

10-19-2015

Real Time Predictive Speed Analysis for High Speed Rail Collision Test

Xian Jian Jiang
University of Windsor

Follow this and additional works at: <https://scholar.uwindsor.ca/etd>

Recommended Citation

Jiang, Xian Jian, "Real Time Predictive Speed Analysis for High Speed Rail Collision Test" (2015). *Electronic Theses and Dissertations*. 5441.
<https://scholar.uwindsor.ca/etd/5441>

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email (scholarship@uwindsor.ca) or by telephone at 519-253-3000ext. 3208.

Real Time Predictive Speed Analysis for High Speed Rail Collision Test

By

Xian Jian Jiang

A Thesis Submitted to the Faculty of Graduate Studies through Computer Science
in Partial Fulfillment of the Requirements for the Degree of Master of Science at
the University of Windsor

Windsor, Ontario, Canada

2015

© 2015 Xian Jian Jiang

Real Time Predictive Speed Analysis for High Speed Rail Collision Test

by

Xian Jian Jiang

APPROVED BY:

Dr. Guoqing Zhang
Department of Mechanical, Automotive and Materials Engineering

Dr. Arunita Jaekel
School of Computer Science

Dr. Dan Wu, Advisor
School of Computer Science

September 2nd, 2015

Author's Declaration of Originality

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Abstract

In a real train collision test, a train locomotive needs to be propelled on a straight, guided path, to a particular speed, at which time the train locomotive is released to coast down towards a barrier where it is required to crash at a desired speed. The current control of the release speed and location is based on theoretical data and previous experience which leads to less accuracy in the actual crash speed. In this research work, the goal is to make improvements in a typical real train collision test that will help obtain a more accurate crash speed and release location by controlling the force release precisely. The contribution of this research work is to implement a solution to simulate the behavior of the propulsion system, and trigger an algorithm to calculate the required release speed and location more accurately and quickly.

Keywords: collision test, crash speed, release speed, release location, coast down, force release.

Dedication

This thesis is dedicated to my wife who has supported me all the way since the beginning of my studies.

Acknowledgements

First of all, I would like to express my deep and sincere gratitude to my supervisor, Dr. Dan Wu, for his invaluable guidance and advices, for his enthusiastic encouragement and his great patience to me. Without his help, the work presented here would not have been possible.

A very special thanks goes out to Mr. Roland Vander Straeten of Anemoi Technologies Inc. Mr. Roland Vander Straeten is the one specialist who truly made a difference in my life. It was under his tutelage that I developed the current application software and became interested in the state-of-the-art test facilities. He provided me with direction, technical support and became more of a mentor and friend, than an employer.

I would also like to thank my external reader, Dr. Guoqing Zhang, and my internal reader, Dr. Arunita Jaekel, for spending their precious time to review this thesis and putting down their comments, suggestions on the thesis work.

Finally, I am very grateful for experts of Anemoi Technologies Inc., Jessica Zheng, Mike Doiron, Slawek Ciurzynski and Bruce Hu. During these two years' study, they gave me great and solid support whenever I met any kind of problems. They always share their valuable knowledge and experiences with me. This is the most valuable asset in my life.

TABLE OF CONTENTS

AUTHOR’S DECLARATION OF ORIGINALITY	iii
ABSTRACT	iv
DEDICATION.....	v
ACKNOWLEDGEMENTS	vi
LIST OF TABLES	ix
LIST OF FIGURES	x
Chapter 1 Introduction	1
1.1 Introduction of High Speed Rail.....	1
1.2 Brief History of HSR	2
1.3 Safety Issues and Technology	4
1.4 Motivation	6
1.5 Contributions.....	6
1.6 Guide to the Thesis.....	8
Chapter 2 Background Knowledge	10
2.1 Regression Analysis	10
2.1.1 Simple Linear Regression.....	11
2.1.2 Multiple Linear Regression	14
2.1.3 Quadratic Relationship between Dependent and Independent Variables	16
2.1.4 Calculation of Regression Coefficients	17
2.1.5 Correlation Coefficient - R^2	18
2.2 Test Phases	20
2.3 Test Facility	21
2.4 Vehicle Motion Analysis	23
2.5 Sample Data Sheet	25
Chapter 3 Proposed Solution - RTPSA	29
3.1 Limitation of Existing Methods.....	29
3.2 Objective	29
3.3 Design Principles	30
3.3.1 Real-Time.....	30
3.3.2 Predictive.....	32
3.3.3 Simulation.....	33
3.4 System Architecture	34
3.5 Software Structure	35
3.6 Main Process	38
3.6.1 The Generation of Simulated Data Stream.....	39
3.6.2 Program Logic in Implementation	39
3.7 Coefficients Calculation Module	41
3.7.1 How to Calculate the Regression Coefficients.....	42
3.7.2 Program Logic in Implementation	44
3.8 Coast-Down Simulation Module.....	45
3.8.1 How to Simulate the Coast-Down Behavior?	46
3.8.2 Program Logic in Implementation	48
3.9 Propulsion Simulation Module	49
3.9.1 How to Simulate the Last Period of Propulsion Phase?.....	50
3.9.2 Program Logic in Implementation	52
3.10 Control/Release Module	53
3.10.1 How to Find the Release Speed and Location?	53
3.10.2 Program Logic in Implementation	54
3.11 Summary	55
Chapter 4 Running Results	57

4.1 The Demonstration Program	57
4.2 Standalone Version of Software	58
4.3 Running Time.....	60
4.4 Run with Different Set of Data	61
4.5 Run with Different Values of Crash Speed.....	62
4.6 Run with Different Values of Barrier Location	62
4.7 R^2 with Different Speed	63
4.8 R^2 with Different Error Ratio	64
4.9 Coast-down Distance with Different Propulsion Forces	65
Chapter 5 Conclusion and Future Work	67
5.1 Conclusion	67
5.2 Future Work.....	68
BIBLIOGRAPHY	69
VITA AUCTORIS	73

LIST OF TABLES

3.1 Leading Time Calculation	51
4.1 Run with Different Set of Data.....	61
4.2 Run with Different Values of Crash Speed.....	62
4.3 Run with Different Values of Barrier Location.....	63
4.4 R^2 for Speed of 60kph.....	63
4.5 R^2 for Speed of 80kph.....	63
4.6 R^2 for Speed of 120kph.....	64
4.7 R^2 for Speed of 150kph.....	64
4.8 R^2 for Different Error Ratios.....	64
4.9 Coast-down Distance with Different Propulsion Forces.....	65

LIST OF FIGURES

1.1 Deadly Collision	4
2.1 Data Distribution of Lean Body Mass and Muscle Strength.....	11
2.2 Simple Linear Regression - Straight Line.....	12
2.3 Different Fitted Lines for A Set of Data.....	13
2.4 Vertical Distance vs Perpendicular Distance.....	14
2.5 Propulsion–Release Force–Coast Down	21
2.6 Train and Track.....	22
2.7 Building Housing the Rigid Wall.....	22
2.8 Rigid Wall.....	22
2.9 Sample Data 1.....	25
2.10 Sample Data 2.....	26
3.1 Small Intervals of Propulsion Curve.....	31
3.2 Start and End Points for Simulations.....	33
3.3 System Architecture.....	35
3.4 Software Structure.....	36
3.5 Intersection Point.....	38
3.6 The Format of Simulated Data Stream.....	39
3.7 Main Process of <i>RTPSA</i>	41
3.8 End Point for Data Collection	43
3.9 Program Logic of Coefficients Calculation.....	45
3.10 Time-Equal Intervals for Coast-Down Simulation.....	47
3.11 Coast-Down Curves	48
3.12 Program Logic of Coast-down Simulation	49
3.13 Start Point for Propulsion Simulation	50
3.14 Program Logic of Propulsion Simulation	53
3.15 Program Logic of Control and Release Processing.....	55
4.1 Screen Shot of Demonstration Program.....	58
4.2 Input Screen of the Standalone Version.....	59

4.3 Output Screen of the Standalone Version.....	60
4.4 Screen Shot of Console - Running Time	60
4.5 Maximum Propulsion Force = 60,000.....	66
4.6 Maximum Propulsion Force = 70,000.....	66
4.7 Maximum Propulsion Force = 80,000.....	66

Chapter 1 Introduction

1.1 Introduction of High Speed Rail

The global high speed rail (HSR) network is one of the great feats of modern engineering, and proves to be one of the best forms of transportation ever invented. The global HSR network is rapidly expanding across continents world wide - delivering fast, efficient mobility to numerous nations every day [35].

HSR is currently in operation in more than twenty countries (including the UK, France, Germany, Belgium, Spain, Italy, Japan, China, Korea, and Taiwan) [34]. HSR is under construction in more than ten countries (including China, Spain, Saudi Arabia, France, and Italy). In the meantime, it is in development in another fourteen countries (including Turkey, Qatar, Morocco, Russia, Poland, Portugal, South Africa, India, Argentina, and Brazil). In Japan, HSR has been in operation for fifty years carrying more than nine billion passengers without a single fatality [31].

HSR delivers many layers of economic benefits. It is one of the solutions to rising gas prices. HSR delivers fast, efficient transportation so riders can save time, energy, and money. It is reliable and operates in almost all weather conditions [12]. HSR is not subject to congestion, so it operates on schedule every day without delay - especially during rush hour and peak travel times.

HSR stimulates the revitalization of cities by encouraging high density, mixed-use real estate development around the stations [35]. It also fosters economic development in second-tier cities along train routes and links cities together into

integrated regions that can then function as a single stronger economy. HSR can generate more job opportunities to labor markets and offers workers a wider network of employers to choose from. It also encourages and enables the development of technology clusters with fast and easy access between locations. In the meantime, it also expands tourism markets while increasing tourists spending.

1.2 Brief History of HSR

High-speed trains were first introduced in the early 19th century when British engineers developed the steam railroad locomotives [1]. These locomotives were the beginning of a long era in which various technologies were pursued in order to achieve higher speeds coupled with operating and fuel efficiencies. Steam propulsion gave maximum speeds of about 160kph (100mph) in the early 20th century. Test speeds of 209kph (130mph) were attained by using diesel-electric and electric propulsion in the 1930s and 1940s. This change in technology provided slightly improved performance and fuel efficiency, but it was not until the 1960s that breakthroughs in suspension, train-track dynamics, and other factors permitted an increase in train speeds.

On November 1st, 1965, the Japanese introduced, as part of the regular train schedule, a standard-gauge railway service that reached the maximum speed of 209kph (130mph) [31]. The average speed of the train was 166kph (103mph). The Pennsylvania railroad developed the “metro-liners” by placing traction motors on each axle [33]. That train achieved the maximum speed of 251kph (156mph) in Princeton, New Jersey, in May 1967. Only a decade ago, “metro-linerlike” coaches pulled by AEM-7(Asea Electro Motive 7000hp) locomotives operating at speeds of up to 201kph (125 mph) in service were the only relatively high-speed rail systems in the United States [35]. At present, high-speed rail trains regularly operate in Japan, France, Germany, and other countries worldwide in service, at speeds between 240 and 300kph (150 and 187mph). The Japanese have been operating their Shinkansen

(bullet train) between Osaka and Tokyo for more than thirty years [31].

The French initiated high-speed service between Paris and Lyon with the TGV (Train à Grande Vitesse - French high speed train) in 1981 [32]. The latest French TGV trains operate in service on several lines (about 25% of the lines are dedicated to their use) connecting Paris with Brussels, London (under the British Channel, called the Eurostar), and cities of the Atlantic and Mediterranean coasts. The TGV achieves peak speeds of 320kph (200mph) and has been tested at a sustained speed of 515kph (330mph). Similarly, the Germans operate their own ICE (Inter-City-Express Train) all over the country (using mainly non-dedicated lines).

HSR in China may refer to any commercial train service in China with an average speed of 200kph (124mph) or higher [29]. By that measure, China has the world's longest HSR network with over 11,028 km (6,852 mi) of track in service as of December 2013, including the world's longest line, the 2,298 km (1,428 mi) Beijing–Guangzhou High-Speed Railway.

Since high-speed rail service in China was introduced on April 18th, 2007, daily ridership has grown from 237,000 in 2007 to 1.33 million in 2012, making the Chinese high-speed rail network the most heavily used in the world [34]. China's high-speed rail network consists of upgraded conventional railways, newly built high-speed passenger designated lines (PDLs), and the world's first high-speed commercial magnetic levitation (maglev) line. Nearly all high-speed rail lines and rolling stock are owned and operated by the China Railway Corporation, the national railway operator formerly known as the Railway Ministry. The Shanghai Maglev Train is owned and operated by Shanghai's city government [29].

China's early high-speed trains were imported or built under technology transfer agreements with foreign train-makers including Alstom, Siemens, Bombardier and Kawasaki Heavy Industries. Chinese engineers then re-designed internal train

components and built indigenous trains that can reach operational speeds of up to 380kph (240mph).

1.3 Safety Issues and Technology

As the speed of trains becomes faster and faster, the severity of accidents can be disastrous. In the city of Wenzhou, China on July 24th, 2011, a deadly collision happened and over forty people were killed in this accident. It was caused when one train, struck by lightning, lost power and a second train crashed into its back [11].

Figure 1.1 shows the rescue effort.



Figure 1.1 Deadly Collision. Source: CBC NEWS

Another accident happened in Bulgaria on July 12th, 2014. A long distance passenger train derailed at Kaloyanovets station, killing the driver and injuring fourteen others. The train passed through turnouts set to diverging track at 104kph.

A large part of the research regarding high speed ground transportation systems has focused on safety issues [36]. Many of these issues deal with crash tests and car designs that will minimize bodily injuries in case of a crash.

While train collisions are extremely rare occurrences, when one happens, heavy casualties are almost unavoidable. For this reason, many European countries and the USA have developed design methods to minimize damage in the event of a train collision [37]. Moreover, in an effort to improve train crashworthiness, these countries have developed safety guidelines for passenger trains. One aspect of such efforts is to apply design to the car and make it crushable, thereby absorbing energy [38]. Those techniques include improved crashworthiness cab car and locomotive structural designs [5, 6], improved crashworthiness coach car structural designs [7], and a variety of interior occupant protection strategies [8]. Since the setup of real train collision test facility requires massive amounts of time and expense, such a technique was rarely used before. Therefore, simulation methods are particularly important in the train collision testing field. So far, crashworthiness has been assessed using four modeling techniques. The first is to sequentially analyze and evaluate the partial structure of a train by using a 3D finite element modeling techniques [1]. The second is to analyze the crushing characteristics of the entire train by applying the crushing characteristics obtained from the first technique to a 1D train model [2]. The third is to use a 2D modeling technique to estimate overriding [3]. The fourth is a technique to predict the behavior of an entire train by using a 3D multi-body dynamics model [10]. Although these methods have been efficiently applied in computer system without much difficulty, they have many restrictions in terms of their ability to predict the nonlinear collision behavior that occur in an actual collision [39]. In other words, the aforementioned evaluation methods all have the shortcoming that they cannot consider interactions sequentially occurring among each component, suspension, bogie and the body [40].

As a matter of fact, the most ideal technique providing the highest reliability in predicting collision behavior is to use a real train collision [41]. Another main purpose of doing the real train collision test is to validate the simulation code because it is much cheaper to run the computer code than crashing an actual train. In the following, the motivation of the thesis is first presented, after which the contributions

of this thesis are highlighted, and the structure of the remaining chapters are outlined.

1.4 Motivation

The most ideal technique providing the highest reliability in predicting collision behavior is to use a real train collision, but this does require massive amounts of time and expense [41]. In the real train collision test, a train locomotive needs to be propelled on a straight, guided path, to a particular speed, at which time the train locomotive is released to coast down to a crash barrier where it is required to crash at a desired speed. That propulsion could be in the form of a cable pulling the train inside a laboratory setting or another locomotive pushing the test object along a long train track. The current control of the release speed and the release location is based on theoretical data and previous experience which leads to less accuracy in controlling the actual crash speed. The resistance forces, including inertia, friction and aerodynamic resistance forces, are playing important roles and significantly influence the behavior of both propulsion and coast-down movement. And due to the lack of a mechanism to collect the real-time data to calculate the resistance force along the test, the current release speed and location is only roughly estimated. Inaccuracy of crash speed means the waste of a great deal of time and money. The motivation in this thesis is to design and implement a solution to simulate the performance data from the propulsion phase, and trigger an algorithm to calculate the release speed and location more accurately and quickly. The critical works in this solution are to implement a proper mathematical model to describe and analyze the behavior of resistance force, calculate the regression coefficients quickly, simulate the coast-down process and find the release speed and location precisely.

1.5 Contributions

This thesis is concerned with the problem of mathematical models to analyze the resistance forces, and the simulation for both coast-down and propulsion

behavior. The contributions of this thesis are as follows.

The primary contribution of this thesis is that a regression analysis data model is applied to analyze the data set obtained from industry experts that is used to calculate and estimate the real time performance of the testing vehicle, such as forces, acceleration, [pulling cable] motor torque, speed and location. In a real testing environment, such data will be collected in a real time basis through an array of sensors. After the calculation based on the provided data set, a new data set is obtained describing the relationship between resistance force and speed. The least squares method is then applied to calculate and determine the regression coefficients which in turn leads to a best fit function of resistance to describe the behavior of the new obtained data set. This resistance function will be used later on to predict the performance of resistance force for coast-down movement as well as for the propulsion behavior.

The second important contribution is that a practical approach is designed and implemented to simulate the coast-down behavior. The coast-down movement starts at the release location where the propulsion force is released and it ends at the position of collision between the test vehicle and the barrier. There are several difficulties in such a simulation. For example, the location and speed of the start position of coast-down are unknown and the whole coast-down movement is a varying acceleration motion. In order to solve this problem, a reverse analysis is designed and provided in this research. The coast-down simulation is done by calculating the speed and location at each time stamp starting from the end position of the coast-down movement instead of the start position, because the speed and location for the end position, which means the crash speed and location of barrier, are predefined and known.

The third contribution of this thesis is that a software application is well designed and developed to carry out the proposed approach. In Chapter 3, the details

of the design and development are presented, such as system architecture, software structure, design principles, program logic and flows chart. This software application is part of the overall control system (Facility Control System) in a real test facility. The data-feed interface in this application can be easily modified to adapt to the real interface to measuring sensors. This software can also be customized as a standalone tool to let the users be able to change and manipulate different parameters to simulate different scenarios.

1.6 Guide to the Thesis

This thesis is organized as follows.

Chapter 2: Background Knowledge. This chapter provides an introduction to the subject that the proposed method builds upon. The theory of regression analysis will be reviewed at first. Multiple linear regression and least squares method are specifically emphasized, since they are the core and foundation of the proposed approach. After the introduction of regression analysis, the idea of real train collision test and some major concepts will be explained.

Chapter 3: Proposed Solution - *RTPSA*. The proposed method is presented and explained in this chapter which covers the major contribution of this thesis to fulfill this research work. The software developed is called *RTPSA (Real Time Predictive Speed Analysis)*. The definition of the problem is described in the beginning of this chapter, followed by objective and design principles of the proposed approach. Then, system architecture and software modules will be introduced in detail. After reading this chapter, readers will be able to understand the advantages of the proposed method and the implementation details in this research work.

Chapter 4: Running Results. In this chapter, the results of experiments are

presented, together with explanation and analysis of the result tables. This leads to the conclusion in following chapter.

Chapter 5: Conclusion and Future Work. This final chapter brings conclusion of the thesis and presents a sketch of possible future work.

Chapter 2 Background Knowledge

This chapter provides the background knowledge on which the proposed method is based. The theory of regression analysis will be reviewed which includes simple linear regression, multiple linear regression, the calculation of regression coefficients and the correlation coefficients. After that, test phases, typical setup in a real train collision test, an in-depth knowledge of the vehicle motion and sample data sheet will be introduced.

2.1 Regression Analysis

Regression analysis is a statistical technique for the investigation of relationships between variables [13]. Usually, the investigator seeks to ascertain the causal effect of one variable upon another—the effect of a price increase upon demand, for example, or the effect of changes in the money supply upon the inflation rate [19]. To explore such issues, the investigator assembles data on the underlying variables of interest and employs regression to estimate the quantitative effect of the causal variables upon the variable that they influence. The investigator also typically assesses the “statistical significance” (correlation coefficient) of the estimated relationships, that is, the degree of confidence that the true relationship is close to the estimated relationship [20].

The goal of regression analysis is to determine the values of parameters for a function that cause the function to best fit a set of data observations provided. In simple linear regression, the function is a linear (straight-line) equation.

2.1.1 Simple Linear Regression

In its simplest form, linear regression shows the relationship between one independent variable X and a dependent variable Y , as in the formula below:

$$Y = a_0 + a_1X + u \quad (2.1)$$

The magnitude and direction of that relationship are given by the slope parameter a_1 , and the status of the dependent variable Y when the independent variable X is 0 is given by the intercept parameter a_0 . An error term u captures the amount of variation not predicted by the slope and intercept terms [14].

An example of simple linear regression is the relationship between muscle strength against lean body mass. Figure 2.1 shows a data distribution about them.

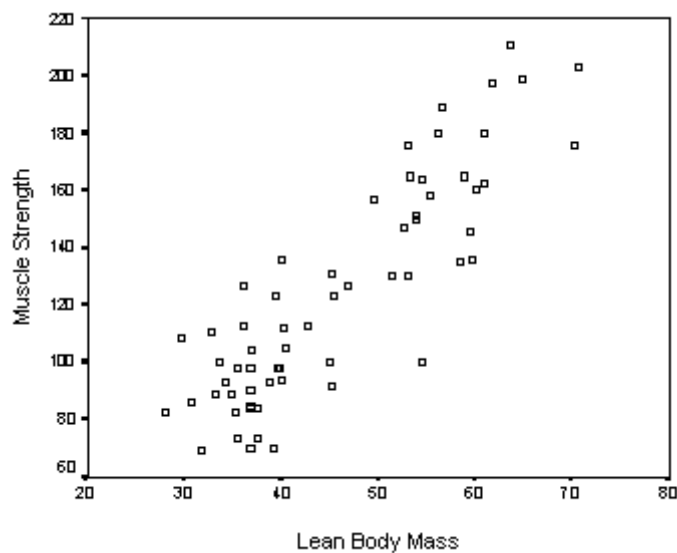


Figure 2.1 Data Distribution of Lean Body Mass (kg) and Muscle Strength (kg/cm²). Source: [46]

The X -axis in Figure 2.1 represents lean body mass while the Y -axis represents muscle strength. Each small square is a pair of data of (x, y) . By observing Figure 2.1, those who have more lean body mass tend to be stronger. Although the

relationship isn't perfect because it's easy to find two people where the one with more lean body mass is the weaker, but in general strength and lean body mass tend to go up and down together. When two variables are displayed in a scatterplot and one can be thought of as a response to the other (here, lean body mass produces strength), standard practice is to place the response on the vertical (or Y) axis. The names of the variables on X and Y axes vary according to the field of application. Some of the common usages are as follows.

X -axis	Y -axis
independent	dependent
predictor	predicted
carrier	response
input	output

The association between lean body mass X and muscle strength Y looks like it could be described by a straight line shown in Figure 2.2.

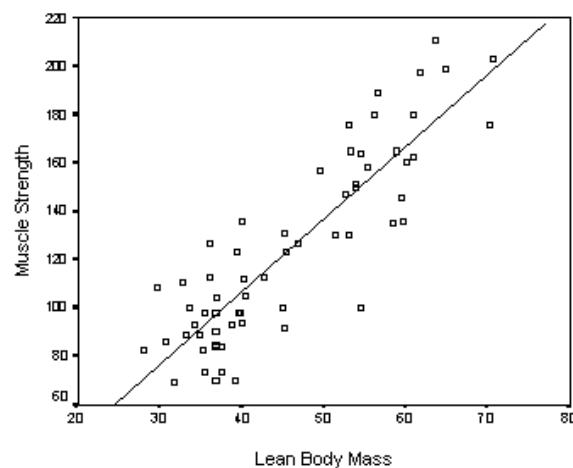


Figure 2.2 Simple Linear Regression - Straight Line. Source: [46]

The data (each small square) are pairs of independent and dependent variable values $\{(x_i, y_i): i = 1, \dots, n\}$. The fitted function could be written as

$$\hat{y} = b_0 + b_1x \quad (2.2)$$

where \hat{y} is the predicted value of the response obtained by using above equation, b_0

and b_1 are regression coefficients. The residuals are the differences between the observed and the predicted values $\{(y_i - \hat{y}_i): i = 1, \dots, n\}$.

There are two primary reasons for fitting a regression function to a set of data. The first is to describe the data and the second is to predict the dependent variable value from the independent variable value. The method of calculation for the fitted regression function will be introduced in Section 2.1.4 and the rationale behind the way the regression line is calculated is best seen from the point-of-view of prediction [21]. A line gives a good fit to a set of data if the points are close to it. Figure 2.3 shows that there could be many different lines to fit the data set. Where the points are not tightly grouped about any lines, a line gives a good fit if the points are closer to it than to any other lines. For predictive purposes, this means that the predicted values obtained by using the line should be close to the values that were actually observed, that is, the residuals should be small. Therefore, when assessing the fit of a line, the vertical distances of the data points to the line are the only distances that matter. Perpendicular distances are not considered because errors are measured as vertical distances, not perpendicular distances. Figure 2.4 shows the relationship among observed value, predicted value and vertical distance.

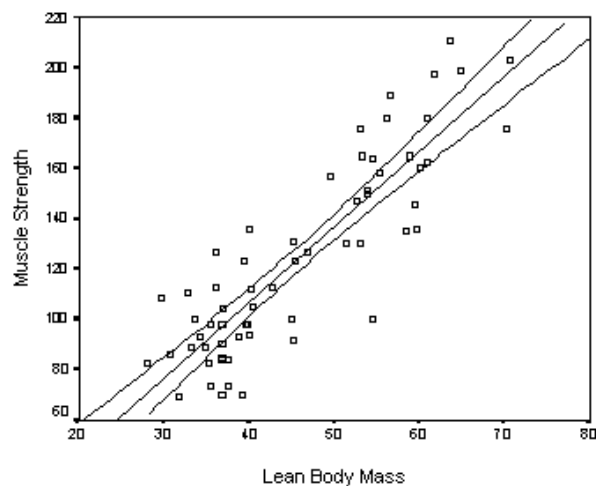


Figure 2.3 Different Fitted Lines for A Set of Data. Source: [46]

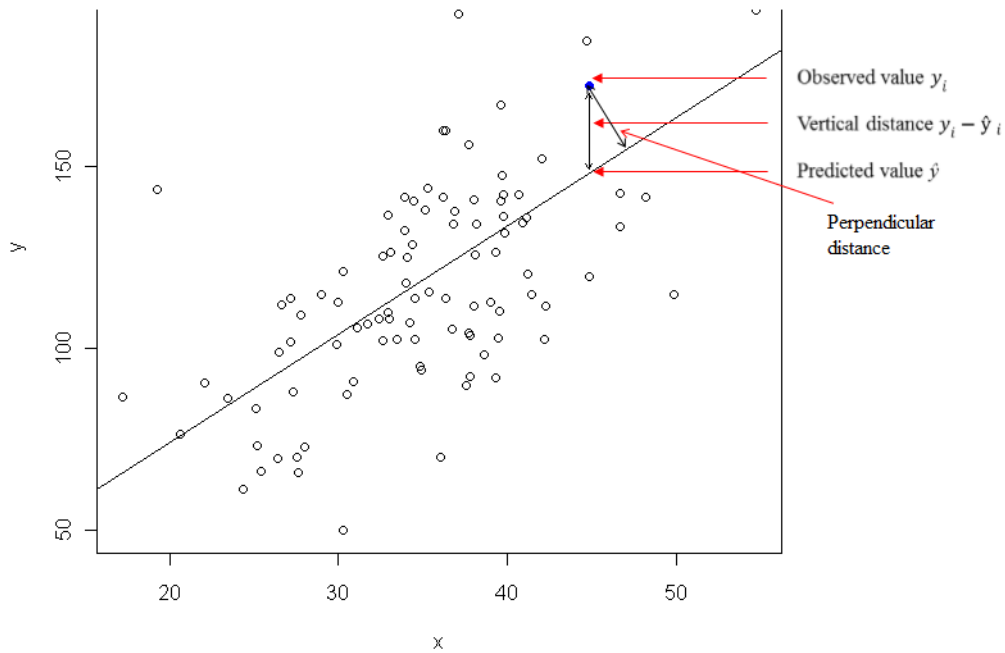


Figure 2.4 Vertical Distance vs Perpendicular Distance. Source: [46]

2.1.2 Multiple Linear Regression

In simple linear regression, there is only one major influencing factor as independent variable to explain the changes of one dependent variable. In the real world, the dependent variable changes are often affected by several important factors, then two or more impact factors as independent variables are used to explain the changes in the dependent variable, which is known as multiple regression. When the relationship between multiple independent variables and the dependent variable is linear, the regression analysis is called multiple linear regression [22].

The general form of multiple linear regression is as follows.

$$Y = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n + u, \quad (2.3)$$

where Y represents the dependent variable, $X_i, i = 1, \dots, n$ are the independent variables, $a_i, i = 0, \dots, n$ are parameters and u is the amount of variation.

Same as simple linear regression, multiple linear regression has similar practical uses. Most applications fall into one of the following two broad categories:

- If the goal is prediction, or forecasting, or reduction, linear regression can be used to fit a predictive model to an observed data set of (x, y) values. After developing such a model, if an additional value of X is then given without its accompanying value of Y , the fitted model can be used to make a prediction of the value of Y .
- Given a variable Y and a number of variables X_1, \dots, X_p that may be related to Y , linear regression analysis can be applied to quantify the strength of the relationship between Y and the $X_j, j = 1, \dots, p$, to assess which X_j may have no relationship with Y at all, and to identify which subsets of the X_j contain redundant information about Y .

One of the major objectives to study multiple linear regression model is to find the fitted function to best describe the data distribution, and then it can be used for prediction [25]. When the number of independent variables is two, the multiple linear regression has the general form as follows,

$$Y = a_0 + a_1X_1 + a_2X_2 + u \quad (2.4)$$

where Y represents the dependent variable, X_1 and X_2 are the independent variables, a_0, a_1 and a_2 are parameters and u is the amount of variation. The fitted function for this general form can be written as

$$\hat{y} = b_0 + b_1X_1 + b_2X_2 \quad (2.5)$$

where \hat{y} is the predicted value of the response obtained by using the equation, and b_0, b_1, b_2 are regression coefficients. It is obvious that the fitted function will be known if the regression coefficients are determined. So the problem of finding the best fit function is equivalent to the determination of regression coefficients.

In the following section a special case will be introduced first which is how to

deal with a quadratic relation between the independent and dependent variables into the multiple linear regression model. And right after that the calculation of regression coefficients will be explained in detail.

2.1.3 Quadratic Relationship between Dependent and Independent Variables

When the dependent variable has a quadratic relation with one independent variable, its general form can be written as

$$Y = a_0 + a_1X + a_2X^2 + u, \quad (2.6)$$

where Y represents the dependent variable, X is the independent variable, a_0 , a_1 and a_2 are parameters and u is the amount of variation.

For example, a data set of pairs of independent X variable values and dependent Y variable values $\{(x_i, y_i) : i = 1, \dots, n\}$ are obtained in a particular application, and it is known that Y has the quadratic relationship with X . Can we apply the multiple linear regression model to this data set and find its fitted function?

The answer is yes. By observing equation (2.6) and substituting X with X_1 , X^2 with X_2 , the equation (2.6) can be re-written as

$$Y = a_0 + a_1X_1 + a_2X_2 + u. \quad (2.7)$$

Equation (2.7) shows that the dependent variable Y has the linear relationship with independent variables X_1 and X_2 . The relationship expressed in equation (2.7) has the same meaning as in equation (2.4). So, by substituting the independent variables, multiple linear regression can be applied to analyze the quadratic relationship.

The above method is a standard procedure in statistic research. Detail explanation about quadratic relationship in multiple linear regression analysis can be found in [45].

2.1.4 Calculation of Regression Coefficients

The method of least squares is a standard approach to calculate the regression coefficients. "Least squares" means that the best fit function minimizes the sum of the squares of the errors (the vertical distance or deviation) made in the results of every single deviation ($(y_i - \hat{y}_i): i = 1, \dots, n$, n is the number of samples). A mathematical procedure for finding the best-fitting curve for a given set of points is done by minimizing the sum of the squares of the deviations ($\sum_{i=1}^n (y_i - \hat{y}_i)^2$) [26].

In practice, the sum of the squares of the vertical deviations V^2 of a set of n data points is defined as

$$V^2(b_0, b_1, \dots, b_n) \equiv \sum_{i=1}^n [y_i - f(x_i, b_0, b_1, \dots, b_n)]^2,$$

from a fitted function f . (b_0, b_1, \dots, b_n) are regression coefficients in f . The data are pairs of independent and dependent variable values $\{(x_i, y_i) : i = 1, \dots, n\}$. According to the definition of "least squares", the minimum of sum of squares of errors needs to be found for its best fit function. The condition for V^2 to take the minimum value is that

$$\frac{\partial(V^2)}{\partial b_i} = 0 \quad \text{for } i = 0, \dots, n.$$

For a simple linear fitted function $f(x, b_0, b_1) = b_0 + b_1 x$

$$V^2(b_0, b_1) \equiv \sum_{i=1}^n [y_i - (b_0 + b_1 x_i)]^2$$

$$\frac{\partial(V^2)}{\partial b_0} = -2 \sum_{i=1}^n [y_i - (b_0 + b_1 x_i)] = 0$$

$$\frac{\partial(V^2)}{\partial b_1} = -2 \sum_{i=1}^n [y_i - (b_0 + b_1 x_i)] x_i = 0.$$

These lead to the equations

$$nb_0 + b_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \quad (2.8)$$

$$b_0 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i. \quad (2.9)$$

By solving above system of equations (2.8) and (2.9), the value of regression

coefficients b_0 and b_1 are obtained. [27]

Similarly, by applying the same approach, for a multiple (with two independent variables) linear fitted function $f(x_1, x_2, b_0, b_1, b_2) = b_0 + b_1 x_1 + b_2 x_2$,

$$V^2(b_0, b_1, b_2) \equiv \sum_{i=1}^n [y_i - (b_0 + b_1 x_{1i} + b_2 x_{2i})]^2.$$

Let

$$\frac{\partial(V^2)}{\partial b_0} = 0,$$

$$\frac{\partial(V^2)}{\partial b_1} = 0,$$

$$\frac{\partial(V^2)}{\partial b_2} = 0.$$

These lead to following three equations.

$$\sum_{i=1}^n y_i = n b_0 + b_1 \sum_{i=1}^n x_{1i} + b_2 \sum_{i=1}^n x_{2i} \quad (2.10)$$

$$\sum_{i=1}^n x_{1i} y_i = b_0 \sum_{i=1}^n x_{1i} + b_1 \sum_{i=1}^n x_{1i}^2 + b_2 \sum_{i=1}^n x_{1i} x_{2i} \quad (2.11)$$

$$\sum_{i=1}^n x_{2i} y_i = b_0 \sum_{i=1}^n x_{2i} + b_1 \sum_{i=1}^n x_{1i} x_{2i} + b_2 \sum_{i=1}^n x_{2i}^2 \quad (2.12)$$

By solving above system of equations (2.10), (2.11) and (2.12), the value of regression coefficients b_0, b_1 and b_2 are obtained [27].

If the independent variables are more than two, the least squares method works in the same way. In this research work, least squares method for one and two independent variables are applied. The least squares method is a standard procedure in statistic research. Detail explanation about least squares method in the calculation of regression coefficients can be found in [45].

2.1.5 Correlation Coefficient - R^2

Variance is a characteristic (or parameter) of a set of data, or a series of answers. It is the sum of the squared differences between each value and the mean of all values.

Let $x_i : i = 1, \dots, n$ denote the values of sample data, μ is the mean of all values, N is the number of values and V is variance. Then

$$V = \frac{1}{N} \sum_{i=1}^n [x_i - \mu]^2.$$

Variance is frequently partitioned into two categories. One is able to be attributed to a specific condition (explained) and another one is assigned to other unmeasured conditions (unexplained). This division is frequently referred to as "explained variance" and "unexplained variance" [47].

R^2 is a statistical measure of how close the data are to the fitted regression function [18]. It is also known as the correlation coefficient. The definition of R^2 is fairly straight-forward. It is the percentage of the response variable variation that is explained by a linear model. Or:

$$R^2 = \frac{\text{Explained variation}}{\text{Total variation}}. \quad (2.13)$$

For the general form of multiple linear regression

$$Y = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n, \quad (2.14)$$

Y represents the dependent variable, y_i is the value of Y , $X_i, i = 1, \dots, n$ are the independent variables, and $a_i, i = 0, \dots, n$ are parameters. The explained variation for the above general multiple linear regression form is the regression sum of squares. Let SS_R denote the regression sum of squares. Then

$$SS_R = \sum_{i=1}^n (\hat{y}_i - \mu)^2,$$

where \hat{y}_i is the modeled values, μ is the mean of all the original value of y_i . The total variation for this general multiple linear regression form is the total sum of squares. Let SS_T denote the total sum of squares. Then

$$SS_T = \sum_{i=1}^n (y_i - \mu)^2,$$

where y_i is the original data value, μ is the mean of all the original value of y_i .

So the equation (2.13) for R^2 can be re-written as follows,

$$R^2 = \frac{SS_R}{SS_T}$$

or

$$R^2 = 1 - \frac{SS_E}{SS_T}$$

where

$$SS_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

SS_E is also called the sum of squared errors.

R^2 is always between 0 and 100%:

- 0% indicates that the model explains none of the variability of the response data around its mean.
- 100% indicates that the model explains exactly all the variability of the response data around its mean.

The above method is a standard procedure in statistic research. Detail explanation about the calculation of R^2 can be found in [45].

2.2 Test Phases

Basically, the whole testing process in a typical real train collision test can be divided into two major phases, propulsion phase and coast-down phase. In propulsion phase, a test vehicle is pulled by a propulsion cart and accelerated on the guided track to a designated high speed in a very short time and then the test vehicle is released from the propulsion cart. From this time on, since the test vehicle is separated from the propulsion cart, it will coast down towards the rigid wall (or a barrier) and hit the rigid wall eventually. The crash speed is measured accurately which is one of the most important indicators for the whole testing. For some special purposes, a barrier (e.g. a car) will be placed close to the rigid wall and to be crashed into the test vehicle.

Under the real test conditions, the test vehicle is not released right away when it reaches the designated speed, it will maintain a certain speed for some time and get released at designated location and speed. In this paper, we only discuss the situation of releasing the test vehicle right away once it reaches the desired speed

and location. In Figure 2.5, it shows that the test vehicle is started from the right-hand side along a guided track. It will be propelled to a certain speed to release the propulsion force. After the propulsion force is released, the test vehicle will coast down towards the barrier (losing speed because of the resistance force), which could be a rigid wall or a car. The crash speed is the speed, at which the test vehicle hits the barrier. The crash speed is a very important indicator for test facilities in this industry and it will be measured according to industry standard.

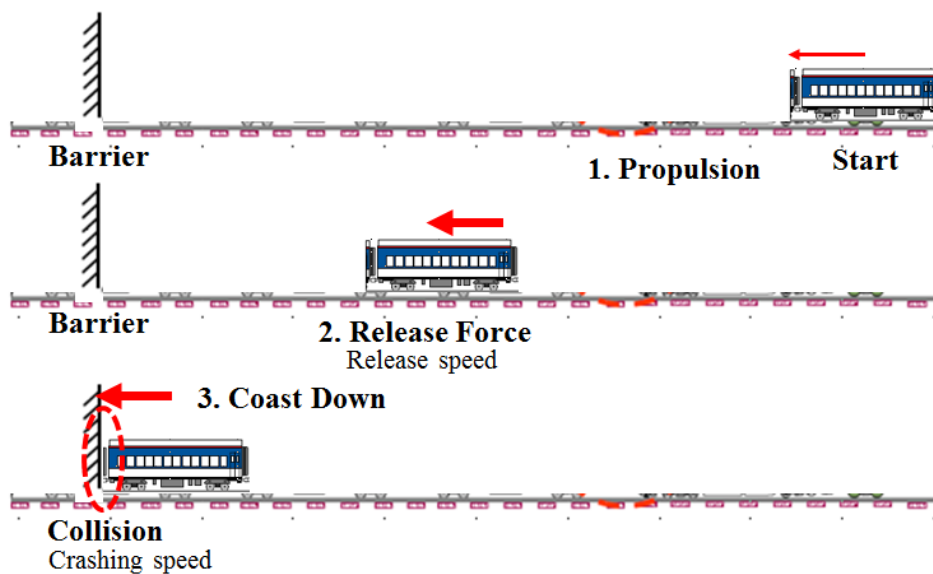


Figure 2.5 Propulsion – Release Force – Coast Down.

2.3 Test Facility

A real train collision test could be a train to truck or car, train to a rigid wall or train to train. In the test facility involved, it will be train to rigid wall and train to train. The track and vehicle are shown in Figure 2.6.

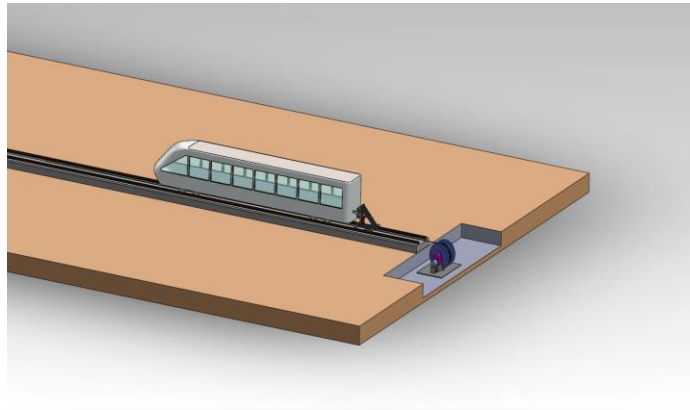


Figure 2.6 Train and Track. Source: Anemoi Technologies Inc.

The rigid wall is built inside a building with well-designed protection facilities. The building is shown in Figure 2.7 which is located at the other end of the track. And in Figure 2.8, it shows the rigid wall inside the building. This rigid wall is built with huge concrete base designed for high-speed collision test.

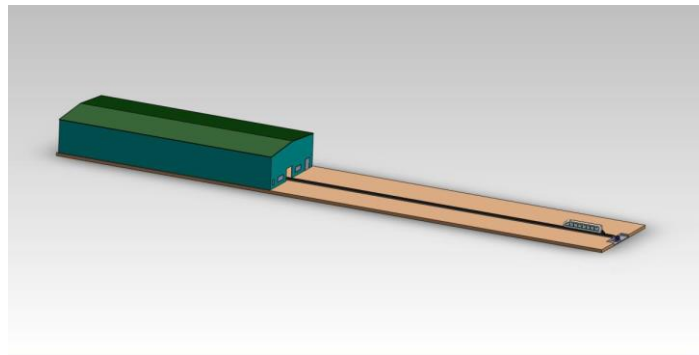


Figure 2.7 Building Housing the Rigid Wall. Source: Anemoi Technologies Inc.

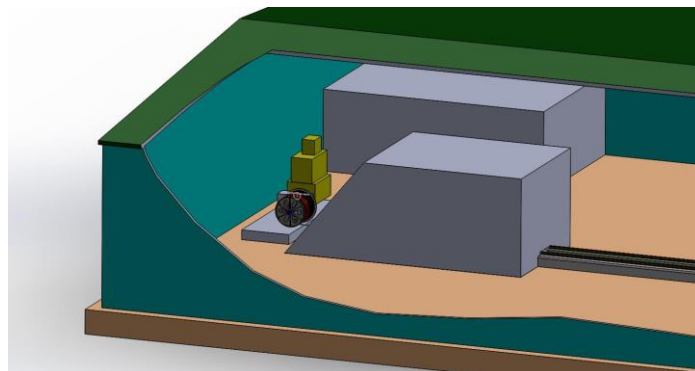


Figure 2.8 Rigid Wall. Source: Anemoi Technologies Inc.

The whole testing process will be done in about 15 to 30 second. A powerful electric motor and a specially-made pulling drum will be used in providing the propulsion force. A powerful braking drum will also be used to brake the propulsion system or the vehicle after the release, in the case of emergency. The highest required crash speed in the test facility in this research is a record high of 150kph for a 15000kg vehicle.

2.4 Vehicle Motion Analysis

During the propulsion phase, a test vehicle needs to be propelled on a straight, guided track, to a particular speed V_r (release speed), at which time the test vehicle is released to coast down towards a barrier where it is required to crash at a speed of V_c (crash speed). The distance between the start of the propulsion and the crash barrier is known. The difference between V_r and V_c is dependent on the coast-down distance (which itself is dependent on where the test vehicle is released from the propulsion mechanism) and the resistance forces during the coast-down, such as friction (both between wheels and rail and also train-internal friction such as bearing frictions etc.) and aerodynamic force.

In general, the equation of motion is

$$F - R = M * A, \quad (2.15)$$

where F is the propulsion force, R is the resistance force, M is the mass of vehicle, and A is the vehicle acceleration. Mass M can be measured and it is constant. Propulsion force F is controlled and measured. The propulsion acceleration A changes but it can be measured during the propulsion phase. Resistance force R is the sum of friction and aerodynamic force. Friction is constant which can be defined by coefficient of friction and mass. Aerodynamic force R_{ae} can be defined as follows

$$R_{ae} = cv + \frac{1}{2}S\rho C_d v^2, \quad (2.16)$$

where c is a skin friction dependency, S is the train frontal area, ρ is the air

density, C_d is train drag coefficient and v is the speed. Because we do not know in advance what those parameters in equation (2.16) are, such as ρ , C_d , etc., it is impossible to obtain the value of R_{ae} by using equation (2.16). But by observing equation (2.15), when F , M and A can be measured, the resistance R is the only unknown variable and can be calculated by equation (2.15). This is the fundamental reason why regression analysis is needed in this research. And the knowledge of how to use multiple linear regression analysis model to deal with the quadratic relationship between dependent variable and independent variables is already introduced in Section 2.1.3.

For the coast-down movement, since the propulsion force is released,

$$F = 0.$$

Because resistance R has the quadratic relationship with speed v , the regression function of resistance R can be expressed in a general form as follows,

$$R = b_0 + b_1v + b_2v^2, \quad (2.17)$$

where b_0, b_1, b_2 are regression coefficients. According to equation (2.15) and (2.17), we have

$$A = \frac{dv}{dt} = -\frac{R}{M} = -\frac{(b_0 + b_1v + b_2v^2)}{M}. \quad (2.18)$$

Therefore, for coast-down movement, the acceleration A can be calculated by equation (2.18) once regression coefficients b_0 , b_1 and b_2 are determined.

Thus, during the propulsion phase, speed v , propulsion force F and acceleration A need to be measured so that the actual propulsion performance could be continually extrapolated over a very short time. Based on the performance data (speed v , acceleration A and propulsion force F) collected, the value of resistance force R can be calculated using equation (2.15). The values of speed v and resistance force R form a set of data, $\{(v, R)\}$, describing the relationship between speed v and resistance force R . Then, following the discussion of regression analysis model in Section 2.1, the quadratic regression analysis model will be applied to this data set $\{(v, R)\}$, where resistance force R is the dependent variable

and speed v is the independent variable. By applying the least squares method discussed in Section 2.1.3 and 2.1.4, the regression coefficients b_0, b_1, b_2 in equation (2.17) can be calculated and determined. Then the regression function of R is determined. This resistance function R is critical in the simulation of coast-down movement and the calculation of release speed and location.

2.5 Sample Data Sheet

	A	D	E	F	G	H	I
8	v (kph)	F_a	F_m	R_a	A_a (m/s ²)	A_m	$R_c = F_m - M \times A_m$
9	0						
10	0.72	99,900	99,525	450	6.629991	6.63190	46.30
11	1.44	99,800	99,881	450	6.623296	6.62347	528.68
12	2.16	99,700	100,000	451	6.616582	6.61838	724.55
13	2.88	99,600	99,593	452	6.60985	6.61186	415.30
14	3.6	99,500	99,151	453	6.603098	6.60560	67.27
15	4.32	99,399	98,632	454	6.596328	6.59679	- 319.96
16	5.04	99,299	98,828	456	6.589539	6.59222	- 54.88
17	5.76	99,199	99,778	458	6.582732	6.58477	1,006.07
18	6.48	99,098	98,328	460	6.575905	6.57866	- 352.39
19	7.2	98,998	98,385	462	6.56906	6.56840	- 140.63
20	7.92	98,898	99,295	465	6.562196	6.56395	835.93
21	8.64	98,797	98,918	467	6.555313	6.55266	628.44
22	9.36	98,697	97,958	470	6.548411	6.54901	- 277.20
23	10.08	98,596	98,803	474	6.541491	6.53923	714.42
24	10.8	98,496	98,178	477	6.534552	6.53148	205.96
25	11.52	98,395	99,096	481	6.527594	6.52895	1,161.78
26	12.24	98,294	98,700	485	6.520617	6.52171	874.08
27	12.96	98,194	97,816	489	6.513621	6.51521	88.22
28	13.68	98,093	98,798	494	6.506607	6.50956	1,155.03

Figure 2.9 Sample Data 1. Source: Anemoi Technologies Inc.

Figure 2.9, part of the sample data sheet, is provided by Anemoi Technologies Inc. This data set is used in the design stage to estimate and simulate the propulsion performance, such as propulsion force, speed and acceleration, in order to better design the whole mechanism. In the real test facility, the propulsion performance data will be collected in a real time manner using sensors. There are 209 rows of records in this data sheet from the speed of zero to speed of 150kph.

There are three kinds of data in this data sheet, actual, measured and

calculated. Actual data is the one output from the propulsion system, or the data applied in theory. It can't be obtained exactly in reality due to mechanical deviations. Measured data is the one collected from the sensors placed along the track or installed on the propulsion cart. Calculated data is the one by calculation based on other known data. The columns in this data sheet are v of speed, F_a of actual force, F_m of measured force by sensor, R_a of actual resistance, A_a of actual acceleration, A_m of measured acceleration by sensor and R_c of calculated resistance.

Figure 2.10 shows another part of sample data sheet which explains the formula to estimate actual force F_a , measured force F_m , actual resistance R_a , actual acceleration A_a , and measured acceleration A_m .

	A	D	E	F	G	H	I
1		101000		fe estimated force		fa actual force (N)	
2		-1000		a0=	0.03	x	450
3		100000		a1=	0	x	
4		-500		a2=	3.025		not dependen
5		-0.5		M	15000		
6		random max	2.00%			0.10%	
7		N	0%	N	m/s2	m/s2	$R_c = F_m - M \times A_m$
8	v (kph)	F_a	F_m	R_a	A_a (m/s2)	A_m	

Figure 2.10 Sample Data 2. Source: Anemoi Technologies Inc.

As shown in Figure 2.10, F_a , cell D_8 , is simulated by a quadratic function of speed v , the expression of F_a can be written as

$$F_a = c_0 + c_1 v + c_2 v^2, \quad (2.19)$$

where c_0 , c_1 and c_2 are coefficients. The value of c_0 , c_1 and c_2 are set in the cell D_3 , D_4 and D_5 . Considering the mechanical deviations between the actual and measured data, the value of F_m , cell E_8 , is the addition of F_a and variance of errors, following the equation

$$F_m = F_a - \text{Max}(F_a) * E_1 + 2 * \text{RAND}() * \text{Max}(F_a) * E_1, \quad (2.20)$$

where E_1 is an error ratio for F_a that is used to describe the noise of initial analogue signal of F_a . The value of E_1 is set in the cell E_6 . The value of $\text{RAND}()$

is ranged from 0 to 1, and $Max(F_a)$ is the maximum value among all the data values of F_a . According to the sample data provided in Figure 2.10, R_a , cell F_8 , is also estimated by using a quadratic function of speed v , its expression can be written as

$$R_a = d_0MG + d_1MGv + d_2v^2, \quad (2.21)$$

where d_0 , d_1 and d_2 are coefficients, M is the mass of test vehicle, and G is the gravity acceleration ($G=9.81$). The value of d_0 , d_1 and d_2 are set in the cell G_2 , G_3 and G_4 . The value of M is set in the cell G_5 . According to equation (2.15), the actual acceleration A_a , cell G_8 , can be obtained by

$$A_a = \frac{F_a - R_a}{M}. \quad (2.22)$$

The value of A_m , cell H_8 , is the addition of A_a and variance of errors, following the equation of

$$A_m = A_a(1 - \frac{1}{2}E_2 + RAND()E_2), \quad (2.23)$$

where E_2 is an error ratio for A_m with its value set in the cell H_6 . The value of $RAND()$ is ranged from 0 to 1 as well.

The last field, cell I_7 , in the sample data sheet of Figure 2.10 is R_c , the calculated resistance force. According to equation (2.15), R_c can be obtained by

$$R_c = F_m - M * A_m. \quad (2.24)$$

Eventually, after implementing the logic and formulas in the developed software based on the data sheet of Figure 2.9 and Figure 2.10, a data set of pairs of speed v and R_c , $\{(v, R_c)\}$, is obtained which describes the characteristics of the relationship between speed and resistance force. This data set $\{(v, R_c)\}$ will be applied later on to determine the regression coefficients for a fitted regression function that can be used to predict the behavior of resistance force.

In this research, the actual data is used only to help estimate the measured data, so that a simulated data set can be generated. In the real collision test environment, the measured data will be obtained from various sensors. In the following sections and chapters, the subscripts for actual, measured and calculated values, such

as F_a , F_m , A_a and R_c , are no longer used. Only the capital letter will be used to express the measured value or the calculated value for each variable. For example, F means the measured force, R means the calculated resistance, A means the measured acceleration and M means the measured mass.

Chapter 3 Proposed Solution - *RTPSA*

In this chapter, details of the proposed approach will be introduced. The software developed is called *RTPSA* – **R**eal **T**ime **P**redictive **S**peed **A**nalysis.

3.1 Limitation of Existing Methods

In a real collision test facility, the propulsion system is the key component. It is a group of mechanical and electrical machinery and equipment and it needs to be calibrated before putting it into operation. To-date such calibration was based on theoretical data and past experience. But during the test, resistance forces, such as friction resistance force and aerodynamic resistance force, play an important role influencing the behavior of test vehicle movement. Resistance forces vary all the time and are subject to the speed of test vehicle, wind speed and direction, temperature and humidity. So every test is different. When the propulsion system is calibrated using old data or parameters, it leads to less accuracy in terms of the release speed and location, and the crash speed as well. The essence of *RTPSA* is that the most up-to-date calibration information can be derived from studying in real time the actual propulsion behavior during the actual test. In order to do this, an efficient algorithm needs to be in place to calculate and predict the movement behavior of test vehicle during propulsion. This algorithm also needs to work in a real-time manner to monitor and control all the activities along the test.

3.2 Objective

The proposed approach in this research work is to address the issues discussed in Chapter 3.1. The objective of this research is to demonstrate a practical solution

by implementing a new design for a real collision test facility. A simulated data set is generated and used to describe the performance in propulsion phase. The quadratic regression analysis model is applied in this implementation to predict the behavior of resistance force and simulate the movement of the test vehicle. The proposed software works in a real-time manner and provides a more accurate release location and speed for the overall control system to fulfill the actual force release.

3.3 Design Principles

There are three design principles for the proposed approach in this thesis. They are real-time, predictive, and simulation.

3.3.1 Real-Time

During the propulsion phase, the propulsion performance (speed, acceleration and propulsion force) are measured on a real-time basis, instead of using old data and past experience, so that the actual performance could be continually extrapolated in a very short period of time. This lets the software calculate the regression coefficients b_0, b_1, b_2 for resistance function (2.17) by inserting this into equation (2.15) and measuring F and A (M is known). And this approach leads to better accuracy in the prediction of resistance force behavior which is the key factor influencing the overall performance.

In this research work, a data set (stream) is generated simulating the behavior of propulsion phase in a real-time manner. When generating the simulated data stream, as shown in Figure 3.1, the whole propulsion phase is divided into many small intervals of speed. The X -axis in Figure 3.1 is time and Y -axis is speed. Each interval has the same speed difference of 0.2m/s. For each of the small speed intervals, propulsion force F (Eq.(2.20)), acceleration A (Eq.(2.23)), and

resistance R (Eq.(2.24)) are all calculated following the rules and formulas discussed in Section 2.5.

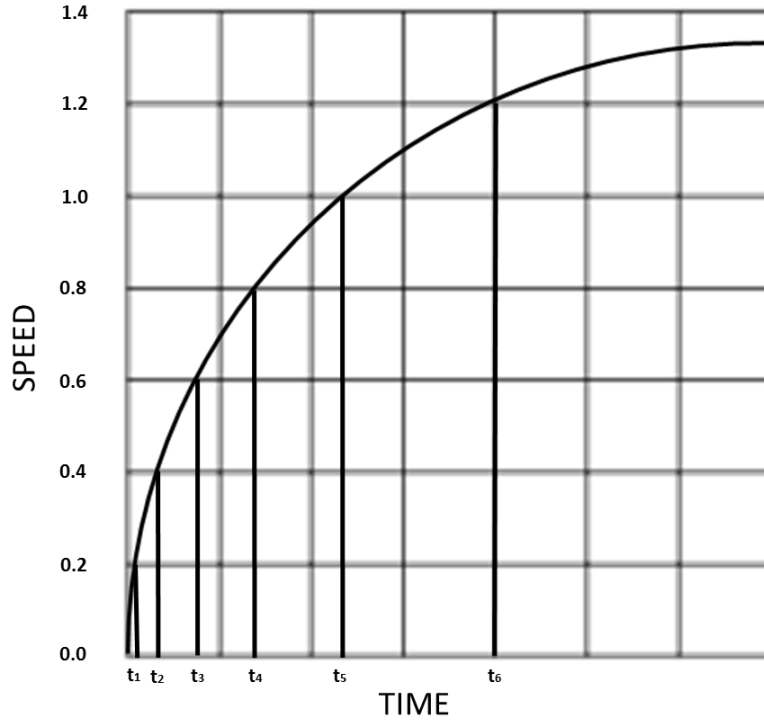


Figure 3.1 Small Intervals of Propulsion Curve

Meanwhile, each small interval is treated as constant acceleration motion. By applying the constant acceleration motion formulas, the time length, distance and location can be calculated. For example, for interval i , let v_{i0} denote the start speed, v_{it} denote the end speed, A denote the acceleration, t denote the time spent, S_{i0} denote the start location, and S_{it} denote the end location, then

$$v_{it} = v_{i0} + At, \quad (3.1)$$

$$S_{it} = S_{i0} + v_{i0}t + \frac{1}{2}At^2. \quad (3.2)$$

From equation (3.1), we have

$$t = \frac{v_{it} - v_{i0}}{A}. \quad (3.3)$$

Since v_{it} and v_{i0} are known for each interval, and acceleration A can be obtained by equation (2.23), so from equation (3.3), time t can be obtained. By substituting the value of t into equation (3.2), the end location S_{it} can be calculated.

Eventually, a data set of details of propulsion performance is obtained after the calculation. Specifically, a data set $\{(v, R)\}$ of pairs of speed v and resistance R is obtained which describes the characteristics of the relationship between speed and resistance force. This data set of speed and resistance $\{(v, R)\}$ will be applied later on to determine the regression coefficients for a fitted regression function that can be used to predict the behavior of resistance force. In conclusion, the whole propulsion phase is monitored and described in detail. The design of small interval makes it possible to monitor the whole propulsion behavior in a real-time manner.

3.3.2 Predictive

The prediction capability in *RTPSA* software is based on the real-time performance data that comes from the propulsion phase. One of the results of simulated data stream is a data set of pairs of speed v and resistance force R , $\{(v, R)\}$. As discussed in Section 2.4, 2.5 and 2.1.3, resistance force R has the quadratic relationship with the speed v and multiple linear regression model can be applied to the quadratic relationship after the substitution of independent variables. The general form of the fitted regression function can be written as

$$R = b_0 + b_1v + b_2v^2, \quad (3.4) \text{ same as (2.17)}$$

where b_0, b_1, b_2 are regression coefficients. By applying the least squares method introduced in Section 2.1.4 into the data set $\{(v, R)\}$, the regression coefficients b_0, b_1, b_2 are calculated. With the determination of regression coefficients b_0, b_1, b_2 , the fitted regression function of resistance R (Eq. (3.4)) will be used to predict the equation of motion during the coast down period, as well as for the last period of propulsion phase. Details of the simulation for the last period of propulsion phase will be introduced in section 3.3.3 and 3.9.

As discussed earlier, resistance force R plays an important role in the whole test phases. With the in-time prediction of resistance force R , details of the behavior for both coast-down (by letting $F=0$) and propulsion can be calculated and described,

that makes it possible to calculate and find the release speed and location.

3.3.3 Simulation

Simulation technique is needed and implemented in *RTPSA* because the algorithm needs to be triggered to calculate and provide the release location and speed to the control system before the test vehicle actually reaches the release point. A leading time of 1-2 second is required for the control system to activate the actual force release. Simulation is a practical way to do this work.

There are two movement phases that need to be simulated in order to figure out the release location and release speed, the last period of propulsion and the coast-down movement.

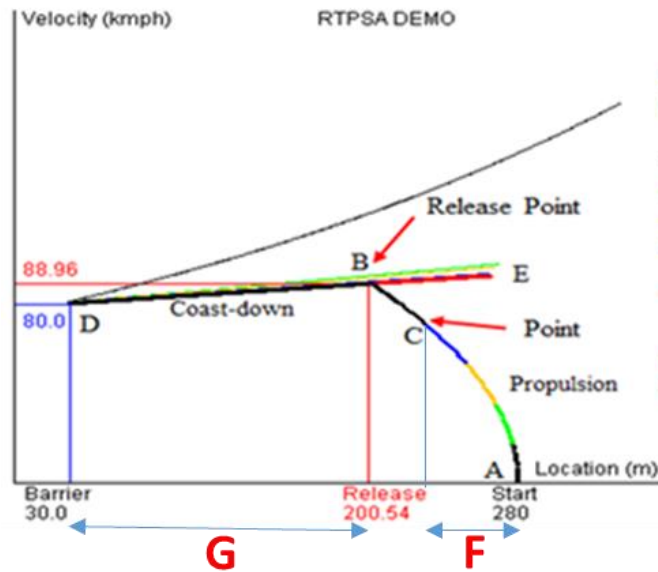


Figure 3.2 Start and End Points for Simulations. F: Calculate Regression Coefficients; G: Use Regression Coefficients to Calculate Coast-down Behavior Backwards

As shown in Figure 3.2, the *X*-axis is location while the *Y*-axis is velocity. The propulsion starts at position *A* (location=280m) and ends at position *B*. At position

B the propulsion force is released and the coast-down phase starts at position B right away. Position D is the end of coast-down movement where the test vehicle hits the barrier. During propulsion, when test vehicle reaches position C , the algorithm is used to calculate the release location and release speed. The movement from C to B is called the last period of propulsion in this research. By observing Figure 3.2, point B is the release point and it is the intersection of CB and BD . Therefore, in order to figure out the release location and release speed at position B , the function of line CB and the function of line BD must be known. Simulation technique is the choice in this research. With the simulation of movements of coast-down BD and the last period of propulsion CB , it becomes possible, at position C , to calculate the release location and release speed.

Details on how to do these two simulations are introduced in Section 3.8.1 and Section 3.9.1.

3.4 System Architecture

In a real test facility, physically, there are an array of sensors to measure the performance, such as propulsion force, acceleration, speed, location, temperature, humidity, wind speed and direction. They are installed on the propulsion cart which pushes the test vehicle, along the guided track and in the building. Sensors are connected to an IP network and controlled by an overall control system, which is called Facility Control System (FCS). PLC (Programming Logic Controller) controllers and DeviceNet/Ethernet converters are also used to help convert signals and do the preliminary calculation. Figure 3.3 shows an overall system architecture of the equipment under control.

SYSTEM ARCHITECTURE

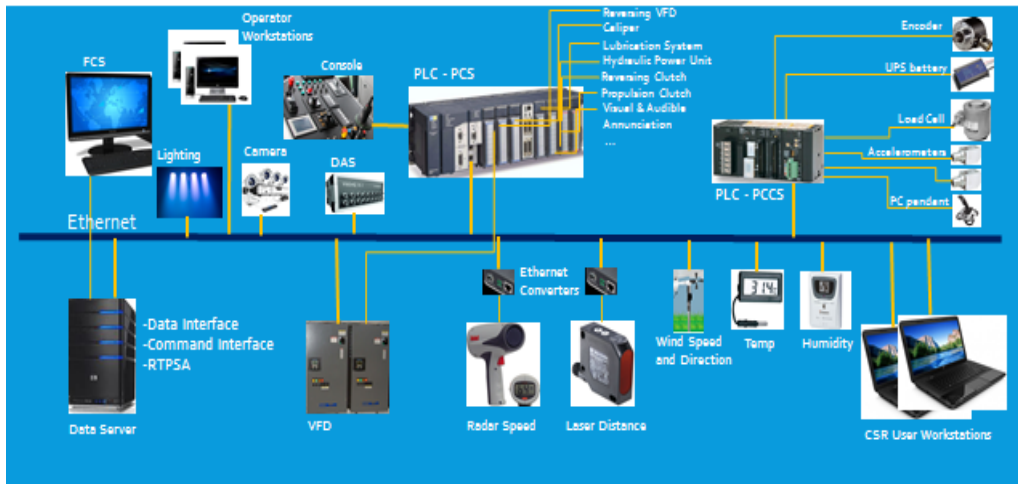


Figure 3.3 System Architecture. Source Anemoi Technologies Inc.

In Figure 3.3, the Facility Control System (FCS) is the core which is installed in a database server located at top left of the figure. All the activities are controlled by FCS. Data interface is designed and used by FCS to collect real-time data from sensors, PLCs and VFD (Vary Frequency Device). Command interface is designed and used by FCS to send control command to PLCs, VFD and sensors. All the equipment, such as DB server, main control monitors, administrative work stations, VFD, PLC controllers and many different sensors, are connected in a high speed Ethernet network. Most of the sensors are connected into Ethernet through PLCs. Some special sensors are connected through Ethernet converters, such as radar speed sensor and laser distance sensor. *RTPSA* is part of FCS and used to collect the performance data in a real-time manner from sensors during propulsion phase and provide input to FCS to actually control the force release.

3.5 Software Structure

The *RTPSA* software works in a real-time manner. It has a main process and four modules.

- Coefficients Calculation module

- Coast-down Simulation module
- Propulsion Simulation module
- Control/Release module.

RTPSA software interacts with other systems especially during propulsion and control/release processing. The input of *RTPSA* system is the continuous data stream generated by sensors during the propulsion phase. In this research a simulated data stream is generated in the main process to simulate the real-time data stream input. One of the outputs of *RTPSA* software is the release speed and release location which is the critical information for the control system to conduct the actual force release. Another output of *RTPSA* software is a demonstration video clip to display the whole simulation process. The average running time for current version (v1.1.9) is 5 milli second. The whole software is developed in *JAVA*.

Software Structure

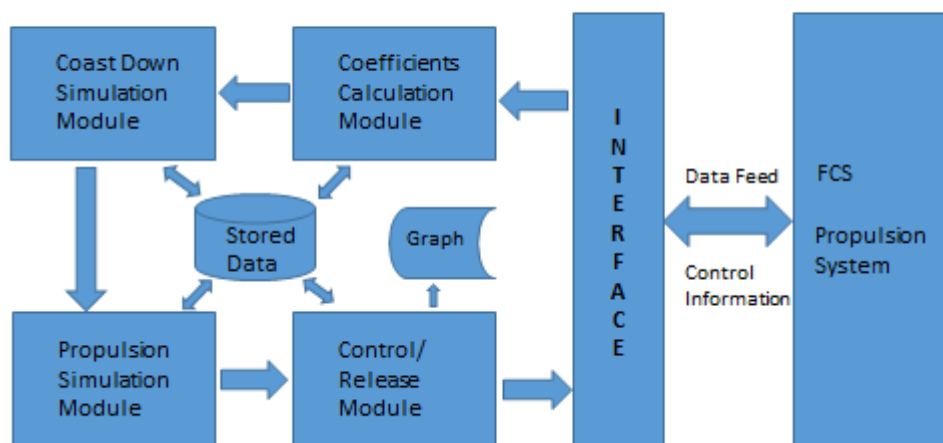


Figure 3.4 Software Structure.

As shown in Figure 3.4, data stream (the values of propulsion force, acceleration, resistance, speed and time stamp) feeds from propulsion system in the real test facility into Coefficients Calculation Module as input. In Coefficients Calculation Module, these data is stored and processed, and then a data set of pairs of speed and

resistance $\{(v_i, R_i) : i = 1, \dots, n\}$ is generated describing the relationship and distribution of speed and resistance force. The quadratic regression analysis model is applied to this data set $\{(v_i, R_i) : i = 1, \dots, n\}$ and the regression coefficients for the fitted function of resistance are calculated accordingly and quickly. Those regression coefficients determine a fitted function to best describe the resistance force along the test vehicle movement.

As discussed earlier, for the coast-down movement, there is no propulsion force but only resistance force existed. So the regression function of resistance force (Eq. (3.4)) will be used in Coast-Down Simulation Module to predict and describe the behavior of resistance force. Based on the prediction of resistance force, acceleration, speed and location are calculated automatically in computer following motion formulas discussed in Section 2.4. At the end of Coast-Down Simulation Module, a data set of coordinates of location and speed $SET_{CD}: \{(S_i, v_i) : i = 1, \dots, n\}$ is obtained which describes the details of whole coast-down behavior.

In a very similar way, in the Propulsion Simulation Module, simulation technique will be applied for the last period of propulsion phase. The reason for this simulation is introduced in Section 3.3.3. Details of simulation for the last period of propulsion phase will be discussed later in Section 3.9. By the end of Propulsion Simulation Module, a data set of coordinates of location and speed $SET_{LP}: \{(S_i, v_i) : i = 1, \dots, n\}$ is also obtained which describes the behavior details of the last period of propulsion phase.

Once the behavior of both coast-down movement and the last period of propulsion movement are known, the simple linear regression model is applied to each of the data set of coordination, SET_{CD} and SET_{LP} , in Control/Release Module. After calculating the regression coefficients b_0, b_1 of equation (2.8) for the data sets of SET_{CD} and SET_{LP} , and using least squares method by solving the equations (2.8) and (2.9), two fitted functions are determined accordingly. These two

functions have the same form of

$$v = e_0 + e_1 S, \quad (3.5)$$

where e_0 , e_1 are regression coefficients, v is the speed, S is the location.

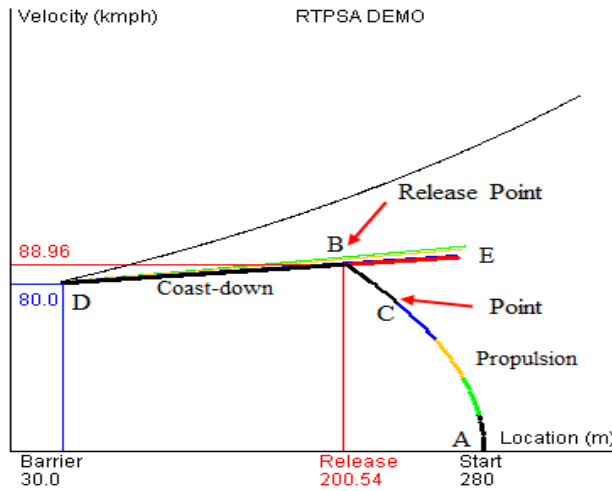


Figure 3.5 Intersection Point.

As shown in Figure 3.5, line DE is the fitted function of coast-down simulation, and line CB is the fitted function of the simulation for the last period of propulsion. Both line DE and line CB have the same form of equation (3.5) for their fitted functions, but the values of coefficients e_0, e_1 for each of them are different. The intersection B is the solution of this research. By solving those two equations of coast-down (line DE) and last period of propulsion (line CB), in Control/Release Module, release speed and location are obtained. The Control/Release Module will send the result to FCS to conduct the actual force release.

From Section 3.6, the details of the implementation will be introduced.

3.6 Main Process

Once $RTPSA$ software get started, the main process of the software is running and proceeding. The major task of this main process is the generation of the

simulated data stream according to the rules and formulas provided in the sample data sheet in Section 2.5. This simulated data stream is designed to describe the performance of propulsion movement. The format of the simulated data stream is as follows.

8	v (m/s)	F _a	F _m	R _a	A _a (m/s ²)	A _m	R _c = F _m -M x A _m	time	ds=v*dt	X from PS
9	0							0	0	280
10	0.2	99,900	99,286	450	6.629991	6.63071	- 174.92	0.030166	0.003017	279.997
11	0.4	99,800	99,104	450	6.623296	6.62419	- 258.91	0.060362	0.012076	279.9879
12	0.6	99,700	100,068	451	6.616582	6.61426	854.38	0.090589	0.027189	279.9728
13	0.8	99,600	100,321	452	6.60985	6.60871	1,190.36	0.120847	0.04837	279.9516
14	1	99,500	100,345	453	6.603098	6.60582	1,257.72	0.151136	0.07563	279.9244
15	1.2	99,399	100,059	454	6.596328	6.59587	1,120.49	0.181456	0.108981	279.891
16	1.4	99,299	99,290	456	6.589539	6.58711	483.61	0.211807	0.148438	279.8516
17	1.6	99,199	98,681	458	6.582732	6.58294	- 63.17	0.24219	0.194012	279.806
18	1.8	99,098	98,397	460	6.575905	6.57557	- 236.96	0.272604	0.245716	279.7543
19	2	98,998	99,627	462	6.56906	6.56592	1,138.45	0.30305	0.303563	279.6964

Figure 3.6 The Format of Simulated Data Stream.

3.6.1 The Generation of Simulated Data Stream

When generating the simulated data stream, the whole propulsion phase is divided into many small intervals. For each of the small intervals, propulsion force, acceleration, resistance, time length, distance and location are all calculated and stored. The format of the data stream is shown in Figure 3.6. The rules and formulas for such calculation is discussed in Section 2.5 and Section 3.3.1.

Eventually, two data sets are generated. One is the data set of pairs of speed and resistance force, $\{(v_i, R_i): i = 1, \dots, n\}$, which will be used later to determine the regression coefficients. Another data set is pairs of coordinates of location and speed, $\{(S_i, v_i): i = 1, \dots, n\}$, which will be used to draw the simulation curve of propulsion movement in the demo program.

3.6.2 Program Logic in Implementation

In Figure 3.7, the flow chart for the main process is presented describing the logic in the program developed. As shown in Figure 3.7, the simulated data stream

is prepared and generated (*A*) at first and then invokes other modules in sequence, such as Coefficients Calculation, Coast-down Simulation, Propulsion Simulation and Control/Release. By the end of this main process, the value of release speed and location are sent to control system to conduct the force release (*I*).

When calling the Coefficients Calculation Module (*B*), the generated data stream is transferred and shared. The Coefficients Calculation Module uses the data set of $\{(v_i, R_i): i = 1, \dots, n\}$ and calculate the regression coefficients b_0, b_1, b_2 for the fitted resistance function (Eq. (3.4)) accordingly. As a result, the regression function of resistance (Eq. (3.4)) is determined after exiting the Coefficients Calculation Module (*C*).

After that, the main process calls on the Coast-down Simulation Module (*D*). When calling Coast-down Simulation Module, the coefficients of regression function of resistance b_0, b_1, b_2 are transferred and shared. The Coast-down Simulation Module uses the regression coefficients b_0, b_1, b_2 in its calculation of behavior details for the whole coast-down movement. As a result, a data set of coordinates of location and speed, $SET_{CD}: \{(S_i, v_i): i = 1, \dots, n\}$, is obtained (*E*).

Then the Propulsion Simulation Module is called (*F*) by main process with the coefficients of regression function of resistance b_0, b_1, b_2 transferred and shared. The Propulsion Simulation Module uses the regression coefficients in its calculation of behavior details for the last period of propulsion phase which is defined and discussed in Section 3.3.3. As a result, a data set of coordinates of location and speed, $SET_{LP}: \{(S_i, v_i): i = 1, \dots, n\}$, is obtained describing the behavior details for the last period of propulsion (*G*).

The Control/Release Module is called at last (*H*) by main process. When calling the Control/Release Module, the behavior data set of coast-down $SET_{CD}: \{(S_i, v_i): i = 1, \dots, n\}$ and data set of the last period of propulsion $SET_{LP}: \{(S_i, v_i): i =$

$1, \dots, n$ are transferred and shared. The Control/Release Module applies simple linear regression analysis model to SET_{CD} and SET_{LP} , calculate the fitted functions separately for both data sets, and find the intersection of those two linear functions. The coordinates of the intersection (S_r, v_r) are the release location S_r and release speed v_r , which is the solution of this research work. The last step in this main process is to send information of the release location and speed to the control system (I).

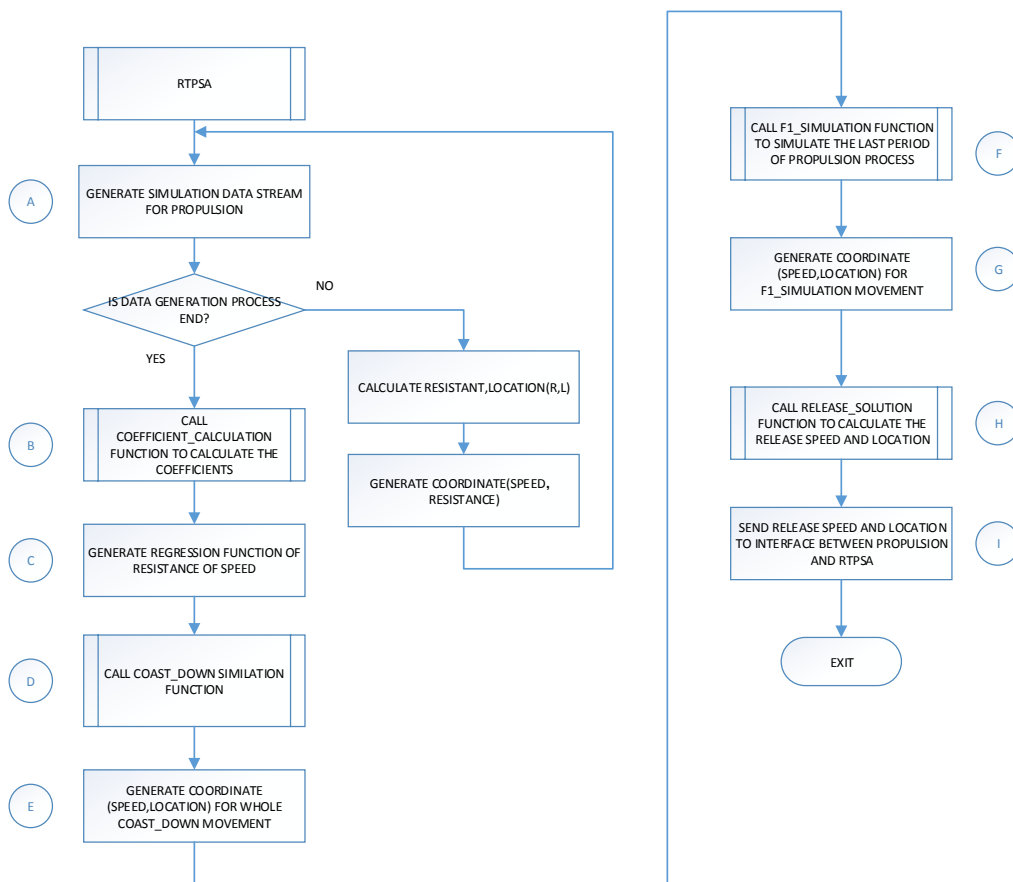


Figure 3.7 Main Process of *RTPSA*.

3.7 Coefficients Calculation Module

The Coefficients Calculation Module is called by main process of *RTPSA*. When it is called, the simulated data stream is also transferred from the main process. The major work in Coefficients Calculation Module is to apply multiple linear

regression model and least squares method to the data set of $\{(v_i, R_i): i = 1, \dots, n\}$ to calculate the regression coefficients b_0, b_1, b_2 for its fitted function of resistance. In such calculation, the correlation coefficient R^2 is also calculated accordingly.

3.7.1 How to Calculate the Regression Coefficients

As shown in Figure 3.3, there are many different sensors installed in the real collision test environment measuring performance details. The calculation of regression coefficients should be done in a real-time manner based on the real-time collected data. The implementation and integration of sensors, *PLC* controllers and computers is a complex project. In this research work, as discussed in Section 3.6, a simulated data stream is generated which provides the performance details of propulsion in a similar way. When applying this software in commissioning, data interface needs to be adjusted to adapt to the real interface to sensors.

Coefficients Calculation Module is a real-time reading and calculating process. Once the propulsion phase starts, this module will read performance data from the propulsion system. In this research, the data is read from the simulated data stream automatically generated by computer. While the propulsion mechanism keeps working, those data feed into this module continuously.

Since the process in this module is a real-time and interactive processing in nature, how to control this process becomes critical. There are two key points in the coefficients calculation module, start point and end point.

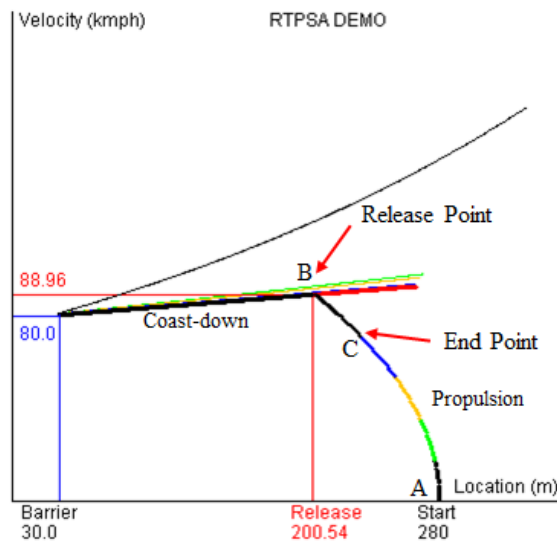


Figure 3.8 End Point for Data Collection.

The original start point of the propulsion phase is the point when the speed of test vehicle equals to zero. But the data of those where speed is less than 36kph (10m/s), are found out after several experiments, to be useless to the accuracy of coefficients. They negatively impact the accuracy of the coefficients. That is why a new start point for the calculation of coefficients is defined in the implementation, instead of simply using the original start point. As shown in Figure 3.8, curve *AB* is the propulsion movement. Position *A* is the original start point where speed equals to zero. Position *B* is the end of propulsion phase where the propulsion force is released. An end point for the data collection (position *C*) is also needed before the calculation because for every collision test, the release speed might be different due to different desired crash speed. The end point for the data collection must be defined at somewhere before the release point and at that point, the *RTPSA* software triggers the algorithm to calculate the release location and release speed. Enough leading time is also needed for the overall control system to fulfill the force release action. In the current version of *RTPSA* software, the predefined crash speed is used as the reference speed and set as the end point *C*. Because the release speed is always greater than the crash speed, it is safe to use the crash speed as the end point for the data collection.

3.7.2 Program Logic in Implementation

In Figure 3.9, the flow chart for Coefficients Calculation Module is presented describing the logic in the developed program. As shown in Figure 3.9, the simulated data stream is transferred and shared when Coefficients Calculation Module is invoked (A). The general form of multiple linear regression is built at (B). The start point and end point for coefficients calculation are set at (C). A stop flag is set as the end point for the data set used in Coefficients Calculation Module which indicates the end of data collection for coefficients calculation. This stop flag is also the point where the simulation for the last period of propulsion get started.

The original data stream is re-organized into different data distribution arrays at (D). Each of the arrays is a subset of the whole data set. For each subset, least squares method is applied to calculate the regression coefficients. The detail of calculation following least squares method is done at (E). As discussed in Section 2.1.4, following equations are used to calculate the regression coefficients.

$$\begin{aligned} \sum_{i=1}^n y_i &= n b_0 + b_1 \sum_{i=1}^n x_{1i} + b_2 \sum_{i=1}^n x_{2i} \\ \sum_{i=1}^n x_{1i} y_i &= b_0 \sum_{i=1}^n x_{1i} + b_1 \sum_{i=1}^n x_{1i}^2 + b_2 \sum_{i=1}^n x_{1i} x_{2i} \\ \sum_{i=1}^n x_{2i} y_i &= b_0 \sum_{i=1}^n x_{2i} + b_1 \sum_{i=1}^n x_{1i} x_{2i} + b_2 \sum_{i=1}^n x_{2i}^2 \end{aligned}$$

where b_0, b_1, b_2 are the regression coefficients, n is the total number of sample data, $x_i, i = 1, 2, \dots, n$ are the values of speed, and $y_i, i = 1, 2, \dots, n$ are the values of resistance force. After each of those subsets is applied to above equations, three equations are determined accordingly (F). By solving the system of equations at (G), the regression coefficients b_0, b_1, b_2 are obtained at (H). At last, the fitted function for each subset is obtained at (I). The correlation coefficient R^2 is calculated accordingly as well. In the demo program of *RTPSA*, the regression coefficients and

expression for each of the fitted function are displayed together with R^2 . So users can see the different curves of each function and the improvement of R^2 .

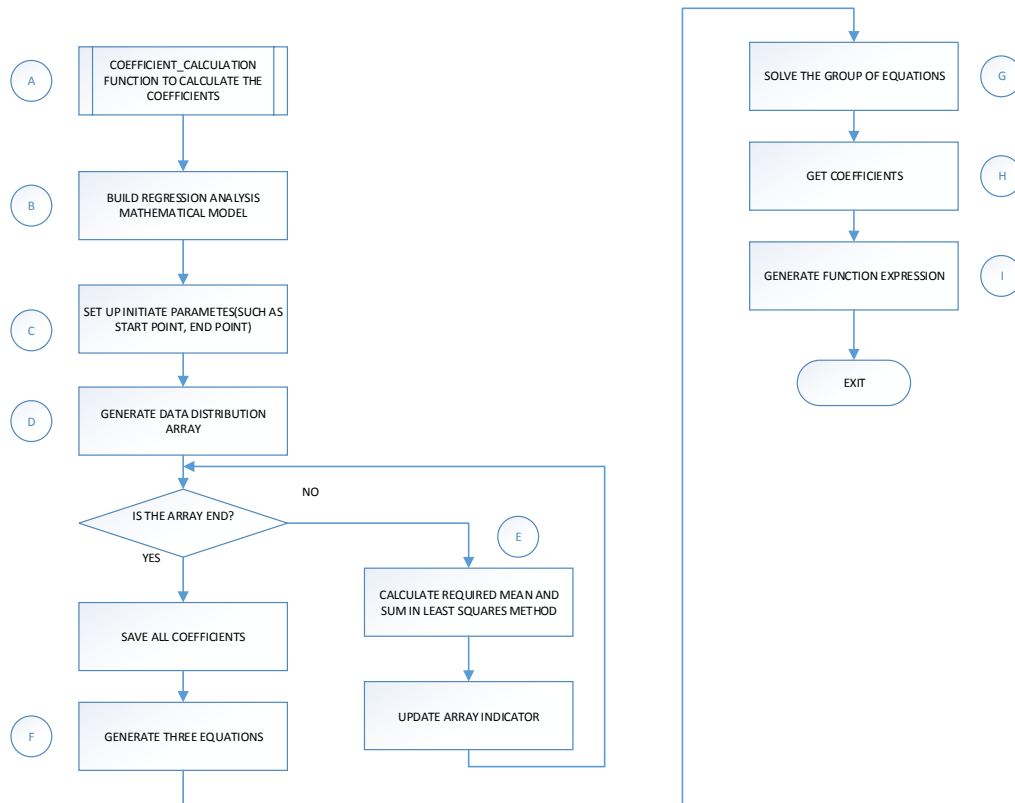


Figure 3.9 Program Logic of Coefficients Calculation.

3.8 Coast-Down Simulation Module

The Coast-Down Simulation Module is invoked by main process of *RTPSA*. When it is invoked, the regression coefficients for the function of resistance force is transferred and shared from the main process. The major task in Coast-Down Simulation Module is to use the resistance regression function to predict the behavior of resistance along the coast-down movement. With the determination of resistance force, all the other details of behavior of coast-down movement can be calculated and described, such as speed, location and acceleration.

3.8.1 How to Simulate the Coast-Down Behavior?

The start of coast-down movement is the position where the propulsion force is released and the end of coast-down movement is the collision between the test vehicle and barrier. The sole objective of this research work is to find the release speed and location quickly and precisely. So it is impossible to simulate the coast-down behavior from the start point where nothing of it is known. But everything about the end point is predefined and known, such as the crash speed and the location of barrier. Another observation is that, during coast-down movement, the propulsion force equals to zero. The resistance (including friction and aerodynamic force) is the only force existed which is a quadratic function of speed. In another word, the coast-down movement is a varying acceleration movement.

The solution of how to simulate the coast-down movement is introduced as follows. At first, the resistance force can be predicted by regression function which is already determined by the Coefficients Calculation Module. With the regression coefficients determined, regression function of resistance force can be used to predict the behavior of resistance force in coast-down movement. Secondly, since the crash speed and the location of barrier are predefined and known, a reverse analysis approach is implemented to calculate the detail behavior in this Coast-Down Simulation Module. The performance details are calculated from the end point towards the start point. The last obstacle is how to calculate the behavior details for a varying acceleration movement. The way of solving this problem is to divide the whole coast-down phase into many small time-equal intervals. Because such intervals can be made small enough, so each of them can be treated as a constant acceleration motion. Then formulas for constant acceleration motion can be applied to each interval and all the details of the behavior can be calculated.

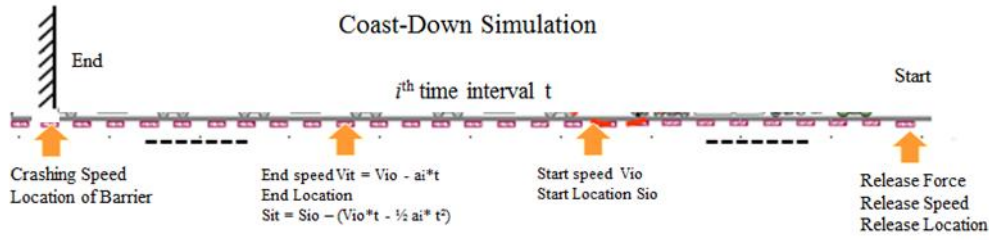


Figure 3.10 Time-Equal Intervals for Coast-Down Simulation.

As shown in Figure 3.10, the start of coast-down phase, on the right hand side, is the position where the propulsion force is released. Release speed and release location are two key parameters describing the start position. The end of the coast-down phase, on the left hand side, is the time when test vehicle hits the barrier. The crash speed and the location of barrier are known and are two key parameters describing the end position. An acceleration a_i is assigned for the i^{th} time interval. Formulas for constant acceleration motion are applied to calculate the speed and location for the i^{th} time interval. The major formulas are

$$v_{it} = v_{i0} - a_i * t,$$

$$S_{it} = S_{i0} - (v_{i0} * t - \frac{1}{2} a_i * t^2),$$

where v_{i0} is the start speed of interval i , v_{it} is the end speed of interval i , t is the time length for each of the interval, S_{i0} is the start location of interval i , and S_{it} is the end location of interval i . In this way, after the calculation for all the intervals in sequence, the behavior details of the whole coast-down movement is obtained. One of the results of the coast-down simulation is a data set of pairs of location and speed

$$\{(S_i, v_i) : i = 1, \dots, n\},$$

where n is the total number of time-equal intervals.

In Figure 3.11, it displays several simulated coast-down curves A, B, C, D and E according to different size of data set used. The data sets for A, B, C and D are all from the original beginning, where the speed equals to zero. Curve A is the first one in black using data records from row 9 to row 32 of Figure 3.6. As the propulsion mechanism keeps working, curve B is the second one in green using

data records from row 9 to row 57 of Figure 3.6. Curve *C* is the third one in yellow using data records from row 9 to row 82 of Figure 3.6. Curve *D* is the fourth one in blue using data records from row 9 to row 106 of Figure 3.6. Curve *E* is the best one in red using data records from row 50 to row 125 of Figure 3.6, in which the speed is greater than 36kph (10m/s) and less than the crash speed (in Figure 3.11, the crash speed is 90kph or 25m/s). On the right side of Figure 3.11, the values of correlation coefficient R^2 are also displayed accordingly. The improvement of R^2 can be easily noticed when the size of the data set becomes larger and larger, from *A* to *D*.

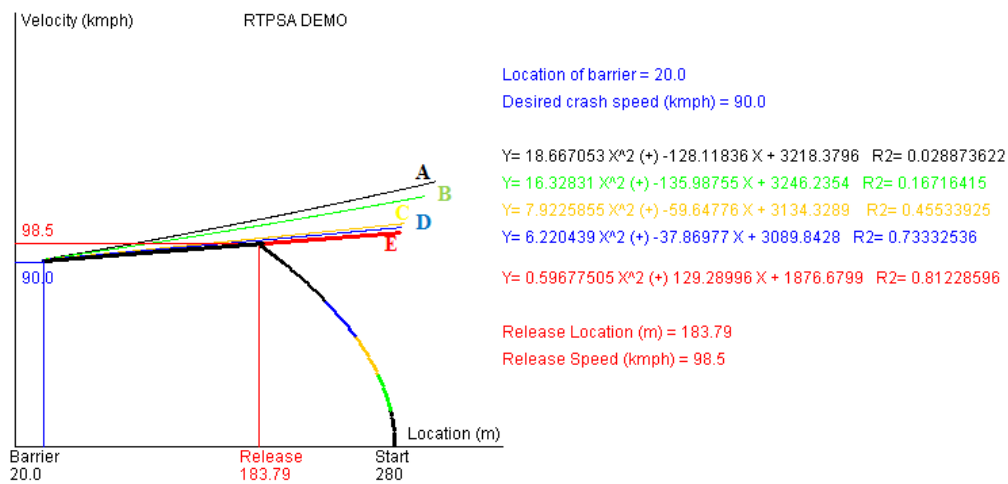


Figure 3.11 Coast-Down Curves.

3.8.2 Program Logic in Implementation

In Figure 3.12, the flow chart for Coast-Down Simulation Module is presented describing the logic in the developed program. As shown in Figure 3.12, the Coast-Down Simulation Module is invoked by main process of *RTPSA*. When it is invoked, regression coefficients for the function of resistance force is transferred and shared from the main process (*A*). The general parameters used in coast-down simulation are defined at (*B*). The whole coast-down movement is divided into many small time-equal intervals (*C*). As discussed earlier, the coast-down behavior is

subject to the crash speed, location of barrier and mass of the vehicle. Those values are defined at (D). The calculation of the behavior details for each interval is processed at (E) and (F) following the constant acceleration motion formulas. At the end of Coast-Down Simulation Module, a data set of coordinates of location and speed is obtained which describes the coast-down behavior in detail (G).

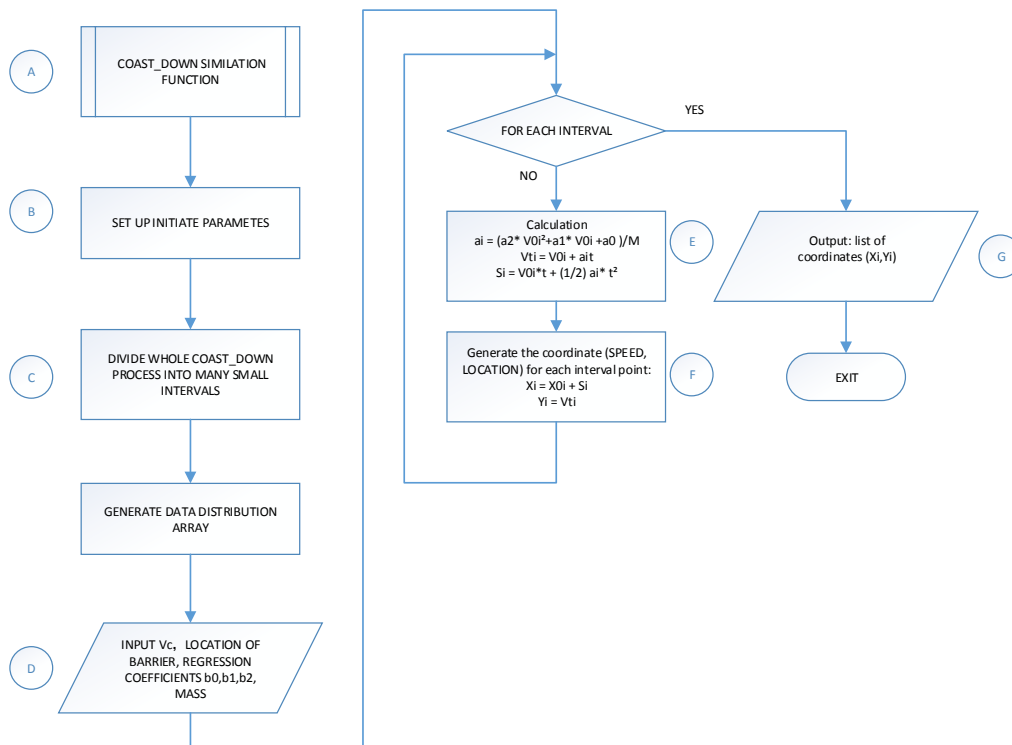


Figure 3.12 Program Logic of Coast-down Simulation.

3.9 Propulsion Simulation Module

The Propulsion Simulation Module is invoked by main process of *RTPSA*. When it is invoked, the regression coefficients for the function of resistance force is transferred and shared from the main process. The major task in Propulsion Simulation Module is to use the resistance regression function to predict the behavior of resistance for the last period of propulsion movement. With the determination of resistance force, all the other details of behavior can be calculated and described, such as speed, location and acceleration.

3.9.1 How to Simulate the Last Period of Propulsion Phase?

In *RTPSA*, the whole propulsion phase is divided into two parts, the regular propulsion phase from *A* to *C* as indicated in Figure 3.13 and a simulation phase from *C* to *B*. The Propulsion Simulation Module deals with the last period of propulsion phase from *C* to *B*, which is started at the time when the data collection process for the calculation of coefficients ended. The position of *C* is varying and is subject to how long the leading time the overall control system needs to actually release the vehicle at desired location and speed.

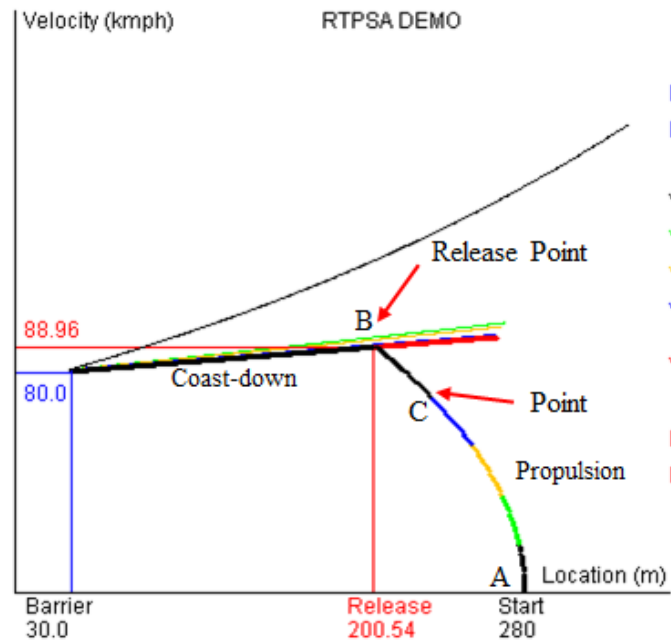


Figure 3.13 Start Point for Propulsion Simulation.

Position *C* is also the time when the algorithm is triggered to calculate the release speed and location. The leading time that the overall control system has equals to the time between position *C* and position *B* along the propulsion movement, denoted as $Time_{BC}$, minus the total calculation time of the *RTPSA* software, which is denoted as $Time_{RTPSA}$.

$$Leading\ Time = Time_{BC} - Time_{RTPSA}$$

Let's assume that the crash speed is set to 80kph, location of barrier is located at 30m. Then after running *RTPSA*, the release speed is obtained which is 88.96kph, and it takes 0.006s for the running. The information is shown in Table 3.1.

<i>Speed (kph)</i>	<i>Propulsion Time (s)</i>	<i>Calculation Time (s)</i>
80.00	6.019	0.006
88.96	6.623	
Leading Time	0.598	

Table 3.1 Leading Time Calculation.

In this case, if the propulsion simulation starts at the time of 6.019s, with the speed of vehicle is 80kph, then by the time of 6.025s (6.019s+0.006s), *RTPSA* finishes the calculation and the overall control system has been notified with the value of release speed which is 88.96kph. The propulsion system will keep working anyway until the propulsion force get released. From the second column of Table 3.1, it shows that at the time of 6.623s, the test vehicle will be propelled and reaches the speed of 88.96kph (the release speed), which means at that time, the test vehicle will be released. By comparing 6.025s and 6.623s, the time of 0.598s is left for the overall control system to prepare and conduct the force release. The time of 0.598s is called the leading time. Regularly the leading time needs to be one to two second. 0.598s is too short for overall control system to fulfill the task. To improve this, the propulsion simulation process needs to be started earlier (the position of *C*), so that the leading time will be left longer.

The equation of motion is

$$F - R = M * A, \quad (3.6) \text{ same as (2.15)}$$

where *F* is the propulsion force, *R* is the resistance force, *M* is the mass of vehicle, and *A* is the vehicle acceleration. With the regression coefficients determined by the Coefficients Calculation Module, the regression function of resistance will be used to predict the value of resistance force *R*. Mass *M* is constant

and known. The propulsion force F can be estimated by using equation (2.19) and equation (2.20). Therefore, the acceleration A is the only unknown variable and can be calculated by equation (3.6).

The propulsion movement is also a varying acceleration motion. As implemented in Coast-Down Simulation Module, similarly, the last period of propulsion movement is also divided into many small time-equal intervals, and each of the small intervals is treated as constant acceleration movement. Formulas for constant acceleration motion will be applied to calculate the speed and location for every time interval along the movement. Eventually a data set of coordinates of location and speed is generated which will be used to calculate the release speed and location later on.

3.9.2 Program Logic in Implementation

In Figure 3.14, the flow chart for Propulsion Simulation Module is presented describing the logic in the developed program. As shown in Figure 3.14, the Propulsion Simulation Module is invoked by main process of *RTPSA*. When it is invoked, regression coefficients for the function of resistance force is transferred and shared from the main process (*A*), as well as the information about the start position. The general parameters used in this simulation are defined at (*B*). The last period of propulsion phase is divided into many small time-equal intervals (*C*). The mass of the test vehicle is defined at (*D*). The propulsion force is calculated following the equation (2.19) and equation (2.20) in Section 2.5 (*E*). The calculation of the behavior details for each interval is processed at (*F*) and (*G*) following the constant acceleration motion formulas. At the end of Propulsion Simulation Module, a data set of coordinates of location and speed is obtained which describes the behavior of last period of propulsion movement in detail (*H*).

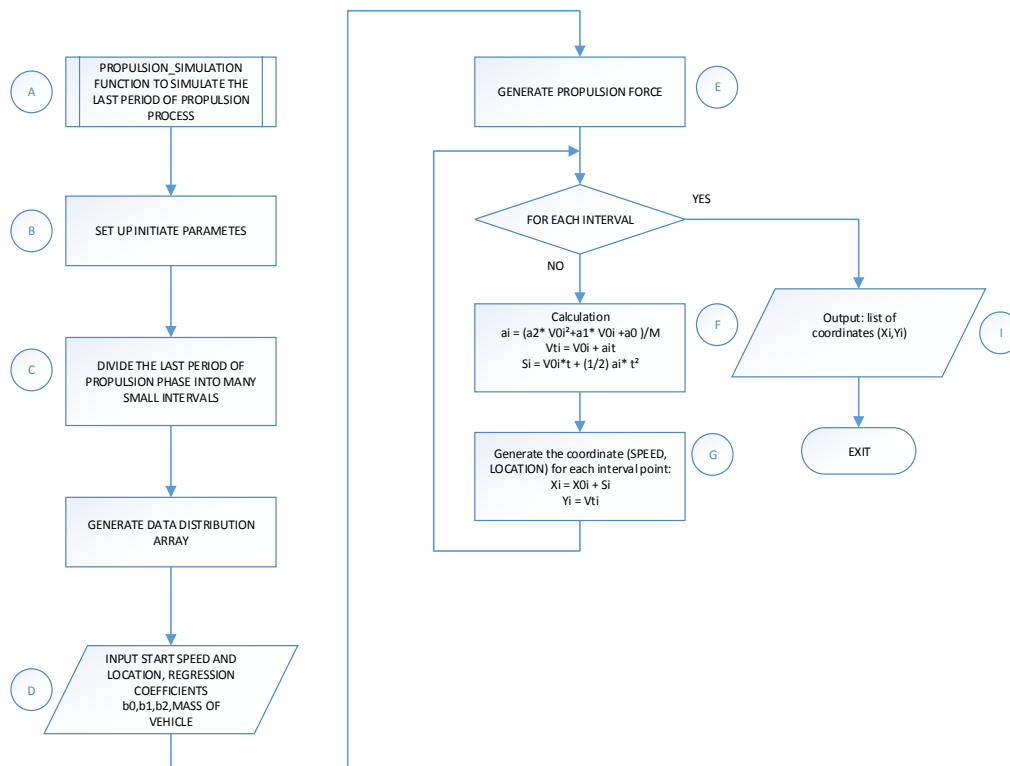


Figure 3.14 Program Logic of Propulsion Simulation.

3.10 Control/Release Module

The Control/Release Module is invoked by main process of *RTPSA*. When it is invoked, both the data set of behavior for coast-down and data set of behavior for the last period of propulsion are transferred and shared from the main process. Those data sets are coordinates of location and speed for both coast-down and the last period of propulsion. The major task in Control/Release Module is to find the release location and speed.

3.10.1 How to Find the Release Speed and Location?

As discussed in Section 3.5, in order to find the release location and speed, both the behavior of coast-down and behavior of the last period of propulsion need to be known and the intersection of coast-down curve and the curve of the last period of propulsion is the solution. The coordinates of this intersection are the release location

and speed. Therefore, based on the data sets shared by main process, simple linear regression analysis model is applied and regression coefficients are calculated. Eventually two fitted functions are obtained. By solving the system of these two fitted functions, the intersection can be obtained.

The reason why applying simple linear regression model to each of above data sets is based on following observations. At first, by observing a great amount of running results, the speed difference between the release speed and the crash speed falls into a very small range. Second, the last period of propulsion is a very short phase. Since the required leading time is 1-2 second, which means the total time for the last period of propulsion is 1-2 second as well. Furthermore, using simple linear regression model can save the running time of the algorithm.

3.10.2 Program Logic in Implementation

In Figure 3.15, the flow chart for the Control/Release Module is presented describing the logic in the developed program. As shown in Figure 3.15, the Control/Release Module is invoked by main process of *RTPSA*. When it is invoked, both the data set of behavior for coast-down and data set of behavior for the last period of propulsion are transferred and shared from the main process (*A*). The initial parameters used in this module are defined at (*B*) and (*C*). From (*D*) to (*E*), least squares method is applied to calculate the regression coefficients for the coast-down movement. And the fitted function for coast-down behavior is obtained at (*E*). From (*F*) to (*G*), least squares method is applied to calculate the regression coefficients for the last period of propulsion. And the fitted function for the last period of propulsion is obtained at (*G*). The software then just solves the two equations (*H*) and finds the coordinate of the intersection which is the solution of release speed and location (*I*). At the end of Control/Release Module, the value of release speed and location are returned to the control system (*J*).

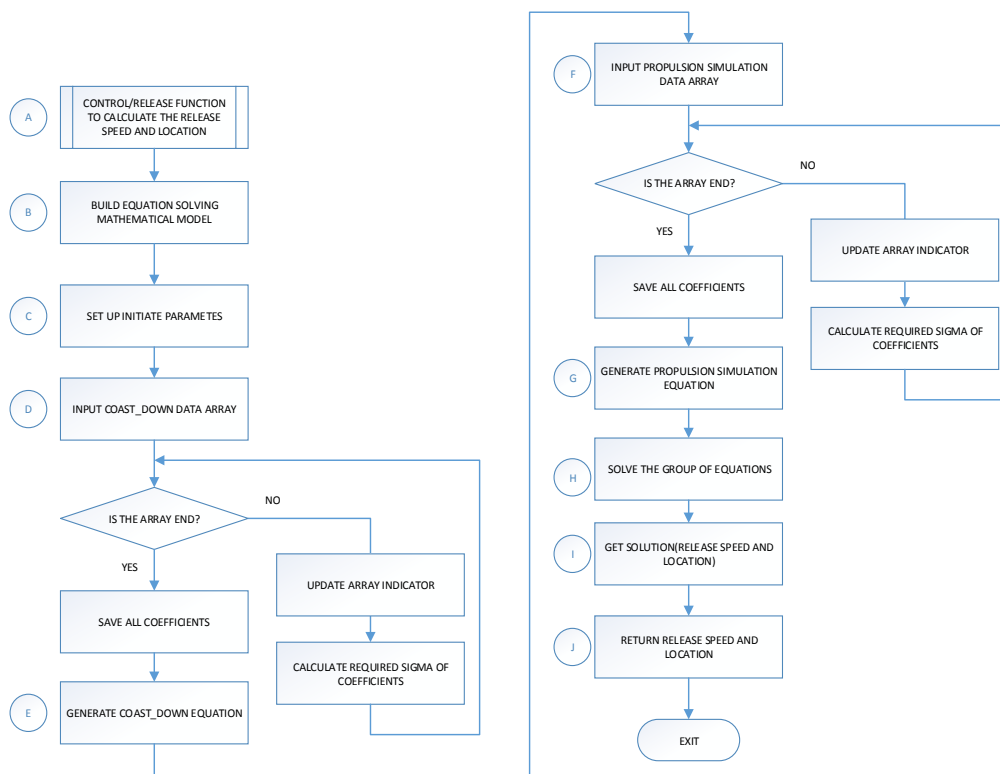


Figure 3.15 Program Logic of Control and Release Processing.

3.11 Summary

In Chapter 3, the proposed solution is introduced which explains in detail how to solve the major issues in the real train collision test. Furthermore, an application software is well designed and developed to fulfill the task. The main process and four major modules in *RTPSA* are explained after the introduction of design principles, system architecture and software structure. The software developed is part of the overall control system (Facility Control System). Section 3.7.1 explains in detail how to calculate the regression coefficients and the result is demonstrated in the demo video which shows that the proposed solution is designed for real-time data collection and calculation. It also shows that *RTPSA* can use the regression function of resistance force to predict the behavior of resistance in both coast-down movement and the last period of propulsion phase. Section 3.8.1 explains in detail how the coast-down simulation is done. The method of dealing with the varying acceleration issue is proved to be a practical method in solving similar problems.

In Chapter 4, running results for different scenarios will be presented and explained.

Chapter 4 Running Results

In this chapter, several outputs of the software developed will be presented, such as a demo program, a standalone version of software, the running time and running results.

4.1 The Demonstration Program

A demo program is developed as part of this research work. In Figure 4.1, which is a screen shot of the demo, the movement of the test vehicle is simulated in a location versus velocity plane. The X -axis is location and Y -axis is velocity. The test vehicle starts at the very end of the right side, moving towards the barrier which is located at the other end of the plane. The software developed takes the data from propulsion phase in a real-time manner, calculates the regression coefficients first and then displays multiple simulation curves for coast-down movement accordingly. Different colors are used to represent different time frames of propulsion phase, display different regression functions and coast-down curves based on the data set of each time period. For example, the first black curve is based on the data set from records row 9 to row 32 of the sample data sheet in Figure 3.6. The regression function is displayed in black on top right. The second green curve is based on the data set from records row 9 to row 57 in Figure 3.6 and its regression function is displayed in green color under the first function. The third yellow curve is based on the data set from records row 9 to row 82 and the fourth blue curve is based on the data set from records row 9 to row 106. The last red curve shown is based on the data set records where the speed is greater than 36kph (10m/s) and less than the designated crash speed.

By observing the multiple coast-down curves, the more the data used, the much

closer the curve approaches the target. And the correlation coefficient R^2 is improved as well when the data set becomes larger. The coefficients are also kept changing along with the propulsion movement. These observation explains again that why the software is designed to run in a real-time manner and can provide a better solution comparing to others.

In Figure 4.1, the location of barrier is 100m away from the left. The designated crash speed is set to 80kph. The test vehicle start at the location of 280m on the right side of the plane, moving towards the barrier. The propulsion force is released at the location of 210.95m. The release speed is 86.23kph. The correlation coefficient R^2 is 0.7379.

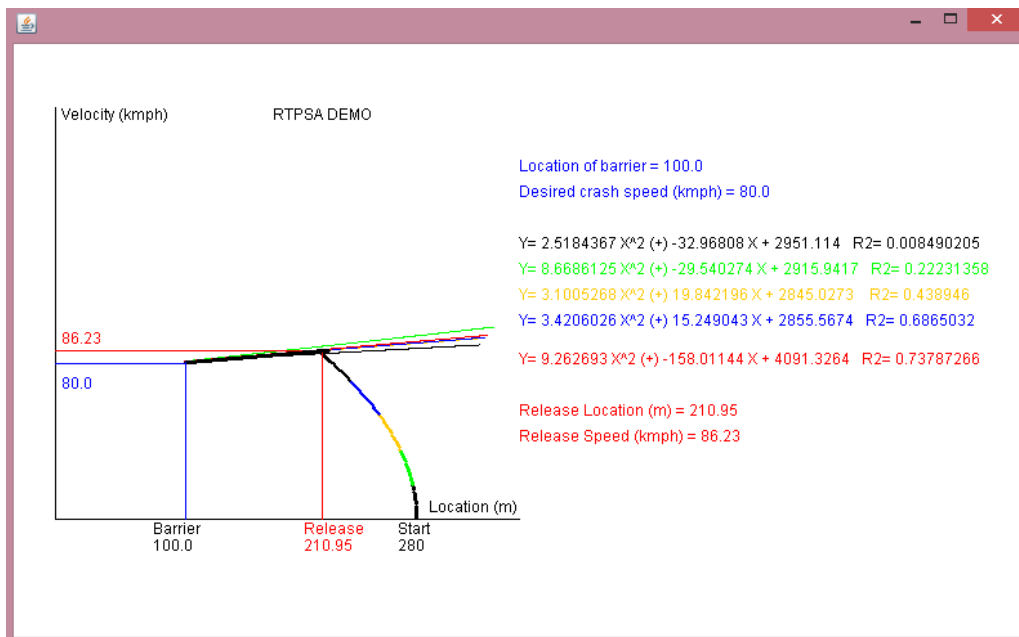


Figure 4.1 Screen Shot of Demonstration Program.

4.2 Standalone Version of Software

One of the result of this research work is a standalone version of software which can be used by users to manipulate different parameters and see different result accordingly. Figure 4.2 is the input screen of this standalone version.

Input Key Parameters For Your Test

Crashing Speed (kph):

Barrier_location (m):

Mass (kg):

$R_a = a_0 * M * G + a_1 * M * G * V + a_2 * V^2$

Ra_a0: Ra_a1:

Ra_a2: Am_ErrorRatio:

$F_a = a + b * V + c * V^2$

Fa_a: Fa_b:

Fa_c: Fa_ErrorRatio:

ACCEPT DEFAULT

Figure 4.2 Input Screen of the Standalone Version.

As shown in Figure 4.2, users can type in different parameters and then run the program. For example, on the top of Figure 4.2, users can input the crash speed, the location of barrier and the mass of the test vehicle. In the middle of this input parameters screen, users can type in parameters to determine the value of resistance force R_a . The error ratio for acceleration A_m can also be determined here. Then users can input the parameters for propulsion force F_a . The formula for calculation of F_a is shown on top of the input fields. And the error ratio for F_a can be determined at last which is used to calculate the noise of F_a . Those are the major parameters leading to different running results. With this tools, users can manipulate the parameters and generate expected results accordingly.

Figure 4.3 is the screen shot of output result of the standalone version. As shown in Figure 4.3, those parameters together with the results are all displayed for

recording.

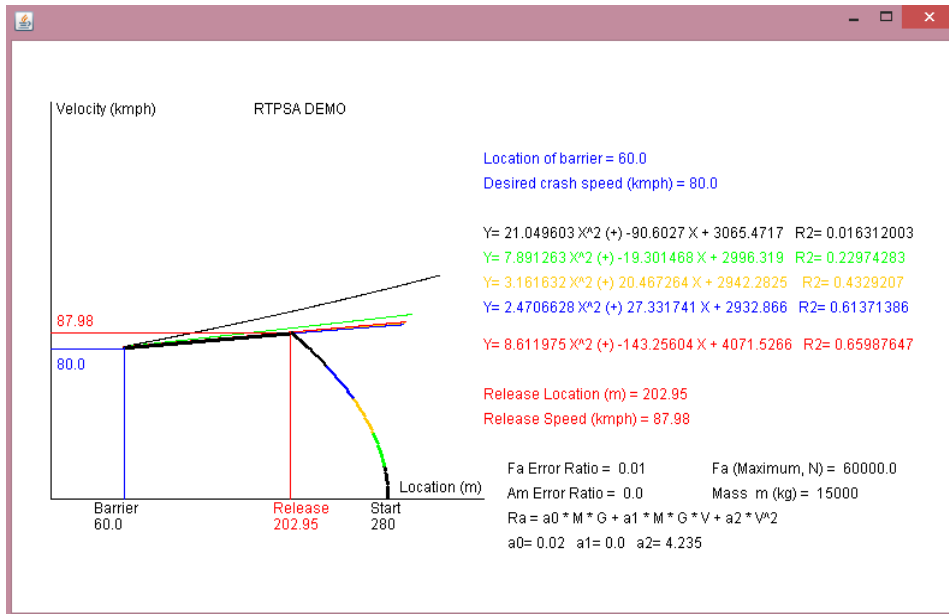


Figure 4.3 Output Screen of the Standalone Version.

4.3 Running Time

The running time that *RTPSA* software takes can be seen in Figure 4.4 which is a screen shot of *JAVA* console. It shows that the program is started at 2015/5/30/4:2:15.312 and ended at 2015/5/30/4:2:15.315 which means that the running time is 3 milli second.

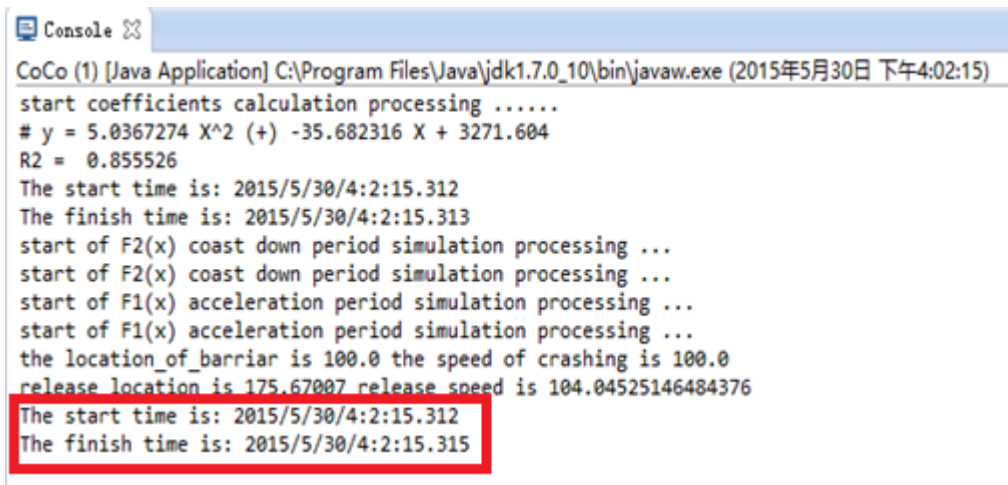


Figure 4.4 Screen Shot of Console - Running Time.

4.4 Run with Different Set of Data

The following test results are based on the generated simulation data set. When generating the simulation data set, there is an error ratio for the calculation of measured force F_m from actual force F_a . Because the initial signal for F_a is analogue, there is noise on the signal which is a constant. So an error ratio is used to describe this noise.

In running result Table 4.1, the error ratio of F_a is set to 1%, the crash speed is set to 80kph and the barrier is located at 30m. As shown in Figure 4.1, different sets of data are taken to calculate the coefficients accordingly, as well as the correlation coefficient R^2 . The information of coefficients and function expression are put into this table and the improvement of R^2 can be seen from the last column on the right. For example, when the data set from item 1 to item 24 are used, the value of R^2 is 0.0554 which means the function f is too bad to fit the data distribution. When the data set from item 1 to item 49 are used, an improved value of R^2 is obtained as 0.1632. When the data set from item 1 to item 74 are used, a further improved R^2 is obtained as 0.3797. And at last when the data set from item 1 to item 98 are used, a much better value of R^2 is obtained as 0.7552. The conclusion is, the data that is collected closer to the release point leads to better regression result.

Running Result 1						
Data Set Range	Barrier/Crashing Speed	Coefficient a_0	Coefficient a_1	Coefficient a_2	$f = a_0 + a_1v + a_2v^2$	$R^2 = 1 - SS_{res}/SS_{tot}$
1 - 24	30m / 80kph	3202	-219	44	$f = 3202 - 219v + 44v^2$	0.0554
1 - 49	30m / 80kph	3150	-118	15	$f = 3150 - 118v + 15v^2$	0.1632
1 - 74	30m / 80kph	3036	-37	6	$f = 3036 - 37v + 6v^2$	0.3797
1 - 98	30m / 80kph	2992	-18	5	$f = 2992 - 18v + 5v^2$	0.7552

Table 4.1 Run with Different Set of Data (Error Ratio of $F_a = 1\%$).

4.5 Run with Different Values of Crash Speed

In running result Table 4.2, the barrier is set at 30m and different values of crash speed are used in the experiment. The data set used for this test is from 36kph (10m/s) to the crash speed. The release location, release speed, distance of coast-down and correlation coefficient R^2 are calculated and displayed accordingly. As shown in Table 4.2, when the crash speed is 60kph a very poor R^2 of 0.2584 is obtained. As the speed increases, a better R^2 is obtained gradually. At last, when the crash speed is set to 150kph, the highest R^2 of 0.9724 is obtained.

Running Result 2					
Barrier Location (m)	Crashing Speed (kph)	Release Location (m)	Release Speed (kph)	Distance of Coast Down (m)	$R^2=1-SS_{res}/SS_{tot}$
30	60	229.72	72.3	199.72	0.2584
30	70	216.46	80.98	186.46	0.513
30	80	202.56	89.26	172.56	0.5997
30	90	184.85	98.17	154.85	0.7406
30	100	165.26	107.34	135.26	0.8736
30	110	144.28	116.31	114.28	0.9016
30	120	122	125.07	92	0.9169
30	130	97.74	133.8	67.74	0.947
30	140	72.4	142.39	42.4	0.9624
30	150	45.91	150.8	15.91	0.9724

Table 4.2 Run with Different Values of Crash Speed (Error Ratio of $F_a = 1\%$).

By observing the result in running result Table 4.2, it can be seen that the developed algorithm generates a better R^2 when the crash speed becomes higher.

4.6 Run with Different Values of Barrier Location

In Table 4.3, the crash speed is set to 100kph and the barrier is located at different positions. The data set used for this test is from 36kph (10m/s) to the crash speed. All the calculation results are recorded down such as release location, release speed, distance of coast-down and the correlation coefficient R^2 .

Running Result 3					
Crashing Speed (kph)	Barrier Location (m)	Release Location (m)	Release Speed (kph)	Distance of Coast Down (m)	$R^2=1-SS_{res}/SS_{tot}$
100	0	160.8	108.74	160.8	0.8549
100	20	163.3	108.03	143.3	0.8715
100	40	166.69	106.94	126.69	0.8472
100	60	170.14	105.83	110.14	0.8463
100	80	172.36	105.15	92.36	0.8567
100	100	175.36	104.07	75.36	0.8541
R^2 Average					0.85516667

Table 4.3 Run with Different Values of Barrier Location (Error Ratio of $F_a = 1\%$).

By observing the result in Table 4.3, it can be seen that, the coast-down distance varies from minimum 75.36m to maximum 160.8m, while the value of R^2 varies from 0.8463 to 0.8715. The R^2 is always close to its average 0.8551 no matter how the position of barrier is moved, which means the coast-down distance, in the fifth column of Table 4.3, does not influence the value of R^2 very much. In another word, the R^2 is much more sensitive and subjected to the speed.

4.7 R^2 with Different Speed

Furthermore, in following experiments, different values of crash speed are set first and then the test gets repeated using different values of barrier location. The purpose is to see the difference of R^2 . The results are put into following tables.

Error ratio of $F_a = 1\%$

Running Result 4					
Crashing Speed (kph)	Barrier Location (m)	Release Location (m)	Release Speed (kph)	Distance of Coast Down (m)	$R^2=1-SS_{res}/SS_{tot}$
60	0	229.04	72.25	229.04	0.3677
60	20	229.5	72.31	209.5	0.2943
60	40	232.58	70.41	192.58	0.3135
60	60	233.72	70.35	173.72	0.3566
60	80	235.64	69.77	155.64	0.3089
60	100	238.11	68.64	138.11	0.3307
R^2 Average					0.328616667

Table 4.4 R^2 for Speed of 60kph.

Error ratio of $F_a = 1\%$

Running Result 5					
Crashing Speed (kph)	Barrier Location (m)	Release Location (m)	Release Speed (kph)	Distance of Coast Down (m)	$R^2=1-SS_{res}/SS_{tot}$
80	0	199.31	90.38	199.31	0.6872
80	20	201.09	89.84	181.09	0.6889
80	40	204.1	88.65	164.1	0.6217
80	60	206.08	87.95	146.08	0.7273
80	80	208.71	86.84	128.71	0.7362
80	100	211.04	86.23	111.04	0.6401
R^2 Average					0.683566667

Table 4.5 R^2 for Speed of 80kph.

Error ratio of $F_a = 1\%$

Running Result 6					
Crashing Speed (kph)	Barrier Location (m)	Release Location (m)	Release Speed (kph)	Distance of Coast Down (m)	$R^2=1-SS_{res}/SS_{tot}$
120	0	116.63	126.53	116.63	0.9449
120	20	120.04	125.56	100.04	0.9318
120	40	123.55	124.62	83.55	0.9248
120	60	126.74	123.75	66.74	0.9406
120	80	130.38	122.77	50.38	0.9339
120	100	134.08	121.86	34.08	0.9347
R ² Average					0.935116667

Table 4.6 R^2 for Speed of 120kph.

Error ratio of $F_a = 1\%$

Running Result 7					
Crashing Speed (kph)	Barrier Location (m)	Release Location (m)	Release Speed (kph)	Distance of Coast Down (m)	$R^2=1-SS_{res}/SS_{tot}$
150	0	39.07	152.23	39.07	0.9698
150	20	43.78	151.29	23.78	0.9739
150	40	47.85	150.32	7.85	0.9736
150	60	52.53	149.35	-7.47	0.9707
150	80	56.36	148.43	-23.64	0.9709
150	100	60.81	147.49	-39.19	0.9673
R ² Average					0.971033333

Table 4.7 R^2 for Speed of 150kph.

By observing above result tables, it can be seen that the R^2 is improved from an average of 0.3286 when speed is 60kph, to a much better average 0.6836 when speed is 80kph, and then it is increased to 0.9351 when speed is 120kph. Finally, an average value of 0.9710 is obtained when speed is 150kph.

4.8 R^2 with Different Error Ratio

All the above test results, from Table 4.1 to Table 4.7, are based on 1% error ratio of F_a . As explained earlier, this error ratio is used to describe the noise of initial analogue signal of F_a . For example, if the maximum value of F_a is 60,000N and the error ratio is 1%, then the noise is 600N. In following experiments, different error ratios of F_a will be used.

Running Result 8							
Maximum F_a (N)	Error Ratio (%)	Noise (N)	Crashing Speed(kph)/Barrier Location (m)	Release Location (m)	Release Speed (kph)	Distance of Coast Down (m)	$R^2=1-SS_{res}/SS_{tot}$
60,000	1	600	60/30	229.72	72.3	199.72	0.2584
60,000	0.75	450	60/30	230.95	71.61	200.95	0.3836
60,000	0.5	300	60/30	230.83	71.7	200.83	0.6652
60,000	0.25	150	60/30	230.83	71.62	200.83	0.8343
60,000	0.1	60	60/30	231.08	71.44	201.08	0.9782
Average				230.682	71.734	200.682	

Table 4.8 R^2 for Different Error Ratios.

As shown in Table 4.8, the crash speed and barrier location are set to 60kph and 30m. Different error ratios of F_a , from 0.1% to 1%, are used for the experiments. By observing the result in Table 4.8, the values of release location, release speed and coast down distance are kept relatively stable around their averages accordingly. But the value of R^2 varies from 0.2584 to 0.9782. It can be concluded that the error ratio of F_a significantly influences the R^2 .

4.9 Coast-down Distance with Different Propulsion Forces

As shown in red in Table 4.6 and 4.7, when coast-down distance is less than 50m, it is too close between the release point and the crash barrier that makes it dangerous during the real test. It is suggested that the barrier has to be located at the very end of the track, location = 0 or 20m, when the crash speed is more than or equal to 120kph, in order to have a relatively safe coast-down distance. Another way to solve this issue is to apply higher propulsion torque in propulsion mechanism to shorten the propulsion distance so that a longer distance could be left for the coast-down movement. As shown in Table 4.9, different parameters of F_a are used, and a longer coast-down distance is obtained.

Running Result 9										
F_{a_a}	F_{a_b}	F_{a_c}	Maximum F_a (N)	Error Ratio (%)	Noise (N)	Crashing Speed(kph)/Barrier Location(m)	Release Location (m)	Release Speed (kph)	Distance of Coast Down (m)	$R^2=1-SS_{res}/SS_{tot}$
60,000	0	0	60,000	1	600	150/0	35.71	152.02	35.71	0.9721
60,000	0	0	60,000	1	600	150/20	39.88	151.06	19.88	0.9708
70,000	0	0	70,000	1	700	150/0	69.14	154.13	69.14	0.9598
70,000	0	0	70,000	1	700	150/20	72.98	153.11	52.98	0.9637
80,000	0	0	80,000	1	800	150/0	95.35	155.74	95.35	0.9457
80,000	0	0	80,000	1	800	150/20	98.83	154.66	78.83	0.9608

Table 4.9 Coast-down Distance with Different Propulsion Forces.

The demo of running results with different propulsion forces, when location of barrier is zero, are shown as follows.

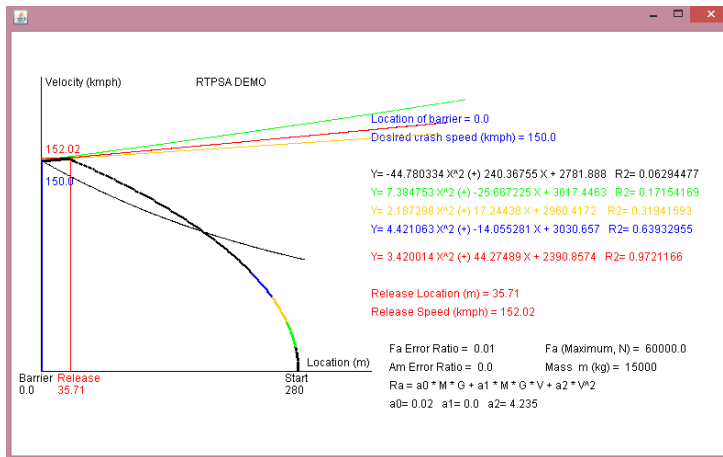


Figure 4.5 Maximum Propulsion Force = 60,000N.

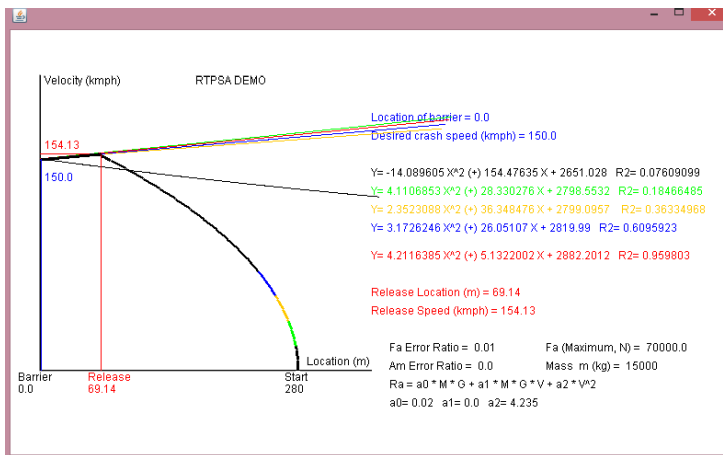


Figure 4.6 Maximum Propulsion Force = 70,000N.

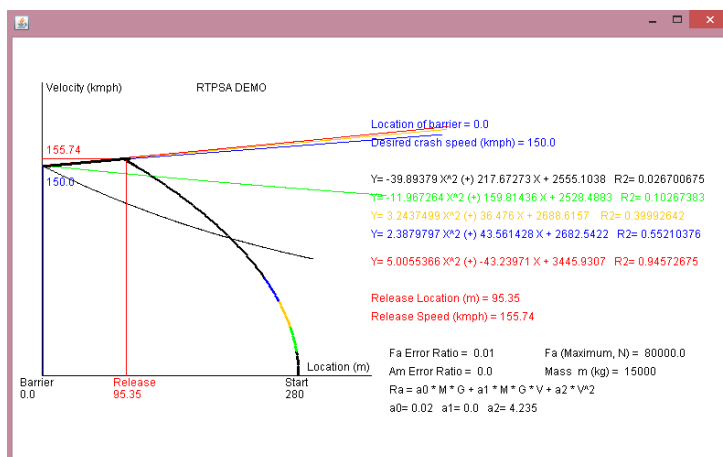


Figure 4.7 Maximum Propulsion Force = 80,000N.

Chapter 5 Conclusion and Future Work

5.1 Conclusion

Based on the running results in Chapter 4 it can be concluded as follows. At first, the error ratio of F_a significantly influences the regression performance. As shown in Table 4.8, a lowest R^2 of 0.2584 is obtained when the error ratio of F_a is set to 1%, and a highest R^2 of 0.9782 is obtained when the error ratio of F_a is set to 0.1%. So, in order to generate a reliable result of this algorithm, the determination of the real error ratio of F_a is one of the major task during the commissioning.

Secondly, given an error ratio of F_a , the value of R^2 is quite subjected to the data set used for the calculation. This can be seen in the result of Table 4.1. A lowest R^2 of 0.0554 is obtained when data set of item 1-24 is used, and a highest R^2 of 0.7552 is obtained when data set of item 1-98 is used. It shows that when the data that we used for R^2 calculation is more and collected closer to the release point, it will lead to a better regression result. That is why the most up-to-date data needs to be collected and used in order to obtain a more accurate release speed and location.

Thirdly, given an error ratio of F_a , the current algorithm can generate a better result of R^2 when crash speed is set higher. This can be seen from Table 4.2. A lowest R^2 of 0.2584 is obtained when the crash speed is set to 60kph, and a highest R^2 of 0.9724 is obtained when the crash speed is set to 150kph. One reason for this is because of the method we use to calculate the R^2 . As discussed earlier in Section 3.7.1, the data of those where speed is less than 36kph (10m/s) are found out after several experiments, to be useless to the accuracy of coefficients. They negatively impact the accuracy of the coefficients. So, in our coefficients calculation, the data set we use is those are greater than 36kph and less than the value of predefined crash

speed. When the crash speed is low, e.g. 60kph, the data set for calculation is those from 36kph to 60kph. There are only 34 records of sample data used for such calculation. This is the reason for the lower value of R^2 when the crash speed is set lower. When the crash speed is set higher, a higher value of R^2 is obtained because there are more sample data used in coefficients calculation. One solution for this issue in low speed situation is to divide the small speed interval of sample data set to 0.1m/s or even 0.01m/s, instead of current 0.2m/s, which will generate more sample data for coefficients calculation.

5.2 Future Work

There are several possible future works that will be helpful to the current research work. Firstly, a new simulated data set for testing with smaller interval is needed which will provide much more sample data and make the algorithm have better performance in lower speed situation. This can improve the performance of this algorithm in high speed situation as well because of more data available.

Secondly, when the real test facility is ready, the current input interface needs to be modified and adjusted to adapt to the real testing environment.

Finally, in the real test facility, the test vehicle is not released right away when it reaches the designated speed, it will maintain at a certain speed for some time and get released at designated location and speed. In this paper, we only discuss the situation of releasing the test vehicle right away once it reaches the desired speed and location. The software developed should be amended in the future to meet the requirement for variable releasing strategies.

BIBLIOGRAPHY

- [1] The Ministry of Construction and Transportation. The Guide of Safety Reference of Railway Train. No. 2007-278(2007).
- [2] H. S. Lee, S. H. Park and D. C. Han. Crashworthiness Analysis of Korean High Speed Train Trailer. *Fall Conference Proceedings of Korea Society for Railway*, pp. 511518, (1999).
- [3] J. S. Koo, H. J. Cho, D. S. Kim and Y. H. Youn. An Evaluation of Crashworthiness for the Full Rake KHST Using 1-D Dynamic Model. *Journal of Korea Society for Railway*. Vol. 4, pp. 94-101, (2001).
- [4] K. S. Rho and J. S. Koo. An Evaluation of Crashworthiness for the final design of the KHST power-car. *Spring Conference Proceedings of Korea Society for Railway*. pp.3846, (2002).
- [5] R. Stringfellow and P. Llana. Detailed Modeling of the Train-to-Train Impact Test. *Final Report*. DOT/FRA/ORD-07/20(2007).
- [6] R. Mayville, R. Stringfellow and E. Martinez. Development of Conventional Passenger Cab Car End Structure Designs for Full Scale Testing. *Final Report*, DOT/FRA/ORD 06/20(2006).
- [7] B. Rickle and R. Walker. Passenger Train Grade Crossing Impact Tests: Test Procedures, Instrumentation, and Data. *Test Report*, FRA/ORD-06/16(2003).
- [8] David Tyrell. Passenger Rail Train-to-Train Impact Test Volume I: Overview and Selected Results. *Final Report*, DOT/FRA/ORD-03/17. I (2003).
- [9] David Tyrell. A Train-to-Train Impact Test of Crash rail_cw_2006_09. *6th International Symposium* (4-5 December 2006),
- [10] S. R. Kim and J. S. Koo. The Development of Virtual Testing Model for Korea High Speed Train. 2014
- [11] URL: <http://www.cbc.ca/news/world/china-train-crash-that-killed-39-caused-by-signal-flaws-1.1083426> CBC NEWS 06/03/2015
- [12] URL: <http://www.ushsr.com/hsr/hsrworldwide.html> 06/03/2015
- [13] Alan O. Sykes*. An Introduction to Regression Analysis. *The Inaugural Coase Lecture*.
- [14] Dan Campbell and Sherlock Campbell. Introduction to Regression and Data Analysis. Statlab workshop Series 2008
- [15] URL: http://en.wikipedia.org/wiki/Linear_regression 06/03/2015

- [16] URL: http://en.wikipedia.org/wiki/Least_squares 06/03/2015
- [17] URL: Cram101 Textbook Reviews - 2013 - Education 06/03/2015
- [18] URL: <http://blog.minitab.com/blog/adventures-in-statistics/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit> 06/03/2015
- [19] Douglas C. Montgomery, Elizabeth A. Peck and G. Geoffrey Vining. Introduction to Linear Regression Analysis. 5th Edition, ISBN: 978-0-470-54281-1, April 2012, ©2012
- [20] Clapham C. and Nicholson, J.. Simple Linear Regression. *Oxford University Press*. 2009. ISBN: 9780199235940; E-ISBN: 9780191727122
- [21] J. P. Onyango and A. M. Plews. A Textbook of Basic Statistics. *East African Publishers*. 1987
- [22] Scott, J. and Marshall, G.. Multiple Linear Regression. *Oxford University Press*. 2009. ISBN: 9780199533008; E-ISBN: 9780191726842
- [23] Keith A. McNeil, Francis J. Kelly 1925- and Judy McNeil. Testing research hypotheses using multiple linear regression. *Carbondale: Southern Illinois University Press*. 1975. ISBN0809307324
- [24] Esa I. Uusipaikka. Confidence Intervals in Generalized regression models. *Boca Raton, FL: CRC Press*. c2009. ISBN9781420060270; ISBN1420060279
- [25] C. Radhakrishna., Rao, Shalabh., Helge Toutenburg., Christian Heumann., and Springer Link. Linear Models and Generalizations Least Squares and Alternatives. Third Extended Edition. *Berlin, Heidelberg: Springer Berlin Heidelberg*. 2008. ISBN3-642-09353-1; ISBN3-540-74226-3
- [26] Law, J. and Smullen, J. R squared. A Dictionary of Finance and Banking. *Oxford University Press*. 2008. Source: Oxford Reference (Oxford University Press). ISBN: 9780199229741; E-ISBN: 9780191726668
- [27] Richard M. Heiberger, Erich Neuwirth, SpringerLink. R Through Excel A Spreadsheet Interface for Statistics, Data Analysis, and Graphics. Series: Use R. New York, NY: Springer New York. 2009. ISBN1-4419-0051-9
- [28] James T. McClave, Frank H Dietrich and Terry Sincich. Statistics. Edition: 7th ed. Englewood, N.J.: Prentice Hall. c1997. ISBN013471542X
- [29] Chi-Keung Leung, University of Chicago. Dept of Geography. China, railway patterns and national goals. Series: Research paper (University of Chicago. Dept. of Geography); no.195. Chicago, Ill.: University of Chicago, Dept. of Geography. 1980. ISBN0890651027:
- [30] Robert Cervero, Michael Bernick 1953-; University of California, Berkeley. Institute of Urban & Regional Development.; University of California, Berkeley. National Transit Access Center.; California Intercity High Speed Rail Commission. High-speed rail and development of California's Central Valley: comparative lessons and public policy considerations. Series: Working paper (University of

California, Berkeley. Institute of Urban & Regional Development); no. 675. Berkeley, Calif.: University of California at Berkeley, Institute of Urban and Regional Development. 1996

[31] Mamoru Taniguchi, University of California, Berkeley. Institute of Urban & Regional Development. High speed rail in Japan: a review and evaluation of the Shinkansen train. Series: Working paper (University of California, Berkeley. Institute of Urban & Regional Development); no. 557. Berkeley: Institute of Urban and Regional Development, University of California at Berkeley. 1992

[32] CANADA. PARLIAMENT. HOUSE OF COMMONS. STANDING COMMITTEE ON TRANSPORT, R.A CORBETT. HIGH SPEED RAIL: THE CANADIAN CONCEPT - SECOND REPORT OF THE STANDING COMMITTEE ON TRANSPORT. *QUEBEC-WINDSOR CORRIDOR*. 1992

[33] Daniel Leavitt, Erin Vaca; Peter Hall; University of California, Berkeley. Institute of Urban & Regional Development. Revenue and ridership potential for a high-speed rail service in the San Francisco/Sacramento-Los Angeles corridor. Series: Working paper (University of California, Berkeley. Institute of Urban & Regional Development); no. 609. Berkeley, Calif.: Institute of Urban and Regional Development, University of California at Berkeley. 1994

[34] Tracking slower; high-speed rail in China. (Brief article). *The Economist* (US), July 2, 2011, Vol.400 (8740), p.32(US). Source: Cengage Learning, Inc. ISSN: 0013-0613

[35] high-speed train. *Oxford University Press*. 2004. Source: Oxford Reference (Oxford University Press). ISBN: 9780198606239; E-ISBN: 9780191728006

[36] Train Collision Avoidance System on test. (News). *International Railway Journal*, Dec, 2012, Vol.52 (12), p.4(1). Source: Cengage Learning, Inc. ISSN: 0744-5326

[37] Transportation Safety Board of Canada. CN North America collision, freight train No. 457-12, Mile 86.6, Fraser Subdivision, Longworth, British Columbia, 13 August 1993. Series: Railway occurrence report; R93V0155. Hull, Quebec: Transportation Safety Board of Canada. 1994. ISBN066223460X

[38] Rolling stock damage without derailment or collision, Canadian National freight train number M-30451-11, mile 238.30, Redditt subdivision, Dugald, Manitoba, 14 January 2009. ISBN1-100-16595-9

[39] Milho, J F, Ambrósio, J A C, Pereira, M F O S. Validated multibody model for train crash analysis. *Woodhead Publishing*. Source: Taylor & Francis, Taylor & Francis Group. ISSN: 1358-8265 ; E-ISSN: 1754- 2111 ; DOI: 10.1533/ijcr.2003.0242

[40] Stringfellow, R., Rancatore, R., Llana, P. and Mayville, R. Analysis of colliding vehicle interactions for the passenger rail train-to-train impact test. *ASME/IEEE Joint Rail Conference*. 2004 2004, pp.11-18. *IEEE Conference Publications*. ISBN: 0-7918-4163-4

[41] Tyrell, D., Jacobsen, K. and Parent, D. Preparations for a train-to-train impact

test of crash-energy management passenger rail equipment. *Proceedings of the 2005 ASME/IEEE Joint Rail Conference*, 2005 2005, pp.107-116. *IEEE Conference Publications*. ISBN: 0-7918-3752-1 ; DOI: 10.1109/RRCON.2005.186063

[42] Validated multibody model for train crash analysis. (Author abstract). *International Journal of Crashworthiness*. Jan, 2003. Vol.8 (1), p.339 [Peer Reviewed Journal]
Source: Cengage Learning, Inc. ISSN: 1358-8265

[43] Parent, D. , Tyrell, D. and Perlman, A B. Crashworthiness analysis of the Placentia, CA rail collision. *International Journal of Crashworthiness*, 2004, Vol.9 (5), p.527-534. *WoodHead Publishing*. ISSN: 1358-8265; E-ISSN: 17542111; DOI: 10.1533/ijcr.2004.0308

[44] Jacobsen, K., Tyrell, D. and Perlman, B. Impact test of a crash-energy management passenger rail car. *ASME/IEEE Joint Rail Conference*, 2004 2004, pp.19-26. *IEEE Conference Publications*. ISBN: 0-7918-4163-4

[45] John O. Rawlings., Sastry G. Pentula., David A. Dickey. *Applied regression analysis: a research tool*. 2nd ed. 1998. *Springer texts in statistics*. ISBN: 0-387-98454-2

[46] Gerard E. Dallal. *Introduction to Simple Linear Regression*. *Tufts University*. 2000. URL: <http://www.jerrydallal.com/LHSP/slr.htm>.

[47] URL: <http://blog.minitab.com/blog/adventures-in-statistics/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit>. 06/03/2015

VITA AUCTORIS

NAME: Xian Jian Jiang

PLACE OF BIRTH: Beijing, China

YEAR OF BIRTH: 1964

EDUCATION: No. 7 Middle School, Beijing, China
1976-1981

Beihang University, Beijing, China
1981-1985 B.Eng

Beihang University, Beijing, China
1985-1988 M.Eng

University of Windsor, Windsor, Ontario
2013-2014 B.Ed

University of Windsor, Windsor, Ontario
2013-2015 M.Sc