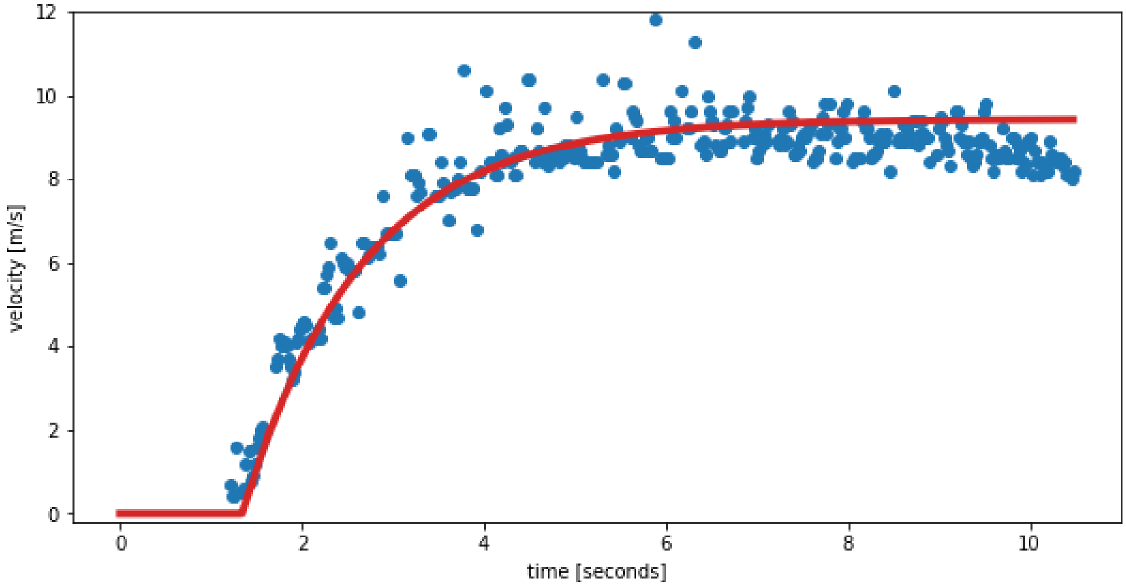
processing 211112-LNOIX_2.rda
Maximum velocity of 11.05 m/s achieved at 48.1 meters, 6.11 s after start.
[2.937668829535689, 3.5811996610165187, 4.705484640152255, 5.712964828158745]
[10.6854875 1.72453209 1.23948407]



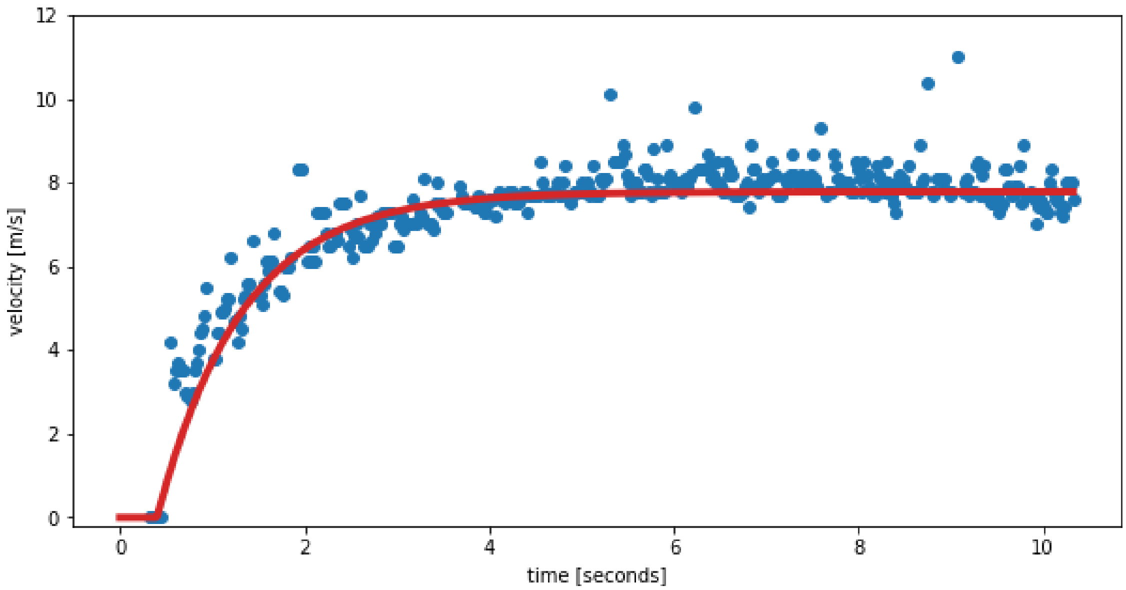
processing 211112-MBXYE_1.rda
Maximum velocity of 9.91 m/s achieved at 46.2 meters, 6.11 s after start.
[5.719709863233543, 6.407445560174157, 7.567625082671959, 8.646253548102003]
[9.75499524 4.49135816 1.14085167]



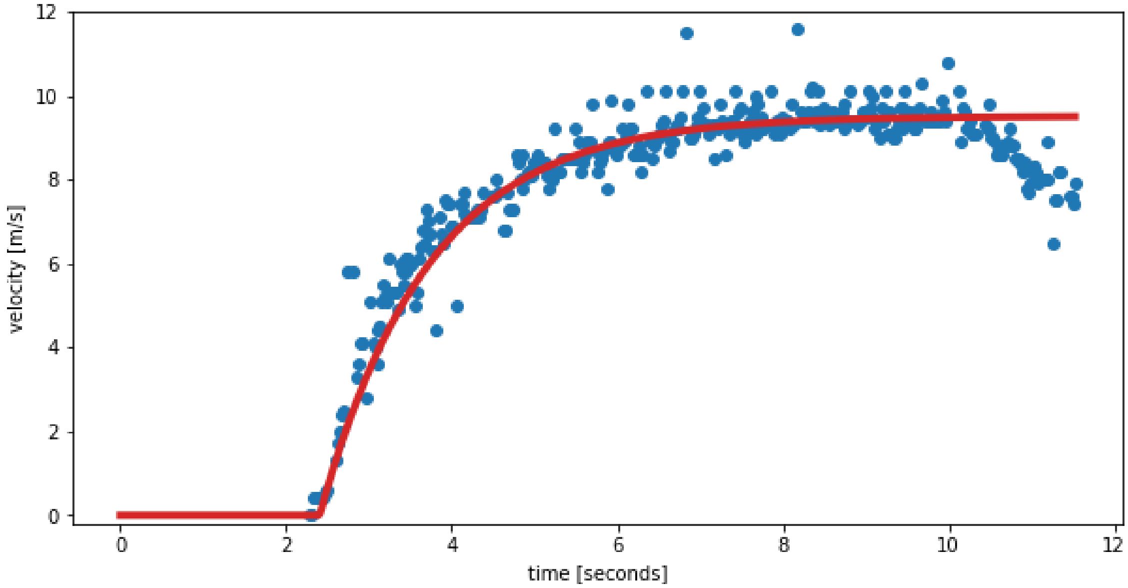
processing 211112-EWPAR_1.rda
Maximum velocity of 9.69 m/s achieved at 35.4 meters, 5.15 s after start.
[2.7123838428737304, 3.4383911728108183, 4.637744580955108, 5.769779009313127]
[9.42965918 1.34279601 1.3128012]



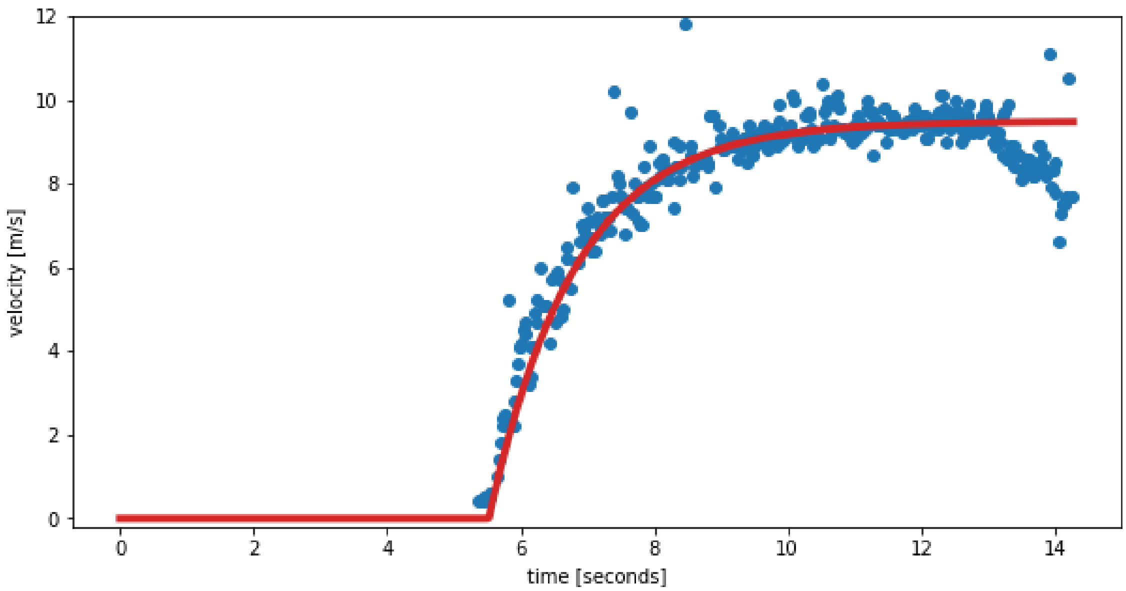
processing 211112-RNAEW_2.rda
Maximum velocity of 8.52 m/s achieved at 32.5 meters, 5.17 s after start.
[1.7354552836377397, 2.4922026201003047, 3.875388843219274, 5.175099808320937]
[7.78177447 0.40096733 0.92126755]



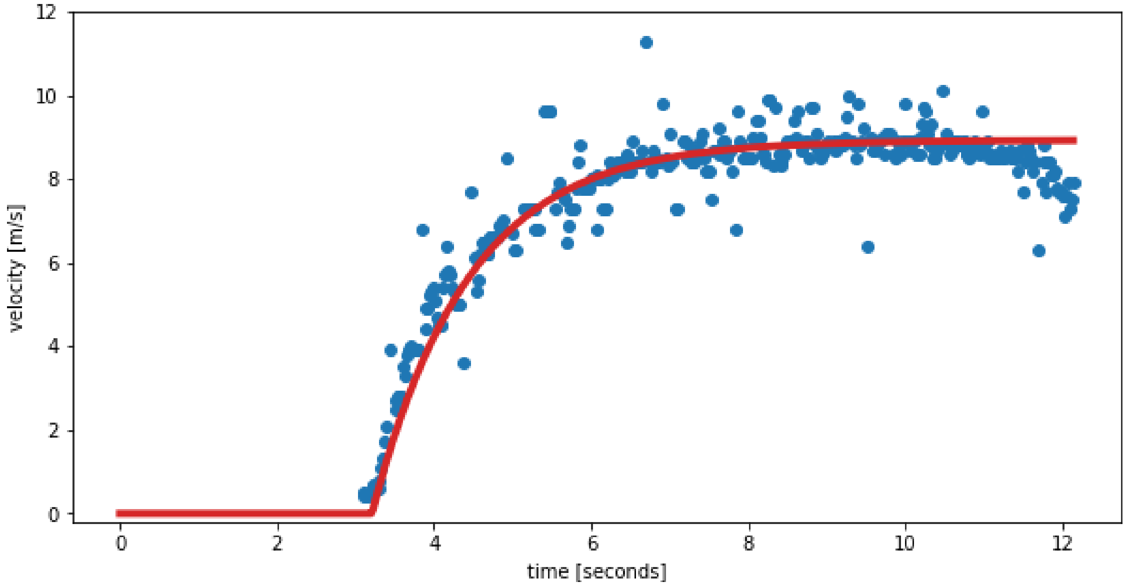
processing 211112-GQOTI_1.rda
Maximum velocity of 9.82 m/s achieved at 44.9 meters, 6.14 s after start.
[3.757155441575515, 4.489360161097147, 5.727359431346788, 6.847681096312514]
[9.5127101 2.40596811 1.33003974]



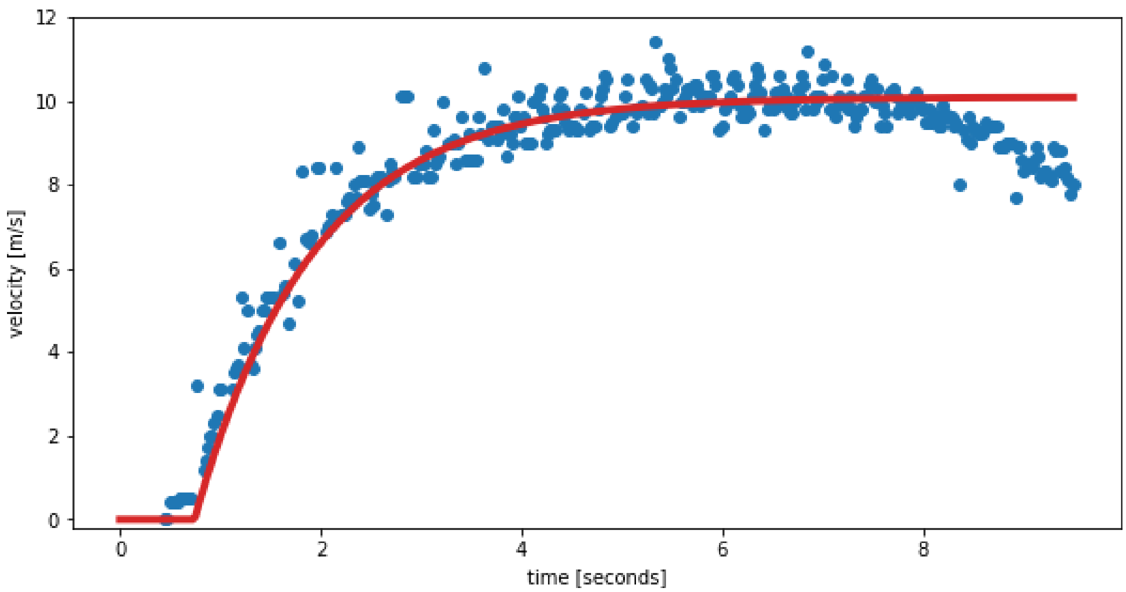
processing 211112-GQOTI_2.rda
Maximum velocity of 9.65 m/s achieved at 37.9 meters, 5.39 s after start.
[6.878661292299847, 7.589523083419797, 8.809193008773226, 9.926469026159026]
[9.48641986 5.52286686 1.29549176]



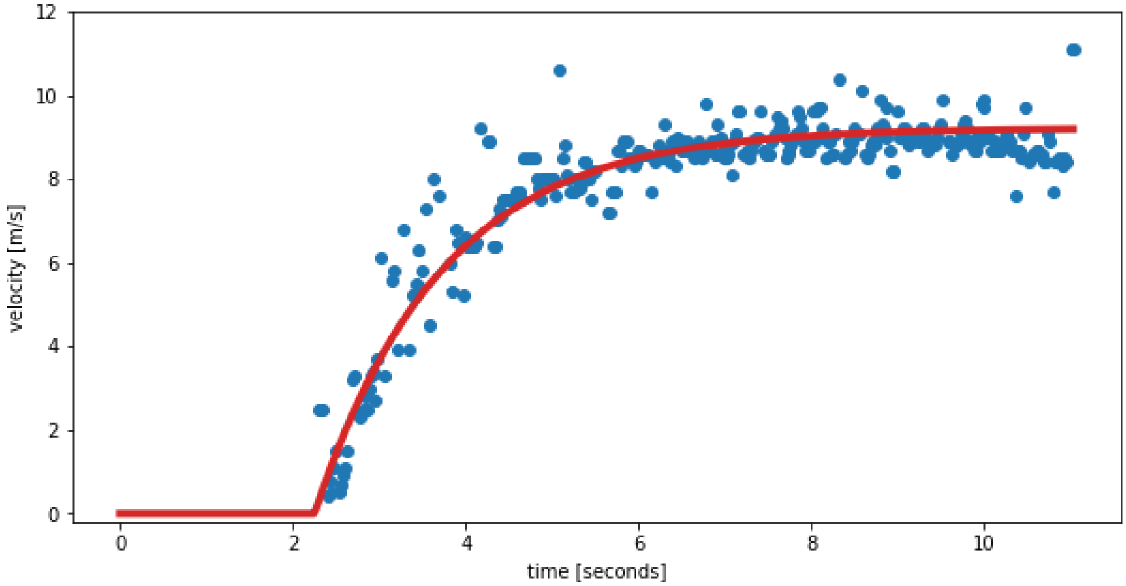
processing 211112-DXHEL_1.rda
Maximum velocity of 9.12 m/s achieved at 45.3 meters, 6.39 s after start.
[4.580354085332063, 5.332514888789014, 6.595204824226405, 7.763639176870458]
[8.92546887 3.2079473 1.23311266]



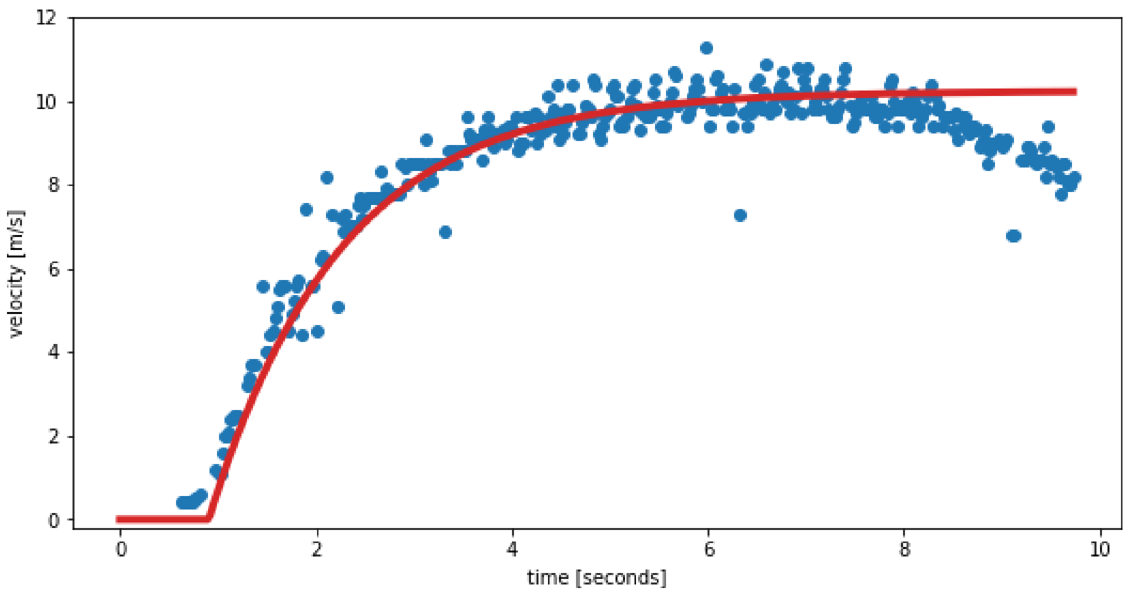
processing 211112-XXRRC_2.rda
Maximum velocity of 10.39 m/s achieved at 37.0 meters, 5.11 s after start.
[2.006472535694758, 2.663036834559628, 3.797145278573544, 4.848919809754241]
[10.08627269 0.74079358 1.18184778]



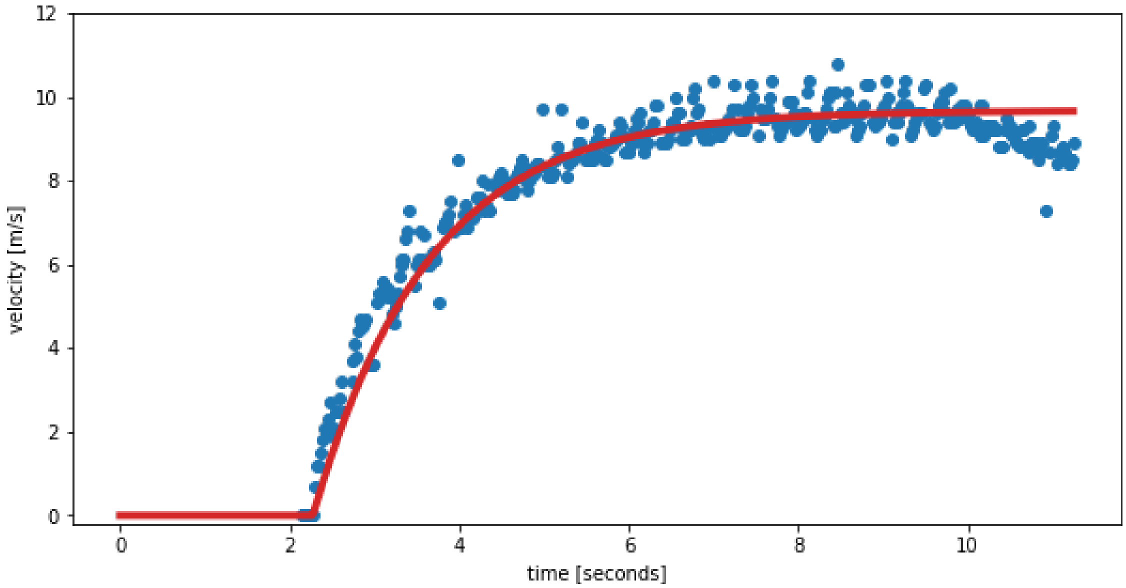
processing 211112-DXHEL_2.rda
Maximum velocity of 9.22 m/s achieved at 40.9 meters, 5.83 s after start.
[3.736660456411752, 4.476199242567029, 5.732958633184076, 6.903177558996769]
[9.21939496 2.25026503 1.47649012]



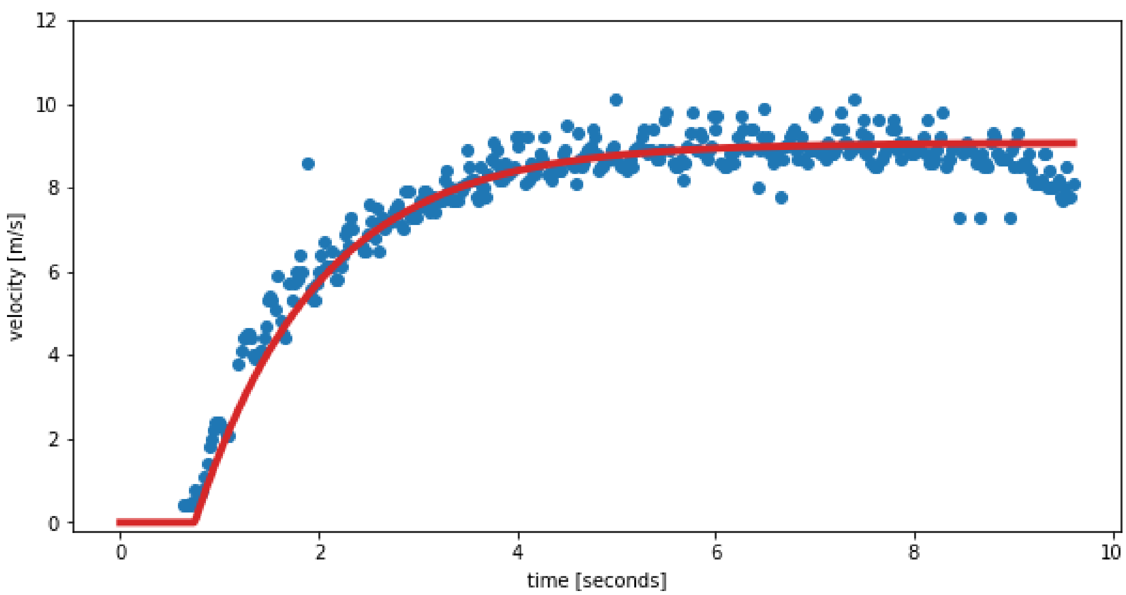
processing 211112-XXRRC_1.rda
Maximum velocity of 10.25 m/s achieved at 40.3 meters, 5.53 s after start.
[2.2114388528290165, 2.8996382643429577, 4.054015831481461, 5.102778965356603]
[10.23218561 0.9047202   1.34675378]



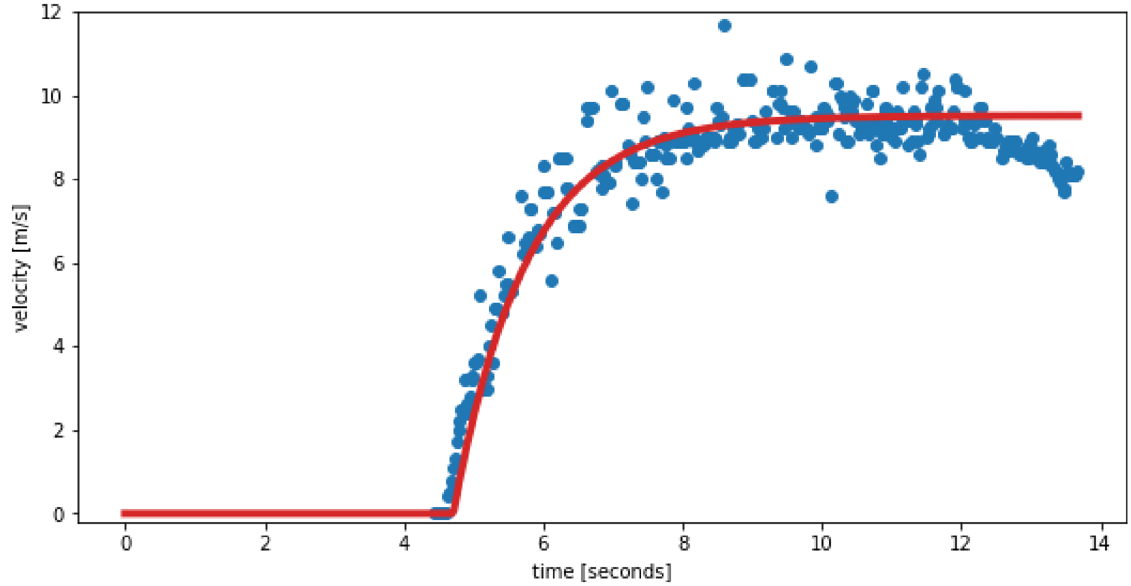
processing 211112-TZGGY_2.rda
Maximum velocity of 9.84 m/s achieved at 48.3 meters, 6.48 s after start.
[3.6354950774302552, 4.3674060648017035, 5.58567934665807, 6.694699761144205]
[9.67087876 2.27510414 1.3703628]



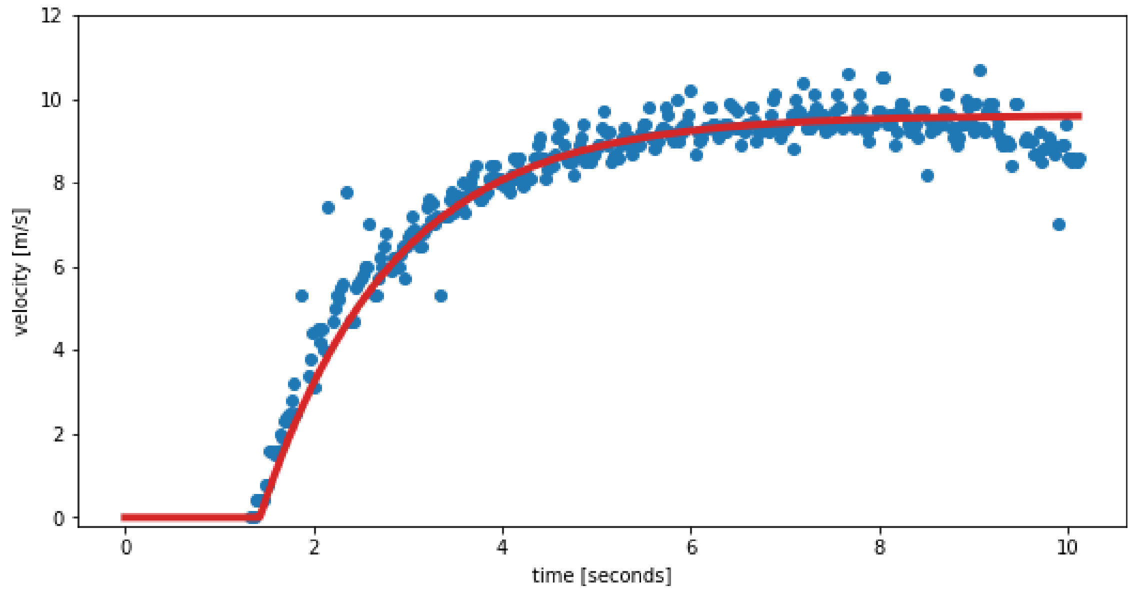
processing 211112-PAWAN_1.rda
Maximum velocity of 9.26 m/s achieved at 50.0 meters, 6.85 s after start.
[2.1016275247831406, 2.8528037952513206, 4.111654974953095, 5.261521885521887]
[9.0695586 0.7525935 1.24675761]



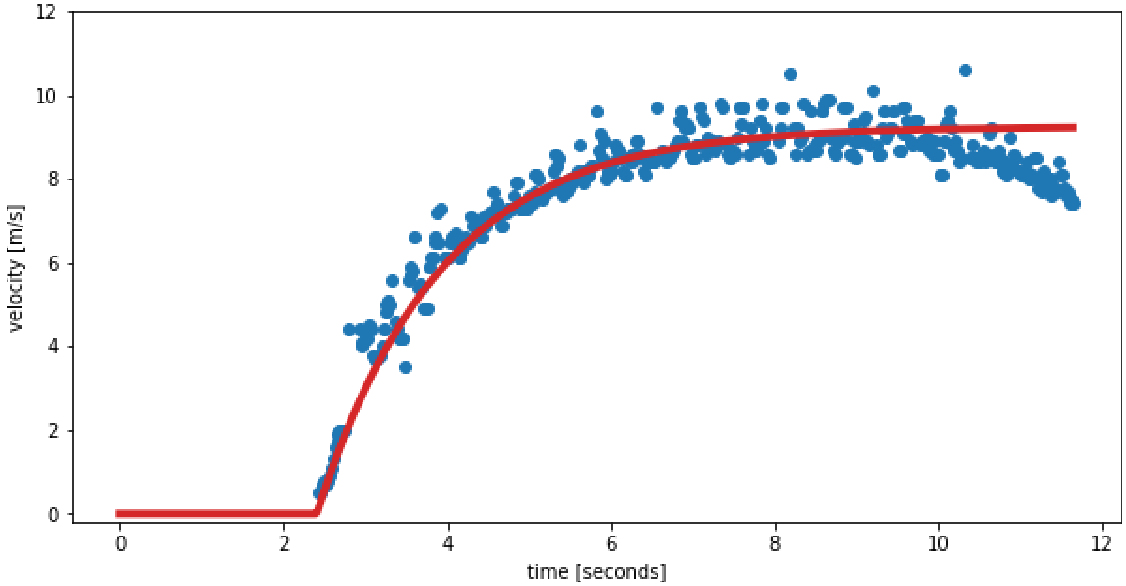
processing 211112-ENCPU_1.rda
Maximum velocity of 10.1 m/s achieved at 35.8 meters, 5.05 s after start.
[5.9559610387951185, 6.630896432222747, 7.79249252442202, 8.894973962536112]
[9.50965086 4.70899773 1.04862792]



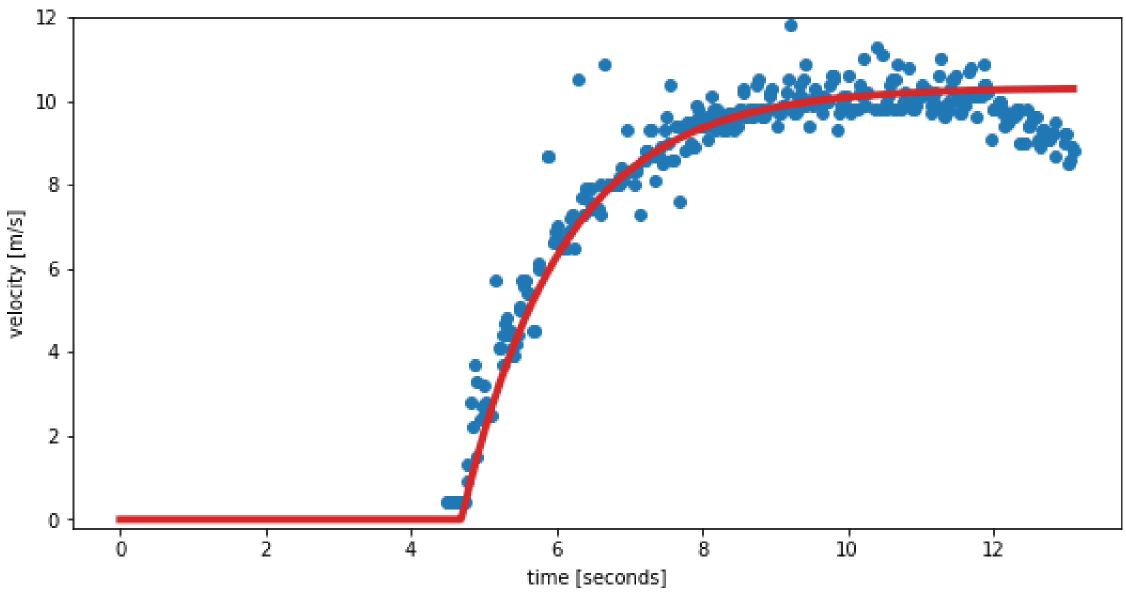
processing 211112-TZGGY_1.rda
Maximum velocity of 9.76 m/s achieved at 48.0 meters, 6.48 s after start.
[2.8010603674963495, 3.5557536999310675, 4.785946737752799, 5.900039718048335]
[9.61229582 1.42852189 1.40992553]



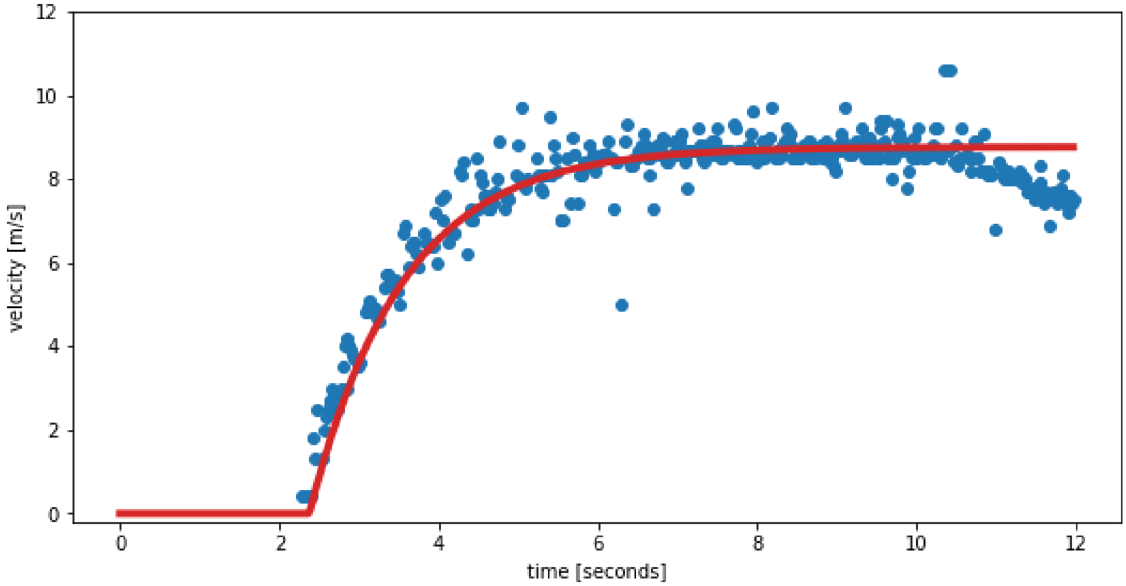
processing 211112-PAWAN_2.rda
Maximum velocity of 9.53 m/s achieved at 40.2 meters, 5.81 s after start.
[3.8803282215640107, 4.646980107570564, 5.937947213558101, 7.100103788437853]
[9.24791956 2.40621305 1.52242909]



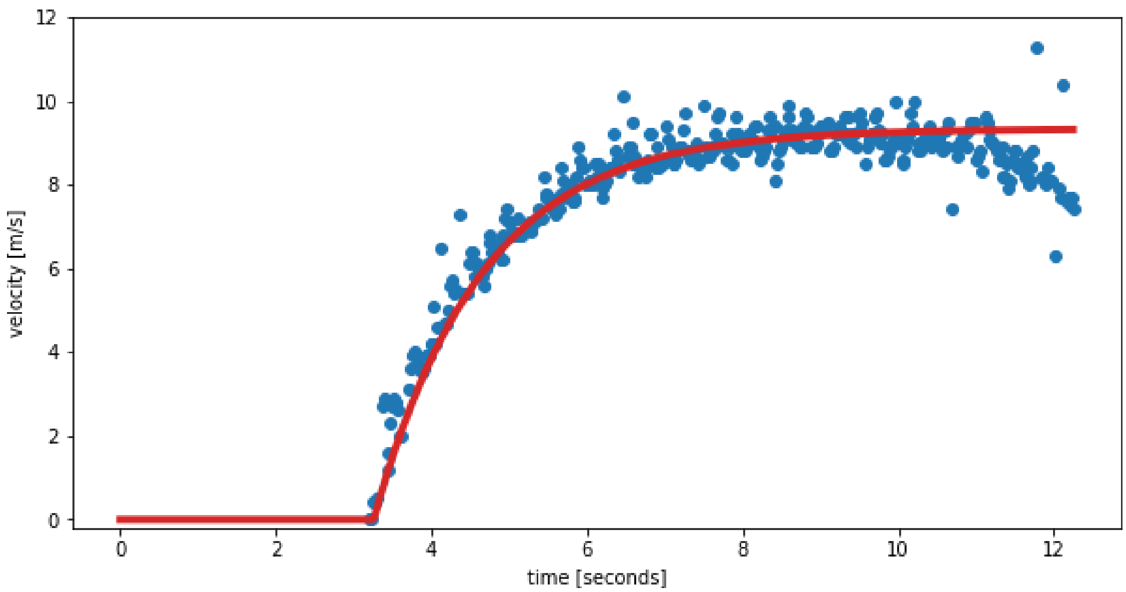
processing 211112-DKISE_2.rda
Maximum velocity of 10.44 m/s achieved at 45.9 meters, 6.01 s after start.
[6.016966939525824, 6.698986883350136, 7.8697876156939985, 8.915812178496441]
[10.31849683 4.68905847 1.39877898]



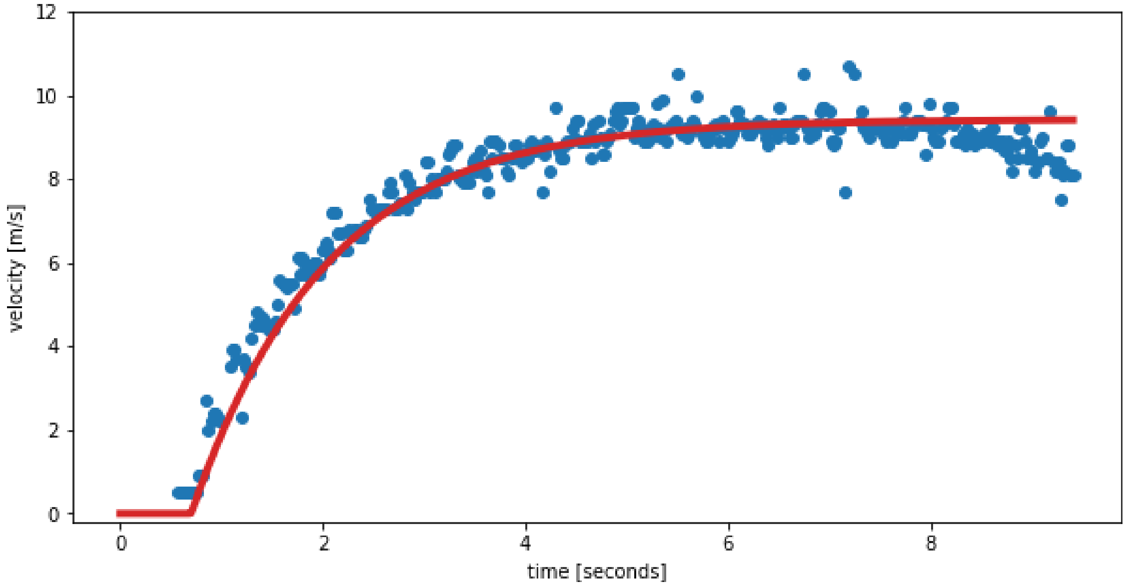
processing 211112-NFJWX_1.rda
Maximum velocity of 9.14 m/s achieved at 61.0 meters, 8.22 s after start.
[3.744828870362312, 4.483630987939909, 5.745528808424101, 6.943802344790834]
[8.76172329 2.37503114 1.1763516]



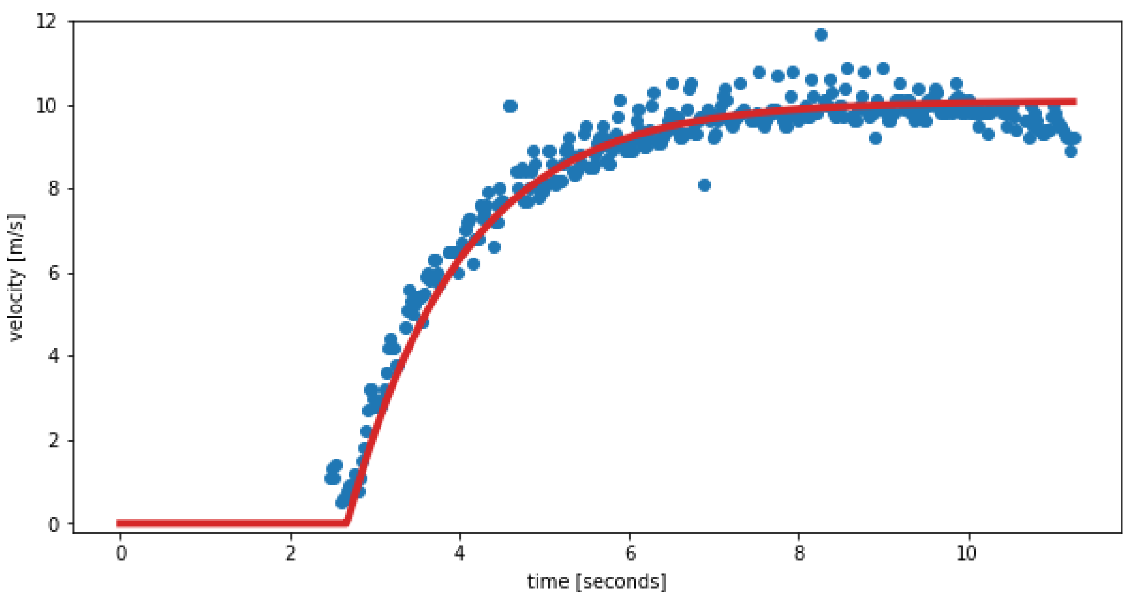
processing 211112-KLRXJ_2.rda
Maximum velocity of 9.32 m/s achieved at 47.7 meters, 6.55 s after start.
[4.6680713120371236, 5.425247205469428, 6.667336057267617, 7.804492615193252]
[9.32594544 3.25766216 1.39782616]



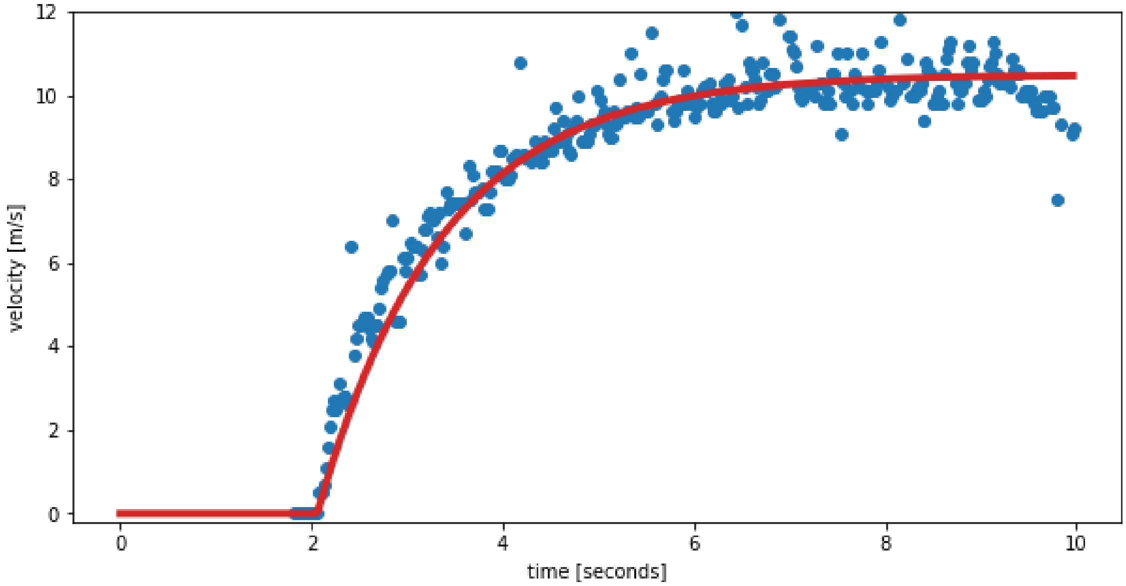
processing 211112-UJNPL_2.rda
Maximum velocity of 9.52 m/s achieved at 51.2 meters, 6.86 s after start.
[2.0724198361444266, 2.8053956724671822, 4.036741748872743, 5.156975109565728]
[9.42566441 0.70209478 1.33554567]



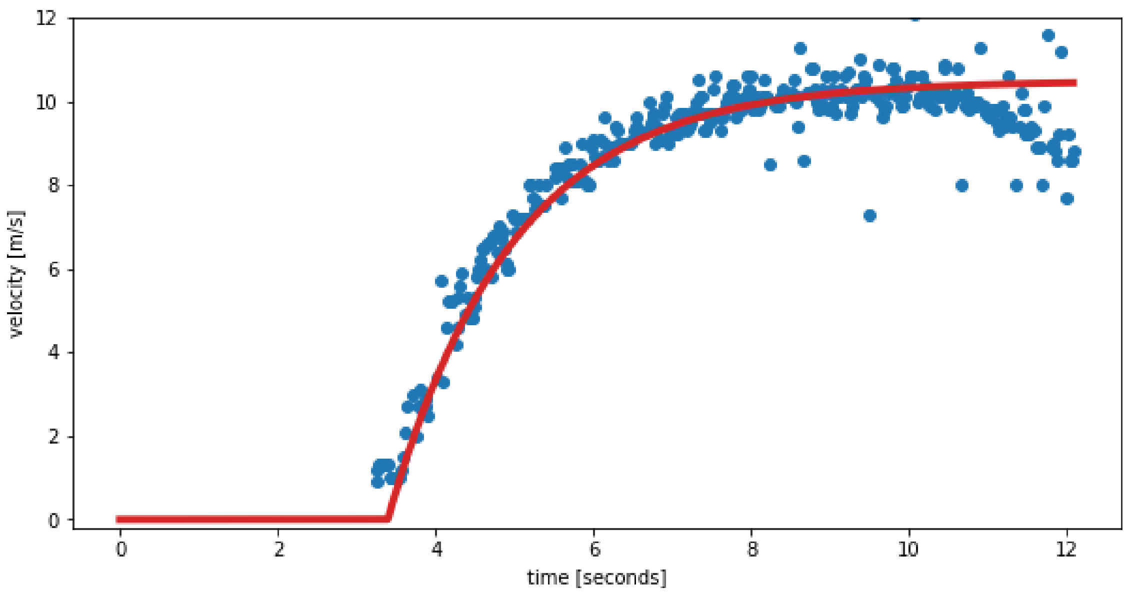
processing 211022-WAPPF_1.rda
Maximum velocity of 10.22 m/s achieved at 44.7 meters, 5.94 s after start.
[3.974350995506551, 4.680397758539567, 5.853832157816944, 6.920639348057943]
[10.08125402 2.67707283 1.35347547]



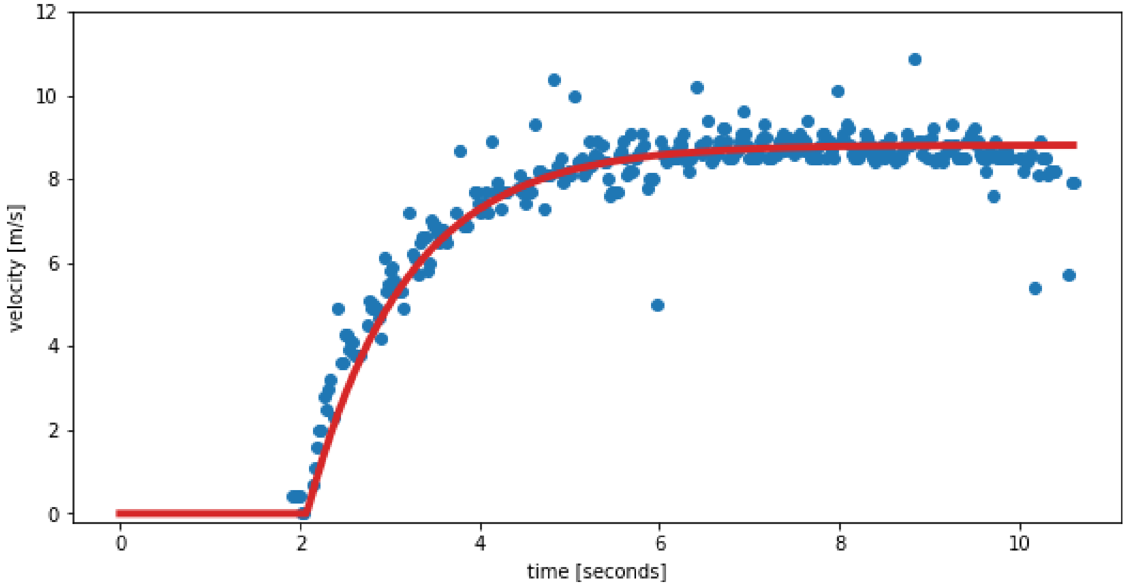
processing 211112-ERQZA_2.rda
Maximum velocity of 10.91 m/s achieved at 39.2 meters, 5.26 s after start.
[3.3313004231834134, 4.012856422554265, 5.148409160305343, 6.164657216693242]
[10.50083294 2.05983208 1.30561924]



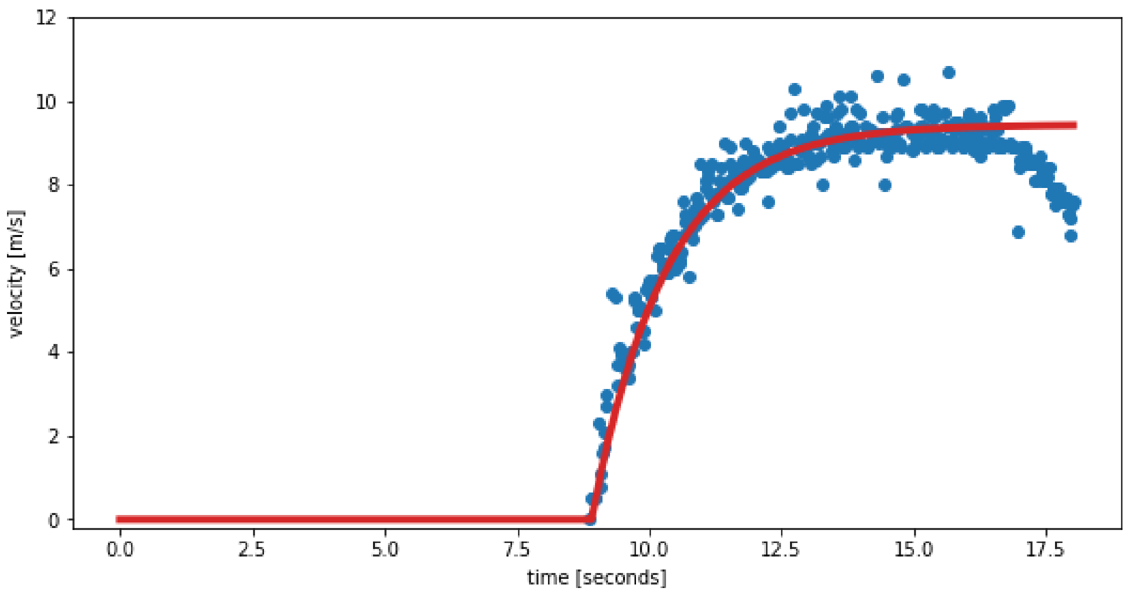
processing 211112-DKISE_1.rda
Maximum velocity of 10.46 m/s achieved at 55.5 meters, 6.97 s after start.
[4.774960715347518, 5.488091342073136, 6.653299061281595, 7.699048907160856]
[10.48400422 3.39125806 1.58346493]



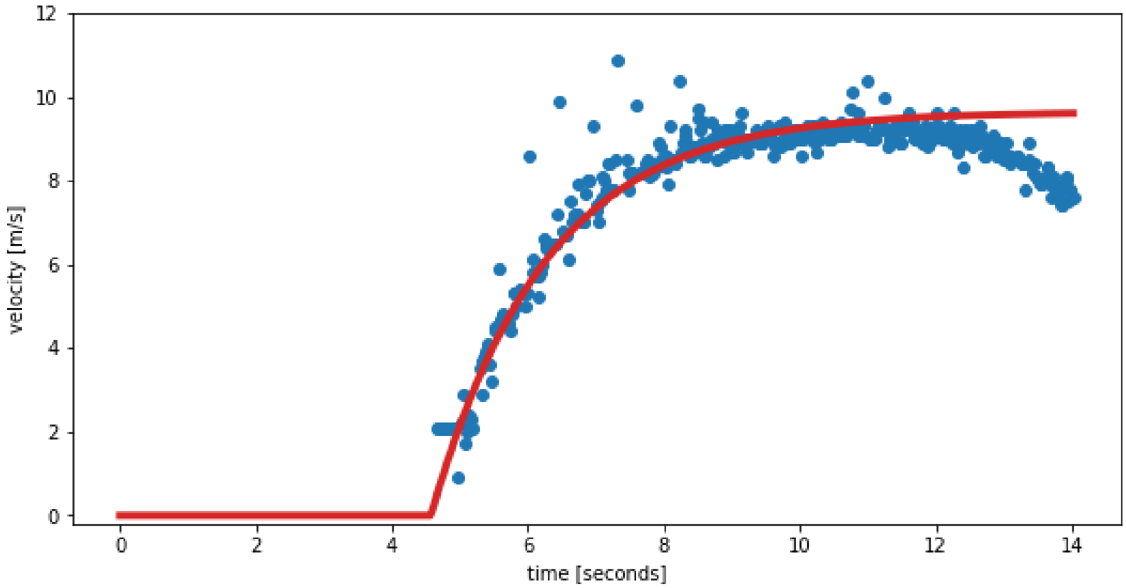
processing 211112-NFJWX_2.rda
Maximum velocity of 9.02 m/s achieved at 51.9 meters, 7.12 s after start.
[3.399201525739007, 4.124944914207513, 5.36380811377175, 6.55499674045761]
[8.81459449 2.07700832 1.09859102]



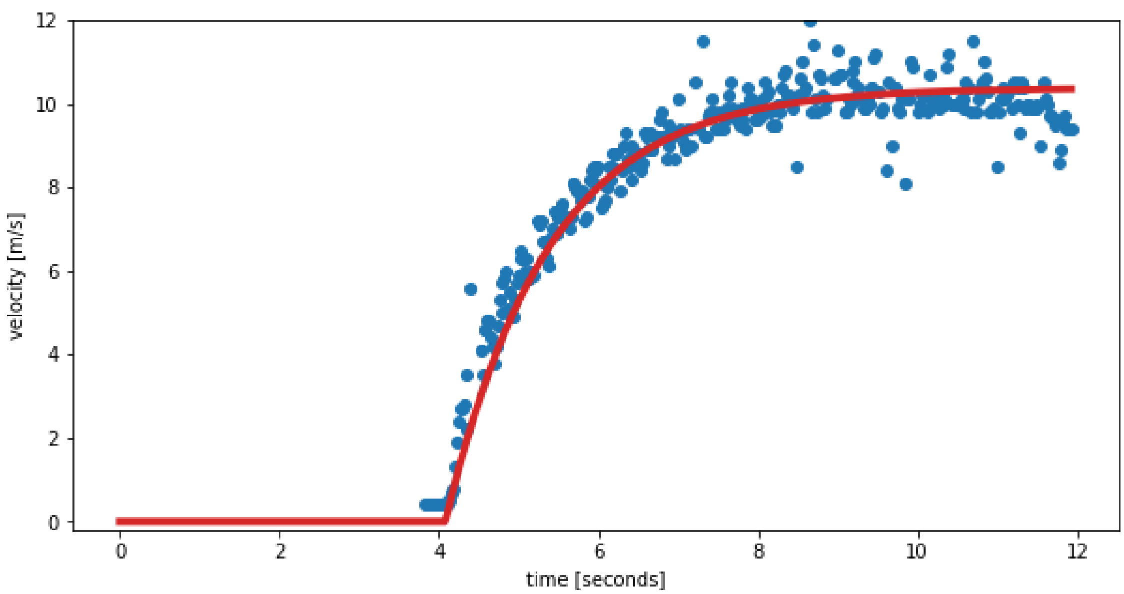
processing 211112-KLRXJ_1.rda
Maximum velocity of 9.48 m/s achieved at 52.4 meters, 6.99 s after start.
[10.322487941002226, 11.075881044797898, 12.304956266597012, 13.434343530380202]
[9.43339851 8.90842997 1.41321468]



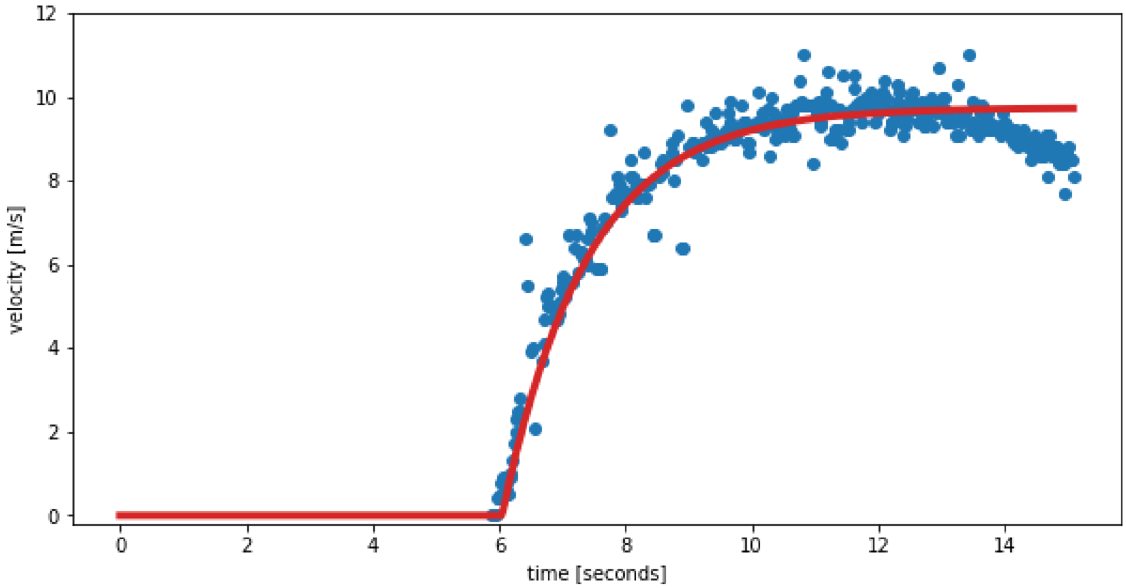
processing 211112-UJNPL_1.rda
Maximum velocity of 9.45 m/s achieved at 46.2 meters, 6.34 s after start.
[6.1288764986496584, 6.88426268618016, 8.107487956052664, 9.230950199891376]
[9.64687273 4.5711618 1.69517946]



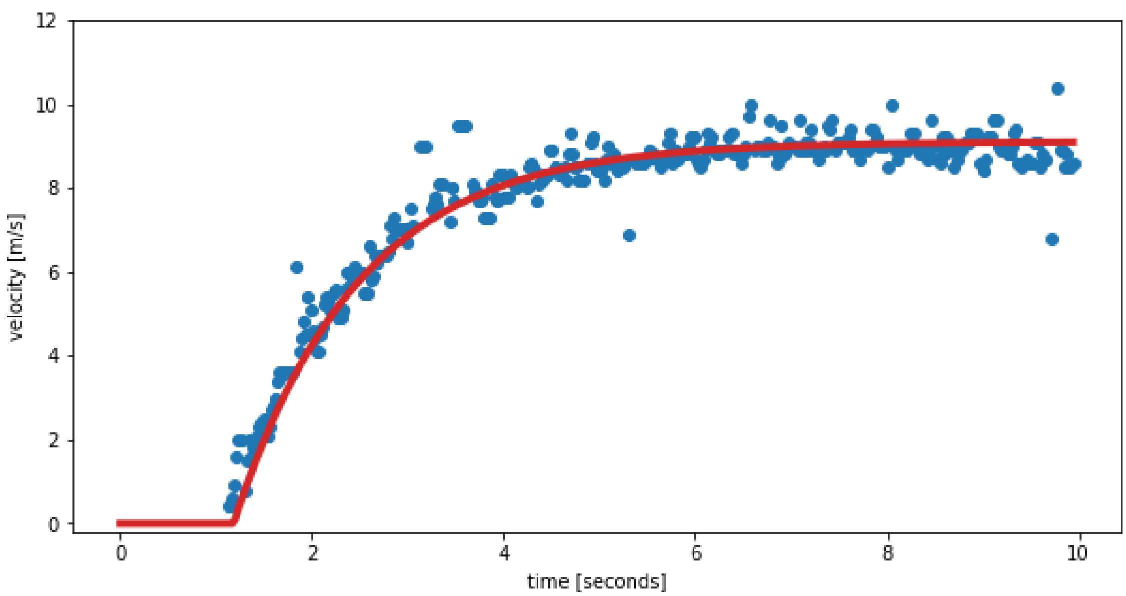
processing 211112-ERQZA_1.rda
Maximum velocity of 10.58 m/s achieved at 36.0 meters, 4.98 s after start.
[5.345752570617026, 6.029900274551726, 7.171764717272298, 8.199715372521183]
[10.37768951 4.07974492 1.2987348]



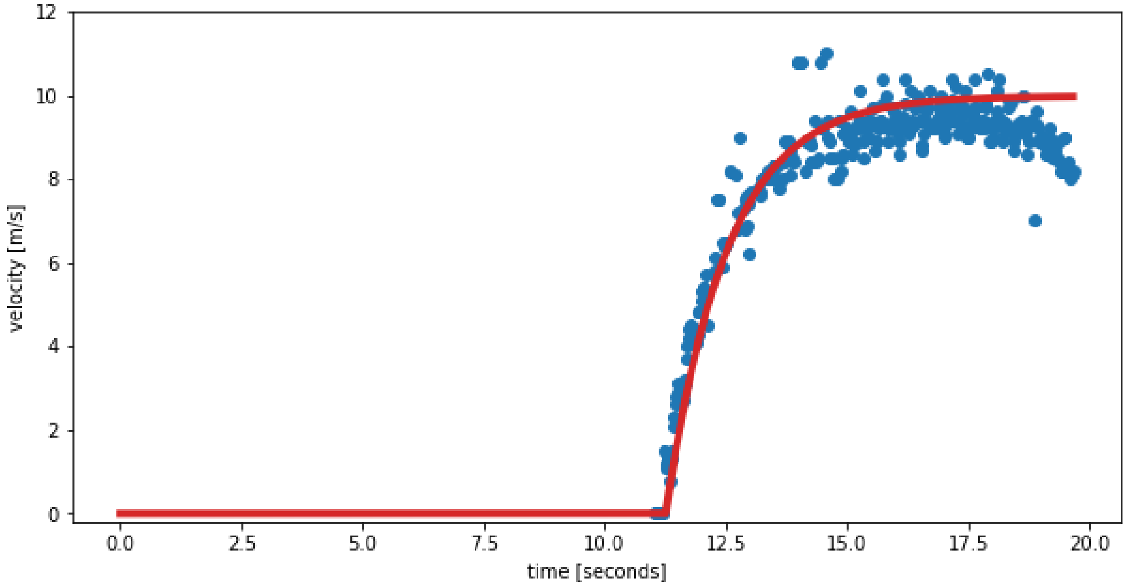
processing 211112-KTAWO_1.rda
Maximum velocity of 9.9 m/s achieved at 35.0 meters, 5.07 s after start.
[7.395735472299407, 8.117004007775202, 9.330626975717884, 10.416839588162157]
[9.73486891 6.04352897 1.35381693]



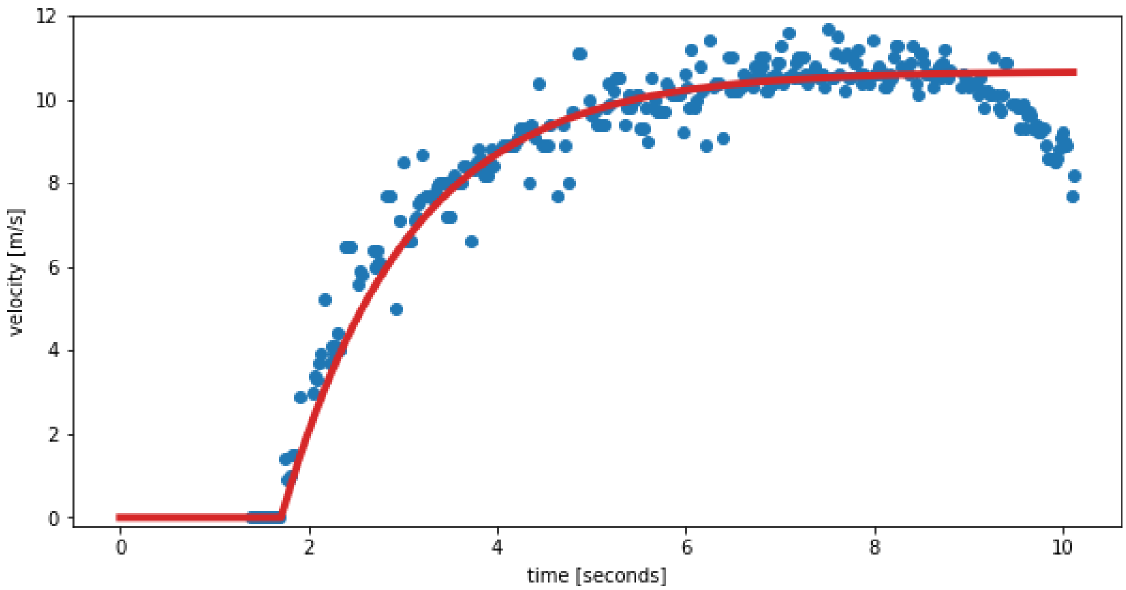
processing 211112-XZFMU_2.rda
Maximum velocity of 9.17 m/s achieved at 46.9 meters, 6.49 s after start.
[2.596650532474062, 3.3309531875632286, 4.55810953281864, 5.7296300934276525]
[9.09996643 1.18446988 1.29815658]



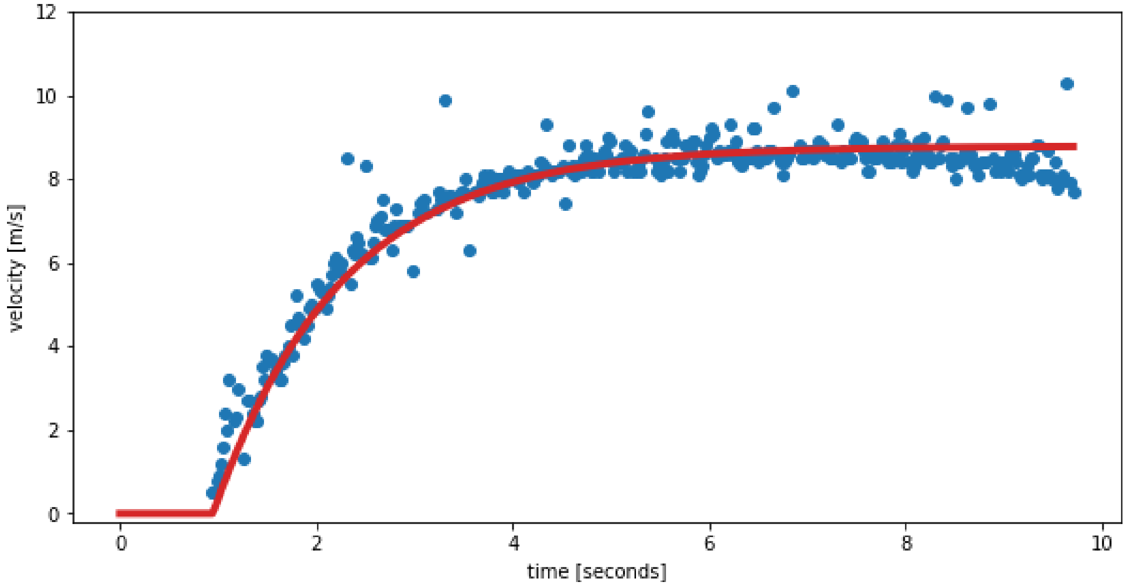
processing 211112-FGUWH_2.rda
Maximum velocity of 9.88 m/s achieved at 18.1 meters, 3.16 s after start.
[12.570370236929195, 13.258264619266267, 14.416877770367106, 15.529726883827601]
[9.98692083 11.26338035 1.26868068]



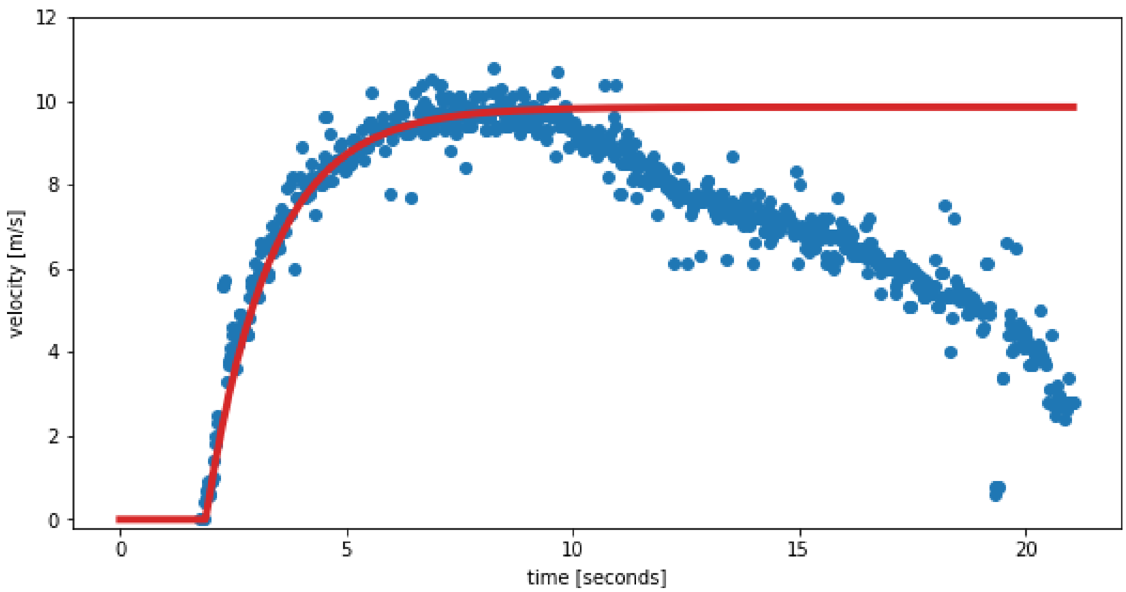
processing 211112-PZWON_2.rda
Maximum velocity of 11.01 m/s achieved at 58.3 meters, 7.12 s after start.
[2.978909755890016, 3.6504334108930316, 4.804347438654537, 5.82698241156559]
[10.67388843 1.70963236 1.36062595]



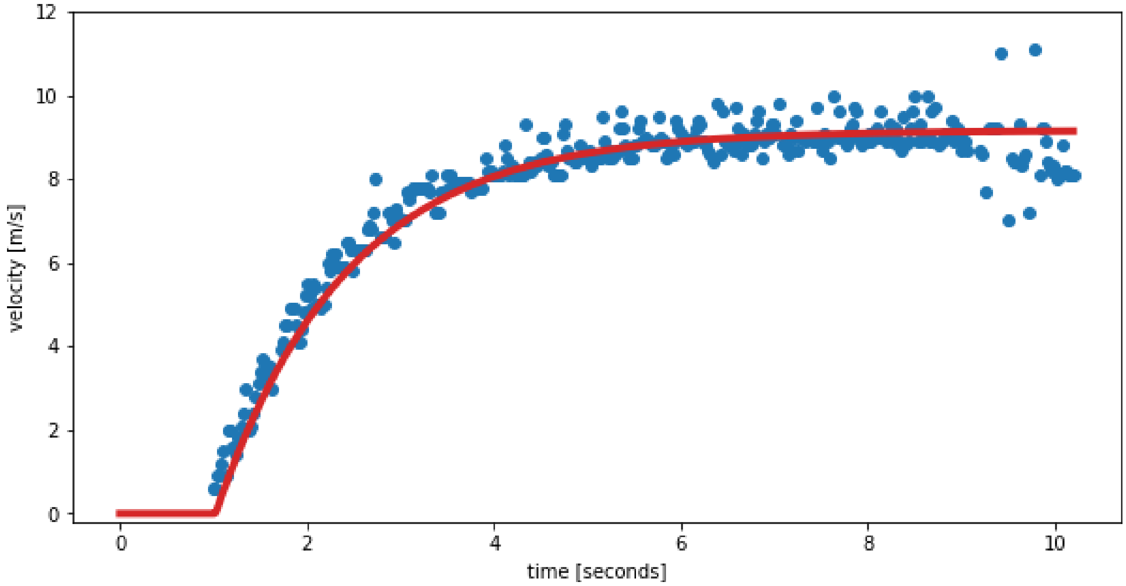
processing 211112-XHCYA_2.rda
Maximum velocity of 8.88 m/s achieved at 54.1 meters, 7.47 s after start.
[2.4020317979419468, 3.143744904041107, 4.4281940969103575, 5.615961550074844]
[8.78200291 0.94558788 1.3133506]



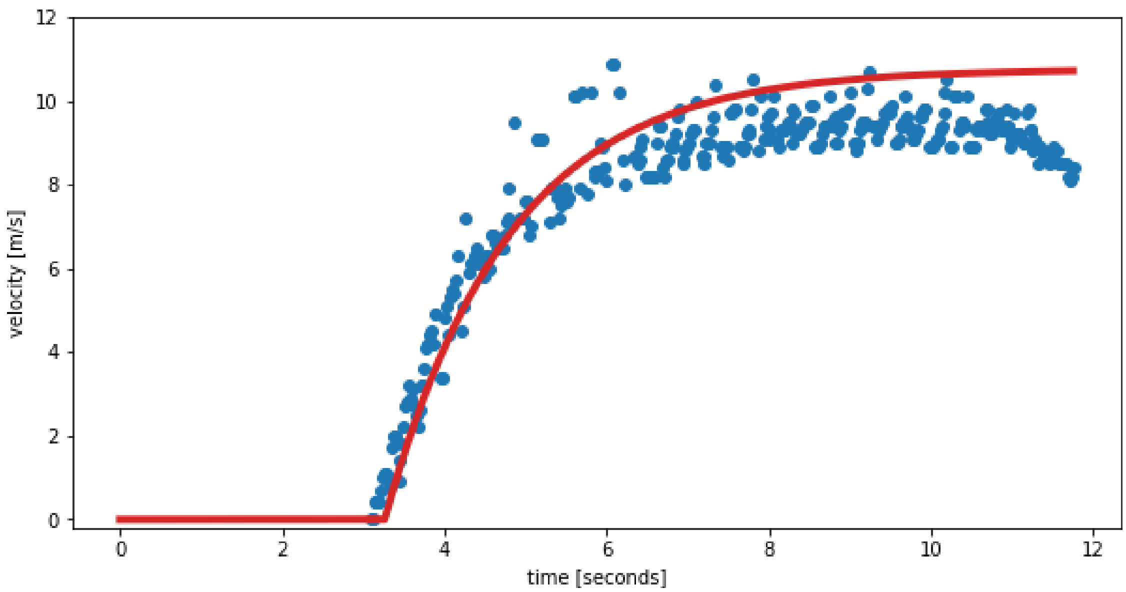
processing 211112-KTAWO_2.rda
Maximum velocity of 9.88 m/s achieved at 50.9 meters, 6.71 s after start.
[3.2771522221368374, 4.002191371687475, 5.207137647098042, 6.300266603295088]
[9.84954449 1.88960643 1.43654259]



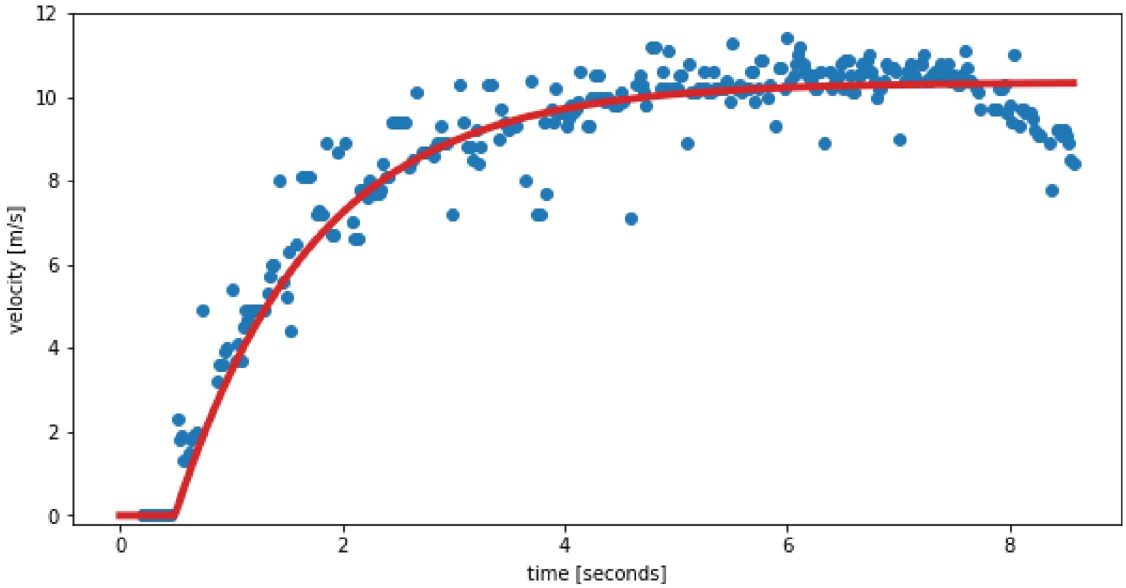
processing 211112-XZFMU_1.rda
Maximum velocity of 9.32 m/s achieved at 58.0 meters, 7.75 s after start.
[2.4703040771059728, 3.215120510411785, 4.475903991058107, 5.63793369701895]
[9.15686517 1.02127709 1.40609239]



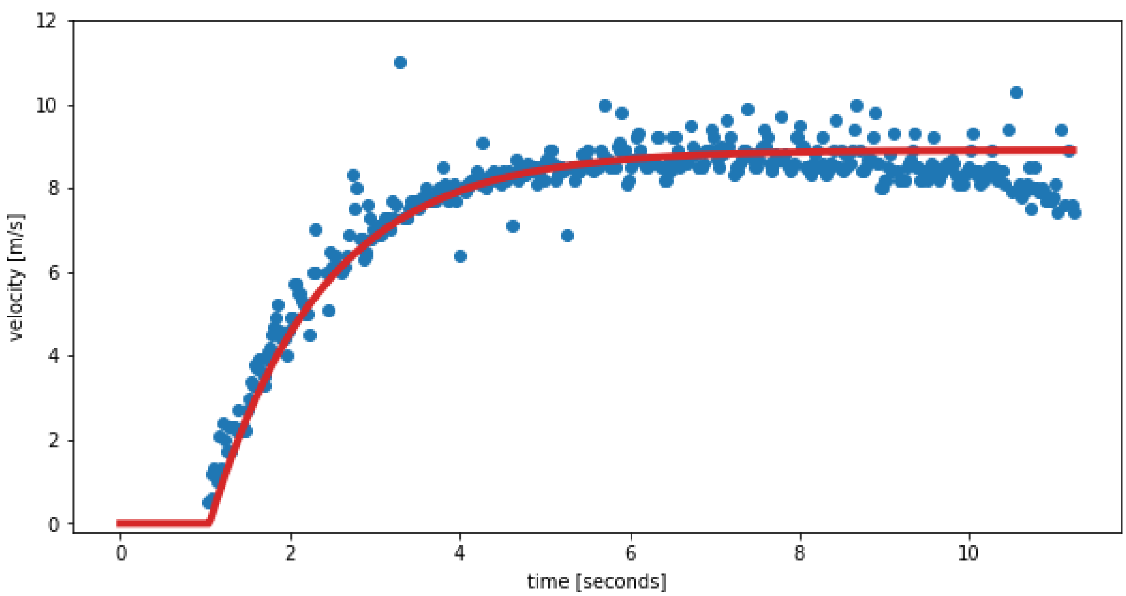
processing 211112-FGUWH_1.rda
Maximum velocity of 9.72 m/s achieved at 18.5 meters, 3.19 s after start.
[4.615536961045948, 5.30823201211422, 6.457061017265076, 7.581006808820216]
[10.76225712 3.26340987 1.53574604]



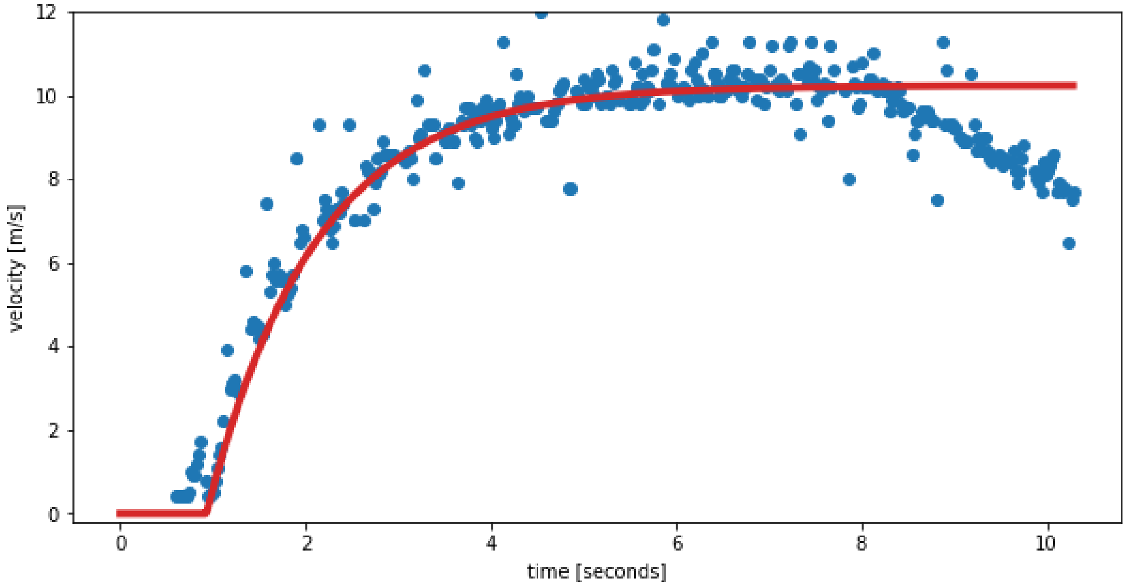
processing 211112-PZWON_1.rda
Maximum velocity of 10.73 m/s achieved at 46.2 meters, 6.0 s after start.
[1.7773739313208892, 2.434516038446674, 3.5535375106866556, 4.614069359124394]
[10.35023079 0.49317694 1.25651788]



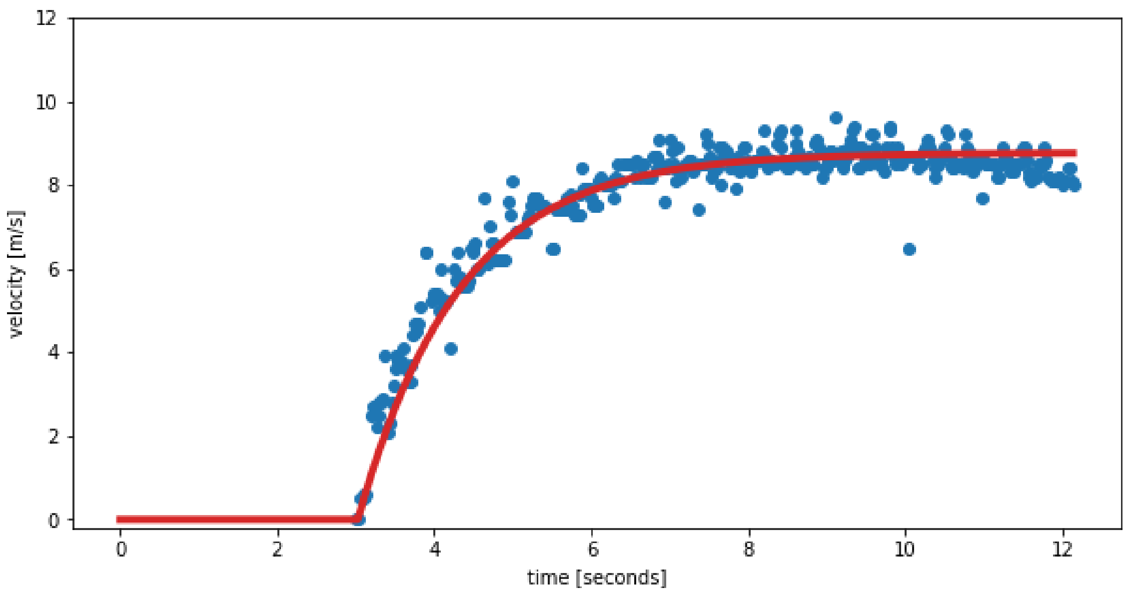
processing 211112-XHCYA_1.rda
Maximum velocity of 9.0 m/s achieved at 43.2 meters, 6.18 s after start.
[2.4951169277677754, 3.2432197896163637, 4.513800201145656, 5.706706277395009]
[8.90312952 1.05272359 1.33463261]



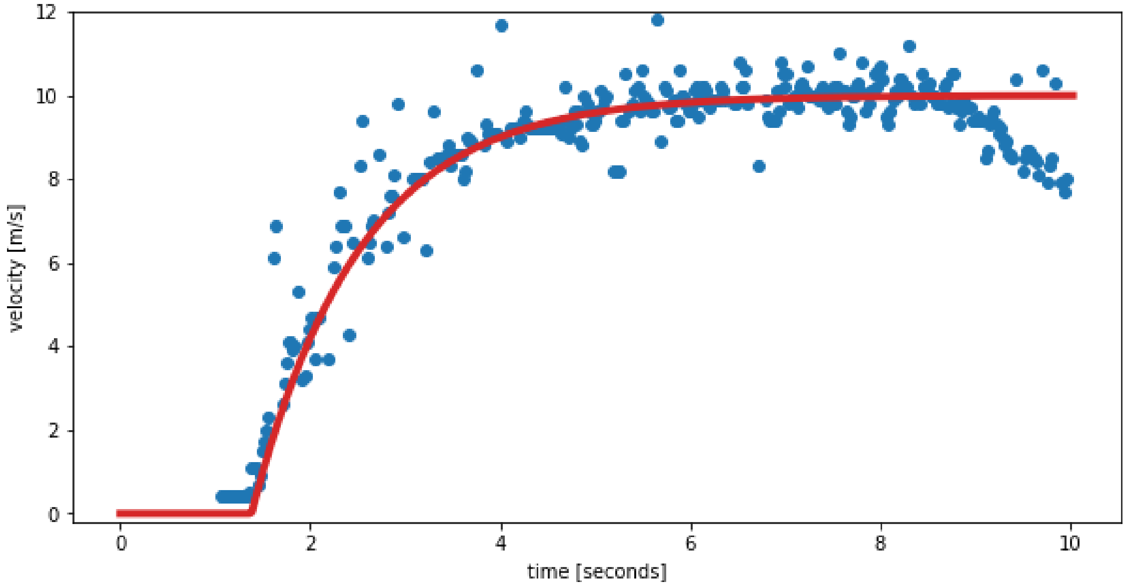
processing 211112-HLTFM_2.rda
Maximum velocity of 10.56 m/s achieved at 40.3 meters, 5.4 s after start.
[2.150991672243661, 2.8169092813261547, 3.9463655586435094, 4.978969799555469]
[10.23189287 0.93253051 1.17058732]



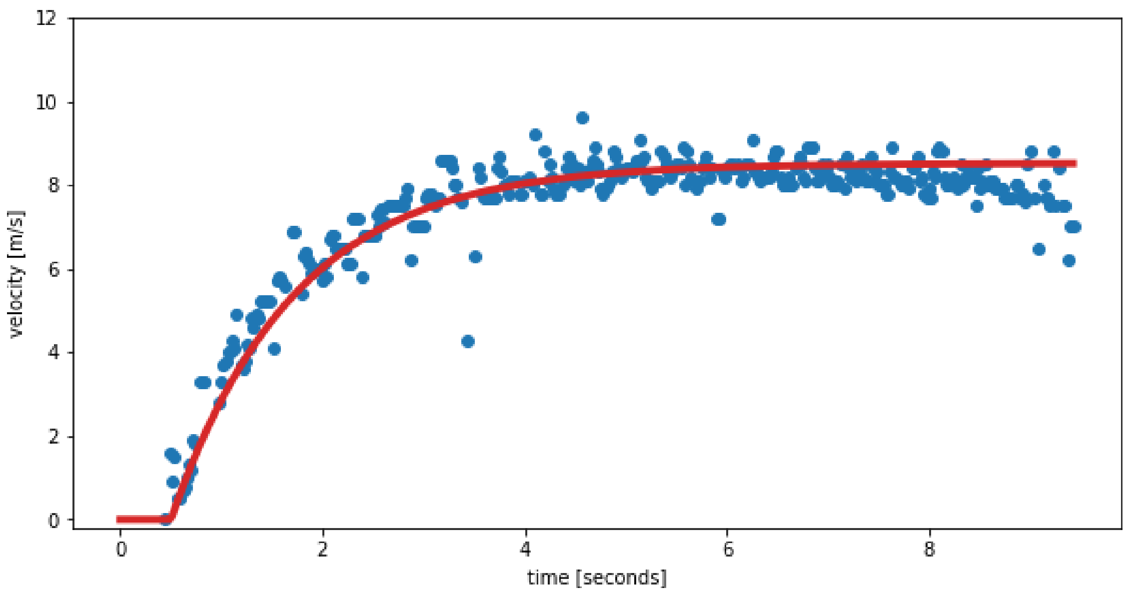
processing 211112-UMYXJ_2.rda
Maximum velocity of 8.9 m/s achieved at 44.5 meters, 6.38 s after start.
[4.446285291347663, 5.220635385447389, 6.53370369480335, 7.715600470105855]
[8.76673683 3.02810586 1.30846633]



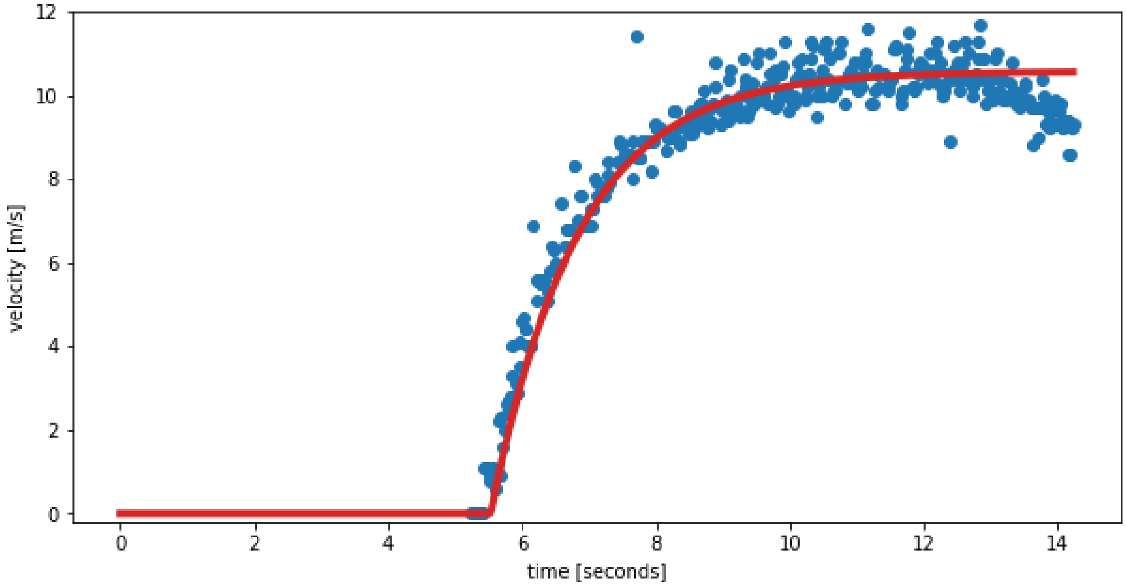
processing 211112-VQRBJ_2.rda
Maximum velocity of 10.27 m/s achieved at 55.5 meters, 6.99 s after start.
[2.6196022754011823, 3.2815723481681984, 4.407468253813848, 5.479208226627172]
[9.99900331 1.38299615 1.13602424]



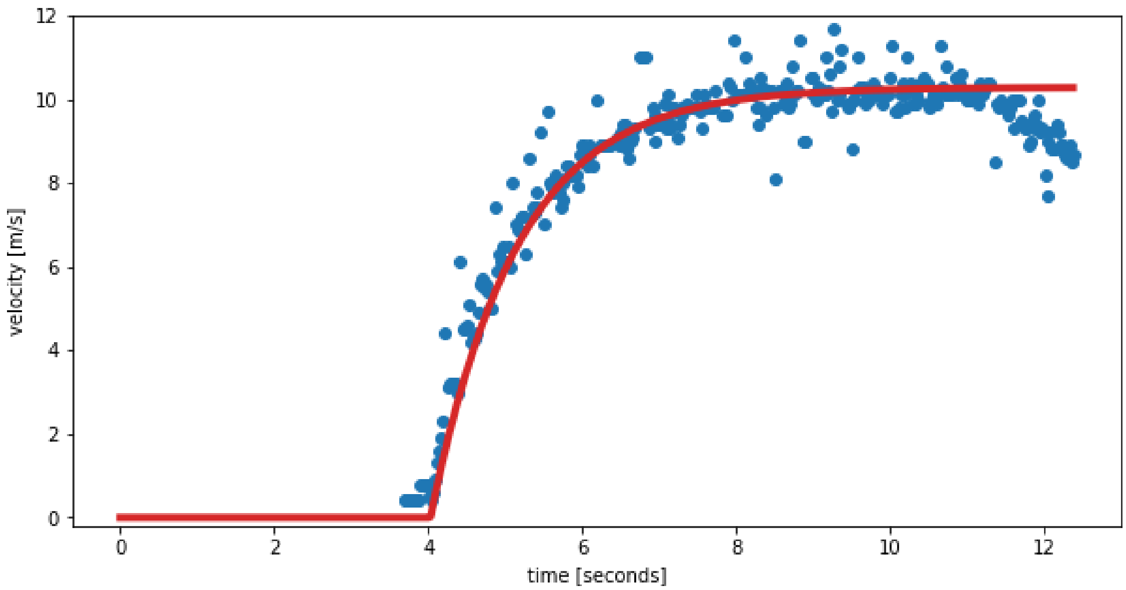
processing 211112-WMWBL_1.rda
Maximum velocity of 8.5 m/s achieved at 29.6 meters, 4.71 s after start.
[1.910572579885949, 2.684131390961542, 3.9973665182554083, 5.20840767210962]
[8.51954956 0.5031616 1.22371905]



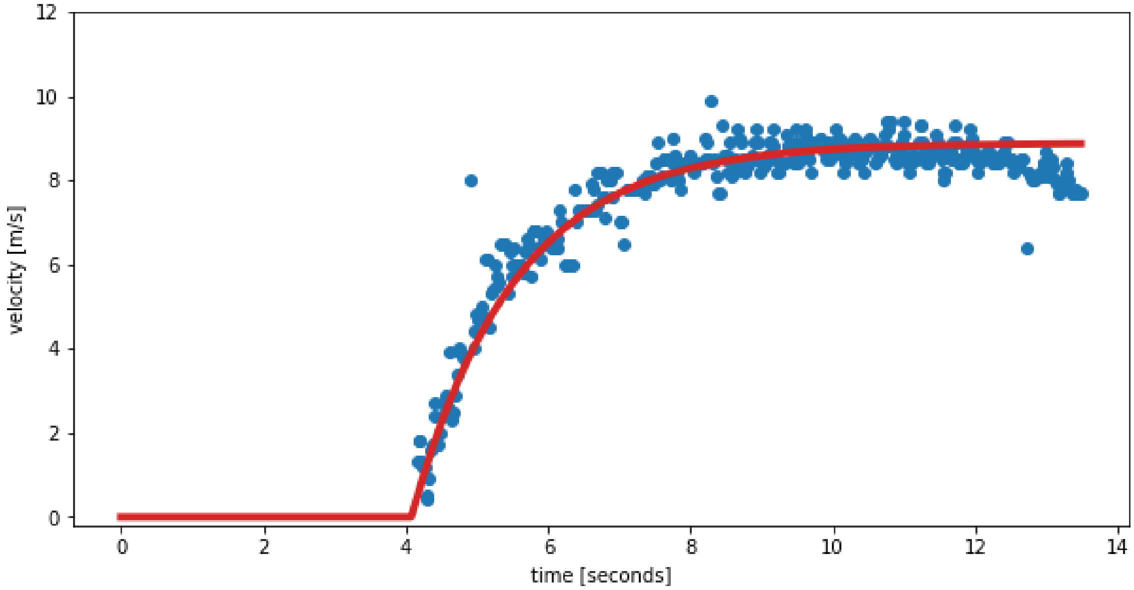
processing 211112-SWZKP_2.rda
Maximum velocity of 10.76 m/s achieved at 63.9 meters, 7.62 s after start.
[6.8025829695509215, 7.477094553255726, 8.596841450617784, 9.621131625560981]
[10.57170266 5.53426861 1.30573625]



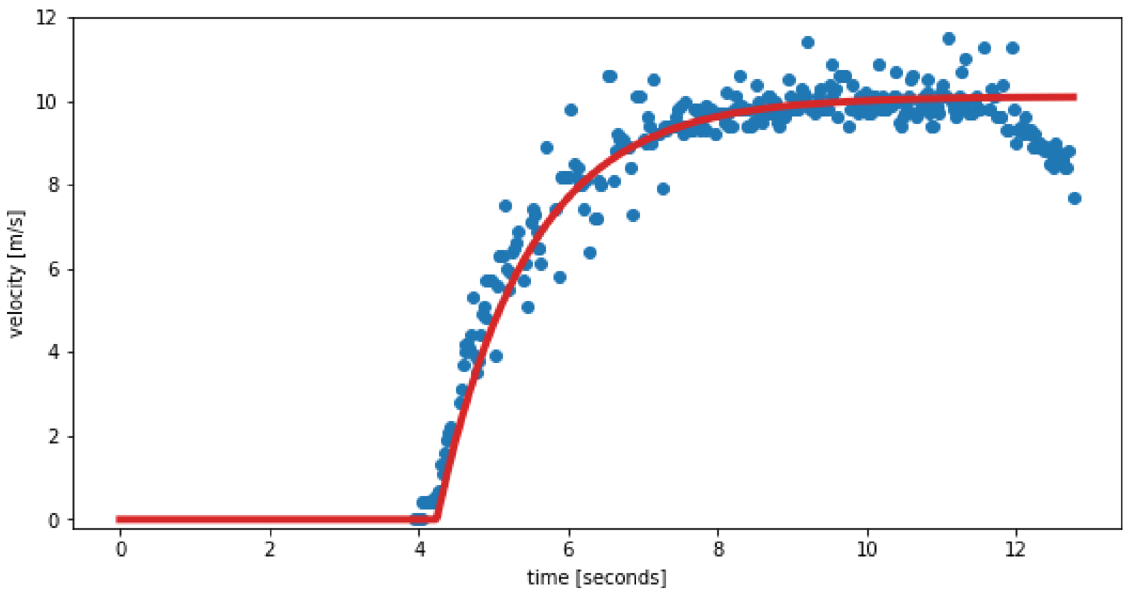
processing 211112-HLTFM_1.rda
Maximum velocity of 10.5 m/s achieved at 44.1 meters, 5.72 s after start.
[5.233883014953983, 5.885629335634326, 6.996252787998691, 8.021333480716136]
[10.2848172   4.02958526 1.12986079]



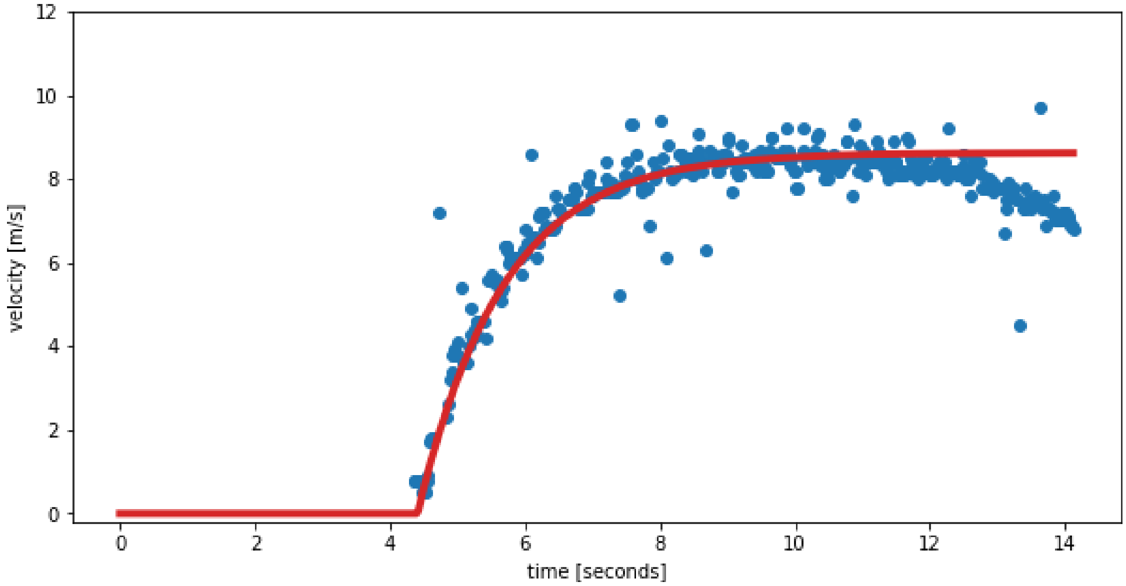
processing 211112-UMYXJ_1.rda
Maximum velocity of 8.95 m/s achieved at 48.8 meters, 6.86 s after start.
[5.5539396141036645, 6.3400200037506735, 7.6640015792923695, 8.847757080634327]
[8.88222923 4.08133926 1.45297209]



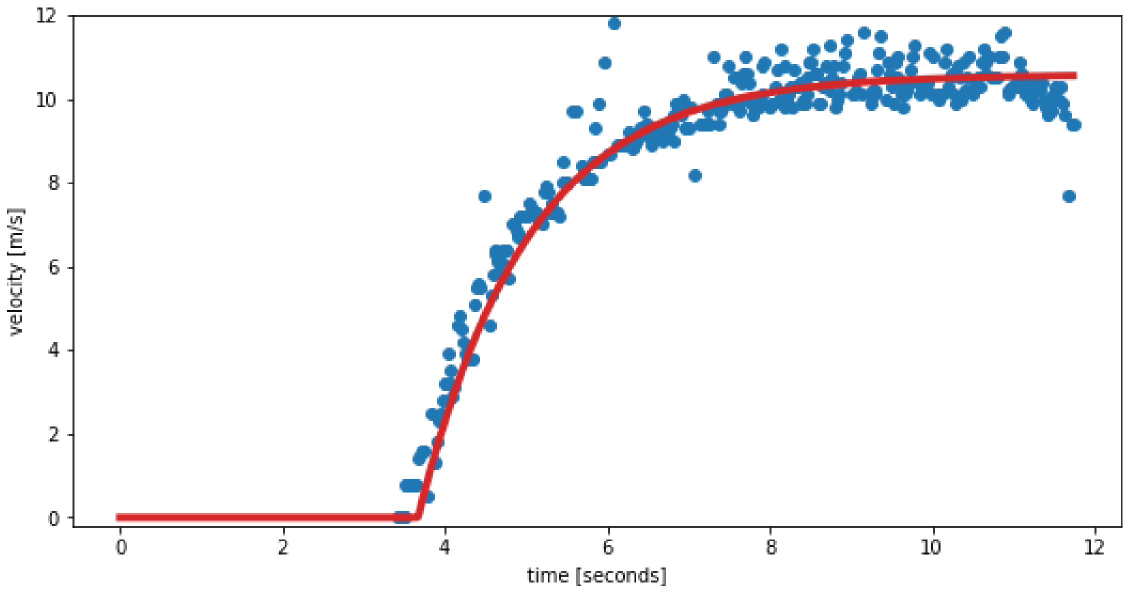
processing 211112-VQRBJ_1.rda
Maximum velocity of 10.32 m/s achieved at 43.3 meters, 5.79 s after start.
[5.500848148890777, 6.187997111712147, 7.336669452128021, 8.383494117747071]
[10.09940142 4.23823486 1.22813609]



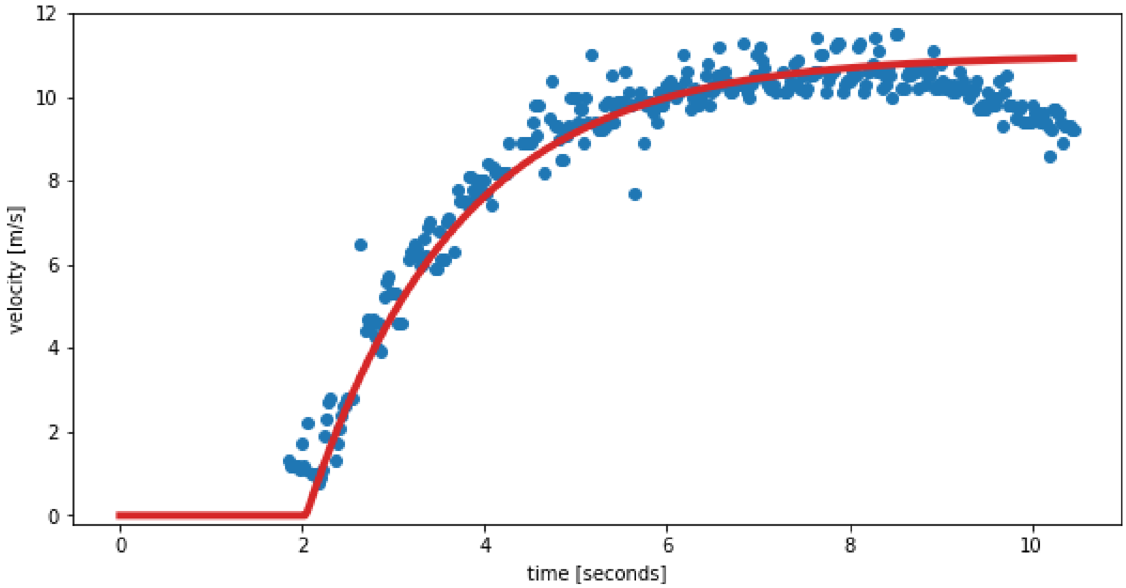
processing 211112-WMWBL_2.rda
Maximum velocity of 8.68 m/s achieved at 36.7 meters, 5.54 s after start.
[5.837151389206292, 6.597264060843008, 7.894806471096155, 9.107017781569606]
[8.6264398 4.40885184 1.26931467]



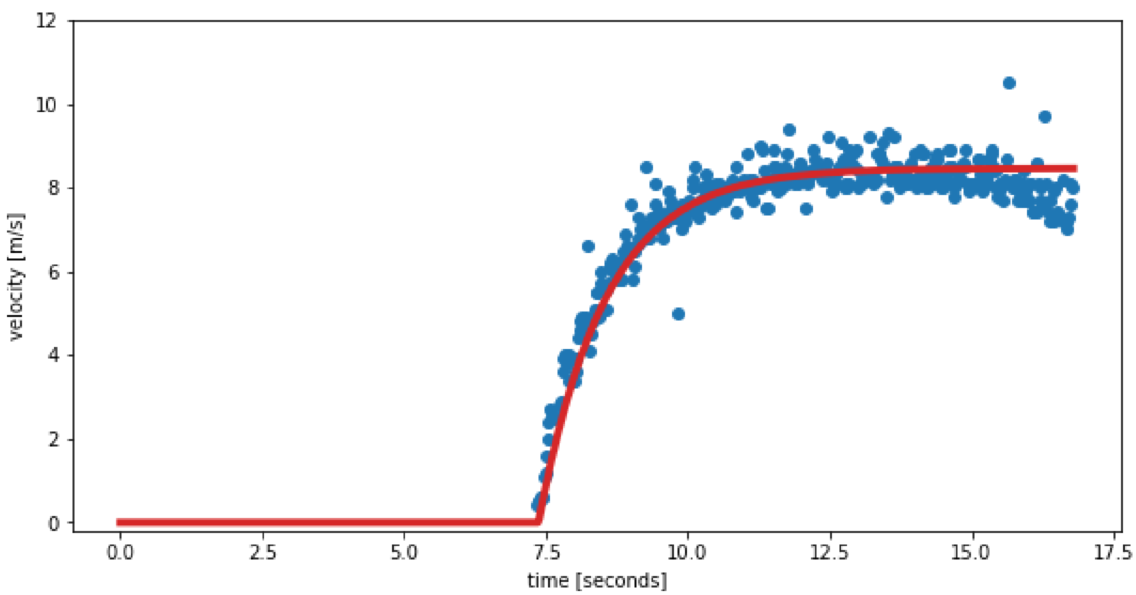
processing 211112-SWZKP_1.rda
Maximum velocity of 10.91 m/s achieved at 62.5 meters, 7.48 s after start.
[4.963095458139136, 5.638784571363809, 6.750565182838687, 7.786155878176948]
[10.58295084 3.6678623   1.35290597]



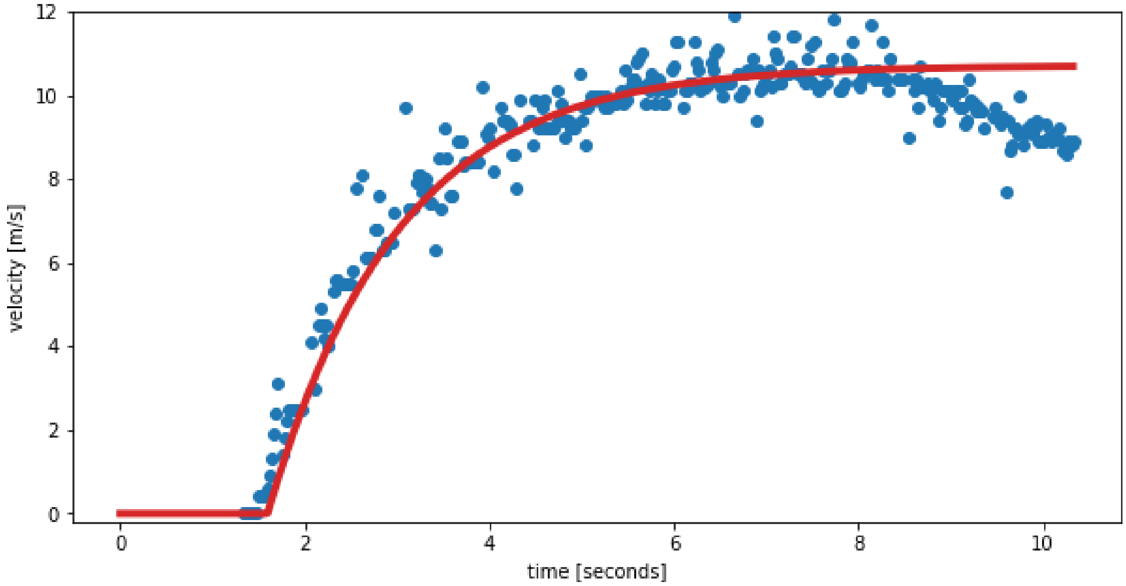
processing 211112-OBTQR_1.rda
Maximum velocity of 10.83 m/s achieved at 47.2 meters, 6.07 s after start.
[3.401150895643042, 4.114594677331559, 5.233618954838265, 6.2573553761553775]
[10.99698478 2.04051923 1.65824716]



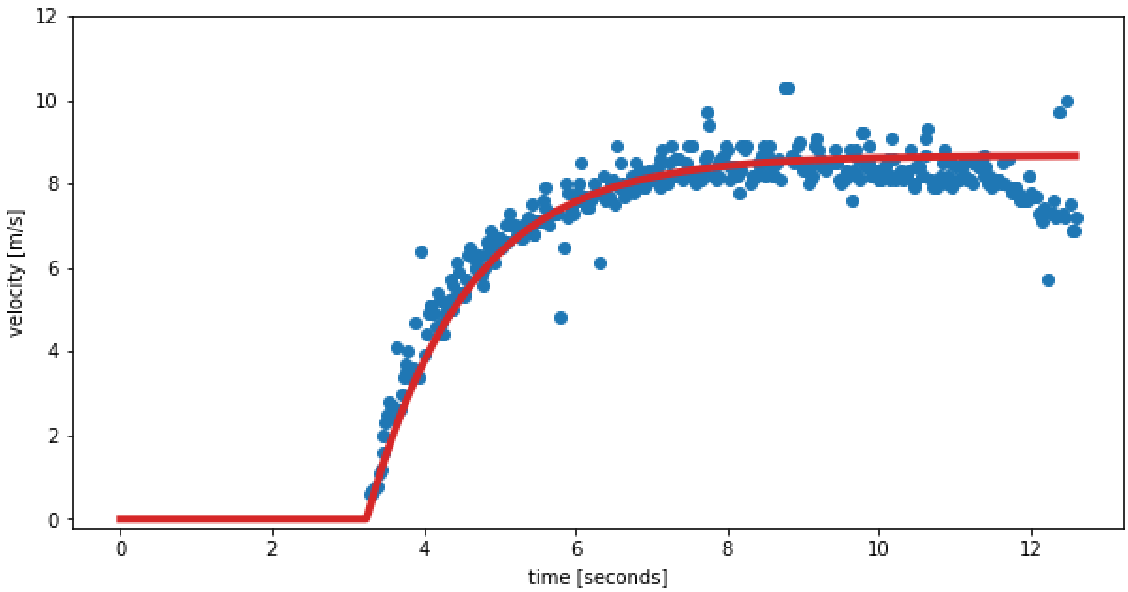
processing 211112-FQKKY_1.rda
Maximum velocity of 8.66 m/s achieved at 43.2 meters, 6.3 s after start.
[8.770559627996322, 9.523671787007352, 10.844200583961735, 12.066156212266852]
[8.4633761 7.36795261 1.18483705]



processing 211112-OBTQR_2.rda
Maximum velocity of 10.88 m/s achieved at 58.1 meters, 7.07 s after start.
[2.8980067176203366, 3.576933354205149, 4.7095574681972225, 5.735470455822051]
[10.71779553 1.59827129 1.41232351]



processing 211112-FQKKY_2.rda
Maximum velocity of 9.08 m/s achieved at 38.6 meters, 5.72 s after start.
[4.692221395318831, 5.459814803943199, 6.797596005178877, 8.002970776262755]
[8.66732955 3.24755692 1.33265311]



Output.to_excel("Sprint_Split_Project.xlsx", index = False)

Bibliography

Arthur-Cameselle, J., Sossin, K., & Quatromoni, P. (2017). A qualitative analysis of factors related to eating disorders onset in female collegiate athletes and non-athlets. *Eating Disorders*, *25*(3).

Bell, R., Ratzlaff, S. E., & Murray, S. R. (2004). The Impact of the HIPAA Privacy Rule on Collegiate Sport Professionals. *The Sport Journal*, *24*.

Clark, K. P., Rieger, R. H., Bruno, R. F., & Stearne, D. J. (2019). The National Football League combine 40-yard dash: How importnat is maximum velocity? *Journal of Strength and Conditioning Research*, *33*, 1542–1550.

Cohen, B., Diamond, J., & Beaton, A. (2019, December 19). The decade when numbers broke sports. *Wall Street Journal*.

Crenna, F., Battista Rossi, G., & Berardengo, M. (2021). Filtering Biomechanical Signals in Movement Analysis. *Sensors*, *21*(13). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8271607/

Dolan, T. G. (2003). PTs respond to HIPAA: The Real World Experience. *PT Magazine of Physical Therapy*, *11*(7), 52–56.

Edwards, T., Banyard, H. G., Piggott, B., Haff, G. G., & Joyce, C. (2022). Reliability and Minimal Detectable Change of Sprint Times and Force-Velocity-Power Characteristics. *Journal of Strength and Conditioning Research*, *36*(1), 268–272. https://doi.org/10.1519/JSC.0000000000004161

Furusawa, K., Hill, A. V., & Parkinson, J. L. (1927). The Dynamics of "Sprint" Running. *Proceedings of the Royal Society of London, Series B, Containing Paper of a Biological Character*, *102*(713), 29–42.

Futterman, M. (2019, October 10). Nike to Shut Down Salazar's Oregon Project. *The New York Times*. https://www.nytimes.com/2019/10/10/sports/nike-oregon-project-salazar.html

Gardner, D. (n.d.). *How Much of Your Sport is Mental?* Maxpreps.Com. https://www.maxpreps.com/news/NSRCCf85gU68n1uBllox3w/how-much-of-your-sport-is-mental.htm

Goe, K. (2021, October 25). *Women athletes allege body shaming within Oregon Ducks track and field program* [Newspaper]. Oregonlive.Com. https://www.oregonlive.com/trackandfield/2021/10/women-athletes-allege-body-shaming-within-oregon-ducks-track-and-field-program.html

Greenleaf, C., Petrie, T. A., Carter, J., & Reel, J. J. (2009). Female Collegiate Athletes: Prevalence of Eating Disorders and Disordered Eating Behaviors. *Journal of American College Health*, *57*(5), 489–496.

Guo, F., Li, Y., Kankanhalli, M., & Brown, M. (2013). An Evaluation of Wearable Activity Monitoring Devices. *Proceedings of the 1st International Workshop on Personal Data Meets Distributed Multimedia*, 31–34.

Hartzell, H. (2018, December 4). Understanding the Nike Oregon Project's Vexed History. *Women's Running*.

Healy, R., Kenny, I. C., & Harrison, A. J. (2022). Profiling elite male 100-m sprint performance: The role of maximum velocity and relative acceleration. *Journal of Sport and Health Science*, *11*(1), 75–84.

Hill, D. (2003). A Matter of Privacy. *Athletic Management*, *15*(2), 37–42.

Jarvis, D., & Westcott, K. (2020, December 7). The Hyperquantified Athlete: Technology, measurement, and the business of sports. *Deloitte Insights*. https://www2.deloitte.com/xe/en/insights/industry/technology/technology-media-and-telecom-predictions/2021/athlete-data-analytics.html

Karkazis, K., & Fishman, J. (2017). Tracking U.S. Professional Athletes: The Ethics of Biometric Technologies. *American Journal of Bioethics*, *17*(1), 45–60.

King, N. M. P., & Robeson, R. (2007). Athlete or guinea pig? Sports and enhancement research. *Studies in Ethics, Law, and Technology*, *1*(1), 1–17. http://dx.doi.org/10.2202/1941-6008.1006

King, N. M. P., & Robeson, R. (2013). Athletes are guinea pigs. *American Journal of Bioethics*, *13*(10), 13–14. http://dx.doi.org/ 10.1080/15265161.2013.828126.

Lewis, M. (2004). *Moneyball: The Art of Winning an Unfair Game*. W.W. Norton & Company.

Manning, J. (2021, October 12). *Distance runner Mary Cain sues former coach Alberto Salazar, Nike over alleged abuse* [Newspaper]. Oregonlive.Com. https://www.oregonlive.com/business/2021/10/distance-runner-mary-cain-sues-former-coach-alberto-salazar-nike-over-alleged-abuse.html

Morgulev, E., Azar, O. H., & Bar-Eli, M. (2020). Searching for Momentum in NBA triplets of free throws. *Journal of Sports Science*, *38*(4), 390–398. https://doi.org/doi: 10.1080/02640414.2019.1702776

Morin, J. (2019). *Spreadsheet for Sprint acceleration force-velocity power profiling*. ResearchGate.

Morin, J. B. (2017, December 13). *A spreadsheet for Sprint acceleration Force-Velocity-Power profiling*. Jbmorin.Net. https://jbmorin.net/2017/12/13/a-spreadsheet-for-sprint-acceleration-force-velocity-power-profiling/

Passfield, L., & Hopker, J. G. (2016). A Mine of Information: Can Sports Analytics Provide Wisdom From your Data? *International Journal of Sports Physiology and Performance*, *12*(7), 1–17. https://doi.org/10.1123/ijspp.2016-0644

Pitz, S. M. (2003). HIPAA and the sports media: Separating fiction from reality. *Nevada Lawyer*, *11*(8), 12–14.

Poehling, R. A. (2018). *Monitoring Explosive Performances in Relation to Training Load Accumulation in Adolescent Female Soccer Players*. University of British Columbia.

Reel, J. J., Petrie, T. A., SooHoo, S., & Anderson, C. M. (2013). Weight pressures in sport: Examining the factor structure and incremental validity of the weight pressures in sport — Females. *Eating Behaviors*, *14*(2), 137–144. https://doi.org/10.1016/j.eatbeh.2013.01.003

Rothman, K. (2008). BMI-related errors in the measurement of obesity. *International Journal of Obesity*, *32*.

Samozino, P., Rabita, G., Dorel, S., Slawinski, J., Peyrot, N., Saez de Villarreal, E., & Morin, J.-B. (2016). A Simple Method for Measuring Power, Force, Velocity Properties, and Mechanical Effectiveness in Sprint Running. *Scandinavian Journal of Medicine and Science in Sports*, *26*(6), 648–658. https://doi.org/10.1111/sms.12490

Schmidt, C. (2020, August 17). *Eight former UA athletes detail "rotten culture" in track and field program* [Newspaper]. Tuscon.Com. https://tucson.com/sports/arizonawildcats/eight-former-ua-athletes-detail-rotten-culture-in-track-and-field-program/article_26ead223-11c5-5fc1-8c17-73ef134ed973.html

Shipman, F. M. (2009). Blending the real and virtual in games: The model of fantasy sports. *Proceedings of the 4th International Conference on Foundations of Digital Games*. https://doi.org/10.1145/1536513.1536547

Spaaij, R., & Thiel, A. (2017). Big data: Critical questions for sport and society. *European Journal for Sport and Society*, *14*(1), 1–4. https://doi.org/10.1080/16138171.2017.1288374

Sports Analytics Market- Growth, Trends, COVID-19 Impact, and Forecasts (2022-2027). (2021, December). Mordor Intelligence. https://www.mordorintelligence.com/industry-reports/sports-analytics-market

Stice, E., Marti, C. N., Shaw, H., & Jaconis, M. (2009). An 8-year longitudinal study of the natural history of threshold, subthreshold, and partial eating disorders from a community sample of adolescents. *Journal of Abnormal Psychology*, *118*(3), 587–597. https://doi.org/10.1037/a0016481

The Stalker Acceleration Testing System. (n.d.). Stalker.Sport. https://stalker.sport/ats-ii/

Vermeulen, E. (2018). Big data in sport analytics: Applications and risks. *South Africa*, 11.

Zucker, T. (n.d.). *Eating Disorders vs. Disordered Eating: What's the difference?* [Blog]. National Eating Disorders Association. Retrieved December 10, 2021, from https://www.nationaleatingdisorders.org/blog/eating-disorders-versus-disordered-eating