



ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL

SCIENCE

**Computational Modeling and Analysis of Traffic Crash and Traffic Volume
in Addis – Adama Expressway**

Ayana Murad

**A Thesis Submitted to the Department of Computational Science in Partial
Fulfillment for the Degree of Master of Science in Computational Science**

Addis Ababa, Ethiopia

April, 2021

ADDIS ABABA UNIVERSITY
COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE

Ayana Murad

Advisor: Solomon Gizaw (PhD)

This is to certify that the thesis prepared by Ayana Murad, titled: computational modeling and analysis of traffic crash and traffic volume in Addis – Adama expressway and submitted in partial fulfillment of the requirements for the degree of Master of Science in Computational Science complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Name	Signature	Date
<u>Solomon Gizaw (PhD)</u> Advisor	_____	_____
<u>Addisu Gezahegn (PhD)</u> Internal Examiner	_____	_____
<u>Fekade Getahun (PhD)</u> External Examiner	_____	_____

Abstract

Transportation has a major contribution in the development of the human civilization. The accessibility of highway transportation has given many focal points that contribute to a high standard of living. However, many issues related to the highway mode of transportation exist. These issues incorporate highway related accidents, parking troubles, clog, natural risks (carbon emissions, clamor contamination, etc.) and delay. To solve these problems building expressways is one of the solutions. Even though building express way is a good solution for solving problems related to highway traffic, Data collected from Ethiopia Toll Road Enterprise indicated that, on average, about 417 road crashes were reported since September 2014 to February 2016 that leads around 672 traffic accidents. Road traffic crashes (RTCs) are globally acknowledged as increasing threat to society, because they can affect many lives when they result in severe injury or fatality. Ethiopia is among the leading countries in road traffic accident. The recent road safety record of Addis Ababa- Adama expressway is also alarming the severity of the situation and calling for an integral effort of all pertinent stakeholders to reverse the trend. In this research we modeled traffic crashes and traffic volumes in Addis – Adama express way with ordinary differential equation using interpolation methods (i.e. newton DVD and Lagrange interpolations). We solved the ordinary differential equations we got after modeling using Euler method and Runge – Kutta method. We observed if there is any relation between traffic crash and traffic volume. We analyzed the traffic crash data parameters i.e. vehicle type, vehicle type with weekdays and direction to observe the factors causing traffic accident. The data we used for modelling and analyzing is collected from ETRE. Using the mathematical model of traffic crash, we were able to predict 2020 number of traffic crash. The finding shows that traffic crash and traffic volume have linear relationship. From the analysis we observed that Small automobiles are causing the highest traffic crash, the highest number of traffic crash occurred in Friday, and vehicles heading to mojo are causing the highest number of traffic crash. Therefore, ETRE should take restrict monitoring on small automobiles, vehicles heading to mojo, and in weekends.

Keywords: computational modeling, numerical analysis, interpolation, newton divided difference, Lagrange interpolations, Euler method, Runge – Kutta method, ODE

Acknowledgement

First and foremost I praise and adore the almighty Allah for his inexpressible love and mercy. I would like to express my gratitude to my adviser Dr. Solomon Gizaw for his invaluable guidance and encouragement throughout the thesis work.

I am also very grateful to my families for their prayer, encouragement and continuous support. I would like to thank my husband for staying with me at late nights. This research might be impossible without their support.

Finally, I would like to thank my work colleagues and classmate for their support and sharing ideas.

Contents

List of Tables	vi
List of Figures	vii
List of Acronyms	viii
Chapter one.....	1
1. Introduction.....	1
1.1. Background.....	1
1.2. Motivation	3
1.3. Problem statement.....	4
1.4. Objective of the study	5
1.4.1. General objective	5
1.4.2. Specific objectives	5
1.5. Methods.....	5
1.6. Scope and limitation of the study	6
1.7. Application of results	6
1.8. Thesis Organization	6
Chapter two.....	7
2. Literature review	7
2.1. Traffic volume	7
2.2. Traffic crash.....	7
2.3. Express way.....	8
2.3.1. Lagos–Ibadan Expressway in Nigeria.....	8
2.3.2. Express way in India	8
2.3.3. Express way in china.....	9

2.3.4.	Express way in Ethiopia	9
2.4.	Computational modeling and simulation	10
2.4.1.	What is modeling?	10
2.4.2.	What is simulation?.....	10
2.5.	Modeling with differential equation	11
2.5.1.	Differential equation	11
2.5.2.	Ordinary differential equation	11
2.5.3.	Modeling with differential equation	12
2.6.	Numerical modeling.....	12
2.6.1.	Interpolation	12
2.6.2.	Newton divided difference	13
2.6.3.	Lagrange interpolation	16
2.7.	Numerical methods for ordinary differential equations	17
2.7.1.	Euler’s Method	18
2.7.2.	Runge–Kutta method	19
2.8.	Data analytics	21
2.8.1.	Data analysis with python	21
2.9.	Summary	22
Chapter three.....		25
3.	Methodology of the research	25
3.1.	Data source and Method of collection	25
3.2.	Modeling methods	25
3.2.1.	Newton divided difference	25
3.2.2.	Lagrange interpolation	27

3.2.3.	Euler method	28
3.2.4.	Runge – kutta method	29
3.3.	Analysis method	30
3.3.1.	Analysis tool.....	30
3.4.	Software used in the research	31
3.4.1.	Matlab	31
3.4.2.	Anaconda.....	32
Chapter four	33
4.	Thesis design framework.....	33
4.1.	The frameworks	33
4.2.	Modelling framework.....	36
4.2.1.	Data Collection and Preparation.....	36
4.2.2.	Data preprocessing.....	36
4.2.3.	Identifying data points	36
4.2.4.	Modelling the data	36
4.2.3.	Solving ODE	37
4.2.4.	Evaluation	37
4.3.	Analysis Framework	38
4.3.1.	Data collection and preparation.....	38
4.3.2.	Data preprocessing.....	38
4.3.3.	Analysis.....	38
4.3.4.	Presentation	38
Chapter five	39
5.	Results and Discussion	39

5.1. Dataset used.....39

5.2. Description of the study area40

5.3. Data modelling.....41

 5.3.1. Traffic crash modelling using newton divided difference41

 5.3.2. Traffic crash modelling using Lagrange interpolation.....43

5.4. Solving ordinary differential equations of traffic crashes46

 5.4.1. Solving ordinary differential equations using Euler method.....46

 5.4.2. Solving ODE using Runge-Kutta method.....48

5.5. Modelling traffic volume.....50

 5.5.1. Modelling traffic volume using newton divided difference50

 5.5.2. Modelling traffic volume using Lagrange method52

5.6. Solving ordinary differential equations of traffic volume55

 5.6.1. Solving ODE of traffic volume using Euler method55

 5.6.2. Solving ODE of traffic volume using Runge-Kutta method.....56

5.7. Correlation of traffic crash and traffic volume58

5.8. Analysis of traffic crash60

 5.8.1. Analysis of traffic crash based on vehicle type60

 5.8.2. Analysis of traffic crash based on vehicle types and week days61

 5.8.3. Analysis of traffic crash based on direction62

5.9. Discussion.....64

 5.9.1. Traffic crash modelling64

 5.9.2. Traffic volume modelling.....64

 5.9.3. The correlation between traffic crash and traffic volume64

 5.9.4. The effect of vehicle type on traffic crash.....65

5.9.5. The effect of vehicle type with weekdays on traffic crash65

5.9.6. The effect of direction on traffic crash65

Chapter six.....66

6. Conclusion and Recommendation.....66

6.1. Conclusion.....66

6.2. Contributions67

6.3. Recommendation67

Reference68

Annex A: Sample code for data preprocessing.....73

Annex B: Sample code for Lagrange interpolation.....74

Annex C: Sample code for Runge – Kutta method.....75

Annex D: Sample code for analyzing traffic crash data parameter vehicle type.....76

List of Tables

Table 2.1: Newton divided difference.....	15
Table 5.1: Number of traffic crashes per year	39
Table 5.2: Number vehicle passed per year.....	40
Table 5.3: The output of Euler method	47
Table 5.4: Runge-Kutta method output.....	48
Table 5.5: The output of Euler method in traffic volume	55
Table 5.6: Runge-Kutta method output.....	56
Table 5.7: Output of traffic crash.....	58
Table 5.8: Output of traffic volume	58
Table 5.9: Vehicle types caused traffic crash with number	60
Table 5.10: Traffic crashes occurred in week days.....	61
Table 5.11: traffic crash occurred by direction.....	62

List of Figures

Figure 2.1: Euler method.....	19
Figure 2.2: Runge-Kutta method	20
Figure 4.1: Thesis design framework of Modeling section.....	34
Figure 4.2: Thesis framework of Analysis section	35
Figure 5.1: Location of Addis Ababa - Adama expressway	41
Figure 5.2: Plot of no. of traffic crashes per year using newton divided difference.....	43
Figure 5.3: Plot of no. of traffic crashes per year using Lagrange	45
Figure 5.4: Newton divided difference vs Lagrange.....	46
Figure 5.5: Plot of traffic crash with Euler method	47
Figure 5.6: Plot of traffic crashes with Runge-Kutta method	49
Figure 5.7: Traffic volume per year using newton divided difference	51
Figure 5.8: Traffic volume per year using Lagrange method.....	53
Figure 5.9: Traffic volume per year using newton DVD and Lagrange	54
Figure 5.10: Plot of traffic volume using Euler method	56
Figure 5.11: Plot of traffic volume using Runge-Kutta method.....	57
Figure 5.12: Traffic volume vs Traffic crash	59
Figure 5.13: Number of traffic crashes happened by vehicles type.....	61
Figure 5.14: Traffic crash happened by vehicle types within week days	62
Figure 5.15: Traffic crashes occurred by direction	63

List of Acronyms

DA	Data Analytics
DVD	Divided Difference
EC	Ethiopian Calendar
ERA	Ethiopian Road Authority
ETRE	Ethiopian Toll Road Enterprise
FB	Full Bayesian
GDP	Growth and Development Plan
ODE	Ordinary Differential Equation
RTA	Road Traffic Accident
RTC	Road Traffic Crash

Chapter one

1. Introduction

1.1. Background

Transportation systems are necessary to economic and the security of countries. The transportation system effectiveness relies on its capacity to support the constant movement of people, vehicles, and goods from place to place [1]. Transportation is a basic human need during their lifetime. From ancient times, everyone travels either for necessity or enjoyment. A closely related need is the transport of raw materials to a production unit or end products for consumption purposes. Transportation satisfies these essential needs of humanity. Transportation has a major contribution in the development of the human civilization [2].

The accessibility of highway transportation has given many focal points that contribute to a high standard of living. However, many issues related to the highway mode of transportation exist. These issues incorporate highway related accidents, parking troubles, clog, natural risks (carbon emissions, clamor contamination, etc.) and delay [2]. To solve these problems building expressways is one of the solutions. Expressway is a highway particularly arranged for high-speed traffic, ordinarily having few in case any convergences, limited points of access or exit, and a divider between paths for traffic moving in inverse bearings. Frequently expressways have been constructed over totally new routes, passing close but not through expansive centers of populace, on more or less coordinate lines between wanted ends. Their advantages incorporate high speed, more noteworthy security, comfort and convenience for drivers and travelers, and lower vehicle working costs [3, 4].

Ethiopian Road Authority (ERA) is in charge of all public roads construction in Ethiopia, they proposed the new and enhanced quality of road in safety and working cost at Oromia region called Addis Ababa– Adama Expressway, which is the first advanced road for Ethiopia and East Africa which completed and officially opened in September 2014. The new six-lane, the 12m-wide road was constructed with a different line but was placed adjacent to and 3.5km from the west side of the old Addis Ababa Adama road. The new road is 20km shorter than the old Addis Ababa Adama road, while the route is fenced on either side for protection from pedestrians and animals [5].

Even though building express way is a good solution for solving problems related to highway traffic, Data gathered from Ethiopia Toll Road Enterprise showed that, on average, about 417 road crashes were reported since September 2014 to February 2016 that leads around 672 traffic accidents [6]. Road traffic crashes (RTCs) are causing a serious risk to modern communities, as they result in injuries, disabilities and even death, along with important economic and social problems [7]. Ethiopia is one of the developing countries that traffic accidents happened regularly and it causes death, serious injury and property destruction to the nation, this shows the traffic accident rate is at a threatening stage [5]. Therefore, it is necessary to model and analyze traffic crashes in the express way in order to resolve the problems observable now and improve future works. Transportation planning uses the word 'models' broadly. This word is used to describe a series of mathematical equations that are used to denote how options are made when people travel. Models are significant because transportation plans and investments are on the basis of what the models say about future. Transportation modeling is used to develop information to support decision making on the future development and management of transportation systems [8]. In the last two decades transportation systems analysis has appeared as a recognized career. Many governmental organizations, colleges, analysts, experts, and private mechanical groups around the world are getting to be really multi-modal in their introduction and they are deciding an efficient approach to transportation issues [2].

Researcher have tried developing models for detecting causes of accident and analyze the severity of accidents. But many of the researchers use common models such as regression [9,10,11,12]. In a research made in four European countries the variation of accident counts using the model generalized by Poisson regression attribute to randomness weather, exposure, daylight, were calculated by varying speed limits and reporting routines[10]. Also, two researchers (Graham and Glaister) studied the influence of density, urban scale and land use-mix on the incidence of pedestrian casualties [11]. Full Bayesian (FB) and negative binomial models were developed for the spatial analysis of fatal and injury crashes in Pennsylvania by researcher's named Aguero-Valverde and Jovani's [12].

Most of traffic crash models developed by different researchers used the regression techniques i.e. linear, logistic, and multiple [9,10,11,12]. The models developed by most researchers are on traffic crash data. They did not recognize traffic volume data and they did not observe if it has any effect

on traffic crash. We propose to model both traffic crash and traffic volume data using numerical methods. We will use interpolation methods to model the data i.e. newton divided difference and Lagrange methods.

In Ethiopia we found only one model developed in Addis – Adama expressway which was developed on the traffic crash data. The research used Multinomial Probit Regression model to assess the determinants of accident on Addis Ababa – Adama expressway. This means, empirical regression method is used to identify agents contributing for road traffic accident [6]. The other research was analysis on 1,137 road traffic accidents between September 4, 2007 and April 30, 2010(E.C). Based on the data collected from ETRE the distribution of traffic accident characterized into distribution of RTAs by severity, distribution of RTAs with time of occurrence, distribution of RTA with week of a day, distribution of RTA with month and driving direction, distribution of RTAs with types of Accident, and distribution of RTAs with types of the vehicle involved [5]. According to the literature, we can see that there are not many studies done on the express way[5, 6]. The researches done are on traffic crash and they used statistical methodologies. In this research we propose to model traffic crashes and traffic volumes in Addis – Adama expressway with ordinary differential equation using interpolation methods. We will observe if there is any relation between traffic crash and traffic volume. We will analyze the traffic crash data to observe the factors causing traffic accident. We will use interpolation methods i.e. newton divided difference and Lagrange methods to model the data. We will solve the ordinary differential equations attained by modeling process using Euler and Runge – Kutta method.

1.2. Motivation

In the past few years, particularly since the early 1990s, the increment in road traffic have caused serious congestion, delays, accidents and environmental problems [13]. The recent road safety record of Addis Ababa- Adama expressway is also alarming. The seriousness of the circumstance calls for an integral effort of all pertinent stakeholders to invert the trend. Data collected from Ethiopia Toll Road Enterprise showed that, on average, about 417 road crashes were reported in the past one and half years that leads around 672 traffic accidents [6]. As a result, it is found necessary to model and analyze the traffic crashes in Addis Ababa - Adama expressway.

The purpose of this research is

1. Observe current traffic crashes,
2. Predict future traffic crashes,
3. Observe the relation between traffic crashes and traffic volume and,
4. Suggest better traffic management.

1.3. Problem statement

Road crash prediction models are very useful instruments, given their potential for deciding the crash frequency occurrence. Crash frequency refers to the prediction of the number of crashes that would happen on a particular road section or intersection in a time extent [14]. Statistical methodologies have been applied to model crash data and findings gained from survey for a better and simple perception. The most common models used are the regression techniques (linear, logistic, multiple) and few authors use regression techniques for finding goodness of fit and then model the equations and coefficients into multinomial logit models. The models mainly applied the regression technologies. They are good but very old and conventional [9].

Regression techniques are one of the most popular statistical methods applied for predictive modeling and data mining functions. A linear regression is a method in which the dependent variable is continuous in nature. In logistic regression, the dependent variable is binary in nature (having two categories) [15]. Multiple regression involves a single dependent variable and two or more independent variables [16]. Several Researchers have made different researches using different methods using a generalized poisson regression model. Fridstrom measured the contribution of randomness, exposure, weather and daylight to the variation in road accident counts [10]. Also, two researchers (Graham and Glaister) studied the influence of density, urban scale and land use-mix on the incidence of pedestrian casualties [11]. Full Bayesian (FB) and negative binomial models were developed for the spatial analysis of fatal and injury crashes in Pennsylvania by researcher's named Aguero-Valverde and Jovani's [12]. Most of the models did not focus on number of vehicles passed and vehicle types affecting traffic crash. In this research we will focus on them.

The purpose of this research is to model number of traffic crashes, to predict future traffic crashes, to model traffic volume, to observe the relation of traffic crash with number of vehicles passed

(traffic volume) and vehicle types and, to suggest better traffic management. We will use ordinary differential equation to model the data. This research makes it different from previous researches that we used traffic volume, we will observe its relation with traffic crash, and the methodology we used is computational methodologies.

1.4. Objective of the study

1.4.1. General objective

This research has a great advantage for Ethiopian road authority and also for Ethiopian toll road enterprise to improve their policy and is a good reference while building new expressways. The main aim of this thesis is to computationally model traffic crash and traffic volume data in Addis – Adama express way and to do analysis on these data.

1.4.2. Specific objectives

The specific objectives of this thesis is given below

- To model the data collected from Ethiopian toll road enterprise using ordinary differential equation.
- To simulate the model.
- To do analysis on the data.
- To predict future traffic crash in the expressway.

1.5. Methods

We will use the following methods in order to achieve the objective and specific objectives of the research

- **Literature review:** articles, books, dissertation, relevant literatures, and studies (which are available online) will be reviewed to get a better understanding on the previous studies regarding computational modelling.
 - We will review numerical methods i.e. newton divided difference and Lagrange interpolation methods.
 - We will examine numerical solution methods i.e. Euler method and Runge – Kutta method.
 - We will review data analysis tools.

- **Data collection:** the data we will use for modelling and analyzing is collected from ETRE.
- **Tools:** we will use Matlab for modelling the data and we will use anaconda, Jupyter notebook, and python for analyzing the data.

1.6. Scope and limitation of the study

The scope of this study is limited to Addis Ababa – Adama express way. The study is to model and analyze traffic crash and traffic volume in Addis Ababa – Adama express way. The model is developed based on interpolation methods.

The data collected from Ethiopian Toll Road Enterprise is from 2015 up to 2019. It does not include 2020 data. The way they record the data has been the main challenge.

1.7. Application of results

The applications of this research are listed below.

- For predicting the future traffic crashes and estimate future safety.
- To know the relationship of traffic crash and traffic volume
- To identify the contributing factors for traffic crash and to know the level of their harm.

1.8. Thesis Organization

The first chapter of this thesis gives a brief introduction about transportation, expressway, modeling, and analysis. Chapter two is review of relevant literatures. Chapter three gives the methodology of the thesis. Chapter four explains the thesis framework and its components. Chapter five encompasses results and discussion. Conclusion and recommendation are included in chapter six.

Chapter two

2. Literature review

In this chapter we will discuss about traffic volume and traffic crash in the first two sections. Then, we will explain what modeling and simulation are. We will use ODE to model the data so modeling with differential equation will be discussed in this chapter. Interpolation and its two methods i.e. newton divided difference and Lagrange methods will be explained briefly. We will solve the ODE using Euler method and Runge-Kutta method. This two methods will be included here. In the last sections we will explain data analysis. Finally we will present summary of literatures.

2.1. Traffic volume

Traffic Data Collection is basic precondition for transport planning. Traffic Data establishes an integral part of national economics and such knowledge is important in formulating a rational transport policy for movement of passengers and goods by both government and the private sectors. Traffic Volume Count is counting of number of vehicles passing through a road over a period of time [17]. For planning, designing and operation of transportation system the primary and preeminent requirement is traffic volume. Traffic volume is simply the number of vehicles passing a section of a roadway [18]. The term traffic volume study can be termed as traffic flow survey or simply the traffic survey. It is defined as the method to decide mainly volume of traffic moving on the roads at a specific section during a specific time [19]. The purposes of traffic volume study are to set up relative significance of any route or road facility, to choose the need for change and extension of a road and to distribute the funds accordingly, to plan and design the existing and new facilities of traffic operations on the road, and to do analysis of traffic trends and pattern on the road [20].

2.2. Traffic crash

Road traffic crashes (RTCs) are universally recognized as expanding threat to society, since they can influence many lives when they result in serious injury or fatality [7]. Road traffic crash data are valuable instruments to support the improvement, execution, and evaluation of highway safety programs that tend to decrease road traffic crashes [21]. Road traffic death rates in low- and middle-income nations are more than double those in high-income nations. Road traffic deaths and

injuries in low- and middle-income nations are assessed to cause financial losses of up to 5% of GDP [5]. Ethiopia is among the leading nations in road traffic accident. A RTC can be caused by different factors related to road conditions, traffic density, vehicle conditions and human mistakes [7].

2.3. Express way

Expressway is described as a major arterial divided highway that structures two or more traffic lanes in each direction, with opposing traffic detached by a median strip; exclusion of grade crossings; controlled entries and exits; and advanced designs eliminating steep grades, sharp curves, and other hazards and inconveniences to driving. Expressways are also called motorway, thruway, throughway, freeway or superhighway. Often expressways are being constructed over completely new routes, passing near but not through large centers of population with less direct lines between desired stations. They have high benefits of high speed, better safety, comfort and suitability for drivers and passengers, and lower vehicle operating costs[4].

2.3.1. Lagos–Ibadan Expressway in Nigeria

The Lagos–Ibadan Expressway is the largest road network in Africa and the oldest expressway in Nigeria which was bespoken in August 1978. It connects the capital of Oyo State Ibadan and the Nigeria's largest city Lagos and it is a 127.6-kilometre-long (79.3 mi). This expressway is used as a major route to eastern, northern and southern part of Nigeria. The expressway is the busiest interstate route in Nigeria which handles more than 250,000 PCUs daily. It was built during the military era under the supervision of Major-General Olusegun Obasanjo who resigned in October 1, 1979. It is part of the Federal Roads Maintenance Agency (FERMA) projects, concerned with road improvement and connectivity between the States of Nigeria[22].

2.3.2. Express way in India

Indian road, network highways are the standard grade of roads. At the location of intersection usually provide with grade, separators and they are 6-8 lane Controlled access highways. In India approximately 1455.4 km of expressways operational currently. In India, an expressway is an arterial highway for motorized traffic, with divided carriageways for high-speed travel, with full control of access and provided with grade separators. In India expressways are different from other roads in the country which traffic consists of only cars and trucks where bicycle, three wheelers

and two-wheelers are not allowed to play. By the year 2022, the government of India aims to expand the highway network and plans to add additional 18,637 kms of expressways. In measuring the growth of a country, a well-developed network of roadways is a key factor. India ranks 3rd in the list of the countries by road networks next to United States and China [23].

2.3.3. Express way in china

The expressway network of China, with the national-level expressway system officially known as the National Trunk Highway System, is an integrated system of national and provincial-level expressways in China [24]. With the construction of the Shenyang–Dalian Expressway began between the cities of Shenyang and Dalian on 7 June 1984, the Chinese government took an interest in an expressway system. The first modern at-grade China National Highways is the Shanghai–Jiading Expressway, opened in October 1988. The early 1990s saw the start of the country's massive plan to upgrade its network of roads. On 13 January 2005, China's Minister of Transport introduced the 7918 network, later renamed the 71118 network, composed of a grid of 7 radial expressways from Beijing, 9 north–south expressways (increased to 11), and 18 east–west expressways that would form the backbone of the national expressway system. By the end of 2019, the total length of China's expressway network reached 149,600 kilometers, the world's largest expressway system by length, having surpassed the overall length of the American Interstate Highway System in 2011. Planned length is 168,478 kilometers by 2020. Many of the major expressways parallel routes of the older China National Highways [24].

2.3.4. Express way in Ethiopia

Ethiopian Road Authority (ERA) is in charge of all public roads construction in Ethiopia, they proposed the new and enhanced quality of road in safety and working cost at Oromia region called Addis Ababa– Adama Expressway, which is the first advanced road for Ethiopia and East Africa which completed and officially opened in September 2014. The new six-lane, the 12m-wide road was constructed with a different line but was placed adjacent to and 3.5km from the west side of the old Addis Ababa Adama road. The new road is 20km shorter than the old Addis Ababa Adama road, while the route is fenced on either side for protection from pedestrians and animals [5].

2.4. Computational modeling and simulation

2.4.1. What is modeling?

Most papers give similar meaning to modeling as the process of constructing a model. Model is just the representation of the building of some system of interest. Model is similar but simpler to the system it represents. Models should enable the analyst to predict the effect of changes to the system they should also be close approximation to the real system and integrate most of its noticeable features. And still models should not be complex good model is sensible transaction between realism and simplicity[25].

When we come to computational modeling it's the use of computers to simulate and study complex systems using mathematics, physics and computer science. Numerous variables that characterized the system are being studied in computational model. These models allow researchers and scientists to conduct thousands of simulated experiments by computer [26]. Other authors also describe computational models as a mathematical model in a computational science that requires widespread computational resources to study the behavior of a complex system by computer simulation [27].

Most problems studied under this computation are nonlinear and intuitive analytical solutions are not readily available. As a result, experiment with the model is done by fine-tuning the parameters of the system rather than delivering mathematical analytical solution. Also, by studying the difference in the outcome of the experiments. Operation theories of the model can be derived from these computational experiments [27].

Models can be used to predict/forecast, explain/explore of future scenarios, understanding theory of design, illustrate or visualize and analogy [28].

As expertise in simulation recommend, we can increase the difficulty of a model iteratively. Model validity is very important. There are different model validation techniques which includes comparing model outputs with known inputs. Mostly, a model planned for a simulation study is a mathematical model developed with the help of simulation software [25].

2.4.2. What is simulation?

Simulating a system means operating the model of the system. Models can be reconfigured and experimented with; frequently, this is impossible, too expensive or impractical to do in the system it represents [25].

Simulation is a means to evaluate the performance of an existing proposed configuration of a given system. Operation of the model can be studied, and hence, properties concerning the behavior of the actual system or its subsystem can be inferred [25].

2.5. Modeling with differential equation

2.5.1. Differential equation

It's a known fact that Differential equation is an equation that involves an unknown function $y = f(x)$ and one or more of its derivatives. A solution to a differential equation is a function $y = f(x)$ that satisfies the differential equation when f and its derivatives are substituted into the equation [29]. Differential equation is a vital tool in a wide range of applications. The reason for this is that many phenomena can be modeled by a relationship between a function and its derivatives [30].

There are quite a few types of differential equations. These are ordinary differential equation, partial differential equation, linear differential equation, non-linear differential equation, homogenous differential equation, and non-homogenous differential equation.

We will use ODE in this thesis because

- ✓ It involves one independent variable,
- ✓ to compute good approximation to the solution through numerical algorithm, and
- ✓ There are several mathematical software to solve ODE.

2.5.2. Ordinary differential equation

Among one of the differential equation's ordinary differential equation (also abbreviated as ODE), in Mathematics, is an equation which contains more than one functions of one independent variable along with their derivatives as explained earlier. A differential equation is an equation that encompasses a function with one or more derivatives. But in the case of ODE, the word ordinary is used for derivative of the functions for the single independent variable [31]. In mathematics, ODE is an equation that contains only one independent variable and one or more of its derivatives with respect to the variable. In other words, the ODE is represented as the relation having one independent variable x , the real dependent variable y , with some of its derivatives [31].

2.5.3. Modeling with differential equation

In the scientific world when we want build scientific knowledge, we start by exploring the truth through observations, collecting data, and finding rules inside or among them. These rules are mathematical based that's why they are called mathematical models and these models are ordinary differential equations which also describes the relations between variables and their derivatives [32].

Modelling with differential equations is popular because equations are solved effectively. In occasional times it's possible to find an explicit formula for the unknown functions. But mostly it is likely to compute good approximations to the solution through numerical algorithms [30].

Steps for modeling with differential equation is given below [29]

1. Identify the problems variables: What are the dependent and independent variables we need to include?
2. Construct the differential equation: What is the rate of change of the dependent variable given by the problem?
3. Solve the differential equation: Use one of the standard methods to find the solution.
4. Interpret the mathematical solution: Does the solution make sense in relation to the original problem? Are there any particularly interesting elements of the solution?

In this research we used ordinary differential equation for modeling. It explains the relation between variables and their derivatives. Its solution is not time consuming. There are several mathematical software which can be used to solve ODEs.

Interpolation is the process of deriving a simple function from a set of discrete data points. This function or equation is usually ordinary differential equation.

2.6. Numerical modeling

2.6.1. Interpolation

Interpolation is a method of building new data points within the range of a discrete set of known data points in the mathematical field of numerical analysis. We have a number of data points, obtained by sampling or experimentation, which represent the values of a function for a limited

number of values of the independent variable. It is often required to interpolate, which is to estimate the value of that function for an intermediate value of the independent variable [33].

A closely related problem is the approximation of a complicated function by a simple function. Suppose the formula for some given function is known, but too complicated to evaluate efficiently. A few data points from the original function can be interpolated to produce a simpler function which is still fairly close to the original. The resulting gain in simplicity may outweigh the loss from interpolation error [33].

Basically Interpolation is described as the method of deriving a simple function from a set of discrete data points so that the function passes through all the given data points or else it reproduces the data points exactly and it can be used to estimate data points in-between the given ones. It plays a great role in dealing with discrete experimental data's in engineering and science. Mainly we use interpolation to simplify complicated functions by sampling method. Since polynomials are easy to evaluate, differentiate and integrate. We usually use them for interpolations known as polynomial interpolation. It can be proven that for a given $n+1$ data points it is always possible to find a polynomial of order/degree n to pass through/reproduce the $n+1$ point [34]. Other authors also describe interpolation as the process of finding a value between two points on a line or curve. In easy words 'inter,' as meaning 'enter,' which recaps us to look 'inside' the data we originally had. Interpolation, is not only beneficial in statistics, but is also very important and useful in science, business and in all fields that need to predict values that fall within two existing data points [35].

2.6.2. Newton divided difference

Rather than solving the interpolation problems all at once, considering solving it for smaller sets of support points first and then update these solutions to obtain the solution to the full interpolation problem is the best alternative. This primes to the idea of “divided differences”. Which is fundamentally a way of writing the interpolating polynomial in Newton form [36].

Let $p_n(x)$ denote the interpolating polynomial of degree n (or less) that interpolates $f(x)$ at x_0, \dots, x_n . Suppose we write the polynomial as

$$p_n(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \dots + a_n(x - x_0)(x - x_1)\dots(x - x_{n-1}) \quad (1)$$

Now you can notice that the last term on the right vanishes at x_0, \dots, x_{n-1} . Therefore, the previous n terms must comprise the interpolating polynomial of degree $n-1$ (or less) that agrees with $f(x)$ at x_0, \dots, x_{n-1} . In other words,

$$p_{n-1}(x) = a_0 + a_1(x - x_0) + \dots + a_{n-1}(x - x_0)(x - x_1) \dots (x - x_{n-2}) \quad (2)$$

with same $\{a_i\}$ coefficients as for $p_n(x)$. So, we can build $p_n(x)$ from $p_{n-1}(x)$, which we can build from $p_{n-2}(x)$, and so on.

Note that the evaluation of (2) for $x = \xi$ may be done recursively using Horner scheme:

$$p_n(\xi) = (\dots(a_n(\xi - x_{n-1}) + a_{n-1})(\xi - x_{n-2}) + \dots a_1)(\xi - x_0) + a_0 \quad (3)$$

While this requires n multiplications and $2n$ additions, assuming that we already have the coefficients a_0, a_1, \dots, a_n . Alternatively, we may evaluate each of the individual terms of (2) from left to right; with $2n-1$ multiplications and $2n$ additions we thereby calculate all of the values $p_0(\xi), p_1(\xi), \dots, p_n(\xi)$. This indicates whether or not an interpolation process is converging [25].

It remains to determine the coefficients a_i . In principle, they can be calculated successively from

$$\begin{aligned} f_0 &= p_n(x_0) = a_0 \\ f_1 &= p_n(x_1) = a_0 + a_1(x_1 - x_0), \\ f_2 &= p_n(x_2) = a_0 + a_1(x_1 - x_0) + a_2(x_2 - x_0)(x_2 - x_1), \end{aligned} \quad (4)$$

We can do this with n divisions and n^2 multiplications. There is, however, a better way, which requires only $n(n+1)/2$ divisions and produces useful intermediate results [25].

Let us define $f[x_i, x_{i+1}, \dots, x_m]$ to be the leading coefficient of the polynomial of degree $m-i$ (or less) that agrees with $f(x)$ at the distinct points x_i, \dots, x_m . Then by construction, a_k the leading coefficient of $p_k(x)$, is $f[x_0, \dots, x_k]$, called the k -th divided difference of f at the points x_0, \dots, x_k , and

$$p_n(x) = f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \dots + f[x_0, x_1, \dots, x_n](x - x_0)(x - x_1) \dots (x - x_{n-1})$$

Observe that

$$\begin{aligned}
 f[x_0] &= f(x_0) \\
 f[x_0, x_1] &= \frac{f(x_0) - f(x_1)}{x_0 - x_1}
 \end{aligned}
 \tag{5}$$

More generally, we will see that

$$f[x_0, \dots, x_k] = \frac{f[x_0, \dots, x_{k-1}] - f[x_1, \dots, x_k]}{x_0 - x_k}
 \tag{6}$$

The calculation of divided differences can be expressed in the following tableau, called the divided-difference scheme:

Table 2.1: Newton divided difference

x_0	$f(x_0)$			
		$f[x_0, x_1]$		
x_1	$f(x_1)$		$f[x_0, x_1, x_2]$	
		$f[x_1, x_2]$		$f[x_0, x_1, x_2, x_3]$
	$f(x_2)$		$f[x_1, x_2, x_3]$	$f[x_0, x_1, x_2, x_3, x_4]$
x_2		$f[x_2, x_3]$		$f[x_1, x_2, x_3, x_4]$
	$f(x_3)$		$f[x_2, x_3, x_4]$	
x_3		$f[x_3, x_4]$		
	$f(x_4)$			
x_4				

2.6.3. Lagrange interpolation

Here we take a different method and assume that the interpolation polynomial is given as a linear combination of $n+1$ polynomials of degree n . Here we set the coefficients as the interpolated values, $\{f(x_j)_{j=0}^n\}$, while the unknowns are the polynomials [37]. We thus let

$$Q_n(x) = \sum_{j=0}^n f(x_j)l_j^n(x), \quad 7$$

where $l_j^n(x)$ are $n+1$ polynomials of degree $\leq n$. We use two indices in these polynomials: the subscript j enumerates $l_j^n(x)$ from 0 to n and the superscript n is used to remind us that the degree of $l_j^n(x)$ is n . Note that in this particular case, the polynomials $l_j^n(x)$ are precisely of degree n (and not $\leq n$). However, $Q_n(x)$, given by (7) may have a lower degree [26]. In either case, the degree of $Q_n(x)$ is n at most. We now require that $Q_n(x)$ satisfies the interpolation conditions

$$Q_n(x_i) = f(x_i), \quad 0 \leq i \leq n \quad 8$$

By substituting x_i for x in (8) we have

$$Q_n(x_i) = \sum_{j=0}^n f(x_j)l_j^n(x_i), \quad 0 \leq i \leq n.$$

In view of (8) we may conclude that $l_j^n(x)$ must satisfy

$$l_j^n(x_i) = \delta_{ij}, \quad i, j = 0, \dots, n, \quad 9$$

where δ_{ij} is the Kronecker delta, defined as

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases}$$

Each polynomial $l_j^n(x)$ has $n+1$ unknown coefficients. The conditions (9) provide exactly $n+1$ equations that the polynomials $l_j^n(x)$ must satisfy and these equations can be solved in order to determine all $l_j^n(x)$'s. fortunately there is a shortcut [26]. An obvious way of constructing polynomials $l_j^n(x)$ of degree n that satisfy (9) is the following:

$$l_j^n(x) = \frac{(x-x_0)\cdots(x-x_{j-1})(x-x_{j+1})\cdots(x-x_n)}{(x_j-x_0)\cdots(x_j-x_{j-1})(x_j-x_{j+1})\cdots(x_j-x_n)}, \quad 0 \leq j \leq n. \quad (10)$$

The uniqueness of the interpolating polynomial of degree $\leq n$ given $n+1$ distinct interpolation points suggests that the polynomials $l_j^n(x)$ given by (9) are the only polynomials of degree $\leq n$ that satisfy (9). Here the denominator in (10) does not vanish since we assume that all interpolation points are distinct [37].

In the Lagrange form of the interpolation polynomial is the polynomial $Q_n(x)$ given by (8), where the polynomials $l_j^n(x)$ of degree $\leq n$ are given by (10). A compact form of rewriting (10) using the product notation is [37]

$$l_j^n(x) = \frac{\prod_{\substack{i=0 \\ i \neq j}}^n (x-x_i)}{\prod_{\substack{i=0 \\ i \neq j}}^n (x_j-x_i)}, \quad j=0, \dots, n. \quad 11$$

2.7. Numerical methods for ordinary differential equations

Most of the mathematicians have studied the nature of differential equations for hundreds of years and there are many well-developed solution techniques. All systems undergoing change can describe nearly as Differential equations. They are universal in science and engineering as well as economics, social science, biology, business, health care, etc. Sometimes systems labeled by differential equations are so complex or large that pure analytical solution to the equations is not tractable. That's why we use computer simulations and numerical methods for complex systems [38].

Methods used to find numerical approximations to the solutions of ordinary differential equations (ODEs) are called Numerical methods for ordinary differential equations. Their use is also known as "numerical integration", although known as computation of integrals [39].

2.7.1. Euler’s Method

Here we consider that the differential equation is [30]

$$y' = f(x, y), y(a) = y_0$$

and our major aim is to compute a sequence of approximations $(x_k, y_k)_{k=0}^n$, to the solution, where $x_k = a + kh$. The initial condition provides us with a point on the true solution, so (x_0, y_0) is also the natural starting point for the approximation. To obtain an approximation to the solution at x_1 , we compute the slope of the tangent at (x_0, y_0) as $y' = f(x_0, y_0)$. This gives us the tangent $T_0(x_0) = y_0 + (x - x_0)y_0'$ to the solution at x_0 . As the approximation y_1 at x_1 we use the value of the tangent T_0 which is given by [30]

$$y_1 = T_0(x_1) = y_0 + hy_0' = y_0 + hf(x_0, y_0) \tag{12}$$

This gives us the next approximate solution point (x_1, y_1) . To advance to the next point (x_2, y_2) , we move along the tangent to the exact solution that passes through (x_1, y_1) [30].

The derivative at this point is $y_1' = f(x_1, y_1)$ and so the tangent is

$$T_0(x) = y_1 + (x - x_1)y_1' = y_1 + (x - x_1)f(x_1, y_1) \tag{13}$$

The approximate solution at x_2 is therefore

$$y_2 = y_1 + hf(x_1, y_1) \tag{14}$$

If we continue in the same way, we can compute an approximation y_2 to the solution at x_3 , then an approximation y_4 at x_4 , and so on. From this description we see that the crucial idea is how to advance the approximate solution from a point (x_k, y_k) to a point (x_{k+1}, y_{k+1}) [30].

Geometric Interpretation of Euler Method

This process is shown graphically below for a single iteration of the process

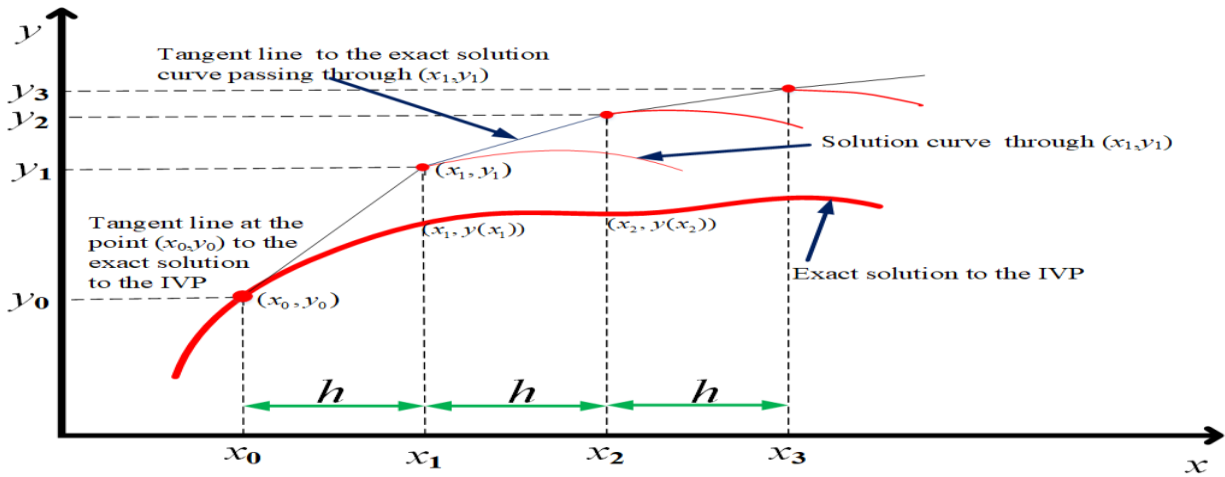


Figure 2.1: Euler method

2.7.2. Runge–Kutta method

The Runge–Kutta methods were developed around 1900 by the German mathematicians C. Runge (1856–1927) and M.W. Kutta (1867–1944). are an important family of iterative methods for the approximation of solutions of ODE’s We start with the consideration of the explicit methods [40]. The Runge Kutta method (as it is usually called) enjoys two conspicuous advantage:

- (1) No special devices are required for starting the computation.
- (2) The length of the step can be modified at any time in the course of the computation without additional labor [41].

The **Runge-Kutta method** of order $N = 4$ is most popular. It is a good choice for common purposes because it is quite accurate, and easy to program. Most authorities proclaim that it is not necessary to go to a higher-order method because the increased accuracy is offset by additional computational effort. If more accuracy is required, then either a smaller step size or an adaptive method should be used.

Fourth Order Runge-Kutta method

The **Runge-Kutta method** of fourth order (also known as *classical Runge-Kutta method*) gives greater accuracy and is most widely used for finding the approximate solution of first order ordinary differential equations. In this method, in each step first it computes four auxiliary

quantities k_1, k_2, k_3, k_4 and then the new value y_{n+1} . The method is well suited to the computer because it needs no special starting procedure, makes light demand on storage, and repeatedly uses the same straightforward computational procedure. It is numerically stable.

A particularly popular example from this family is:

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \tag{15}$$

Where

Geometrically

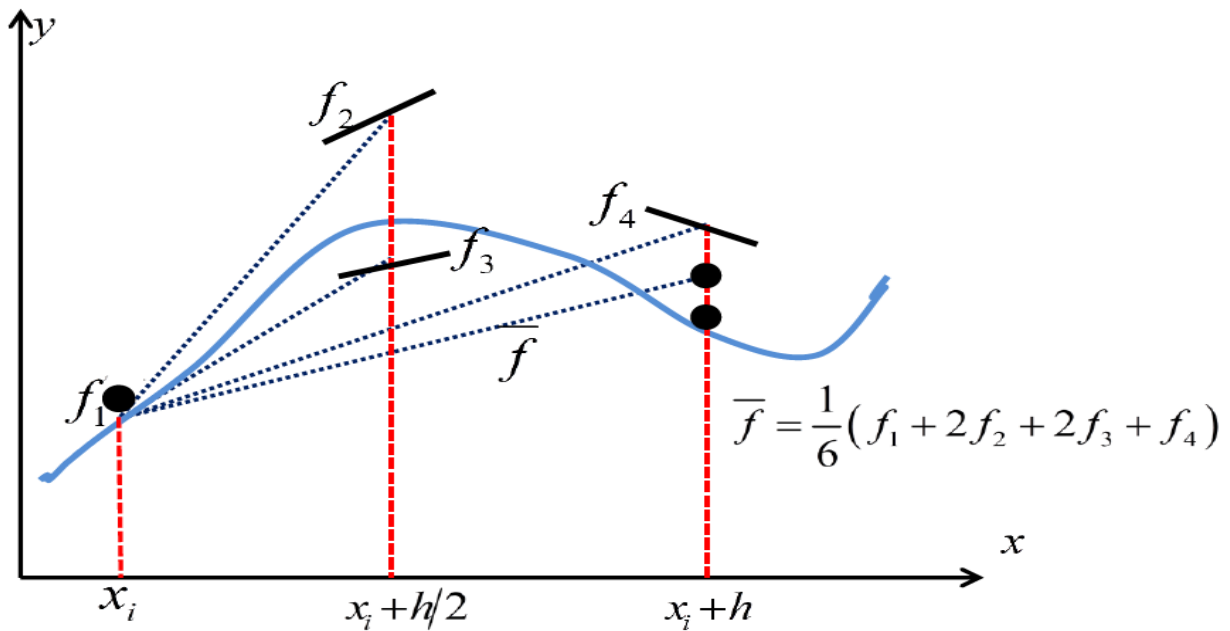


Figure 2.2: Runge-Kutta method

2.8. Data analytics

The extraction of information from large datasets, also analyzing and solving problems by modeling, simulation, and optimization, removing the computational skills that make solving the complex problems through modelling and visualization are the main focuses of computational modeling and data analytics [44].

As explained data analytics is a very vast term which consists many diverse techniques and processes of drawing insights from historical data over time. Mostly data analytics is used to designate the analysis of large volumes of data and/or high-velocity data, which presents unique computational and data-handling challenges [45]. The discipline of examining raw data with the process of drawing conclusion about that information is also called Data analytics (DA). DA is applied to make better business decisions in many companies and industries using the science and art to verify existing theories and models [46]. Now the most significant role played by covering every industry in the market like finance, business, ecommerce, government, and education is by data analysis. Currently most organizations are playing a 360-degree role in analyzing the behavior and interest of their customers to take decisions in favor of them [47].

2.8.1. Data analysis with python

Data analysis is done using programming languages like python, one of the most versatile programming languages which allows to do enormous things [47].

Although there are many more domain-specific open source and commercial programming languages and tools in wide use, such as R, MATLAB, SAS, Stata, and others. Python will certainly be the best for data analysis interactive, exploratory computing and data visualizations. Through years python's improved library support (mainly pandas) has become very good choice for data manipulation tasks. therefore python has become preferable for a data centric applications and analysis [48]. Python is the preferred language for data analytics due to this conditions [49].

- Unlike other languages like R, Python is very easy to learn the programming language and achieve results in fewer lines of code. This makes it strong to handle complex scenarios with minimal code which decreases confusion on the flow of the program.
- It can be perfectly used in a multi-environment setup easily since it's a cross platform code works in multiple environments.

- Python execution time is faster compared to R and MATLAB [49].
- Python can manage very large volume data transformation, slicing, Dicing and visualization gracefully since it has excellent memory management capability
- Python has got a very large collection of libraries which serve as special purpose analysis tools. For instance, the NumPy package deals with scientific computing and its array needs much less memory than the conventional python list for managing numeric data. And the number of such packages is continuously growing.
- Whenever it gives a better result, Python has packages which can be used directly to use the code from other languages like Java or C just to optimize the code performance by using existing code of other languages.

2.9. Summary

In the last decades road crashes have been in an increasing trend. As road crashes have become a major problem which led researchers to find possible causes and precaution measures to prevent crashes from happening which guided researchers to develop and discover new models that predicts road crashes precisely [9].

Several researchers have made different researches using different methods using a generalized poisson regression model, Fridstrom measured the contribution of randomness, exposure, weather and daylight to the variation in road accident counts. The difference in accident counts in 4 different European countries were calculated into parts attributable to randomness, exposure, weather, daylight, or changing reporting routines and speed limits. Set of specialized goodness-of-fit measures were also developed in small accident counts (fatal accidents per country per month) pure randomness was seen to explain the major variations while exposure was the dominant systematic determinant. Connections between exposure and injury accidents seems to be almost relative but not proportional to fatal accidents or death victims. Randomness and exposure Collectively hold for 80% to 90% of the visible variations in the datasets [10]. Other researchers (Graham and Glaister) inspected the part of urban scale, density and land-use mix on the incidence of road pedestrian casualties. The researchers carried out analysis using census of English districts as the spatial unit of the study and developed negative binomial models. Their study concludes in residential areas the prevalence of pedestrian casualties and Killed and

Seriously Injured (KSI) were higher than that of in business areas. Also, relationship between urban density and pedestrian casualties were found to be in quadratic form compared to incidents that is reduced in highly populated districts [11]. Spatial analysis of fatal and injury were carried out in Pennsylvania and Full Bayesian (FB) and negative binomial models were developed by Agüero-Valverde and Jovanis using counties as the spatial unit. Their study concluded that there is a significant increase in crash risk that counties with a higher percentage of the population under poverty level, higher percentage of their population in age groups 0-14, 15-24 and over 64, and increased road mileage and road density. They also suggested that it was important to consider spatial correlation in road-segment and intersection-level accident models from their study[12]. On the urban roads of Aman Jordan road traffic accidents at 28 hazardous locations were studied by Obaidat and Ramadan. By these study logarithmic and linear models were the most significant and realistic models were found that can be used to predict the relationship between the accident characteristic's as a dependent variable and the other studied variables as independent variables. The succeeding variables are found to be the most important contributors to the traffic accidents at hazardous locations average running speed, posted speed, maximum and average degree of horizontal curves (horizontal curve provides a transition between two tangent strips of roadway, allowing a vehicle to negotiate a turn at a gradual rate rather than a sharp cut), number of vertical curves (vertical curves are used in highway and street vertical alignment to provide a gradual change between two adjacent grade lines), median width, type of road surface, lighting (day or night), number of vehicles per hour, number of pedestrian crossing facilities and percentage of trucks [50]. Based on the study and analysis done by Shanker frequency of occurrence of highway accidents on the basis of a multivariate analysis of roadway geometry (horizontal and vertical alignments), weather and other seasonal effects was explored. A negative binomial model of overall accident frequencies was predicted along with models of the frequency of specific accident types based on the accident data collected from the field. As part of the model specifications interactions between weather and geometric variables were proposed. The results of the analysis uncover important determinants of accident frequency [51].

Melaku Tadege applied retrospective research design. 255 records were taken from Finote Selam traffic police office, northwest part of Ethiopia from September 2009 to January 2018 [62]. The statistical analysis was performed by using SPSS version 23 software. Chi-square for association

test and ordinal logistic regression for predictor identification were used. Age of drivers were the responsible causes of fatal road traffic accident. The more experienced drivers decreased the occurrence of fatal traffic accidents. In addition, increasing vehicle service year reduced the occurrence of accidental death. Besides, the occurrence of fatal car accident in autumn season was 0.44 times less than that of in summer season. Additionally, drivers' educational level was played a crucial role in a road traffic accident. For instance, drivers whose educational level was below 12th grade were the most responsible factor for car accident deaths. What is more, it was seen that drivers who drove their vehicles could minimize the occurrence of fatal traffic accident [62]. In Addis – Adama expressway a model was developed on the traffic crash data. The research used Multinomial Probit Regression model to assess the determinants of accident on Addis Ababa – Adama expressway. This means, empirical regression method is used to identify agents contributing for road traffic accident [6].

As we have seen from the above models, all of them used statistical methodologies. They commonly used regression techniques (linear, logistic, multiple). These techniques are old and conventional. In this research we used ODE to model the crash data using numerical modeling. Most of the studies we have seen earlier did not study the relationship between traffic crash and number of vehicles passed and the relationship between traffic crash and vehicle types. But in this study we focused on these ideas.

Chapter three

3. Methodology of the research

This chapter includes the description of the study area, the data collection method, the modelling methods (i.e. algorithms of the methods), analysis methods (i.e. the modules), and software used.

In order to model the traffic flows data, we will use interpolation. Newton divided difference and Lagrange are the methods of interpolation we will use. After modeling the data with ordinary differential equation, we will solve it using Euler's method and runge kutta method. We will analyze the data using one of data analytics tool which is called Python. We will use its libraries for good analysis.

3.1. Data source and Method of collection

Ethiopian toll road enterprise has been recording traffic flow data from customers that use the expressway since 2015. The data they have been collecting comprises of both qualitative and quantitative data. The organization is one of its kind to introduce smart traffic system or intelligent transportation system in Ethiopia. The system includes traffic cameras and variable message signs (VMS) to collect data from customers. The enterprise collects daily data using this system and in this study this data is used for both modeling and analysis.

3.2. Modeling methods

Modeling is a process or cognitive activity in which we simulate how objects of interest or device will behave [52]. For this specific study interpolation is utilized to create model. Newton divided difference and Lagrange are the type of interpolations that are used. In the next section the algorithms of both methods we will use are presented.

3.2.1. Newton divided difference

Newton's divided difference is an interpolation technique used when the interval difference is not same for all sequence of values.

Algorithm for newton divided difference

Step 1: choose the point that you want to find x

Where x is an independent variable (i.e. year)

Step 2: compute

$$\begin{aligned}
 b_0 &= f(x_0) = f[x_0] \\
 b_1 &= \frac{f(x_1) - f(x_0)}{x_1 - x_0} = f[x_1, x_0] \\
 &\quad \frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0} \\
 b_2 &= \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0} = f[x_2, x_1, x_0] \\
 &\quad \vdots \\
 b_n &= \frac{f[x_n, x_{n-1}, \dots, x_1] - f[x_{n-1}, x_{n-2}, \dots, x_0]}{x_n - x_0}
 \end{aligned}$$

Where

$x_0, x_1, x_2, \dots, x_n$ are independent variables of the known data values i.e. the number of traffic crash and traffic volume in the express way

$f(x_0), f(x_1), f(x_2), \dots$ are the dependent variables i.e. the known data values

Step 3: compute

$$f_n(x) = b_0 + b_1(x - x_0) + b_2(x - x_0)(x - x_1) + \dots + b_n(x - x_0)(x - x_1)\dots(x - x_{n-1})$$

Where

x is the independent variable of the unknown data value (i.e. the data value that we are finding)

$f_n(x)$ is the dependent variable (i.e. the value we are interpolating)

3.2.2. Lagrange interpolation

The Lagrange interpolating polynomial is a tool which helps us construct a polynomial which goes through any desired set of points.

Algorithm for Lagrange method

Step 1: choose the point that you want to find x

Where x is an independent variable (i.e. year)

Step 2: compute

$$P(x) = \sum_{j=1}^n P_j(x)$$

$$P_j(x) = f(x_j) \prod_{\substack{k=1 \\ k \neq j}}^n \frac{x - x_k}{x_j - x_k}$$

$$P(x) = f(x_1) \frac{(x - x_2)(x - x_3) \cdots (x - x_n)}{(x_1 - x_2)(x_1 - x_3) \cdots (x_1 - x_n)} + \dots$$

$$+ f(x_2) \frac{(x - x_1)(x - x_3) \cdots (x - x_n)}{(x_2 - x_1)(x_2 - x_3) \cdots (x_2 - x_n)} + \dots$$

$$+ f(x_n) \frac{(x - x_1)(x - x_3) \cdots (x - x_{n-1})}{(x_n - x_1)(x_n - x_3) \cdots (x_n - x_{n-1})}$$

Where

x is an independent variable (i.e. year) of the unknown data value

x_1, x_2, \dots, x_n are independent variables of the known data values

y_1, y_2, \dots, y_n are dependent variables and known data values (the number of traffic flows)

$P(x)$ is the unknown data value i.e. the value we are interpolating

After modeling the data using the above two methods with ordinary differential equation then we will solve it using Euler method and Runge – Kutta method. The algorithms of both methods are presented below.

3.2.3. Euler method

The Euler method (also called forward Euler method) ,in the mathematics and computational science, it is the first-order numerical procedure for solving ordinary differential equations (ODEs) with a given initial value. This method is basic explicit method for numerical integration of ordinary differential equations and is the simplest Runge–Kutta method.

Algorithm for Euler’s method

Let the differential equation $y' = f(x, y)$ be given together with the initial condition $y(x_0) = y_0$, the input value for h , and the number of steps n .

Step 1: set the given initial condition for x_0 and y_0

Step 2: request and get input values for h and for n

Step 3: for $k = 0, 1, \dots, n - 1$

Step 4: calculate the derivative of $y(x)$ at $x = x_0$

Step 5: calculate an approximate value for $y(x_0 + h)$

$$y_{k+1} = y_k + hf(x_k, y_k)$$

Step 6: calculate for the next step size

$$x_{k+1} = x_k + h$$

Step 7: repeat step 3 through 5 until the solution is finished

Where

x_0 is independent variable (i.e. year) and the year we want to give as initial condition to solve the ODE

y_0 is dependent variable and the data value of the year we will give as initial condition

h is incremental change in the x-coordinate

3.2.4. Runge – kutta method

Runge-Kutta method is a numerical technique used to solve ordinary differential equation of the form

$$\frac{dy}{dx} = f(x, y), y(0) = y_0$$

Algorithm for Runge – Kutta

Let the differential equation $y' = f(x, y)$ be given together with the initial condition $y(x_0) = y_0$, the input value for h , and the number of steps n .

Step 1: set the given initial condition for x_0 and y_0

Step 2: request and get input values for h and for n

Step 3: calculate the derivative of $y(x)$ at $x = x_0$

Step 4: compute

$$\begin{aligned} k_1 &= h \times f(x_i, y_i) \\ k_2 &= h \times f\left(x_i + \frac{h}{2}, y_i + \frac{k_1}{2}\right) \\ k_3 &= h \times f\left(x_i + \frac{h}{2}, y_i + \frac{k_2}{2}\right) \\ k_4 &= h \times f(x_i + h, y_i + k_3) \end{aligned}$$

Step 5: based on step 4 compute

$$y_{i+1} = y_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)h$$

Step 6: calculate for the next step size

$$x_{k+1} = x_k + h$$

Step 7: repeat step 3 through 5 until the solution is finished

Where

x_0 is independent variable (i.e. year) and the year we want to give as initial condition to solve the ODE

y_0 is dependent variable and the data value of the year we will give as initial condition

k_1 is the increment based on the slope at the beginning of the interval, using y

k_2 is the increment based on the slope at the midpoint of the interval, using $y + hk_1 / 2$

k_3 is again the increment based on the slope at the midpoint, using $y + hk_2 / 2$

k_4 is the increment based on the slope at the end of the interval, using $y + hk_3$

3.3. Analysis method

Data modeling is a technique of analysis that helps as canopy term for engineering metrics and gives additional insights for value, direction, and context by using a mix of strategies and techniques. In order to succeed having a data modeling mindset is essential by considering how to analyze data, as such, it's an area that is worth exploring in greater detail [53]. In this research we will model traffic crash and number of vehicles data and we will observe their relationship. We will also observe the relationship between vehicle types and traffic crash.

3.3.1. Analysis tool

Data analysis is used in order to find useful information from data to make rational decisions. In this research we will use python for the analysis. Python has diverse visualization options which makes it preferable. And it also have long list of totally free libraries.

Python and its libraries

As we discussed earlier we will use python and its libraries for analysis and it has its own libraries. The libraries which we will use are given below.

1. Pandas

Pandas stand for *Python Data Analysis Library*. It is an open-source Python package that delivers high-performance, user friendly data structures and data analysis tools for the labeled

data in Python programming language. Pandas take data in a CSV or TSV file or a SQL database and create a Python object with rows and columns called a data frame [54].

2. NumPy

The other library mostly used for numerical analysis is NumPy which is an efficient container of generic multi-dimensional data. NumPy is a general-purpose array-processing package. This package provides high-performance multidimensional array objects and tools to work with the arrays [54].

3. Matplotlib

The object-oriented API Matplotlib is the plotting library for Python that provides API for embedding plots into applications. It is a close resemblance to MATLAB embedded in Python programming language [54].

4. Seaborn

The python library seaborn is an extension of Matplotlib with advanced features. It is the data visualization library based on Matplotlib that provides a high-level interface for drawing attractive and informative statistical graphics [54].

3.4. Software used in the research

3.4.1. Matlab

MATLAB its name stands for matrix laboratory. MATLAB was initially written to provide easy access to matrix software developed by the LINPACK and EISPACK projects, which together characterize the state-of-the-art in software for matrix computation [55].

The MATLAB language.

The MATLAB language is a high-level matrix/array language with control flow statements, functions, data structures, input/output, and object-oriented programming features. This language allows both "programming in the small" to rapidly create quick and dirty throw-away programs, and "programming in the large" to create complete large and complex application programs [55].

The MATLAB working environment.

The MATLAB working environment is the set of tools for managing the variables in your workspace and importing and exporting data. It also contains tools for developing, managing, debugging, and profiling M-files, MATLAB's applications [55].

3.4.2. Anaconda

Anaconda is a free and open-source distribution for scientific computing (data science, machine learning applications, large scale data processing, predictive analytics, etc.)The distribution includes data-science packages suitable for Windows, Linux, and macOS. Anaconda uses Python and R programming languages that aims to simplify package management and deployment[56].

The Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda distribution that permits users to launch applications and manage conda packages, environments and channels without using command-line commands. This applications are available by default in Navigator: JupyterLab, Jupyter Notebook, QtConsole, Spyder, Glue, Orange, RStudio, and Visual Studio Code [56].

Jupyter Notebook

Jupyter Notebook which is an open-source web application. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the IPython kernel, which allows you to write your programs in Python, this notebook can be used to create and share documents that contain live code, equations, visualizations, and text Jupyter Notebooks are a spin-off project from the IPython project, which used to have an IPython Notebook project itself. But there are currently over 100 other kernels that you can also use [57].

Chapter four

4. Thesis design framework

In this chapter, we will discuss thesis design frameworks. The components of the frameworks will be presented. In the modelling framework, the data collection and preparation will be explained in the first sections. In the next section, the modelling process will be explained and solving ODE will be discussed later. Evaluation will be presented at the last section. In the analysis framework data collection and preparation will be presented in the first sections. Analysis and presentation will be discussed later.

4.1. The frameworks

In this thesis we have two frameworks i.e. for modelling and analysis.

The modelling framework has six phases: the first phase is data collection and preparation, the second phase is data preprocessing, the third and fourth phase is identifying known data point and unknown data points, the fifth phase is solving ODE, the sixth and the last phase is evaluation of the results.

In the data collection and preparation phase, we will collect the data we will use for modelling.

In the second phase i.e. data preprocessing we will count the number of traffic crashes and traffic volumes within a year (from 2015 to 2019).

In the next phases, we will identify known data point and unknown data points.

In the modelling phase, we will model the data using interpolation methods (i.e. newton DVD and Lagrange methods). We will plot the equations we got from both methods to see how much our model is correct.

In solving phase, we will solve the ODE we got from the modelling using Euler method and Runge – Kutta method.

In the evaluation phase, we will compare the solution of Euler method and Runge – Kutta method with the exact solution to see how much they are correct.

In analysis framework, there are four phases. Data collection and preparation is the first phase. Data preprocessing is the second phase. The third phase is analysis of parameters. Presentation of results is the last phase.



Figure 4.1: Thesis design framework of Modeling section

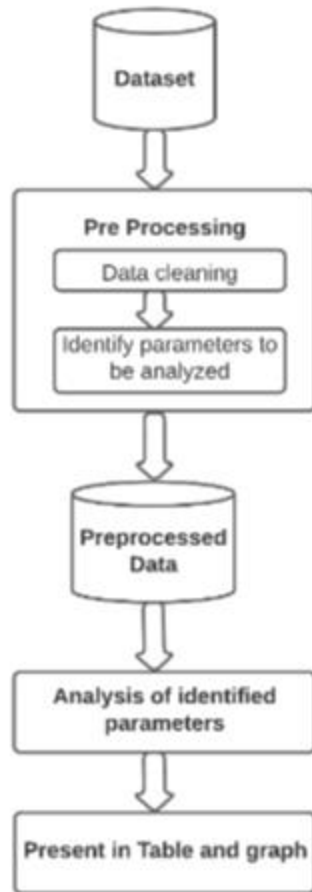


Figure 4.2: Thesis framework of Analysis section

4.2. Modelling framework

4.2.1. Data Collection and Preparation

The data was collected on daily basis. The traffic crash data was registered when a traffic crash happen with the time it happened, the vehicle type, the day it happened, and the way it was heading. The traffic volume data was registered when a vehicle pass in the express way. The vehicle type, the day it passed, the way it was heading and the time it passed are included in the data. Intelligent transportation systems (ITS) including traffic cameras and variable message signs (VMS) were used for collecting data. We divided these data into five years to make it suitable for modelling.

4.2.2. Data preprocessing

The main purpose of this phase is to make the data suitable for modelling. As we have five years data (i.e. 2015 to 2019) and the data is collected in daily basis. We need to count how many traffic crashes happened in a year. Similarly, we also count the number of vehicles passed in the express way within a year. This will give us the traffic volume.

4.2.3. Identifying data points

We have preprocessed data i.e. counted within a year. So, the traffic crash and traffic volume data are divided in five years. As we have five data points, the four data points will be the known data points and the rest one will be the unknown data point. This data point has to be in the middle as we are using interpolation.

4.2.4. Modelling the data

Using the identified data points, we now start the modelling process. We use the two interpolation methods Newton DVD and Lagrange methods.

- Newton divided difference - Instead of solving the interpolation problem all at once, one might consider solving it for smaller sets of support points first and then update these solutions to obtain the solution to the full interpolation problem. This leads to the idea of “divided differences”. It is essentially a way of writing the interpolating polynomial in Newton form [36].
- Lagrange interpolation - In numerical analysis, Lagrange polynomials are used for polynomial interpolation. For a given set of points (x_j, y_j) with no two x_j values equal,

the Lagrange polynomial is the polynomial of lowest degree that assumes at each value x_j the corresponding value y_j , so that the functions coincide at each point [58].

4.2.3. Solving ODE

We model the data using two interpolation methods i.e. Newton DVD and Lagrange interpolations with third order ordinary differential equation. Then, we solve the resulting ODEs using Euler method and fourth order Runge – Kutta method.

- Euler method - In mathematics and computational science, the Euler method (also called forward Euler method) is a first-order numerical procedure for solving ordinary differential equations (ODEs) with a given initial value. It is the most basic explicit method for numerical integration of ordinary differential equations and is the simplest Runge–Kutta method [59].
- Runge – Kutta method - a numerical method, involving successive approximations, used to solve differential equations [60]. In numerical analysis, the Runge–Kutta methods are a family of implicit and explicit iterative methods, which include the well-known routine called the Euler Method, used in temporal discretization for the approximate solutions of ordinary differential equations [61].

4.2.4. Evaluation

We solve the ordinary differential equations using Euler method and Runge – Kutta. Then, we have to evaluate how much the solution we got from the methods is correct. Therefore, we compare the solutions of Euler method and Runge – Kutta method with the exact solution. We got the exact solutions using the derivatives of the equations. To see the difference between the exact solution and the solutions of Euler method and Runge – Kutta method very clearly, we plot both solutions in one graph using Matlab.

4.3. Analysis Framework

4.3.1. Data collection and preparation

The data was collected on daily basis. The traffic crash data was registered when a traffic crash happen with the time it happened, the vehicle type, the day it happened, and the way it was heading.

4.3.2. Data preprocessing

The data are recorded by different people, as a result there are differences in the record. In this phase, we clean the data to correct these differences and make it suitable for analysis. We select or identify parameters that are relevant to our work.

4.3.3. Analysis

In this phase, we have the preprocessed data. We selected vehicle type, day, and direction parameters. These are the parameters we want to focus on this research. We analyze these parameters using python to observe the level of their effect on traffic crash.

4.3.4. Presentation

We present the analysis using table and bar graph to make it very clear and easily understandable.

Chapter five

5. Results and Discussion

In this chapter, the results we got from the interpolation methods i.e. Newton DVD and Lagrange interpolations for both traffic crash and traffic volume are presented. The solutions of Euler method and Runge – Kutta method for both traffic crash and traffic volume are included. The analysis of the selected parameters are presented. Finally, we putted the discussion section.

5.1. Dataset used

The dataset we used for modelling is by counting the number of traffic crashes occurred within a year. So, we classified the data in five parts. And it is presented below in table 5.1. Similarly, for modelling traffic volumes we counted the number of vehicles passed in the expressway within a year. It is presented in table 5.2. For analyzing the traffic crash data, we cleaned some recording errors in the selected parameters columns to avoid mistake in the result.

Table 5.1: Number of traffic crashes per year

Year	i	No. of traffic crash
2015	1	164
2016	2	307
2017	3	359
2018	4	391
2019	5	344

Table 5.2: Number vehicle passed per year

Year	i	No. of vehicles passed
2015	1	4,211,044
2016	2	5,631,228
2017	3	7,338,181
2018	4	7,976,106
2019	5	8,119,378

5.2. Description of the study area

The Addis Ababa–Adama Expressway is a toll road that connects Addis Ababa to Adama. It is the first expressway in Ethiopia. The road was built between 2010 and 2014 by the China Communications Construction Company. The government opened the highway for traffic on September 14, 2014. The Government of Ethiopia covered 43 percent of the road's construction cost, while the remaining 57 percent was supplied by a loan from the Exim Bank of China [39].

The highway, with six lanes on two sides for its 84.7-kilometre (52.6 mi) length, aims to abate the heavy traffic between its two endpoints. It reduces the time required to reach Adama from Addis Ababa to 45 minutes, a reduction of more than 50 percent over the time on previously available routes [39].



Figure 5.1: Location of Addis Ababa - Adama expressway

5.3. Data modelling

The modeling was made on the gathered quantitative data (i.e. by counting the number of traffic crashes happened within a year and by counting the number of vehicles passed in the express way within a year). We used newton divided difference and Lagrange interpolation methods to model the data. We also solved the ODE equations using Euler method and Runge-Kutta method. The analysis was made on some parameters (i.e. vehicle type, year, time interval...) using analysis tool called python. Finally, the relationship between traffic crashes and number of vehicles passed or traffic volume is discussed.

5.3.1. Traffic crash modelling using newton divided difference

Before the modelling was done we counted the number of traffic crashes within a year i.e. for five years from 2015 to 2019. Then, we chose four data points that are the known values and the rest one is the unknown value. As we used one of the interpolation methods this data point has to be in the middle, so we chose the value of 2018 as unknown value. We used four data points so the resulting ODE is third order polynomial.

$$x = 2018 = 4 \quad x_0 = 1 \quad x_1 = 2 \quad x_2 = 3 \quad x_3 = 5$$

$$b_0 = f(x_0) = f(1) = 164$$

$$b_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{f(2) - f(1)}{2 - 1}$$

$$= 307 - 164 = 143$$

$$b_2 = \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_1}$$

$$= \frac{\frac{f(3) - f(2)}{3 - 2} - \frac{f(2) - f(1)}{2 - 1}}{3 - 1}$$

$$= \frac{(359 - 307) - (307 - 164)}{2} = \frac{52 - 143}{2}$$

$$= -45.5$$

$$b_3 = \frac{\frac{f(x_3) - f(x_2)}{x_3 - x_2} - \frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_3 - x_0}$$

$$= \frac{\frac{f(5) - f(3)}{5 - 3} - \frac{f(3) - f(2)}{3 - 2} - \frac{f(2) - f(1)}{2 - 1}}{5 - 1}$$

$$= \frac{\frac{344 - 359}{2} - (359 - 307) - (307 - 164)}{4}$$

$$= \frac{-7.5 - 52 - 143}{4} = -50.6$$

$$b_0 = 164 \quad b_1 = 143 \quad b_2 = -45.5 \quad b_3 = -50.6$$

$$f_3(x) = 164 + 143(x - 1) + (-45.5)(x - 1)(x - 2) + (-50.6)(x - 1)(x - 2)(x - 3)$$

The ODE we got is given below.

$$f_3x = 6.4x^3 - 83.9x^2 + 349.9x - 108.4$$

We plotted the equation and we can see the value of the unknown data point. We were also able to predict the traffic crashes that will happen in the next year. The plot is given below.

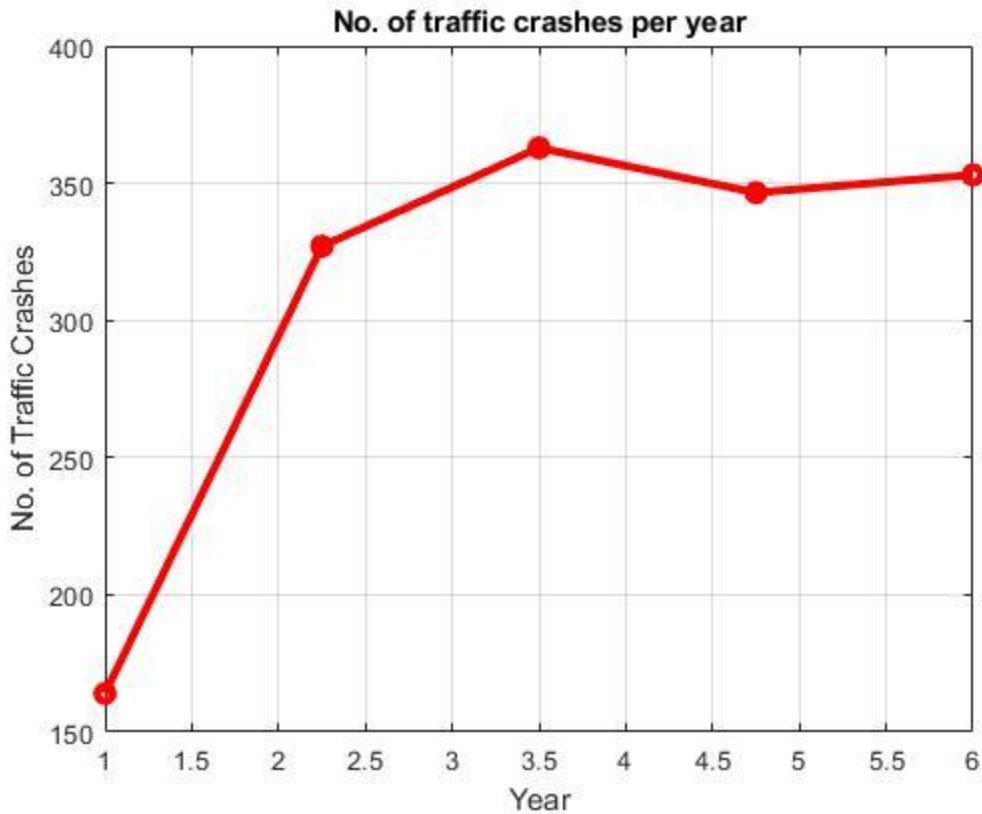


Figure 5.2: Plot of no. of traffic crashes per year using newton divided difference

5.3.2. Traffic crash modelling using Lagrange interpolation

As we did during modelling the data using newton divided difference, we counted the number of traffic crashes happened within a year. And we used the same parameters i.e. we chose 2018 to be the unknown value and the rest ones to be the known values. Similarly, we used four data points, so it is third order polynomial ODE.

$$x_1 = 1 f(x_1) = 164 \quad x_2 = 2 f(x_2) = 307 \quad x_3 = 3 f(x_3) = 359 \quad x_4 = 5 f(x_4) = 344$$

$$\begin{aligned} f(x) &= 164 \frac{(x-2)(x-3)(x-5)}{(1-2)(1-3)(1-5)} \\ &\quad + 307 \frac{(x-1)(x-3)(x-5)}{(2-1)(2-3)(2-5)} \\ &\quad + 359 \frac{(x-1)(x-2)(x-5)}{(3-1)(3-2)(3-5)} \\ &\quad + 344 \frac{(x-1)(x-2)(x-3)}{(5-1)(5-2)(5-3)} \\ &= -20.5(x^3 - 10x^2 + 31x - 30) \\ &\quad + 102.5(x^3 - 9x^2 + 23x - 15) \\ &\quad - 89.75(x^3 - 8x^2 + 17x - 10) \\ &\quad + 14.3(x^3 - 6x^2 + 11x - 6) \end{aligned}$$

The ODE we got after modeling is given below.

$$f_3x = 6.35x^3 - 83.5x^2 + 348.05x - 107.8$$

When we compare this ODE with the one we got using newton divided difference, it is almost the same. As a result, we can ensure that our modeling is correct. Then, we plot this ODE. The plot is given below.

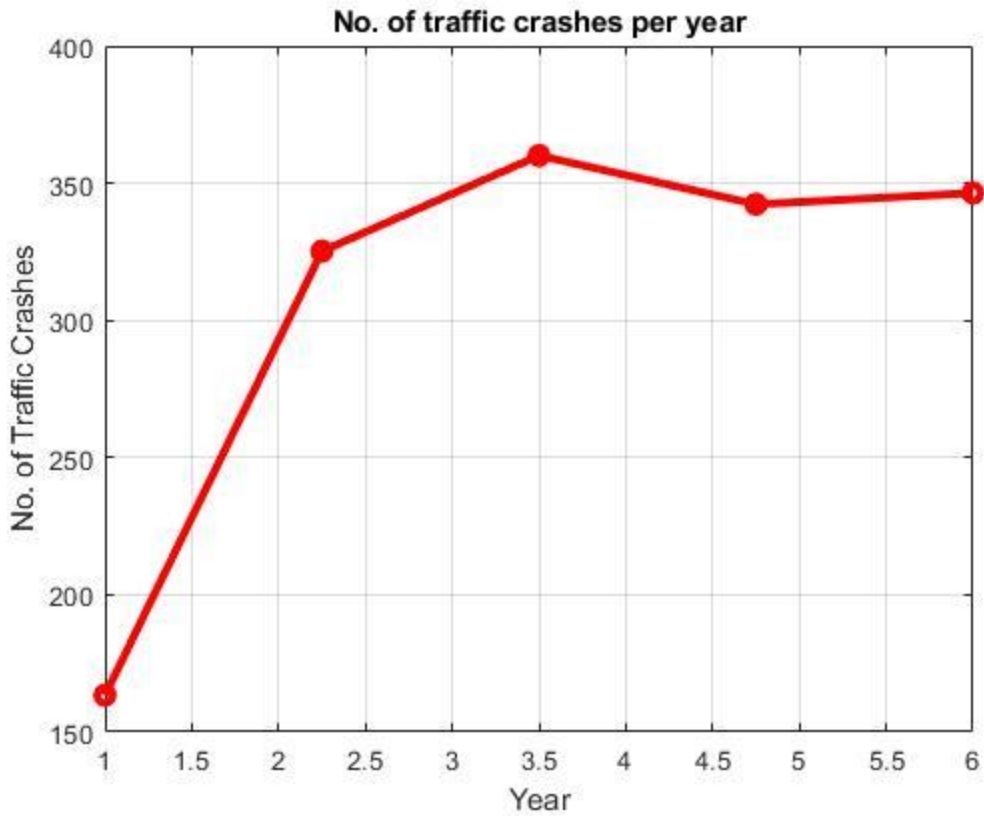


Figure 5.3: Plot of no. of traffic crashes per year using Lagrange

As we can see from Figure 5.3, we got the value of the unknown data point and we were also able to predict the number of traffic crashes that will happen in the next year.

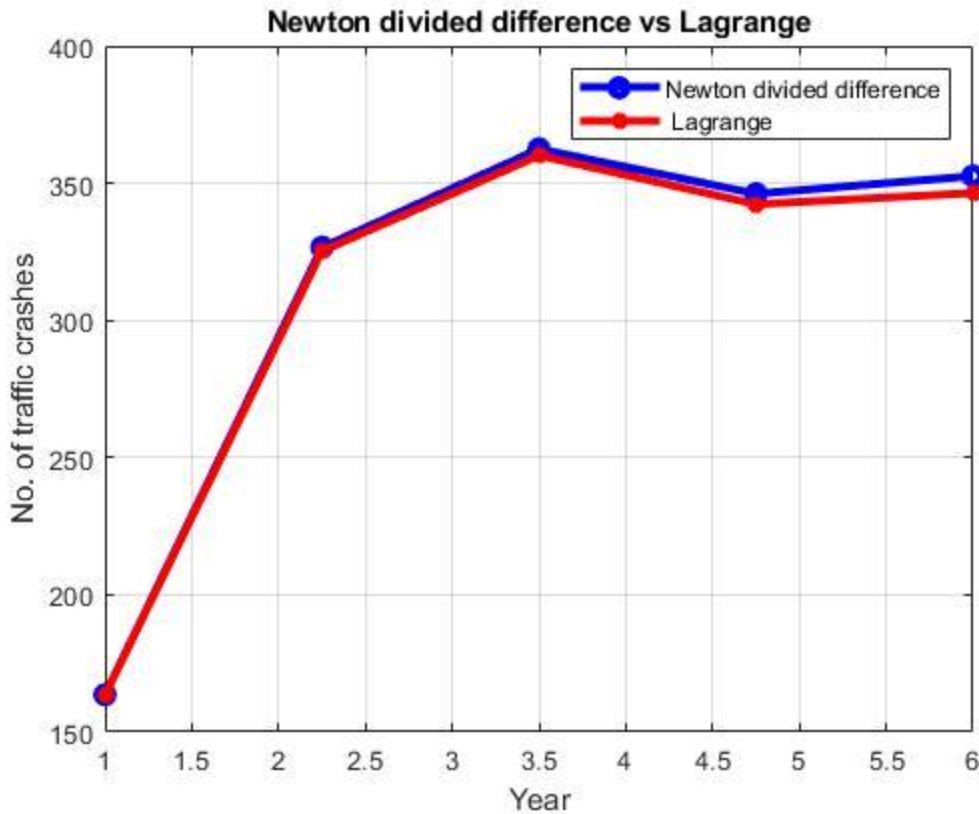


Figure 5.4: Newton divided difference vs Lagrange

We plotted both equations in one graph and we can observe with both methods we got almost same result.

5.4. Solving ordinary differential equations of traffic crashes

5.4.1. Solving ordinary differential equations using Euler method

We solved the ODE we got from the interpolation methods using Euler method. The equation, its derivative, initial value and step size h we used are given below.

$$y = 6.4x^3 - 83.9x^2 + 349.9x - 108.4$$

$$y' = 19.2x^2 - 167.8x + 349.9$$

$$y(1) = 164, h = 1$$

Table 5.3: The output of Euler method

x	Euler	Exact
1.00	164.00000	164.00000
2.00	365.30000	307.00000
3.00	456.40000	359.00000
4.00	475.70000	358.40000
5.00	461.60000	343.60000

Table 5.3 shows the output of Euler method and we can compare the output of Euler method with the exact solution. There is a big difference between them. We can see it graphically below.

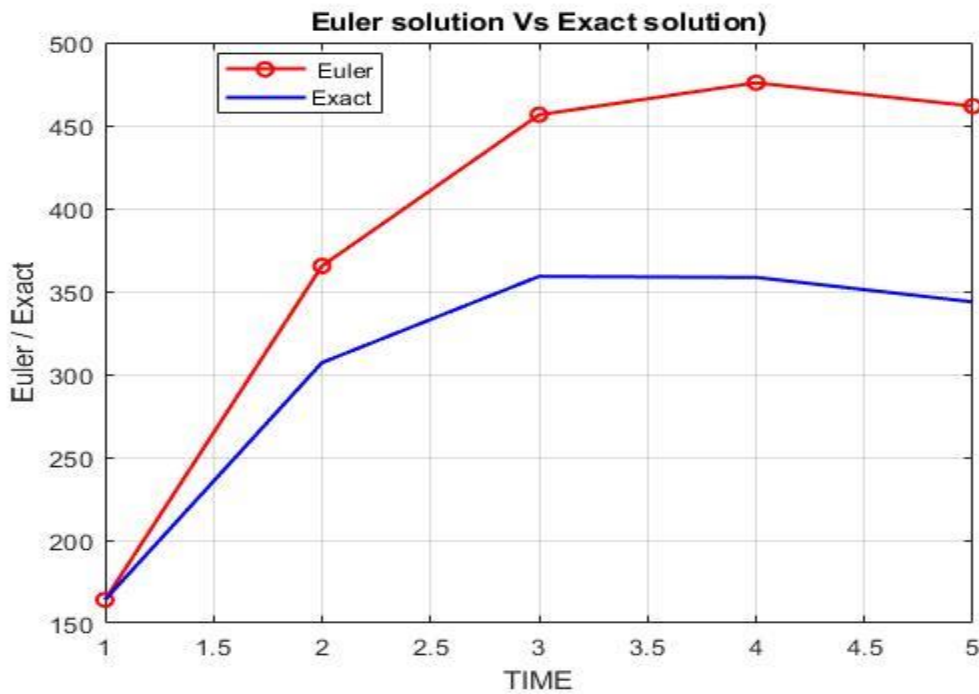


Figure 5.5: Plot of traffic crash with Euler method

We can see the difference between exact solution and Euler solution clearly in Fig. 5.5. As the gap is big we used fourth order Runge-Kutta method to solve the ODE.

5.4.2. Solving ODE using Runge-Kutta method

We solved the ODE using Runge-Kutta method. We used the same equation, initial value and step size h as in the Euler method.

Table 5.4: Runge-Kutta method output

x	y	k1	k2	k3	k4
1.0,	164.0000				
2.00	307.00000	201.30000	141.40000	141.40000	91.10000
3.00	359.00000	91.10000	50.40000	50.40000	19.30000
4.00	358.40000	19.30000	-2.20000	-2.20000	-14.10000
5.00	343.60000	-14.10000	-16.40000	-16.40000	-9.10000

Table 5.4 shows the output of Runge-Kutta method. The solution we got is the same as the exact solution. Let's see it graphically to make the comparison clearer.

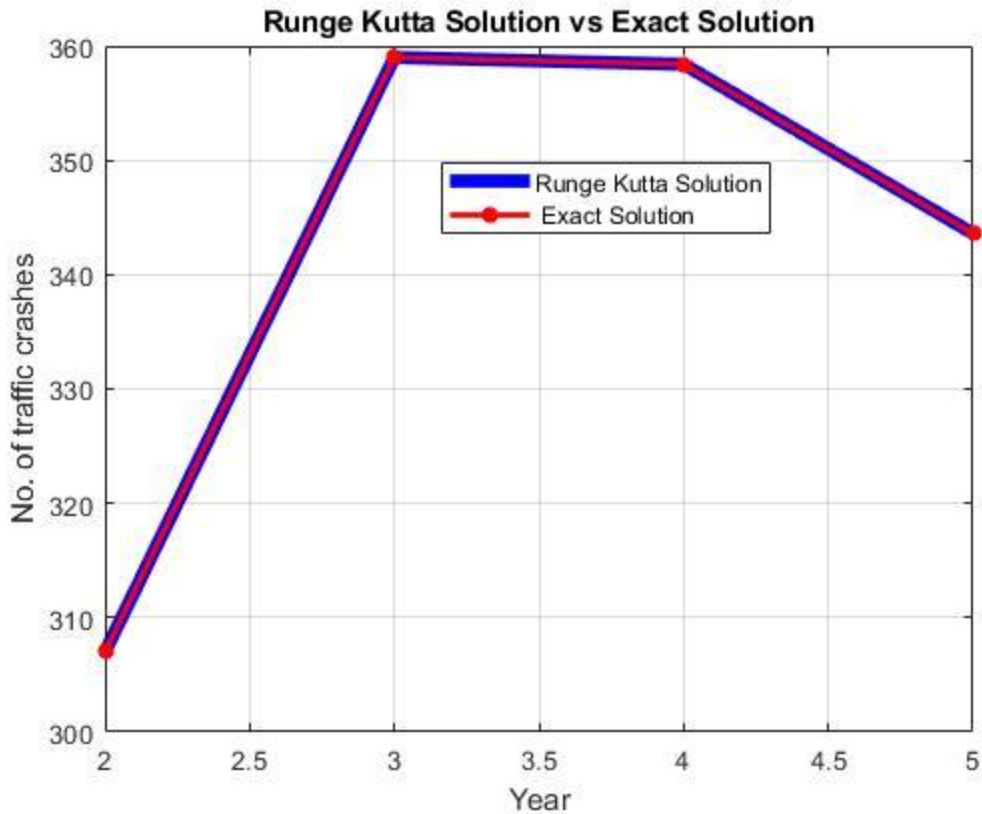


Figure 5.6: Plot of traffic crashes with Runge-Kutta method

In Fig. 5.6 we can see the solution we got using fourth order Runge-Kutta method is the same as the exact solution.

5.5. Modelling traffic volume

5.5.1. Modelling traffic volume using newton divided difference

We counted the number of vehicles passed in the express way within a year i.e. from 2015 to 2019. Then, we chose four data points to do the modelling. The data points are 2015, 2016, 2017, and 2019. We represented these years with 1, 2, 3, and 5 respectively. The reason is to simplify the mathematical complication. 2018 was chosen to be the unknown value and represented by 4. As we chose four data points the ODE we got is third order polynomial.

$$\begin{aligned}
 x &= 2018 = 4 \quad x_0 = 1 \quad x_1 = 2 \quad x_2 = 3 \quad x_3 = 5 \\
 b_0 &= f(x_0) = f(1) = 4211044 \\
 b_1 &= \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{f(2) - f(1)}{2 - 1} \\
 &= 5631228 - 4211044 = 1420184 \\
 b_2 &= \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_1} \\
 &= \frac{\frac{f(3) - f(2)}{3 - 2} - \frac{f(2) - f(1)}{2 - 1}}{3 - 1} \\
 &= \frac{(7338181 - 5631228) - (5631228 - 4211044)}{2} = \frac{1706953 - 1420184}{2} \\
 &= 143384 \\
 b_3 &= \frac{\frac{f(x_3) - f(x_2)}{x_3 - x_2} - \frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_3 - x_0} \\
 &= \frac{\frac{f(5) - f(3)}{5 - 3} - \frac{f(3) - f(2)}{3 - 2} - \frac{f(2) - f(1)}{2 - 1}}{5 - 1} \\
 &= \frac{8119378 - 7338181}{2} - (7338181 - 5631228) - (5631228 - 4211044) \\
 &= \frac{-145542}{4} \\
 &= -145542 \\
 b_0 &= 4211044 \quad b_1 = 1420184 \quad b_2 = 143385 \quad b_3 = -145542 \\
 f_3(x) &= 4211044 + 1420184(x - 1) + 143385(x - 1)(x - 2) + (-145542)(x - 1)(x - 2)(x - 3)
 \end{aligned}$$

$$f_3x = -145542x^3 + 1016637x^2 - 610933x + 3950882$$

After we got the ODE, we plotted it and we observed the value of the unknown data point.

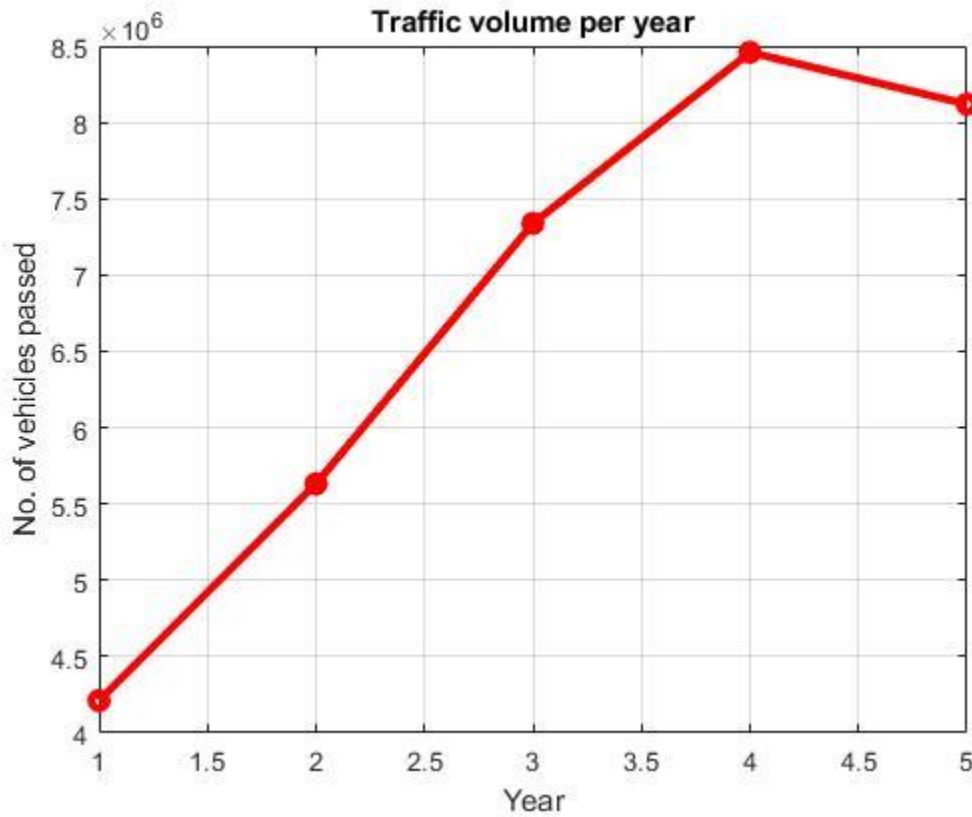


Figure 5.7: Traffic volume per year using newton divided difference

5.5.2. Modelling traffic volume using Lagrange method

As we did in newton divided difference, we counted the number of vehicles passed in the express way within a year. Similarly, we chose four data points (i.e. 2015, 2016, 2017, and 2019). We used interpolation method, so we chose 2018 the middle data point like in the newton divided difference. We got third order polynomial ODE.

$$x_1 = 1 f(x_1) = 4211044 \quad x_2 = 2 f(x_2) = 5631228 \quad x_3 = 3 f(x_3) = 7338181 \quad x_4 = 5 f(x_4) = 8119378$$

$$f(x) = 4211044 \frac{(x-2)(x-3)(x-5)}{(1-2)(1-3)(1-5)} \\ + 5631228 \frac{(x-1)(x-3)(x-5)}{(2-1)(2-3)(2-5)} \\ + 7338181 \frac{(x-1)(x-2)(x-5)}{(3-1)(3-2)(3-5)} \\ + 8119378 \frac{(x-1)(x-2)(x-3)}{(5-1)(5-2)(5-3)}$$

$$= -526381(x^3 - 10x^2 + 31x - 30) \\ + 1877076(x^3 - 9x^2 + 23x - 15) \\ - 1834545(x^3 - 8x^2 + 17x - 10) \\ + 338307(x^3 - 6x^2 + 11x - 6)$$

$$f_3x = -145543x^3 + 1016644x^2 - 610951x + 3950898$$

When we compare this equation with the one we got from newton divided difference, they are almost same. As result, we can ensure that our modelling is correct. We plotted this equation.

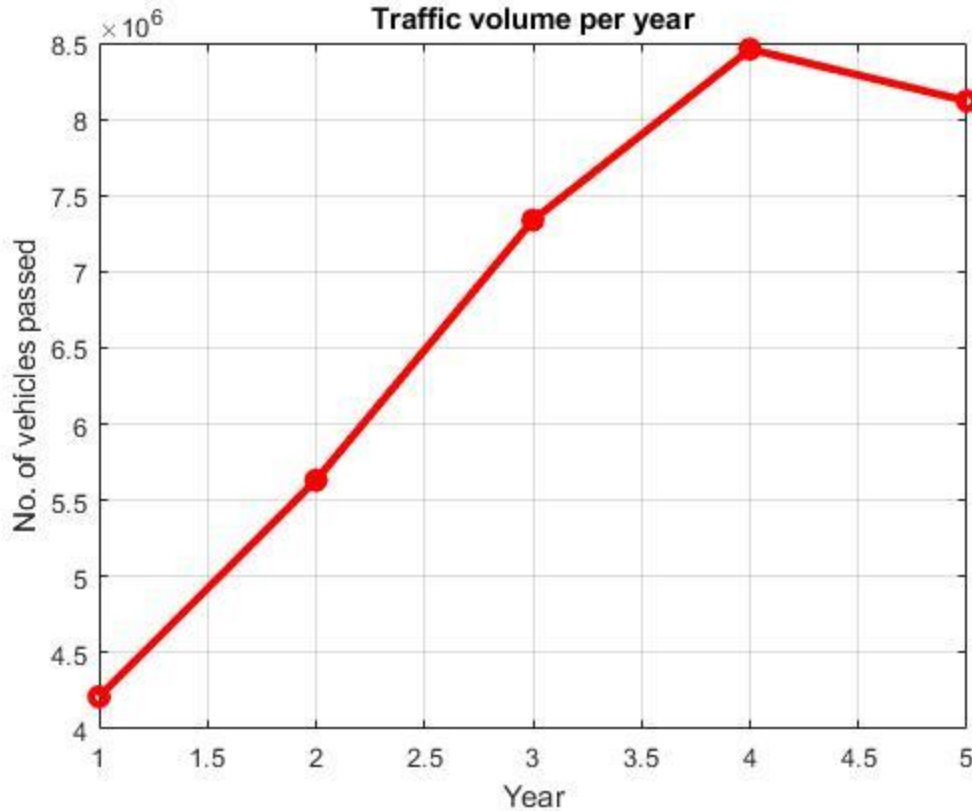


Figure 5.8: Traffic volume per year using Lagrange method

In Fig. 5.8 we can see year and the number of vehicles passed in each year including the value of the year we considered as unknown data point (i.e. 2018). In Fig. 5.9 we plotted both equation we got from the two interpolation methods and we can compare them.

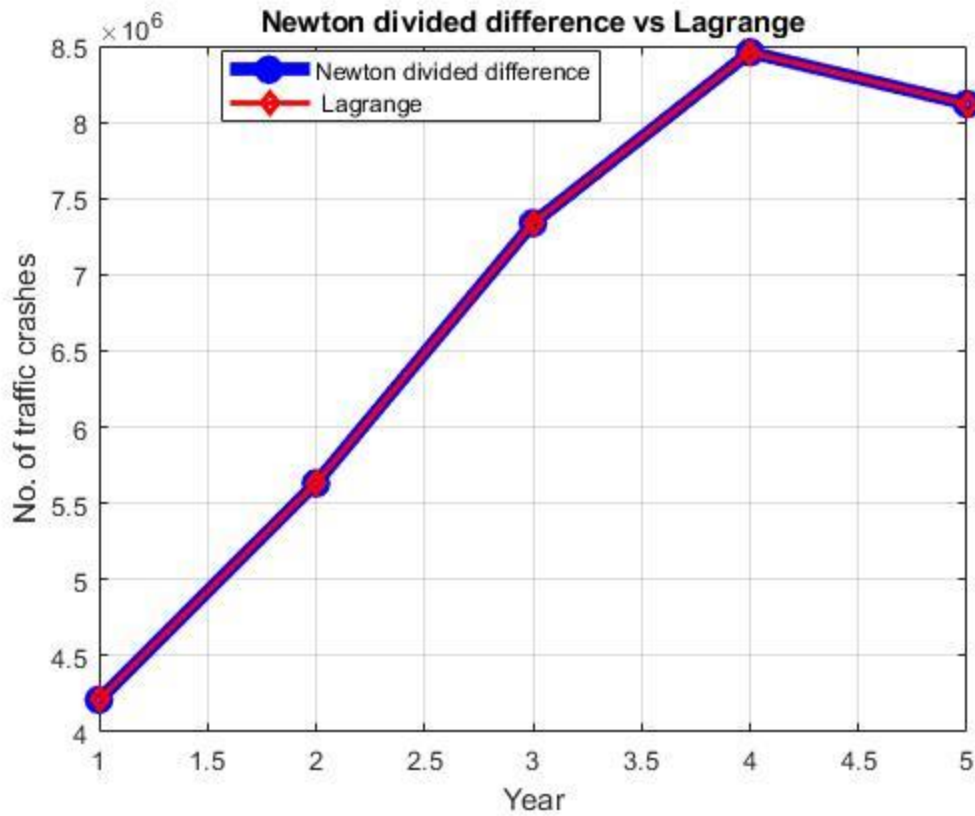


Figure 5.9: Traffic volume per year using newton DVD and Lagrange

As we can see in the graph from both methods we got same model and same result.

5.6. Solving ordinary differential equations of traffic volume

5.6.1. Solving ODE of traffic volume using Euler method

After modeling the data using ODE, we solved the equation using Euler method. We used the equation, its derivative, initial value, and step size h . The values are given below.

$$y = -145543x^2 + 1016644x^2 - 610951x + 3950898$$

$$y' = -436629x^2 + 2033288x - 610951$$

$$y(1) = 4211044, h = 1$$

Table 5.5: The output of Euler method in traffic volume

x	Euler	Exact
1.00	4211044.00000	4211048.00000
2.00	5196752.00000	5631228.00000
3.00	6905861.00000	7338180.00000
4.00	8465113.00000	8458646.00000
5.00	9001250.00000	8119368.00000

Table 5.5 shows that the output we got after solving the ODE using Euler method. There is a difference between the Euler solution and the exact solution.

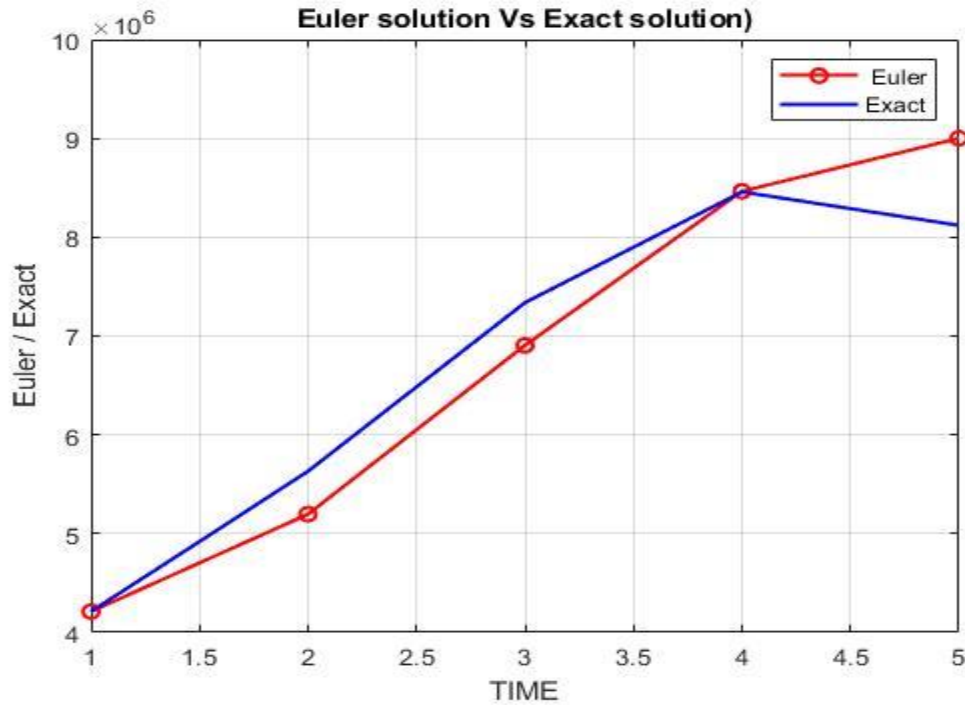


Figure 5.10: Plot of traffic volume using Euler method

In Fig. 5.10 we can see the difference between them clearly. So, we used fourth order Runge-Kutta to find the solution.

5.6.2. Solving ODE of traffic volume using Runge-Kutta method

We solved the equation using the same initial value and step size that we used in Euler method. The output is given below.

Table 5.6: Runge-Kutta method output

x	y	k1	k2	k3	k4
1.0,	4211044.0000				
2.00	5631224.00000	985708.00000	1456565.75000	1456565.75000	1709109.00000
3.00	7338176.00000	1709109.00000	1743337.75000	1743337.75000	1559252.00000

4.00	8458642.00000	1559252.00000	1156851.75000	1156851.75000	536137.00000
5.00	8119364.00000	536137.00000	-302892.25000	-302892.25000	-1360236.00000

Table 5.6 shows the solution we got after solving the equation using Runge-Kutta method. The result is almost the same as the real data.

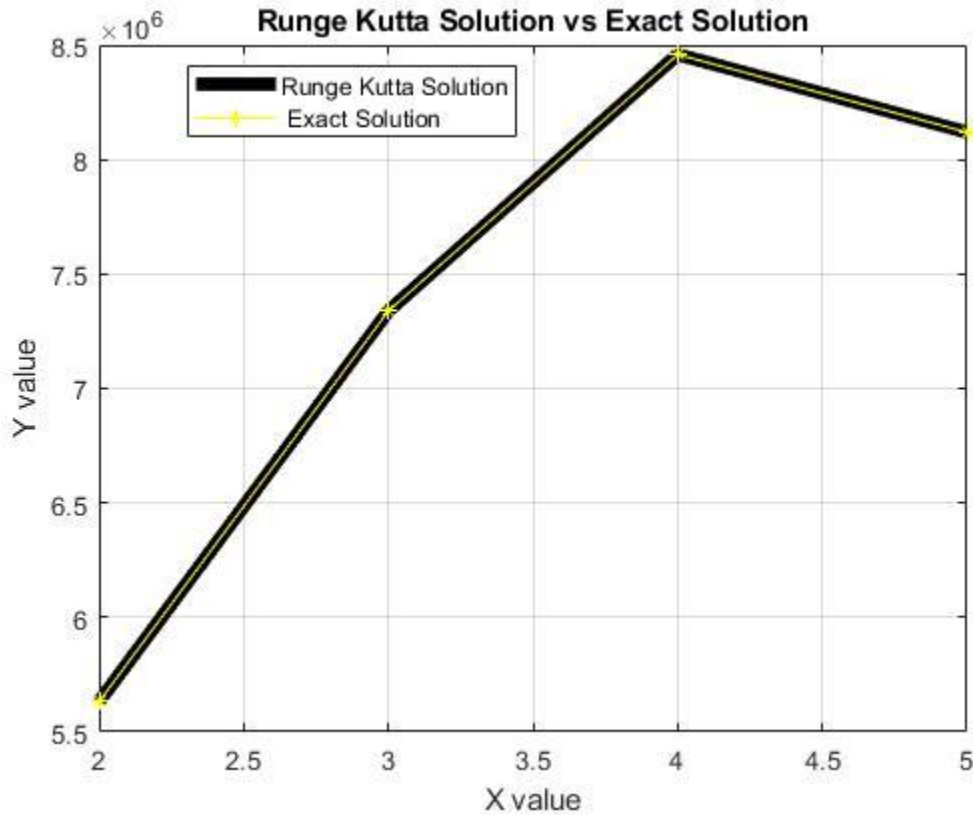


Figure 5.11: Plot of traffic volume using Runge-Kutta method

We can observe the narrow difference in Fig. 5.11.

5.7. Correlation of traffic crash and traffic volume

We modeled both data i.e. traffic crash and traffic volume that we collected from ETRE using newton DVD and Lagrange methods. Then we solved the resulting ODEs using Euler method and Runge-Kutta method. We observed the relationship between traffic crash and traffic volume by comparing their models and outputs.

Table 5.7: Output of traffic crash

x	y
1.0,	164.0000
2.00	307.00000
3.00	359.00000
4.00	358.40000
5.00	343.60000

Table 5.8: Output of traffic volume

x	y
1.0,	4211044.0000
2.00	5631224.00000
3.00	7338176.00000
4.00	8458642.00000
5.00	8119364.00000

Table 5.7 and Table 5.8 show us the outputs we got after modelling the data and solving the equations using Runge-Kutta method. Observing these outputs we can decide whether they have any relationship between them or not. To understand clearly we plotted the equations side by side in Fig. 5.12.

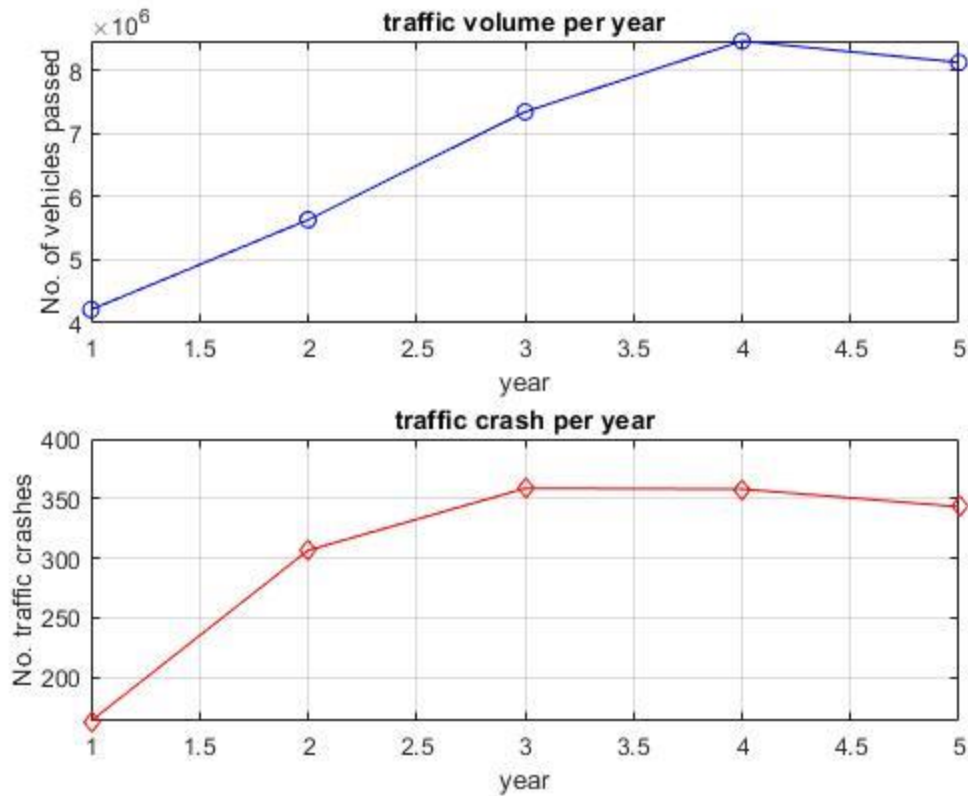


Figure 5.12: Traffic volume vs Traffic crash

In Fig. 5.12 the top graph shows the number of vehicles passed in the express way within five years. The bottom graph shows the traffic crashes happened in the express way within five years. By observing the way the graph goes we can see their relationship. As the number of vehicles passed in the express way increases the number of traffic crashes happened in the express way increases. We observed that their relationship is directly proportional.

Let's do the average of both traffic crash and traffic volume data and then we will do ratio of the results.

$$\text{Average of traffic crash} = \frac{164 + 307 + 359 + 391 + 344}{5} = 313$$

$$\text{Average of traffic volume} = \frac{4211044 + 5631228 + 7338181 + 7976106 + 8119378}{5} = 6655187$$

$$\frac{313}{6655187} \approx \frac{1}{21262}$$

The ratio is 1: 21262

We see that there will be one traffic crash in 21262 traffic volumes. In a country like Ethiopia, in which there are not many vehicles, this number is big.

5.8. Analysis of traffic crash

5.8.1. Analysis of traffic crash based on vehicle type

We analyzed the traffic crash data collected from ETRE. We used vehicle type parameter. We analyzed traffic crash happened by different vehicle types. We separated which type of vehicle is causing more crash. We found that small automobiles are causing more crash than other types of vehicles. Heavy truck trailers and medium buses took the second place in causing accidents. Big size buses and minibus took the third and fourth place respectively.

Table 5.9: Vehicle types caused traffic crash with number

Vehicle_Category	
Small Auto	562
Heavy Truck trailers	172
Medium Bus/Isuzu	169
Big Size Bus	110
Minibus	100
Unidentified	13

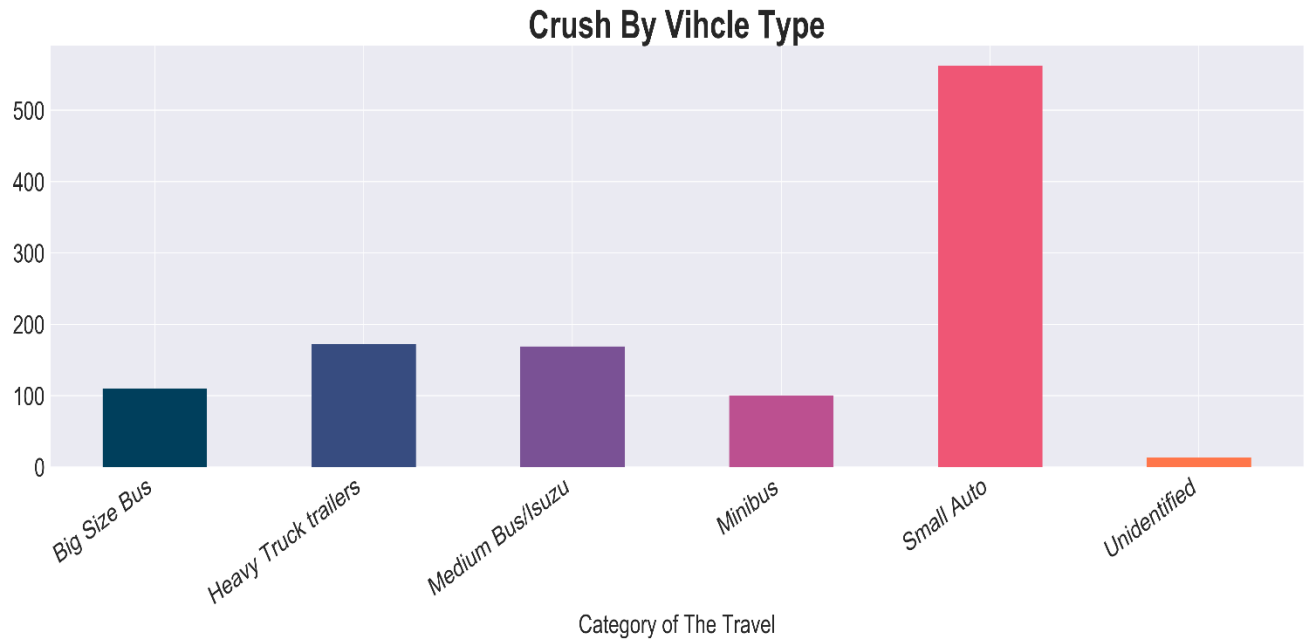


Figure 5.13: Number of traffic crashes happened by vehicles type

5.8.2. Analysis of traffic crash based on vehicle types and week days

We analyzed traffic crash happened by different types of vehicles within week days. Many accidents happened on Friday and Sunday compared to other days. Small automobiles were causing the accidents. Saturday took the second place. Similarly, more accidents were caused by small automobiles. We can see which vehicle types are causing more accidents in which days in Fig. 5.14.

Table 5.10: Traffic crashes occurred in week days

Vehicle_Category	Big Size Bus	Heavy Truck trailers	Medium Bus/Isuzu	Minibus	Small Auto	Unidentified
Day						
Friday	17.0	16.0	29.0	13.0	82.0	4.0
Monday	12.0	16.0	19.0	13.0	70.0	0.0
Saturday	15.0	26.0	18.0	15.0	78.0	1.0
Sunday	7.0	16.0	13.0	16.0	83.0	1.0
Thursday	13.0	30.0	25.0	11.0	77.0	1.0
Tuesday	18.0	24.0	32.0	13.0	69.0	1.0
Wednesday	10.0	26.0	20.0	11.0	57.0	0.0

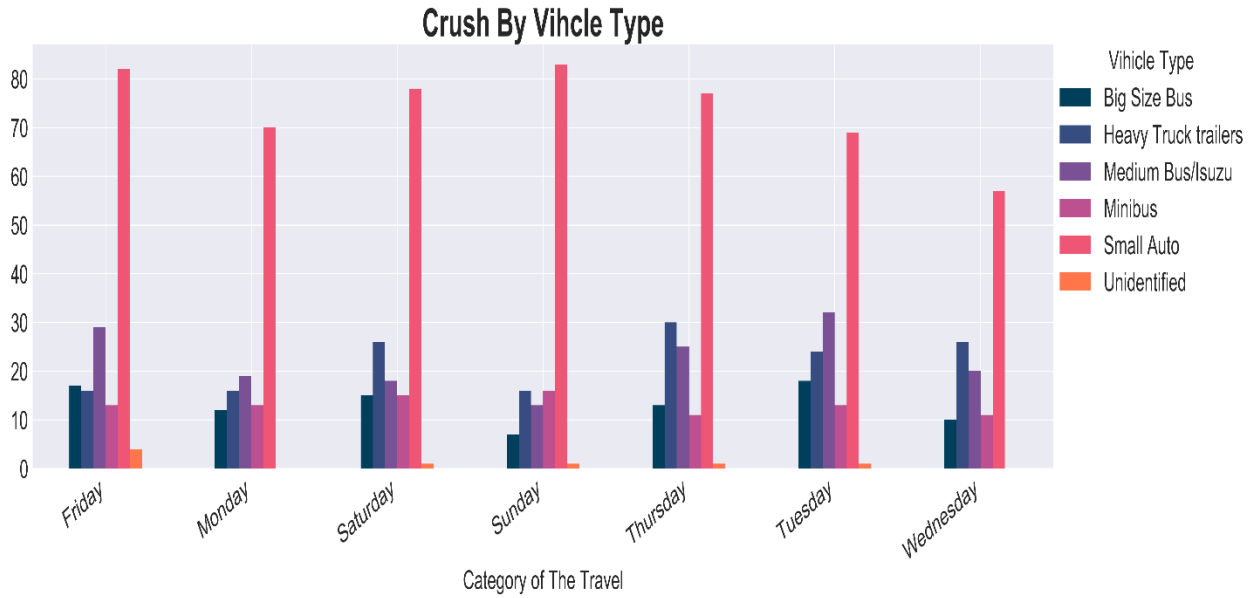


Figure 5.14: Traffic crash happened by vehicle types within week days

5.8.3. Analysis of traffic crash based on direction

We analyzed to which direction is more accidents were happening. As we can see in Fig. 5.15 vehicles headed to mojo were causing more accident than vehicles headed to other direction. Vehicles headed to Addis Ababa were causing the second most accident. Vehicles headed to welenchiti, stasion, and debrezeyt were causing more accidents respectively.

Table 5.11: traffic crash occurred by direction

Direction	
To Addis	438
To Debrezeyt	5
To Mojo	526
To Stasion	19
To Welenchiti	25

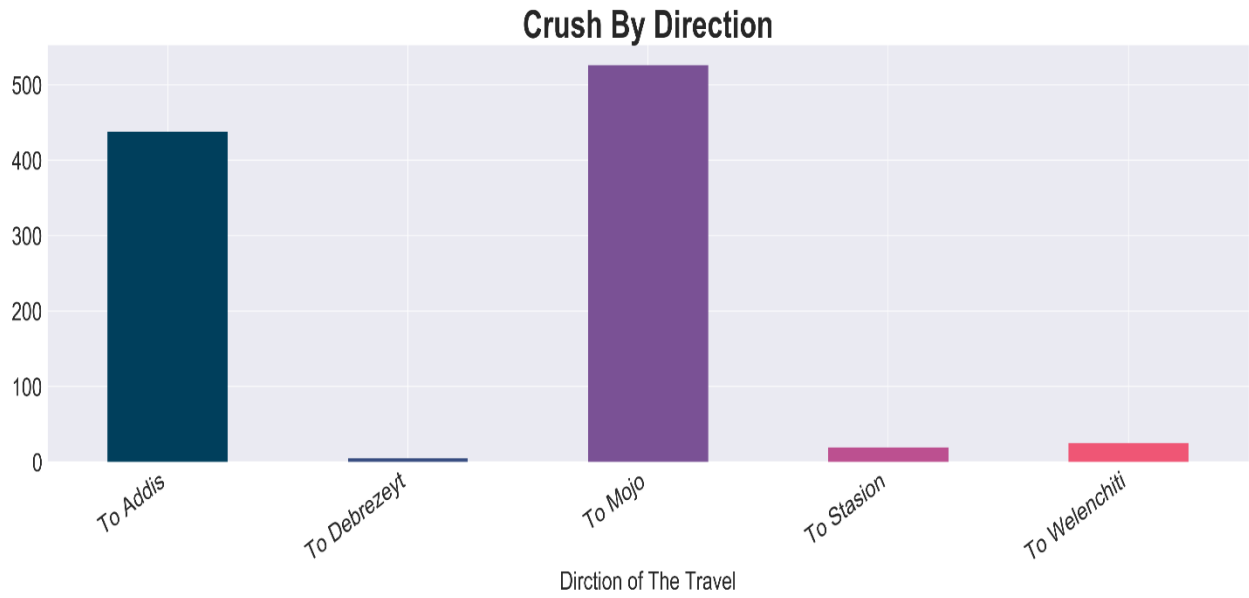


Figure 5.15: Traffic crashes occurred by direction

5.9. Discussion

5.9.1. Traffic crash modelling

Traffic crash data are valuable instruments to support the development, execution, and evaluation of highway safety programs that tend to reduce road traffic crashes. We collected traffic crash data from ETRE. We modeled these data with ODE. Mathematical modelling assists us in making better decision. We used interpolation to model the data. Newton divided difference and Lagrange are interpolation methods we used in the modelling. The ordinary differential equations we got after modelling the data using these two methods are almost the same. This showed us our model is correct. We plotted the equation in Matlab. We were able to predict the next year data (i.e. number of traffic crashes that will happen in 2020). Then, we solved the equation using Euler method and Runge-Kutta method. We compare the solutions from both methods we got with the exact solution. The solution we got from Euler method had a difference with the exact solution. However, the solution we got from Runge-Kutta method is almost the same as the exact solution. We plotted both methods in matlab to understand them clearly.

5.9.2. Traffic volume modelling

Studying traffic volume gives better way for expansion of infrastructures. We counted the number of vehicles in the express way within a year. Then we modeled the data using two methods of interpolation i.e. newton DVD and Lagrange methods. We modeled the data with ODE. The ODEs we got after modelling using both methods are almost the same. This showed the model is good. Then we plotted the ODE in Matlab to observe the equation clearly. Using Euler method and Runge-Kutta method we solved the equation. When we compare the solution we got from Euler method with the exact solution, there is a difference. The result we got from Runge-Kutta method is almost the same as the exact solution. We plotted both solutions with the exact solution in Matlab. It makes it easy to observe their difference with the exact solution.

5.9.3. The correlation between traffic crash and traffic volume

We discussed how we modeled traffic crash data and traffic volume data that we collected from ETRE in the above two sections. The next thing we did is to check if there is any relationship between traffic crash and traffic volume based on the ODE models. We tried to compare the outputs of both ODEs in Table 5.7 and 5.8 and we observe that as the traffic volume increases, the number of traffic crash also increases. Similarly, when traffic volume decreases, the number of

traffic crash also decreases. To be very clear we plotted both ODEs in Matlab (see Fig. 5.12) and we can observe how the plots appear. Both plots are going in the similar way. This indicates that there is a relationship between traffic volume and traffic crash.

5.9.4. The effect of vehicle type on traffic crash

The traffic crash data that we collected from ETRE includes some parameters i.e. vehicle type, day, direction, time and so on. We chose vehicle type to see its effect on the number of traffic crash. The vehicles are classified in five categories i.e. big size bus, heavy truck, medium size bus, minibus, and small automobiles. We analyzed this parameter to observe which vehicle category causes the highest traffic crash. The analysis result indicates that small automobiles are causing the highest traffic crash. Heavy trucks took the second place in causing traffic crash. Medium size bus, big size bus, and minibus took third, fourth, and five places respectively.

5.9.5. The effect of vehicle type with weekdays on traffic crash

We knew which vehicle type is causing high traffic crash in the above section. Then we related it with weekdays. We observed that in which day which type of vehicle is causing high traffic crash. This helps ETRE to focus on the vehicle type causing high traffic crash and take high control and also to monitor vehicles on days with high traffic crash. Small automobiles are causing high traffic crash in all days. Heavy truck trailers took the second place causing high traffic crash on Wednesday, Thursday, Saturday, and Sunday. Big size bus took the second place causing high traffic crash on the rest of days.

5.9.6. The effect of direction on traffic crash

Knowing the direction with higher traffic crash is very essential, so that there will be special monitor on that direction. Mainly there are five direction in the express way. That are Addis, debrezeyt, mojo, welenchiti, and station. Vehicles which are heading to mojo are causing the highest traffic crash. Vehicles which are heading to Addis Ababa took the second place. Vehicles which are heading to welenchiti, station, and debrezeyt took the third, fourth, and fifth place respectively.

Chapter six

6. Conclusion and Recommendation

6.1. Conclusion

This thesis discusses traffic crashes and traffic volume in Addis – Adama expressway. It also studies the relationship between traffic crash and traffic volume. Additionally, it includes analysis on the effects of vehicle type, vehicle type with weekdays, and direction on traffic crash.

We modelled the traffic crash data with ODE using newton DVD and Lagrange interpolation methods. We solved the equation with Euler method and Runge-Kutta method. The result we got from Euler method is far from exact solution but the Runge-Kutta solution is close to the exact solution. We plotted the equation in Matlab to see it clearly. And we were able to predict the next year i.e. 2020 traffic crash.

As we did in the traffic crash data, we did the same thing in the traffic volume data. We modelled it using newton DVD and Lagrange methods. The equation we got from both methods is almost the same. So we continued to solving the equation. Similarly, we used Euler method and Runge-Kutta method. Similar to what happened in solving traffic crash equation the Runge-Kutta solution is close to the exact solution. We plotted the equation for better understanding.

After we modelled both data i.e. traffic crash data and traffic volume data, we observe if there is any correlation between them. We compared the outputs we got after solving the equations. We observe the number of traffic crash increases with the increase of traffic volume. We plotted both equations side by side to observe how the graphs appear. Similarly, we observe the same thing as the outputs. As a result, we can say that there is a relationship between traffic crash and traffic volume.

We did analysis on some parameters of traffic crash data. We analyzed vehicle type and we knew which type of vehicle is causing the highest traffic crash i.e. small automobiles. The highest number of traffic crash is happening on Friday by small automobiles. And also vehicles which are heading to mojo are causing the highest number of traffic crash.

6.2. Contributions

This research has contributions on:

- Improvement of traffic management system in Addis – Adama express way.
- Provide guidance for ERA and ETRE while building new express way.

6.3. Recommendation

This thesis worked on traffic crash and traffic volume of Addis – Adama express way. There are not enough studies on this express way. Even the studies done previously focus only on traffic crash. We suggest that many researches to be done on traffic volume and its relationship with traffic crash.

We recommend that there should be a special monitoring on small automobiles since they are causing the highest number of traffic crash. There should be a strict control on weekends especially on Friday. There should be speed breaker or something else in mojo gate which can help to decrease the number of traffic crash.

Reference

- [1] G. Gebremedhn, “Road Traffic Congestion Analysis and Traffic Flow Modeling at Meskel Square Signalized Control Intersections,” no. June, 2017.
- [2] S. Shanko, “SEGMENTAL ASSESSMENT OF LEVEL OF TRAFFIC CONGESTION ON Kality RR To Dukem Bridge,” p. 125, 2015.
- [3] “Symbiosis _ Definition of Symbiosis at Dictionary.” .
- [4] “Expressway | road | Britannica.” [Online]. Available: <https://www.britannica.com/technology/expressway-road>., 9/2/2020.
- [5] M. A. Abera, “Evaluation of road traffic accident in Addis Ababa – Adama Expressway,” vol. 5, no. 5, pp. 450–456, 2019.
- [6] D. Deme and M. Bari, “Full length research paper Traffic accident causes and its countermeasures on Addis Ababa-Adama expressway,” *J. Sci. Sustain. Dev.*, vol. 2, no. 2, pp. 13–23, 2018, [Online]. Available: <http://www.mwu.edu.et/journal-sites>.
- [7] C. Timmermans, W. Alhajyaseen, A. Al Mamun, M. Qasem, M. Almallah, and H. Younis, “Analysis of road traffic crashes in the State of Qatar,” *Int. J. Inj. Contr. Saf. Promot.*, vol. 26, no. 3, pp. 242–250, 2019, doi: 10.1080/17457300.2019.1620289.
- [8] E. A. Beimborn, “Transportation Modeling Primer, Univ of Wisc-Milw, Center for Urban Transportation Studies (CUTS).” 1995, [Online]. Available: <https://www4.uwm.edu/cuts/primer.htm>., August 24, 2020
- [9] M. Mohanty and A. Gupta, “Factors affecting road crash modeling,” vol. 9, no. 2, pp. 15–19, 2015.
- [10] L. Fridstrøm, J. Ifver, S. Ingebrigtsen, R. Kulmala, and L. K. Thomsen, “Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts,” *Accid. Anal. Prev.*, 1995, doi: 10.1016/0001-4575(94)E0023-E.
- [11] D. J. Graham and S. Glaister, “Spatial variation in road pedestrian casualties: The role of urban scale, density and land-use mix,” *Urban Stud.*, 2003, doi:

- 10.1080/0042098032000094441.
- [12] J. Aguero-Valverde and P. P. Jovanis, “Spatial analysis of fatal and injury crashes in Pennsylvania,” *Accid. Anal. Prev.*, 2006, doi: 10.1016/j.aap.2005.12.006.
- [13] A. Bull, *Traffic Congestion - The Problem and How to Deal with it?* 2004.
- [14] A. Abdulhafedh, “Road Crash Prediction Models: Different Statistical Modeling Approaches,” *J. Transp. Technol.*, vol. 07, no. 02, pp. 190–205, 2017, doi: 10.4236/jtts.2017.72014.
- [15] D. Bhalla, “15 Types of Regression in Data Science,” *Listen Data*. 2018, [Online]. Available: <https://www.listendata.com/2018/03/regression-analysis.html>, December 5, 2020.
- [16] “Multiple Regression - Statistics Solutions,” *Statistics Solutions*. 2019, [Online]. Available: <https://www.statisticssolutions.com/regression-analysis-multiple-regression/>, December 5, 2020.
- [17] “Traffic Volume Count | Need, methods & methods of collection.” [Online]. Available: <https://planningtank.com/survey/traffic-volume-count>, December 5, 2020.
- [18] N. Hoque, S. Halder, R. Bin Alam, and M. Hassan, “Traffic Volume Study,” no. February 2015, pp. 1–47, 2013.
- [19] “Traffic Volume Study - Definition, Methods and Importance of Traffic Volume Study.”, December 5, 2020.
- [20] “Purposes of Traffic Volume Study _ Civil engineering training.”, December 5, 2020.
- [21] A. Abdulhafedh, “Road Traffic Crash Data : An Overview on Sources , Problems , and Collection Methods,” pp. 206–219, 2017, doi: 10.4236/jtts.2017.72015.
- [22] “Lagos–Ibadan Expressway - Wikipedia.”, September 4, 2020.
- [23] V. Patil, T. P. Kurugod, and M. Tech, “Modeling to traffic flow on Indian expressway and urban mid-block section using simulation technique,” *Int. Res. J. Eng. Technol.*, vol. 4, no. 7, pp. 856–864, 2017, [Online]. Available: <https://irjet.net/archives/V4/i7/IRJET->

V4I7149.pdf.

- [24] F. Wikipedia and E. Union, “Expressways of China.” pp. 1–26, 2013, September 4, 2020.
- [25] A. Maria, “Introduction to modeling and simulation,” *Winter Simul. Conf. Proc.*, pp. 7–13, 1997, doi: 10.1145/268437.268440.
- [26] G. R. Liu and S. S. Quek, “Computational Modeling,” *The Finite Element Method*. pp. 1–11, 2014, doi: 10.1016/b978-0-08-098356-1.00001-1, September 6, 2020.
- [27] “Computational Model,” *Encyclopedia of Neuroscience*. pp. 824–824, 2008, doi: 10.1007/978-3-540-29678-2_1164, September 6, 2020.
- [28] “Computational modelling for decision-making_ where, why, what, who and how _ Royal Society Open Science.” , September 6, 2020.
- [29] デヴィッド Burghes, モラグ Borrie, 垣田高夫, and 大町比佐栄, “微分方程式で数学モデルを作ろう (Modelling with Differential Equations) .” p. 222, 1990.
- [30] K. W. Morton and D. F. Mayers, “Numerical solution of partial differential equations: An introduction,” *Numer. Solut. Partial Differ. Equations An Introd.*, pp. 1–279, 2005, doi: 10.1017/CBO9780511812248.
- [31] “Ordinary differential equations,” *Choice Reviews Online*, vol. 50, no. 05. pp. 50-2707-50–2707, 2013, doi: 10.5860/choice.50-2707, September 11, 2020.
- [32] T. Hirano and K. Wakashima, “Mathematical Modeling and Design,” *MRS Bull.*, vol. 20, no. 1, pp. 40–42, 1995, doi: 10.1557/S0883769400048922.
- [33] “Interpolation - Wikipedia.”, September 6, 2020.
- [34] “Introduction to Sociology/Deviance - Wikibooks, open books for an open world.” [Online]. Available: http://en.wikibooks.org/wiki/Introduction_to_Sociology/Deviance.
- [35] “Interpolation in Statistics: Definition, Formula & Example - Video & Lesson Transcript | Study.com.” <https://study.com/academy/lesson/interpolation-in-statistics-definition-formula-example.html>, September 6, 2020.

- [36] T. Bloom, “Polynomial interpolation,” *Bol. da Soc. Bras. Matemática*, vol. 10, no. 2, pp. 75–86, 1979, doi: 10.1007/BF02584632.
- [37] T. A. Osswald, J. P. Hernández-Ortiz, T. A. Osswald, and J. P. Hernández-Ortiz, “Introduction to Numerical Analysis,” *Polym. Process.*, pp. 341–384, 2006, doi: 10.3139/9783446412866.007, September 6, 2020.
- [38] N. Analysis, “Numerical Methods for Differential Equations 1 . Brief overview of PDE problems,” pp. 1–50, 2008, doi: 10.3840/002751, September 12, 2020.
- [39] J. C. Butcher, “Numerical Methods for Ordinary Differential Equations,” *Numerical Methods for Ordinary Differential Equations*. pp. 1–513, 2016, doi: 10.1002/9781119121534, September 12, 2020.
- [40] T. Runge-kutta, “Runge-Kutta Methods,” 1944.
- [41] W. E. Milne, “Note on the Runge-Kutta method,” *J. Res. Natl. Bur. Stand. (1934).*, vol. 44, no. 5, p. 549, 1950, doi: 10.6028/jres.044.049.
- [42] Z. Zlatev, I. Dimov, I. Faragó, and Á. Havasi, “Richardson Extrapolation,” *Richardson Extrapolation*, pp. 1–3, 2017, doi: 10.1515/9783110533002.
- [43] L. Notes, “Lecture Notes 4 Numerical differentiation and integration,” *Notes*, vol. 1, no. x.
- [44] “Computational Modeling and Data Analytics | Virginia Tech.” [Online]. Available: <http://www.admiss.vt.edu/majors/index.php/majors/major/CMDA>, September 12, 2020.
- [45] A. I. Resource and I. Members, “Data analytics,” no. November, 2018.
- [46] S. Mujawar and A. Joshi, “Data Analytics Types , Tools and their Comparison,” vol. 4, no. 2, pp. 488–491, 2015, doi: 10.17148/IJARCCE.2015.42110.
- [47] “(PDF) Introduction to Data Science & Analysis using Python.” .
- [48] W. P.D. and L. C.Y., *Python for Data Analysis*, vol. 344, no. 4. 2012.
- [49] “Python - Data Science Introduction.” ,
https://www.tutorialspoint.com/python_data_science/python_data_science_introduction.ht

m., April 23, 2021

- [50] M. T. Obaidat and T. M. Ramadan, “Traffic accidents at hazardous locations of urban roads,” *Jordan J. Civ. Eng.*, 2012.
- [51] V. Shankar, F. Mannering, and W. Barfield, “Effect of roadway geometrics and environmental factors on rural freeway accident frequencies,” *Accid. Anal. Prev.*, 1995, doi: 10.1016/0001-4575(94)00078-Z.
- [52] J. C. Barbosa, “What is Mathematical Modelling?,” *Math. Model. A W. Life*, pp. 227–234, 2003, doi: 10.1533/9780857099549.5.227.
- [53] “What Is Data Analysis_ Methods, Techniq new.”, October 27, 2020 .
- [54] “Top 10 Python Libraries for Data Science | by Rashi Desai | Towards Data Science.” [Online]. Available: <https://towardsdatascience.com/top-10-skills-for-a-data-scientist-in-2020-2b8e6122a742%0Ahttps://towardsdatascience.com/top-10-python-libraries-for-data-science-cd82294ec266>, December 16, 2020.
- [55] S. S. Yasiran, “What Is Matlab ?,” *Matrix*. pp. 1–23.
- [56] “Anaconda (Python distribution) - Wikipedia.” [Online]. Available: [https://en.wikipedia.org/wiki/Anaconda_\(Python_distribution\)](https://en.wikipedia.org/wiki/Anaconda_(Python_distribution)), December 16, 2020.
- [57] Mike Driscoll, “Jupyter Notebook: An Introduction – Real Python,” *Real Python*. [Online]. Available: <https://realpython.com/jupyter-notebook-introduction/>, December 16, 2020.
- [58] “Lagrange polynomial - Wikipedia.” [Online]. Available: https://en.wikipedia.org/wiki/Lagrange_polynomial, December 14, 2020.
- [59] Wikipedia, “Euler method - Wikipedia,” , December 14, 2020.
- [60] “Runge-kutta method _ Dictionary.” , December 14, 2020.
- [61] “Runge–Kutta methods - Wikipedia.” [Online]. Available: https://en.wikipedia.org/wiki/Runge–Kutta_methods, December 14, 2020.

[62] M. Tadege, “Determinants of fatal car accident risk in Finote Selamtown, Northwest Ethiopia”, 2020

Annex A: Sample code for data preprocessing

```
Vehicle_Type = []

for Type in CrushDf.Vehicle_Type:

    if Type == 0:

        Vehicle_Type.append("Unidentified")

    if Type ==1:

        Vehicle_Type.append("Small Auto")

    if Type ==2:

        Vehicle_Type.append("Minibus")

    if Type ==3:

        Vehicle_Type.append("Medium Bus/Isuzu")

    if Type ==4:

        Vehicle_Type.append("Big Size Bus")

    if Type >=5:

        Vehicle_Type.append("Heavy Truck trailers")
```

Annex B: Sample code for Lagrange interpolation

```

function y=lagrange1(x,pointx,pointy)
clc
x=4;
pointx=[1 2 3 5];
pointy=[4211044 5631228 7338181 8119378];
n=size(pointx,2);
L=ones(n,size(x,2));
if (size(pointx,2)~=size(pointy,2))
    fprintf(1,'\nERROR!\nPOINTX and POINTY must have the same number of
elements\n');
    y=NaN;
else
    for i=1:n
        for j=1:n
            if (i~=j)
                L(i,:)=L(i,:).*(x-pointx(j))/(pointx(i)-pointx(j));
            end
        end
    end
    y=0;
    for i=1:n
        y=y+pointy(i)*L(i,:);
    end
end
end

```

Annex C: Sample code for Runge – Kutta method

```

function rungekutta
clear;clc;
h = 1;
x = 1;
y = 4211044;
fprintf('\n Runge Kutta method Output\n');
format short
disp('-----')
disp('-----')
disp(' x          y          k1          k2          k3          k4');
fprintf('%4.1f, %4.4f\n', x, y);
for i=1:4
k1 = h*f(x,y);
k2 = h*f(x+(h/2), y+k1/2);
k3 = h*f(x+(h/2), y+k2/2);
k4 = h*f(x+h, y+k3);
y = y + (k1+2*k2+2*k3+k4)/6;
yy(i)=y;
x = x + h;
xx(i)=x;
disp( sprintf(' %2.2f %2.5f || %2.5f || %2.5f || %2.5f ||
%2.5f', x, y, k1, k2, k3, k4));
end
disp('-----')
disp('-----')
plot(xx,yy, 'k-', 'linewidth', 5)
hold on
plot (xx,yy, 'y-*')
grid on
title ('Runge Kutta Solution vs Exact Solution')
xlabel ('X value')
ylabel ('Y value', 'LineWidth', 5)
legend ('Runge Kutta Solution', ' Exact Solution');
function f = f(x,y)
f=-436629.*x.^2+2033288.*x-610951;

```

Annex D: Sample code for analyzing traffic crash data parameter vehicle type

```

import matplotlib.pyplot as plt

import matplotlib as mpl

import seaborn as sb

import numpy as np

import pandas as pd

from pandas import DataFrame, Series

plt.style.use("seaborn")

mpl.rcParams['agg.path.chunksize'] = 10000

fig, ax=plt.subplots(figsize=(28,8),dpi=256)

ax.set_title("Crush By Vihcle Type",fontweight="bold",fontsize=36)

colors=["#003f5c","#374c80","#7a5195","#bc5090","#ef5675","#ff764a","#ffa600"]

group_crush=CrushDf.groupby(['Vehicle_Category']) ['Vehicle_Category'].count()

px= group_crush.plot(ax=ax,fontsize=24,color=colors,kind="bar")

plt.xlabel("Category of the travel",fontsize=24)

fig.autofmt_xdate()

```

Declaration

I, the undersigned, declare that this research is my original work and has not been presented for degree in any other university, and that all sources of materials used for the research have been acknowledged.

Declared by:

Name: Ayana Murad

Signature: _____

Date: _____

Confirmed by advisor:

Name: Solomon Gizaw (PhD)

Signature: _____

Date: _____