Spatio-Temporal Modelling, Analysis and Forecasting of Road Traffic Accidents in Oman

Ghanim Al-Hasani

A thesis submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy

in the

Staffordshire University

October 2021

Department of Engineering, School of Digital, Technologies and Arts, Staffordshire University, College Road, Stoke on Trent, ST4 2DE, United Kingdom

Dedication

... to the spirit of my grandmother who passed away when I was doing this study

... to my beloved parents who enlighten my path in life to the hereafter

... to my wonderful wife Fatma and my children Hafsa, Yasir, Aisha, Hind, Moza and Sheikha who were the soul behind this work

... to my dearest siblings, cousins and all members of our big family and tribe

... to all of my teachers, friends and colleagues

Abstract

Road traffic accident (RTA) is one of the most significant reasons for deaths and disabilities globally. Although there have been many attempts over the years, a number of insights yet need to be investigated, which would provide a better understanding of road traffic accidents (RTAs) in countries with a higher number of accidents. Many advanced mathematical models have been developed and applied to uncover a wide variety of issues of RTAs for many countries, for instance, accident frequencies, observed and unobserved factors, unobserved heterogeneity, endogeneity, etc. Oman is one of the wealthy middle eastern countries that experiences a very high volume of RTAs every year. However, little is known about the nature, variation, and causal factors for road traffic accidents in the Sultanate of Oman. Therefore, this study attempts to characterise the temporal, spatial, and combined space-time patterns of RTAs in Oman by mathematical modelling techniques. The RTAs data have been collected for the Sultanate of Oman for this study. Firstly a number of temporal models have been applied to discover the trend, other temporal components and forecast of the road traffic accidents and injuries (RTI) in Oman. The study found that the SARIMA model is the best temporal model with the highest goodness of fit for RTA and RTI. The higher number of RTAs occurred in June-August period, and the peak is in July during the summer holiday. Similarly, the months of May, June and July are expected to have the highest number of RTIs in Oman forecasted by the SARIMA model. Secondly, a variety of spatial models including spatial lag model (SLM), spatial error model (SEM) and geographically weighted Poisson regression (GWPR) model have been applied to capture spatial effects and influencing factors for the RTAs in Oman. One of the significant findings to emerge from this study is that the spatial lag model (SLM) outperformed the spatial error model (SEM) due to the best values in diagnostic indicators. However, the spatial variations in these models (SLM and SEM) are taken into account only through the spatial error structure. Another type of spatial modelling approach that provides a set of local models obtained by calibrating multiple geographical entities simultaneously is the geographically weighted

Poisson regression (GWPR).

The main challenge with the GWPR models is to set appropriate kernel function to give weights for each neighbouring point during the model calibration. Likelihood function, parameter estimation and model selection criteria have been shown in details and model formulation has been applied to the RTAs data in Oman. A GWPR model has been developed for five different kernel weighting functions: box-car, bi-square, tri-cube, exponential and Gaussian weighted function. The study found that GWPR models can substantially capture the heterogeneity of the spatial factors over the regions or spatial units. The crucial finding to emerge from this study is that GWPR model with exponential kernel function and adaptive bandwidth is the most suitable for modelling, fitting and analysing RTA data in Oman. Finally, a combined approach of spatio-temporal modelling have been investigated, which would capture the temporal and spatial effects simultaneously. Spatio-temporal modelling is taken into account of both spatial and temporal correlations dependencies. The main challenge is building up the spatio-temporal model consisting of three components: choosing the perfect techniques to parts for space effects, time effects and the interaction between space and time effects. Spatio-temporal models have been built based on a Bayesian hierarchical framework to generate stable estimation that provides good characteristics to benefits even with low counts of RTAs. Integrated nested Laplace approximation (INLA) is fully Bayes tools used for fitting models, estimating fixed, random or posterior parameters. One major challenge is to model the interaction between spatial effects and temporal effects for count data models. This study develops Bayesian hierarchical spatio-temporal interaction models taking Oman RTA data. This study founds that unseparated spatio-temporal models are playing essential roles to capture unobserved heterogeneity for various factors. The most exciting finding is that the best performance pf a spatio-temporal model is the spatio-temporal interaction typeII model. Spatio-temporal interaction typeII acts by interacting between unstructured spatial effects and structured temporal effects. It suggests considering the second-order random walk (RW2) and Leroux CAR (LCAR) model into the spatio-temporal interaction model with a longer time scale and multivariate levels.

Acknowledgements

Foremost, I would like to praise Almighty Allah for blessing me to accomplish this study. It has also been through the support, help and inspiration of some people that this research has come to fruition to whom I am extremely grateful. First, I am incredibly thankful for the hard work of my principal supervisor, Dr Md Asaduzzaman. During my PhD course, our weekly meetings were instrumental in shaping my vision for this work and quickly became the highlight of my weaks, found ways to overcome challenges and carry on a belief in myself. I particularly appreciate the many conversations we had about 'telling a story' through our writing and translating ideas into words that others, potentially outside the field of mathematics and statistics, would actually want to read. I also found technical support from him either in R software packages or Latex editing and sometimes after university opening hours. I am indebted to him for his motivation, enthusiasm, patience and immense knowledge. Second, I would like to express my most profound appreciation and admiration to my co-supervisor, and my mentor, Dr Abdel-Hamid Soliman, for sharing his knowledge, wisdom, experience, and enthusiasm. Individual conversations from him have inspired me and gave me hope that I really can fill a gap in my area. I would also like to express deepest gratitude to my parents and siblings who were a constant source of encouragement for me throughout this research.

I would also like to acknowledge reviewers of my work in early stage and late stage reviews, Dr Sedky and Dr Awan at Staffordshire University. Also, I would like to acknowledge Professor Ruth Swetnam the viva chair, Dr Faisal Tariq from Glasgow university, the external examiner in the viva and Dr Abdul Wahid, the internal examiner, for their positive discussion and recommendations.I am immensely thankful to my colleagues and friends at Staffordshire University: Masum Billah, Majid Ali, Dr Anas Amjad, Samir Alagab, Wadhah Al Sibani and Sultan Al Bahri. I extend my gratitude to my cousin Dr Ghanim who supported, followed and urged me for accomplishment. My sincere thanks also go to Dr Said Al-Jelihawi, Dr Salim Sakroon, Dr Rashid Al Ghaithi, Dr Hashil Al-Saadi and Dr Mohammed Al Araimi for their expert opinion, and valuable discussions.

My wife and kids have been a pillar of support throughout my studies and I cannot thank them enough for their patience and support. Also, I appreciated some people in Stoke-on-Trent and Birimingham community during I am studying in the UK which they gave me achance to feel Like I live in my home and amember of thier commuinty such as Mahmoud ghrum, Anwar Hussain, Sami Al-Kyumi, Zafar Khan, Qamar husaain and his family...etc. Finally, I would like to acknowledge Royal Omani Police, Ministry of Labour (Oman), and National Centre for Statistics and Information for providing data, Ministry of Education, Ministry of Higher Education, Research and Innovation and Omani culture attache in London for their support and cooperation.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACF	Autocorrelation function
ADF	Augmented Dickey-Fuller test
AIC	Akaike information criterion
AR1	1st order autoregressive model
ARMA	Autoregressive moving average model
ARIMA	Autoregressive integrated moving average model
BIC	Bayesian Information Criteria
BYM	Besag–York–Molli model
EB	Empirical Bayes
EMR	East Mediterranean region
FB	fully Bayes
GCC	Gulf cooperation council
GDP	Gross domestic product
GLM	Generalised linear model
GWD	Geographically weighted deviance
GWPR	Geographically weighted Poisson regression
GWR	Geographically weighted regression
INLA	Integrated nested Laplace approximation
MCMC	Markov chain Monte Carlo
MAPE	Mean absolute percentage error
MASE	Mean absolute scaled error
NCSI	National Centre for Statistics and Information in Oman
OLS	Ordinary least square
PACF	Partial autocorrelation function
RW1	1st order random walk model

RW2	2nd order random walk model
RMSE	Root mean square error
ROP	Royal Omani Police
RTA	Road traffic accident
RTD	Road traffic death
RTDs	Road traffic deaths
RTI	Road traffic injury
RTIs	Road traffic injuries
SARIMA	Seasonal ARIMA models
SEM	Spatial error model
SLM	Spatial lag model
UN	United Nations
WHO	World Health Organisation
WOS	Web of Science

List of Publications

1. Journal paper
    (a) Al-Hasani, G, Asaduzzaman, M and Soliman, AH (2021). Geographically weighted Poisson regression models with different kernels: Application to road traffic accident data, *Communications in Statistics: Case Studies, Data Analysis and Applications, 1-16* , Taylor & Francis, Published.
    (b) Al-Hasani, G, Asaduzzaman, M and Soliman, AH (2021). Bayesian hierarchical modelling for spatio-temporal interaction with road traffic accidents data, Submitted.
2. Book chapter
    (a) Al-Hasani, G, Asaduzzaman, M and Soliman, AH (2020). Time series modelling strategies for road traffic accident and injury data: a case study, Springer Book Chapter: *Advances in Data Science and Information Engineering*, Stahlbock, R., Weiss, G.M., Abou-Nasr, M., Yang, C.-Y., Arabnia, H.R., Deligiannidis, L. (Eds.), Forthcoming.
3. Conference paper
    (a) G. Al-Hasani, M. Asaduzzaman, and A.H. Soliman. Comparison of Spatial Regression Models with Road Traffic Accident Data, In Proceedings of the *International Conference on Statistics: Theory and Applications (ICSTA'19)*, 2019, pp. 31.1-31.4, DOI: 10.11159/icsta19.31
4. Conference presentation
    (a) G. Al-Hasani, M. Asaduzzaman, and A.H. Soliman. Comparison of Spatial Regression Models with Road Traffic Accident Data, International Conference on Statistics: Theory and Applications (ICSTA),13-14th August 2019.

   (b) G. Al-Hasani, M. Asaduzzaman, and A.H. Soliman. Evaluation of generalised geographically weighted models for different kernels with road traffic accident data. Presented at the Young Statisticians Meeting (YSM 2019), University of Leeds, 31st July - 1st August 2019.

   (c) G. Al-Hasani, M. Asaduzzaman, and A.H. Soliman. Evaluation of geographically weighted Regression models in different kernels with road traffic accidents, 7th Postgraduate research students conference (7th PGRS), 2019, Staffordshire University, 12th June 2019.

5. Poster presentation

   (a) G. Al-Hasani, M. Asaduzzaman, and A.H. Soliman. Modelling, analysis and Forecasting of Road Traffic Accidents, 6th Postgraduate Research Conference (6th PGR), 2018, Staffordshire University, 23th March 2018.

   (b) G. Al-Hasani, M. Asaduzzaman, and A.H. Soliman. Spatial-temporal analysis of road traffic accidents in Oman,3rd Omani open day, 2019, Newcastle University, 23rd March 2019.

Chapter 1

Introduction

This chapter presents an overview of the characteristics and constraints of road traffic accidents and road traffic injuries as well as an overview of the statistical modelling in this field. The motivation of this study has also been provided with brief literature. Oman, as a case study and source of data for the implementation of the purpose of the study, was also fully described here. Afterwards, the chapter proceeds towards the aim and objectives of this work. A brief summary of the contributions to knowledge by this research study is also given. Finally, the outline of the rest of the chapters of the thesis is presented.

1.1 Background and rationale

Over the last decades, the transport system has been used rapidly across the countries or over borders. Although the highways were developed to contribute to economic growth with rapid urbanisation and modernisation, the use of motor vehicles increased exponentially across the world. The unprecedented boost in the motor vehicles introduced significant challenges due to the availability of inexpensive private vehicles to mid-income social classes. Consequently, global societies face direct and indirect costs such as road traffic accidents, air pollution, noise etc. The exceeding number of road traffic accidents lead to death, disabilities and many others issues that are threatening the human community.

1.1.1 Road traffic accidents dilemma

Road traffic accident (RTA) is one of the prime reasons for fatalities and disabilities globally. It has been considered as one of the significant health problems in term of death and disability (Islam and Al Hadhrami, 2012; Staton et al., 2016; Boulieri et al., 2017). Over fifty million injuries occur and more than 1.2 million people die in roadway-related accidents yearly worldwide. The RTA is going to be the fifth main cause of death in the world by 2030 (Mannering and Bhat, 2014). Since over 50% of young adults, aged 15 to 44 years die due to RTA, a significant economic impact is discernible through the loss of earning upon their families (Peden et al., 2004). Moreover, road crashes including deaths and injuries cost from 1% to 2% of the gross national product in low and middle-income countries in addition to the total development aid received by these countries.

Based on the statistics, 1.2 million people die annually caused by RTAs globally as well as around millions of others suffer from harms and some form of permanent disabilities that are acting massive impact on the communities (Mannering and Bhat, 2014). Many young persons are killed on the road traffic accidents every day globally. Therefore, RTAs were ranked 11th, and 9th cause of death and disability, respectively in the world in 2008 (Al-Lamki, 2010). Furthermore, it has been found that more than half of the RTA's victims were young adults who are the earning members for a family which has meant other socio-economic challenges in the societies (Peden et al., 2004; Al-Reesi and Al-Maniri, 2014; Hamed, 2016). Therefore, a vast amount of expenditures for safety countermeasures, laws to using roads traffic and many regulations are forcing the automobiles industry as a result of the enormity of road traffic accidents (RTA) on the societies (Mannering and Bhat, 2014). In addition, within low-income and middle-income nations, the burden of RTA is represented around 2% of their gross national product in excess of the full development aid payments from international organisations or governments to support these communities and their governments (Peden et al., 2004). Over the last decade, the East Mediterranean Region (EMR) had the highest average number of RTA mortalities and morbidities (Al-Maniri et al., 2013). Furthermore, road traffic injuries RTIs are the prime reasons of disability-adjusted life years lost in the Sultanate of Oman, Qatar, Saudi Arabia, Kuwait, Bahrain and the United Arab Emirates. This cost governments expenses concerning treatments, medicine, rehabilitates, funeral and the country loses that youngs' production. It has meant the Gulf cooperation council (GCC) countries suffer from losing victim's production since RTA leads to disability or serious hurt and early death mid adults in the GCC. Because of RTA problems, inpatient deaths (died in the hospitals) is number one in the Sultanate of Oman regarding official reports of the Ministry of Health in the as Sultanate (Islam and Al Hadhrami, 2012). Al-Aamri et al. (2017) underlined that the most harm of severe and fatal road traffic injuries RTIs in the Sultanate was virtually referred to driving speed behaviour of the young males between 20 to 29 years old.

1.1.2 Road Traffic accidents in Oman

Oman is situated in the Arabian Peninsula's southeastern tip, and its economy is virtually considered one of the emerging economies. The total land area of Oman is 309,000 square kilometres, and the population is 4,660,153 (NCSI,2018). After the Kingdom of Saudi Arabia, Oman has second size ranking among the Gulf Cooperation Council (GCC) countries (Qatar, Kingdom of Saudi Arabia, Kingdom of Bahrain, Kuwait and the United Arab Emirates). In 2011, the Omani government divided the country into eleven governorates as administrative division according to the Royal Decree No. 114/2011, Figure (1.1). As shown in Oman map [Figure (1.1)], the governorates are 1) Muscat governorate which contains the capital city Muscat; 2) Musandam governorate; 3) Al Buraimi governorate; 4) Al Dakhilya governorate; 5) Al Batinah North governorate; 6) Al Batinah South governorate; 7) Al Sharqiyah North governorate; 8) Al Sharqiyah

South governorate; 9) Al Dhahira governorate; 10) Al Wusta governorate and, 11) Dhofar governorate. In Figure (1.1), governorates from north to south of Oman are numbered as shown above. The country shares its border with the north of Arabian Gulf and Iran, and the West of UAE and KSA while Yemen in the south. Oman sea and Arab sea borders Oman from the East. The regime of Oman is Sultani, and the Sultan is the head of the government and prime minister. Parliament consists of two councils first Al-shura council whose members are elected. Second, Aldwlah council (Council of State) whose member are selected by the Sultan of Oman.

Oman is a Gulf country rich in oil. The oil and the gas were discovered, and due to the boom in prices of oil since the middle of the twentieth century, the economy of the country and lifestyle of the peoples of the Arabian Gulf states has developed, including the Sultanate of Oman, with the Gross domestic product (GDP) and per capita income growing swiftly in these states (Islam and Al Hadhrami, 2012). Oman is ranked as a high-income country by the United Nations organisation, and the GDP was $13,621$ sterling pounds (US$18,080) in 2016 (WHO, 2019).

Since the late 1990s, Sultanate of Oman witnessed greatly improved premium road network, because of the major outlay of the transport sector. Nowadays, roads are dual carriage ways in most of Oman's governates either in the cities or rural areas, lighted up properly at night, with fitted roundabouts and shoulders, well equipped with traffic signs and signals, a standard width on the roads bridges and tunnels. The total length of roads reached $63,000$ km (ROP, 2015). In the last forty years, since the rise in economic growth, modernisation and infrastructure improvement, there has been an enormous upsurge in automobile ownership and vehicles usages in Oman. The average of registering new vehicles in Oman was over $85,000$ annually between the year 2000 and 2009 (Islam and Al Hadhrami, 2012), and the total number of vehicles was 1.37 million vehicles in 2016 (ROP, 2017a). In Oman, roads are the option for transport both in terms of human

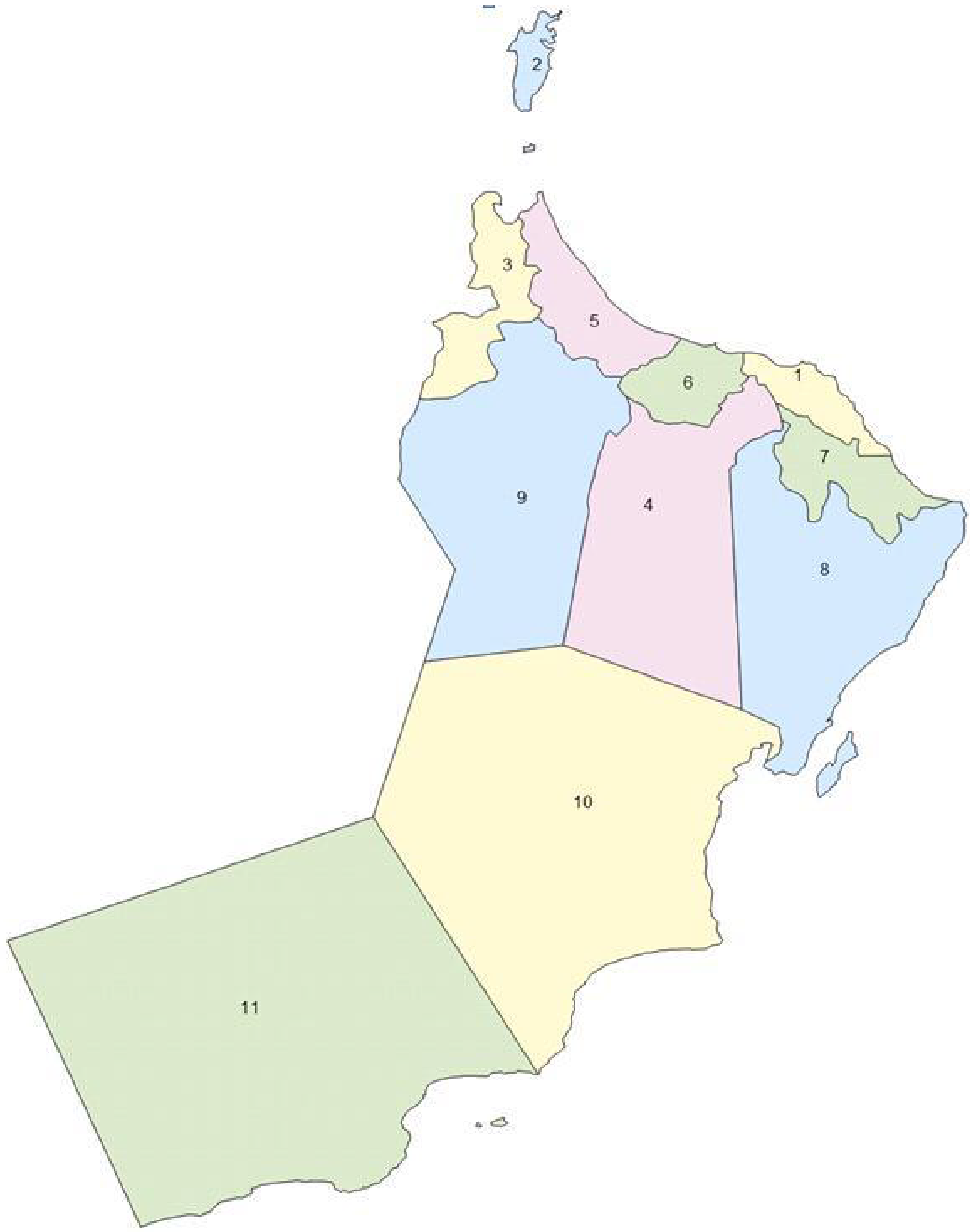


Figure 1.1: Oman governorates

occupants or stuff transport instead of other carriers like rail-way, as a result road traffic accidents (RTAs) emerged as a severe public health problem in the country (Al-Lamki, 2010; Islam and Al Hadhrami, 2012; Al-Maniri et al., 2013).

1.2 Significance and motivation

In the Arab world, Sultanate of Oman has one of the high averages of road traffic deaths (RTD) (Al-Maniri et al., 2013; Al-Reesi and Al-Maniri, 2014; Al-Bulushi et al., 2015b). Developing and emerging economy in the Arabian Gulf States such as Oman acts as a fruitful environment to investigate the primary factor that caused RTA as the country has risen to motorisation heralded with economic growth and with the increase of population (Al-Reesi et al., 2013b). As mentioned earlier, RTA has been treated as one of the significant health problems in terms of death and disability in this country (Al-Lamki, 2010; Islam and Al Hadhrami, 2012; Al-Maniri et al., 2013). Moreover, during the years from 2005 until 2009, 43,167 RTAs were recorded in the Sultanate of Oman, and thus RTAs resulted in 4,072 mortalities and 43,078 morbidities. Figure (1.2) shows the highest ten countries in the rate of road traffic death (RTD) in the world, and it shows that Oman is over the global average with 30.04 deaths per 100,000 population (WHO, 2013). Furthermore, Oman witnessed over 7,700 RTAs in 2011, averaging around one accident each hour and one mortality each 10 hours (Al-Bulushi et al., 2015b). It seems to suggest that Oman needs more efforts and studies to control and reduce traffic accidents and its impact in the society (Al Reesi et al., 2013a; Al-Reesi and Al-Maniri, 2014; Al Reesi et al., 2016).

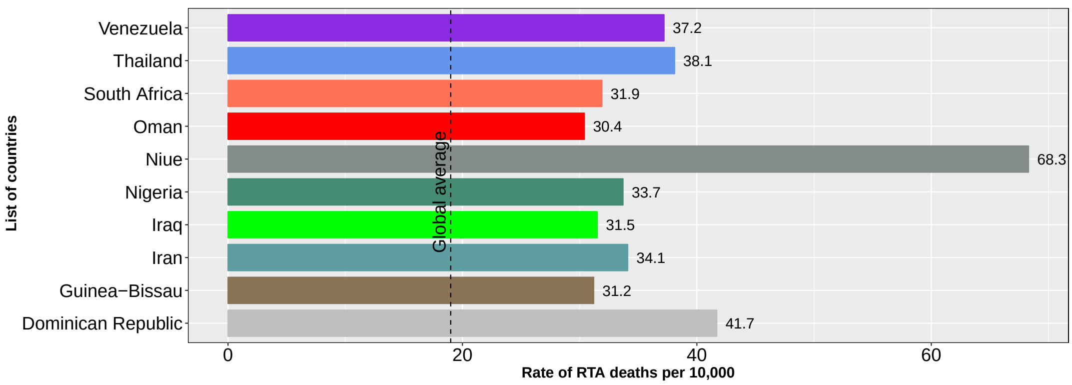


Figure 1.2: Top ten countries in accidents mortality

The safety on the roads is on the top of the priority of the subjects currently that are

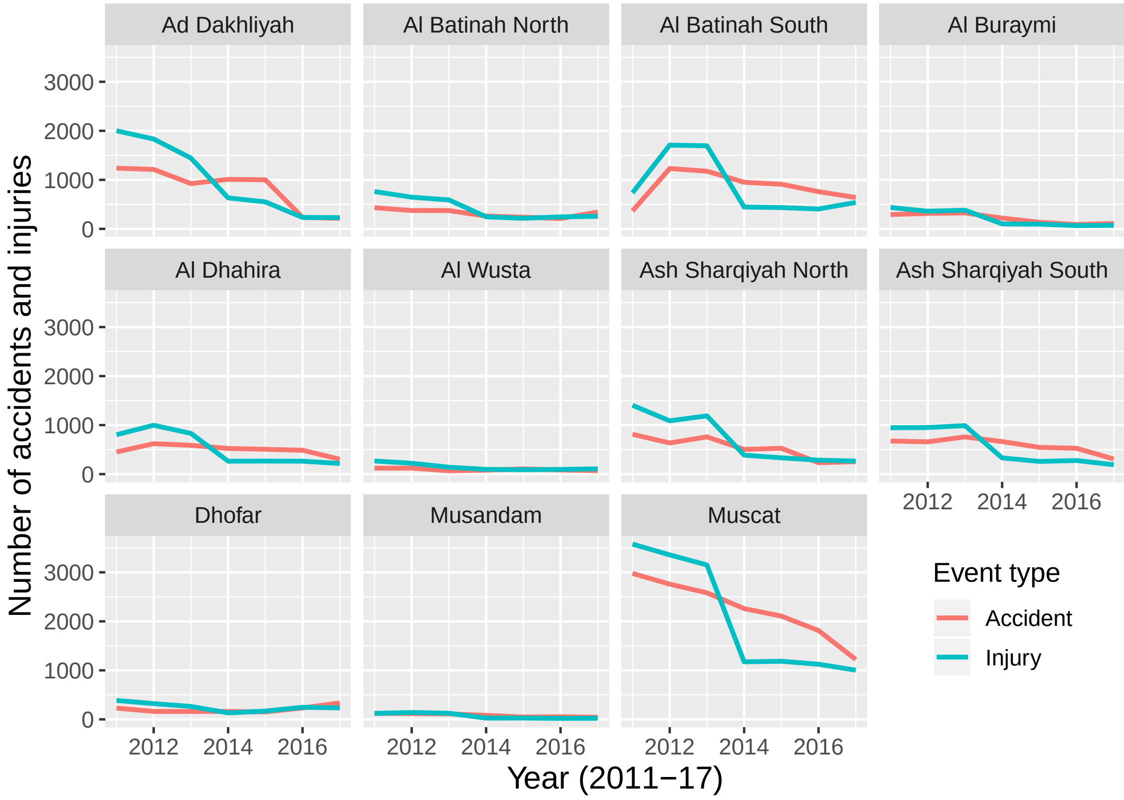


Figure 1.3: Number of traffic accidents and injuries among Oman governorates

deliberated in various countries or with numerous organisation, and so this should better to be considered in Oman (Al-Lamki, 2010). This study will highlight road accidents mortality and morbidity magnitude and compare with GCC countries and other high-income countries. Last decade, many efforts were made to reduce this problem, and this research attempts to define and explore the current situation of RTA in Oman compares to other countries in the world. It will emphasise the prime factors of RTA in Oman in order to use, model and forecast these accidents. Figure (1.3) illustrates the number of accidents and injuries among Oman's governorates from 2011 until 2017. As shown in Figure (1.3), the capital city Muscat has the highest number of RTAs and RTIs between 2011 and 2017. In fact, 2011 is the year when most accidents occurred which resulted in a vast amount of injuries among Oman governorates. However, the question has arisen if the RTA significantly random or clustering in Oman. A more in-depth investigation of RTA in Oman in terms of time and space (governorates) is needed. Also, identifying associated factors and spatial variations are required to support policies makers in the country.

Modelling, analysis and forecasting of road traffic accidents should be considered as a potential guide to allocate resources and provide a foundation for evolving policies and interventions (Zhang et al., 2015). Statistical modelling and analysis have been playing a significant role in road traffic accident research. Many researchers performed accident research to achieve three main overall objectives: (i) quantifying an effect of significant determinants (explanatory variables) on the severity of crashes or likelihood, (ii) using the parameter estimates of the statistical models to forecast accident severities or likelihoods, (iii) evaluating the efficiency of identified safety countermeasure or changing of a factor could influence on the likelihood or severity of crashes (Mannering, 2018). Generally, it appears that two prime types of mathematical models were addressed: count data models or severity models on RTAs (Lord and Mannering, 2010; Savolainen et al., 2011). Recently, several reviewers indicated a new approach which could compile the main both types of models in this field studies. In the future, modellers on RTA research should develop models referring to spatio-temporal correlations, risk compensation, endogeneity and unobserved heterogeneity (Savolainen et al., 2011; Mannering and Bhat, 2014; Mannering, 2018).

Mathematical models play an important role in the maintenance of road safety by the investigators. Besides, among modern sophisticated analysis tools, statistical modelling are becoming a key instrument in this discipline. Therefore, statistical analysis of RTA data has become one of the major concerns for the development and implementation of the standard and diverse safety policies (Mannering, 2018). Statistical modelling analysis and forecasting of RTA would provide a great depth of understanding of the causes, associated factors and implement or revise policies and countermeasures to reduce the number of RTAs, RTIs and, RTDs (Lord and Mannering, 2010). Therefore, this study attempts to characterise the temporal, spatial, and combined spatio-temporal patterns of RTAs in Oman by mathematical modelling techniques.

1.3 Aim and objectives of the study

This study aims to investigate spatial and temporal effects of the road traffic accidents in Oman to inform the government departments on the likelihood, associated factors and variations and help generate policies. The specific objectives are:

1. to explore the current situation, causal factors, and modelling of road traffic accidents (RTA) discipline,
2. to investigate the trend and temporal variation by time series modelling technique,
3. to identify associated factors and variations by spatial and spatio-temporal modelling techniques and,
4. to forecast using the most suitable model and suggest relevant policies to reduce the number of road traffic accidents in Oman.

1.4 Research questions

In this study, based on the above aim, the main question is:

- How could this study develop a suitable spatio-temporal model to investigate RTA data in the Sultanate of Oman?

Also, according to the study objectives, the following specific questions are to be addressed:

1. What is the current situation of RTA? How could the causal factors of the RTA be identified? How could the causal factors of RTA be categorised? How could the RTA data be modelled?

2. What is the best time series model to forecast the number of RTA in the Sultanate of Oman? How could the time series model of RTA data in Oman be assessed and diagnosed?

3. Which would be an appropriate method to model and analyse spatial RTA data in Oman? How could I develop a suitable spatial model for better performance with Oman RTA data?

4. How could I develop a suitable spatio-temporal model for understanding and analysing of RTA data in Oman?

1.5 Research contribution

The original contributions of this work to the knowledge is twofold represented in methodological and statistical aspects as well as the application aspect. The contributions of this research study are briefly described as follows: (1) development of spatio-temporal models considering the four types of space-time interaction effects in the RTA data, which can provide a more accurate accidents estimation, (2) development of a suitable spatial model with an appropriate kernel function for the RTA data,(3) development of a suitable time series model for Oman RTA data and, (4) suggest the most suitable advanced spatial and spatio-temporal models for Oman RTA data to help generating appropriate policies for concerned departments and authorities and interested researchers globally.

The model presented in Chapter (5) would be useful in other count data applications. The space-time interacting model in Chapter (6) could be beneficial for other applications that provide a more accurate estimation for the data. Temporal modelling in Chapter (4) revealed that the trend of RTAs in Oman has a seasonal pattern. The study results show that more number of RTA occurred during the month of June, July and August of every year (2000-2019). Therefore, the concerned authorities in Oman should consider these result for RTA intervention plans. The interested researchers could also take into account seasonal months with RTA data in the Sultanate of Oman. Thus, the outcome of this thesis can help government's departments to effectively handle their budget for improving road safety in high-risk areas and to inform RTA prevention and control programs.

1.6 Organisation of the thesis

The rest of the thesis is organised as follows:

Chapter 2: State of the art in the modelling of road traffic accidents data

This chapter discusses and reviews the existing state of the art in the RTA modelling and analytical method. The chapter has provided a comprehensive survey of literature in this discipline. It has synthesised the whole picture of RTA research overview characteristics to understand the problem domain. Besides, reviewing a piece of literature for the East Mediterranean Region (EMR), Gulf Cooperation Council (GCC) with a focus on Oman to discuss further studies of RTA. The chapter also has presented and categorised the RTA factors and analysis techniques. Also, it has reviewed some basic models used in the RTA field. Finally, the chapter explains the state of arts' critical views for three types of models that temporal, spatial and Spatio-temporal models.

Chapter 3: Data collection and computation packages

This chapter presents study data sets as well as an overview of the statistical software packages. The chapter describes study data and variables for temporal, spatial and, spatio-temporal modelling and analysis. The sources of this secondary data have been explained in this chapter. Also, related software packages utilised for the modelling framework are reviewed in this chapter.

Chapter 4: Temporal modelling and forecasting of road traffic accidents data

In this chapter, we explore and review existing temporal models relevant to the scope of this thesis, namely, time series models in road safety. Time-series data were gathered from January 2000 to June 2019 from secondary sources, and statistical time-series analyses were performed. The diagnostic tools to evaluate the temporal models such as Akaike information criterion (AIC) and Bayesian information criteria (BIC) have been formulated and used. The chapter also has provided suitable temporal models for the traffic accidents and injuries in Oman.

Chapter 5: Spatial modelling of road traffic accidents data

In this chapter, we demonstrate the several types of spatial models including the implementation of those kinds of models in RTA research. A geographically weighted Poisson regression model (GWPR) is spatial count regression model class that has been presented in Chapter (5). GWPR models have been considered for many different kernel functions, including box-car, bi-square, tri-cube, exponential and Gaussian function. Likelihood function, parameter estimation and model selection criteria have been shown in details. We applied the model formulation to the road traffic accident data in Oman with eleven governorates (regions) as the spatial units.

Chapter 6: Spatio-temporal modelling of road traffic accidents data

In this chapter, we extend the current research with spatio-temporal models, where the temporal and spatial trend and the effect of factors will be combined in an inseparable model. This chapter has investigated the application of space and time models to jointly analyses RTA frequency with a fine temporal scale and found that model capturing more comprehensively. This kind of model enables to borrow strength across spatial units and over temporal scale due to underlying unobserved heterogeneity. A more sophisticated modelling approach, Bayesian hierarchical technique, has also been investigated together with the spatial and temporal random effect components in the Chapter (6). This chapter has examined relationships and impactive the factors over the spatial units through time scale. Integrated nested Laplace approximation (INLA) has been used to approximate the RTA data in Oman. In this chapter, we implemented RTA data from Oman between 2013 and 2017 involving eleven spatial units and five time units.

Chapter 7: Conclusion and future work

Finally, conclusions and further research directions are discussed in Chapter (7). This chapter has summarised the thesis and provided the concluding remarks. The future scope of the work have been proposed in this chapter.

Chapter 2

State of the art in the modelling of road traffic accidents data

2.1 Introduction

This chapter reviews the relevant studies and reports in road traffic accidents (RTAs) globally, with more focus in the Gulf Cooperation Council (GCC) countries and Oman. The main focus is to explore the current status of RTAs, their impacts and, causal factors. Also, it is necessary to explore mathematical models that have been implemented and developed in RTA field. The chapter is divided into five sections. The first section (2.2) gives an overview of RTA which included concepts and definitions, RTA situation globally, RTA situation in the Middle East and Arabian Gulf states. Second section (2.3) provides the context of RTA in Oman, where we discuss the current situation and research effort in the Sultanate of Oman. In the third section (2.4), we present the causal factors and classification of the RTA factors. Section four (2.5) explores mathematical models of the RTA and their categorisation, this section divided into three subsections, temporal models, spatial models and spatio-temporal models. At the end of the chapter, a summary (2.6) has been given.

2.2 Overview of Road traffic accidents

2.2.1 Concepts and definitions

There are many terms in the literature to express road traffic accidents (RTAs). The terms which have been used are the following: road traffic crash, road traffic collision, road traffic incident, motor vehicle accident, motor vehicle collision, motor vehicle traffic collision, auto accident, car accident, car smash, car crash, road accident, road traffic accident, and road traffic injury. In this thesis, road traffic accident (RTA), road traffic injury (RTI) and road traffic death (RTD) have been taken into the research terminology.

The road traffic accident term has taken several definitions in the literature. The WHO determined RTA as an event or sequence of events that results or could result in an injury (WHO, 1989). Henrich (1959) defined RTA as the unplanned and uncontrolled event in which the reaction of substance, person, object or radiation resulted in personal injury or the probability thereof. More definition for RTA have been found such as an event involving at least one vehicle, occurring on the road open to public circulation, and in which at least one person is injured or killed" (Ruikar et al., 2013; Al Aamri, 2018). Likewise, Karvonen et al. (1986) explained RTA as a process of parallel and consecutive events leading to a harmful consequence. However, it seems from the preceding definitions that the term RTA is closely related to the term RTI. Those definitions are inaccurate due to the fact that the two concepts RTA and RTI refer to two separate terms (Ahmed and El-Sadig, 2002). Although mutually interrelated phenomena that take the form of cause and effect, action and reaction or exposure and outcome in epidemiological terminology, but those two terms (RTA and RTI) are differnt (Ahmed and El-Sadig, 2002). Indeed, both terms are used in the state of the art methods interchangeably, but apparently from a public health perspective, both phenomena are distinguished in etiological nature. Therfore, it means that both should to be addressed and analysed separately (Ahmed and El-Sadig, 2002). A different perspective on this is provided by several studies that the two terms are not interchangeable in order to they definitely refer to two separate phenomena (Bijur, 1995; Avery, 1995; Andreassen, 1985).

A considerable amount of literature has been published on accidents definition and other terms. These studies argued the description of this area concepts and distinguished among diverse terms. Avery (1995) criticised WHO's determination of the RTA due to restricting RTA with fatalist nature. He suggested an alternative definition explaining the RTA as a sudden event or sequence of events that, for an individual or groups of individuals, is apparently unpredictable and may not result in injury (Avery, 1995). Likewise, Andreassen (1985) underlines the obvious reality that RTA could be leading to RTI (injury), but it possibly leads to other consequences like psychological harm or property damage only. On the other hand, the word injury alone is acting as a result of the process only in which an event, previously referred to as the RTA, plays a central part (Bijur, 1995). Furthermore, an injury may be sustained through RTA, but also as a result, for example, of suicidal attempt or violence may cause injury (Andreassen, 1985; Ahmed and El-Sadig, 2002).

According to cause and effect (or exposure and outcome) principle, road traffic death (RTD), a vital term should be discussed. Definitions of RTD as the result of RTA vary among different organisations and countries. This concept mentions human casualties that sustained injuries which cause death less than 30 days after RTA has occurred (Sheikh, 2009; Al Bulushi, 2017). Likewise, WHO determined RTD as a death that occurs within 30 days of being involved in the RTD (WHO, 2013). However, in some countries, the RTD has been defined as these in which death related to an RTA has occurred between the time of the accident and the closure of the case file in January of the next year (Al-Reesi et al., 2013b; Al Bulushi, 2017).

2.2.2 Road traffic accidents situation worldwide

More recent attention has focused on the provision of road safety. Indeed, most global countries are confronting growing the number of RTAs, RTIs and RTDs, which consequence as a high economic burden. Based on WHO, the RTA are the ninth reason for injuries and deaths around the world (WHO, 2013, 2015). Moreover, current trends expect that if corrective action is not taken, RTA will be the seventh reason for disability and death by 2030 among the world population (Al-Bulushi et al., 2015b; Al Bulushi, 2017). Therefore, the alarming growth of RTA, RTI and, RTD is acknowledged to be a global phenomenon. Concisely, RTA research approaches are grown to determine, guide, and describe road safety actions to overcome this phenomena's burden. In 2010, the Moscow Declaration was declared at the First Global Ministerial conference on the global safety of road taken place in Russia (Al-Lamki, 2010). At that UN general conference, around 1,500 participants were attended, some of the participants are senior ministers of some countries and representatives of UN agencies. The participants in the Moscow Declaration declared a Decade of Action for Road Safety from 2011 to 2020. It became clear at that global conference that the statistics for RTA will be worse than initially thought (Al-Lamki, 2010). The Global Status Report on road safety series have been published by WHO to monitor progress through. These reports were published as series in 2013, 2015 and 2019, giving a better understanding of the current situation of road safety globally. Figure 2.1 shows the trend of RTD globally from 2000 till 2016, which indicated to a bit steadiness in the number and rates of RTD (WHO, 2019). Fundamentals, WHO categorised the world into six regions: African, Eastern Mediterranean, European, Region of the Americas, South-East Asian, and Western Pacific region. Researchers and decision-makers investigated the effective policies in each region to achieve the Decade of Action for Road Safety goals (WHO, 2013, 2015, 2018, 2019).

More recently attention has focused on the provision of road safety. In the last century, the high burden influencing caused by RTA has led to rising primary concern and concentrate on this issue, especially in the motorising countries. Initially, on 30th May 1896 the first accident was recorded in NewYork and then after a few months first RTD occurred in London (Peden et al., 2004; Mastinu and Plöchl, 2014). Since then, RTAs are growing and considered to be the main cause of misfortunes, calamities, and costs and human losses in life and property in human society. Currently, worldwide the RTA is the number one cause of deaths among young adults and children aged 5-29 years (WHO, 2019). In the European region, RTD reached 84,589 people in 2013, that meant more than 230 persons every day are killed on the road (Jackisch et al., 2015). However, WHO reported that Europe made significant progress in the prevention because the RTD rate decreased to 9.3 death per 100,000 population in 2018 while the global rate was 18.2 per 100,000 population (WHO, 2018). Similarly, Americas (15.6 per 100,00) and western Pacific (16.9 per 100,00) have a low RTD rate compared to other regions (Eastern Mediterranean, Africa and South-East Asia) (WHO, 2018).

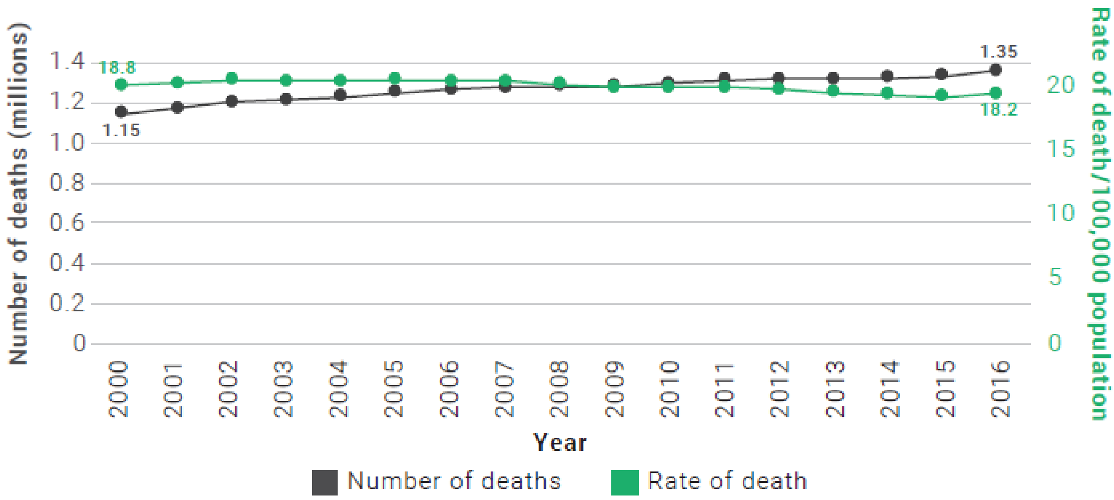


Figure 2.1: Number and rates of RTD in the world between 2000 and 2016 (WHO, 2019)

2.2.3 RTA situation in the Middle East and Arab Gulf

As mentioned earlier, WHO usually reports the trend of RTA and effective safety policies which vary extensively by WHO regions. Based on the WHO reports series on road safety, the Eastern Mediterranean Region (EMR) consists of 22 countries: Afghanistan, Bahrain, Djibouti, Egypt, Iran, Iraq, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Pakistan, Palestine, Qatar, Saudi Arabia, Somalia, Sudan, Syria, Tunisia, United Arab Emirates and Yemen (WHO, 2013, 2015, 2018, 2019). In 2004, the EMR has recorded the highest RTD rate among people aged 15-29 years, this rate was 34.2 deaths per 100,000 population (Al-Maniri et al., 2013; Al-Aamri et al., 2017). Furthermore, the RTD in EMR was approximately 4.5 times higher than RTD rate in those countries which the countermeasure policies in the top in 2010 (Dahdah and Bose, 2013; Al Aamri, 2018). More than 127,000 persons were killed at EMR in 2010, representing 10% of the global RTD (WHO, 2010). Furthermore, WHO ranked the EMR as the second-costliest impacted by RTA worldwide in 2013 (WHO, 2015). Consequently, RTDs cost around GBP 5.6 billion (US$7.4 billion) yearly in rehabilitation or lost production in the EMR countries (Al-Bulushi et al., 2015b). Although world alarmed of RTA significant and announced several recommendations in Mosco declaration, EMR was impacted by RTA more after that declaration. In 2013, WHO (2015) found the EMR has the second-highest RTD rate in the world with 19.9 RTDs per 100,000 population. Recently, the last global status reports by WHO are showing that the RTD rate has decreased in EMR since it reaches 18 RTDs per 100,000 population but still the region ranked third in 2016 (WHO, 2018, 2019).

Within the EMR region, the trouble situation of this phenomenon differs significantly between countries. A growing body of literature has studied the RTA situation in EMR countries. Low- and middle-income countries in EMR have higher RTD rates corresponded to higher-income countries in this region, and this is predicted to rise in the following decades (Al Bulushi, 2017). Likewise, Sengoelge et al. (2018) analysed the EMR countries based on three levels of income and user type, which underlined remarkable differenc between low-income countries and world rate. Hamid and Davoud (2019) indicated that the EMR is the only region where the higher income-countries have higher RTD rate than low-middle income countries. However, Hamid and Davoud (2019) relied on only one-year data and excluded a middle-income country that represents 3.5% of total EMR population. Supported by road safety status report, RTDs are more than twice in middle-low income countries than high-income countries (WHO, 2015; Al Bulushi, 2017). However, there are variations of RTD rates among the EMR countries based on the country's economic level. Therefore, a very high RTD rate in low-income Afghanistan and Yemen, middle income Sudan and Iran, and high-income countries Oman, Qatar, and UAE are found (Sengoelge et al., 2018).

Arab world faces around 4.8% of the number of RTDs even though Arab states represent solely 3.6% of all world population in 2004 (Peden et al., 2004; Al-Aamri et al., 2017). Within the Arab world and the EMR, there are five countries laying on the Arabian Gulf known as the Gulf Cooperation Council (GCC) countries: Bahrain, Kuwait, Oman, Qatar, Saudi Arabia and the UAE. All countries in the GCC are ranked high-income countries in the EMR by WHO (WHO, 2013, 2015, 2018). Unfortunately, GCC countries recorded a high number of deaths due to traffic accidents during the last two decades. In 1995, the RTD rate at GCC was 32.4 people per 100,000 population whereas the RTD rate in the global was 21.5 (Sengoelge et al., 2018). A similar situation of traffic mortality at GCC in the beginning of the current century, since the WHO reports point out that RTD rates in GCC are 28.1, and 23.8 compared to a global rate 21.0 and 18.4 per 100,000 people in the year 2005 and 2015 respectively (Sengoelge et al., 2018; WHO, 2018). Moreover, in Oman and the UAE as such countries in the GCC, the number of RTDs increased in 2016 compared to the year before (Sengoelge et al., 2018).

Since the 1970s, the lifestyle of GCC people has changed remarkably during the booms of oil prices. Indeed, the gross domestic product has been increasing that attracts an enormous number of human force powers from different countries in the world. Therefore, population size expanded that lead to growth of roadway networks and the number of vehicles in the GCC countries (El-Sadig et al., 2002; Al-Reesi et al., 2013b; Dahdah and Bose, 2013). However, all those modernisation among GCC countries play a negative role in term of RTA and consequent high burden in deaths and injuries (Bener et al., 1994; Dahdah and Bose, 2013). Table (2.1) compares the rate of RTD in GCC countries except Bahrain (no data available) with other two high-income countries: the United Kingdom (UK) and United States (USA) (WHO, 2019). Although there is a high number of cars and a higher population in the UK and USA than all GCC states, RTD is under control 3.1 and 12.4 per 100,00 population respectively Table (2.1). Clearly, these high-income countries (UK and USA) built the countermeasures of road safety rely on a scientific approach. Therefore, Arabian GCC states as high-income countries should follow developed and other high-income countries such as the UK and the USA to protect their citizens of the RTA dilemma. Yet, a quantity and quality of the publications about RTA situation are insufficient to face the increasing related public health and economic cost at the GCC states (Butt et al., 2020).

There is a large volume of published studies describing the role of RTA in GCC countries. Ansari et al. (2000) found two primary reasons which caused 65% of RTAs in Saudi Arabia, which are: excessive driving speed and drivers disobeying traffic signals. In the UAE, El-Sadig et al. (2002) indicated the reasons for RTA are careless driving, speedy driving, the changing vehicle mix on the roads and, the standard of immediate care available for victims. Recently, Jamal et al. (2020) attributed the highest number of RTAs in Saudi Arabia to exposure factors including overspeeding, distraction, sleep, collisions with other moving vehicles, sudden deviation, road fences, pedestrians and rainy weather conditions. Besides, from 1971 to 1997 it was estimated one person was killed and four people injured every hour on Saudi's roads (Ansari et al., 2000). In the same country, Nofal et al. (1996) revealed that the major RTI occured in the month of Ramadan (the holy month of fasting for Muslims). In Qatar, Timmermans et al. (2019) indicated an increasing trend of the number of RTAs that resulted in severe RTIs but decreasing trend the number of RTDs between 2010 and 2016. However, several studies indicated the more preparation of RTDs among GCC countries with pedestrians accidents (Nofal et al., 1996; Timmermans et al., 2019; Jamal et al., 2020). Overall, in 2000, Ahmed and El-Sadig (2002) pointed out Kuwait was the lowest and best compared to all GCC countries in the number of RTAs while the UAE has relatively better RTA rates, ranking second behind Kuwait. However, in 2016, Qatar is the lowest country of the RTD rate while Saudi Arabia is the worst with 28.8 people per 100,000 population killed, as shown in Table (2.1) (WHO, 2019). Broadly speaking, RTA cost is 3.9% of the GCC states' annual GDP and Oman has the worst ranking since 7.4% of its annual GDP is for the RTA cost (Dahdah and Bose, 2013; Al Aamri, 2018).

Table 2.1: Comparative road traffic deaths rates of nations in 2016

Nation	GDP(£) per person	Population	RTD/100,000 population	Number of vehicles
Kuwait	31,243	4,052,584	17.6	2,001,940
Oman	13,621	4,424,762	16.1	1,370,913
Qatar	56,714	2,579,804	9.3	1,330,478
Saudi Arabia	16,304	32,275,688	28.8	6,895,799
UAE	30,344	9,269,612	18.1	3,391,125
United Kingdom	31,775	65,788,572	3.1	38,388,214
United States	42,112	322,179,616	12.4	281,312,446

2.3 Context of RTA in Oman

As mentioned earlier, Oman is an Arabian state in the Eastern Mediterranean Region (EMR) which is ranked a high-income country by WHO as well as all GCC states. Oman government commenced utilising the prime revenue such as oil and natural gas for basic construction needs such as health, education, transportation and telecommunication (Al-Risi, 2014). Similar to other GCC states, since 1970s Sultanate of Oman has been transformed to a modern state, including infrastructures and all facilities as well as the living standard which has improved rapidly (Islam and Al Hadhrami, 2012). Therefore, roads traffic networks have been developed remarkably that was leading to rise in the motorisation rate (Al-Lamki, 2010; Islam and Al Hadhrami, 2012; Al-Reesi et al., 2013b). However, since the roads are the sole choice for human transport and logistics stuff alternatively of other kinds of transports, RTAs resulted in a severe public health problem in Oman (Al-Lamki, 2010; Islam and Al Hadhrami, 2012; Al-Maniri et al., 2013; Al-Aamri et al., 2017). There are several attempts being conducted for different aspects of RTA in Oman. However, Butt et al. (2020) reviewed the productive studies of RTA at GCC groups that have been published from the Web of Science (WOS) indexing and abstracting database journals. That study has found only 19 publications related to RTA in Oman from 1981 to 2019 which has meant more research is required in this field in this country (Butt et al., 2020).

A considerable amount of literature has been published on RTA discipline in Oman. As such, it has some studies about the young drivers (Al-Sinawi et al., 2012; Al Reesi et al., 2013a, 2016, 2018), heavy vehicles (Al-Bulushi et al., 2015b,a; Al Bulushi, 2017), the effect of economic growth on RTA (Al-Reesi et al., 2013b), the effect of motorisation on RTA (Islam and Al Hadhrami, 2012), the effect of medical conditions on RTA (Chitme et al., 2018), assessment of livestock-vehicle traffic accidents (Abdo and Al-Ojaili, 2015), causes and factors of RTA (Abdo, 2006; Plankermann, 2014; Al-Aamri et al., 2017; Javid and Al-Neama, 2018; Al-Maimani and Yahia, 2019), etc. Moreover, there are a few studies which analysed and evaluated RTA in a certain governorate in the Sultanate such as Muscat (Ramana et al., 2018; Al-Aamri et al., 2020), Dhofar (Farag et al., 2014; Abdo and Al-Ibrahim, 2016; Farag and Hashim, 2017), while few studies attempted to predict and forecast of RTA, RTI and, RTD (Al-Reesi et al., 2013b; Narasimhan et al., 2017; Al-Hasani et al., 2019b). Moreover, the National Committee for Road Safety in Oman is leading national road safety strategy that targets reduction of RTDs by 25% between 2011 to 2020 (WHO, 2019). Furthermore, the Research Council (TRC) in Oman (former) funded several research proposals under the road safety program and gave scholarships to postgraduate students in RTA research efforts (Al-Lamki, 2010; WHO, 2019).

Even though the efforts mentioned above are undertaken to understand, analyse and, control the number of RTAs and their impacts in Sultanate of Oman, the situation and trend of RTA, RTI and RTD is going upwards over time among various cities in the country. Indeed, Oman has one of the highest numbers of RTA worldwide, which is costing the country socio-economic burden through the last fifty years (Al-Lamki, 2010; Plankermann, 2014). In 2000, a national health survey in Oman demonstrated the first reason for morbidity and mortality is RTA that represented 61% (Islam and Al Hadhrami, 2012; Al Aamri, 2018). Despite the fact that the number of RTDs was reduced in 2010, but it increased again by more than 30% in 2012 (Plankermann, 2014). However, in 2010, the number of RTDs was very high and the country ranked number ten worldwide and 3rd in the EMR since RTDs reached 30.04 people per 100,000 inhabitants (WHO, 2013). It has meant, the cost of lost life, grave, pain. etc is high that need to solved in RTAs hotspot places of Oman governrates. In 2013, RTD rate was 25.4 in Oman compared to the global and EMR rates of 17.4 and 19.9 respectively (WHO, 2015). Recently, the rate of fatalities due to RTA in Oman is 16.1 persons per 100,000 population in 2016 while the global rate is 18.2 and EMR rate is 18 (WHO, 2018, 2019). Therefore, it is crucial to identify patterns, reasons, modelling and forecasting RTA in Oman.

Data from several sources in Oman have identified the increased RTI and RTD associated with RTA. Figure (2.2) displays the number of RTD among all Oman's governorates. Overspeeding is the main cause of RTA in the Sultanate (Al-Maniri et al., 2013; Al-Aamri et al., 2017; Ramana et al., 2018). Plankermann (2014) indicated the high employment rate, which led to increasing driving licenses and vehicle ownerships leading to more RTA in Oman. Several studies attributed the high number of RTA in Oman due to the increased number of vehicles on the highways (Islam and Al Hadhrami, 2012; Al-Reesi et al., 2013b; Javid and Al-Neama, 2018). Clearly, the highest population is in the capital city Muscat which has the highest number of RTD, as shown in Figure (2.2). The question has been raised if the population size and population density are both significant reasons for RTAs in Oman. However, Al-Aamri et al. (2020) concluded that road intersections at Muscat governorate rise the possibilities of occurring RTA than other road geometry features. The generalisability of much-published research on this issue is problematic. There is a dearth of published studies that conduct the spatial and spatio-temporal effects of RTA in the sultanate of Oman or its governorates. Therefore, this study attempts to characterise the temporal, spatial, and combined space-time patterns of RTAs in Oman by the state of the art mathematical modelling techniques.

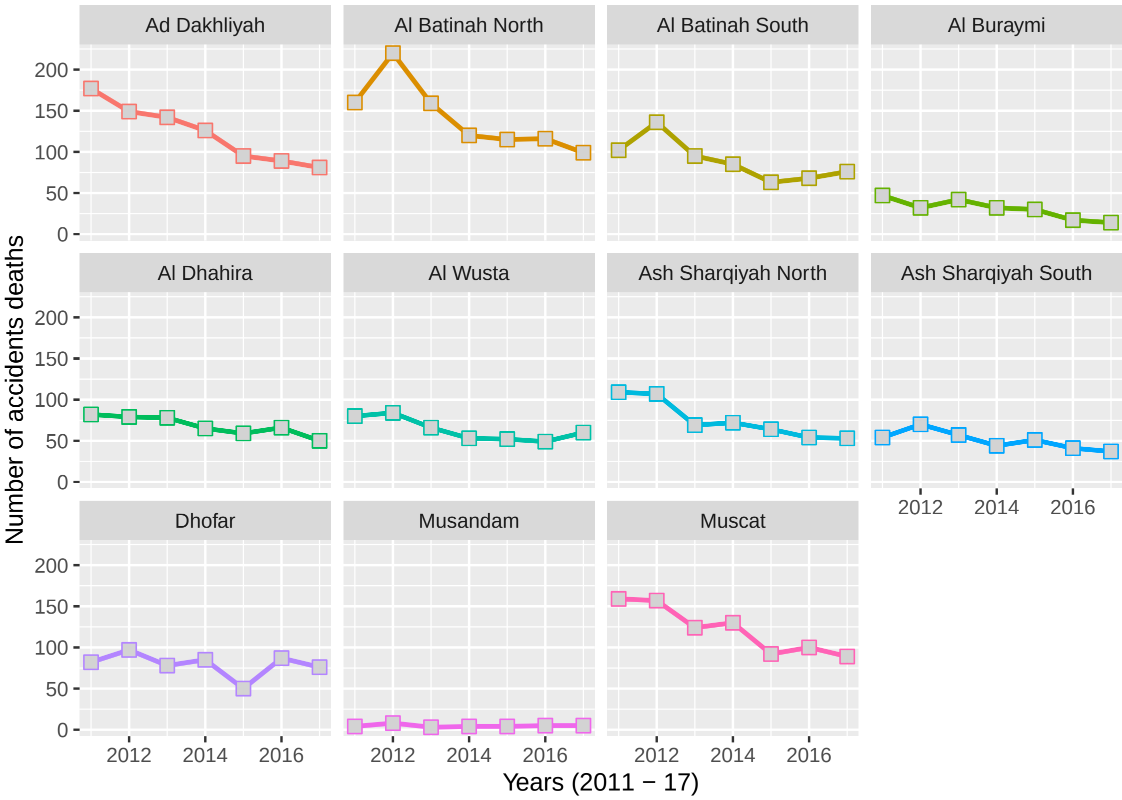


Figure 2.2: Number of road traffic deaths among Oman governorates

2.4 Causal factors for road traffic accidents

With a view to decline the road toll, the origin and nature of road traffic accidents (RTAs) empirically and theoretically investigated (Al Reesi et al., 2018). Intending to decrease the number of traffic accidents with this massive loss to the community resulting from traffic accidents, investigators have carried out research to gain the clear picture of the impact the probability of traffic accidents trusty (Lord and Mannering, 2010). They could fit of better predict the likelihood of accidents and give policies maker new directing (Lord and Mannering, 2010). Investigators carried out many studies on RTA, but they investigated different aspects through many countries or regions data. Therefore, examining the factors that caused the RTA was the first step to analyse, model and forecast RTA by the predictors. However, finding the causal factors for RTA in a particular place is a nontrivial job (Plankermann, 2014; Al Aamri, 2018).

Causal factors thought to be influencing RTA have been explored in several studies. A handful of causal factors were identified by the National Highway Traffic Safety Administration (NHTSA) such as economy, unemployment and improvements in vehicle design (Chekijian et al., 2014). Paradoxically, another study assumed increasing employment and more people obtaining driving licenses is leading to an increasing amount of RTAs (Plankermann, 2014). Numerous studies have attempted to model temporal correlation reflects the affect of varied factors concern to traffic accidents RTA. For example, Liu and Sharma (2017) modelled the following factors: travel demand, environment, law, weather and economy which often exhibit some temporal trends or periodicities. It seems that numerous factors are affecting RTA, and it is challenging to aggregate them in one model. A different classification of the causal factors for RTA have been attempted in significant quantities of kinds of literature in this field. However, this study categorises factors into five groups: human behaviour, socio-economic, environments, demographic factors and spatial or spatio-temporal unobserved factors.

There are many kinds of research which revealed that human behaviours represent the main contributing factors of RTA, RTI and RTD, as shown in Table (2.2). McIlvenny (2006) underlined that the primary factor for RTA is human behaviour factors, especially in developing countries where it is estimated to be 64% to 95%. Likewise, Islam and Al Hadhrami (2012) reported about 98% of the RTAs are referred to as drivers' behaviour factors. Also, this agreed with Al-Lamki (2010) who found that the user's error causes 90% of deadly crashes. However, Plankermann (2014) has conducted the causal factors in the role of human behaviour on occurring of RTA at GCC states concentrating on Oman. Broadly, human behaviour includes pre-crash such as speeding, attitudes, behaviour and pedestrian habits.

Overspeeding and seatbelt were the most influential of human behaviour factors (McIlvenny, 2006; Al-Madani and Al-Janahi, 2018) but Kaplan et al. (2015) reported drowsiness and distraction as the most important human behaviour factors. This is supported by some studies which have showed speed driving as the main human factors. It is easy to access the data or examine this factor within worldwide and in GCC states (Ansari et al., 2000; El-Sadig et al., 2002; Wang et al., 2013a; Plankermann, 2014; Jamal et al., 2020). Similarly, Royal Oman Police have useful data regarding speed as the main cause of RTA, which is available (ROP, 2015, 2017b,a). However, overtaking has appeared as a significant causal factor within drivers behaviour resulted in RTA (Al-Maniri et al., 2013; Plankermann, 2014). Different human behaviour may cause RTA, could be a distraction, sleep, sudden deviation and pedestrian behaviour (Jamal et al., 2020). Even though, Liu and Sharma (2017) did not cover the human factors except mileage but mileage is a compile factor point for experience instead of age. However, this is yet not enough to get a real scenario. Jovanis and Chang (1986) pooled model included vehicle milage and trucks milage with Indiana Toll Road's data. This kind of data, such as mileage are easy to access in a developed country such as the US and UK; however, it is not available in developing countries such as in EMR and GCC countries. Therefore, some studies have attempted to estimate mileage by sampling such as Al Reesi et al. (2013a) who estimated the average of annual mileage to be $20,379$ km in Oman and Ahmed and El-Sadig (2002) who estimated the average of annual mileage to be $25,000$ km in the UAE. Overall, it seems that human behaviour factors may be the prime source of RTA in Oman, includ speed driving (Al-Maniri et al., 2013; Al Reesi et al., 2013a; Ramana et al., 2018).

Some researchers conducted studies on socio-economic factors to measure their effect on traffic accidents such as (Al Reesi et al., 2013a; Al-Maniri et al., 2013; Al-Bulushi et al., 2015b; Machado-León et al., 2016; Al-Aamri et al., 2017; Al Reesi et al., 2018; Al-Madani and Al-Janahi, 2018). Indeed, Machado-León et al. (2016) extended their research to examine socio-economic factors including seven factors as follows: gender, age, occupational status, qualification, household size, net family income level and driving experience. Gender and age were dominating explanatory on social factors for RTA, while Al-Aamri et al. (2017) attempted to disentangle between age and gender in RTA (for Oman data) than other researchers who examined age and gender normalcy. However, some researchers considered age and driving experience as one factor and, they compiled age and experience as a one-factor such as (Hamed, 2016). Since there is a dearth of data in developing countries such as Oman, age is the proxy of other factors such as driving experience (Mannering et al., 2016). Some researchers could consider age as a pointer for an experience which is shown in Ahmed and El-Sadig (2002) where the age of under forty is added as a factor in their models. It appears that a model with the integration of mileage and age is expected to provide more vital outcome and forecasting in case there is no data for drivers' experience as a factor. In Oman, several studies emphasise that young drivers are in more risk than others (Al Reesi et al., 2018, 2016, 2013a; Al-Maniri et al., 2013). However, unemployment rate plays a role in this discipline, and it is considered for models predictors (Li et al., 2013). On the contrary, employment rate can be considered as exposure factor in the RTA model (Plankermann, 2014). It seems that if there is no other transport to commute in a particular country of some regions, this is leading employments people to use private car. This increases the number of vehicles on the road leading to high probabilities of occurring RTA. Therefore, ignoring the economic growth in developing country leads to inaccuracies in the forecast and several studies considered Gross Domestic Product (GDP) as the prime factor. As such, in Oman, there is a shred of substantial evidence that economic growth affected RTA (Al-Lamki, 2010; Islam and Al Hadhrami, 2012; Al-Reesi et al., 2013b; Al-Maniri et al., 2013).

Many researchers considered virtual factors like environments and road defects, as shown in Table (2.2). As such studies, Liu and Sharma (2017) investigated the number of days with the minimum temperature higher than 32°F, rainfall and snowfall with Iowa RTA data. Similarly, Ma et al. (2017) developed the models with a wet surface and November as indicator factor in Colorado. Both studies conducted for the USA and these factors are out of our research scope as these are particular circumstances and weather in the USA. On the other hand, several studies consider roads defects or road geometry by taking a variety of factors into study account. For instance, Li et al. (2013) used traffic patterns and road networks. Likewise, Ma et al. (2017) conducted shoulder width, the number of entering ramp per mile per lane, two-lane indicator and poor pavement indicator. In GCC countries, Jamal et al. (2020) revealed the significant evidence for road fences and rainy weather condition in the east region of Saudi Arabia. In Oman, ROP published general data like the number of accidents caused by weather, vehicle defects and road defects (ROP, 2014).

There are many studies which included the populations and number of vehicles in the specific region, continent, country, territory and city in this discipline in either developed or developing countries. It is important to consider those two factors such as in El-Sadig et al. (2002); Islam and Al Hadhrami (2012); Al-Reesi et al. (2013b); Al Reesi et al. (2016); Al-Aamri et al. (2017); Liu and Sharma (2017). Some of these studies on the GCC group are representatives as significant evidence for considering both factors for modelling, forecasting, and analysing road safety in these states. Moreover, Smeed (1949) investigated population and number of vehicles to estimate the number of fatalities caused by RTA. Besides, this model developed by Smeed (1974) that is utilising the researchers in this discipline. Smeed models could be used to forecast the number of traffic fatalities in both developing and developed countries (Ahmed and El-Sadig, 2002). Al-Reesi et al. (2013b) applied Smeed(1949) model with Oman data for the period from 1985 to 2009.

Nowadays, there is massive amount of literature that takes into account endogeneity spatial and spatio-temporal unobserved factors in many fields. Huang and Abdel-Aty (2010) proposed five levels of causal factors for RTA data based on the macroscopic and microscopic levels. This study refers the RTA to the macroscopic level with spatial units (geographic units) such as regions, counties, cities (Huang and Abdel-Aty, 2010). On the other hand, the study refers to the microscopic level with traffic site (location), traffic crash, driver-vehicle unit and occupant (Huang and Abdel-Aty, 2010). In the same way, Aguero-Valverde and Jovanis (2006) suggested to take into model account the following factors: demographic characteristics, accessibility to hospitals and road infrastructure functional class. Furthermore, Parvareh et al. (2018) posed geographic and environmental factors when they modelled RTA data and forecasted RTA. Recently, in the RTA studies there are evidences indicating that geographically, temporary heterogeneities factors should be considered individually or together separately or jointly (Liu and Sharma, 2018; Hezaveh et al., 2019; Cheng et al., 2020). Given all that has been mentioned so far, one should consider causal factors as predictors in the model with taking into account spatial (geographic) and spatio-temporal unobserved factors.

To summarise, Table (2.2) shows the previous studies which used various factors in five categories for RTA data.

Table 2.2: Various RTA factors in five categories

Type	Study
Human behaviour	Jamal et al. (2020),Al Reesi et al. (2018),Ramana et al. (2018), Ma et al. (2017), Hamed (2016), Kaplan et al. (2015), Al-Bulushi et al. (2015b) Barua et al. (2015),Plankermann (2014), Al Reesi et al. (2013a), Wang et al. (2013a), Islam and Al Hadhrami (2012), Al-Lamki (2010), McIlvenny (2006), El-Sadig et al. (2002), Ansari et al. (2000), Jovanis and Chang (1986)
Socio-Economic	Al-Madani and Al-Janahi (2018), Al Reesi et al. (2018), Al-Aamri et al. (2017), Boulieri et al. (2017), Liu and Sharma (2017), Machado-León et al. (2016), Al Reesi et al. (2016), Al-Bulushi et al. (2015b), Al Reesi et al. (2013a), Wang et al. (2013a), Al-Reesi et al. (2013b),Li et al. (2013) Islam and Al Hadhrami (2012), Al-Lamki (2010), El-Sadig et al. (2002).
Environments and road defects	Jamal et al. (2020), Parvareh et al. (2018), Ma et al. (2017), Liu and Sharma (2017), Hamed (2016), Kaplan et al. (2015), Barua et al. (2015), Li et al. (2013), Aguero-Valverde (2013), Al-Lamki (2010), Jovanis and Chang (1986).
Demographic	Parvareh et al. (2018), Boulieri et al. (2017), Al-Aamri et al. (2017), Liu and Sharma (2017), Al Reesi et al. (2016), Al-Reesi et al. (2013b), Islam and Al Hadhrami (2012),El-Sadig et al. (2002).
Spatial and spatio-temporal factors	Al-Aamri et al. (2020),Cheng et al. (2020),Li et al. (2019), Hezaveh et al. (2019),Wen et al. (2019), Liu and Sharma (2018), Parvareh et al. (2018), Boulieri et al. (2017), Rhee et al. (2016),Xu and Huang (2015), Pirdavani et al. (2014),Li et al. (2013), Huang and Abdel-Aty (2010), Quddus (2008a), Aguero-Valverde and Jovanis (2006).

2.5 Statistical modelling in RTA overview

Road traffic accident (RTA) investigators have applied an immense diversity of methodological techniques over the years in order to gain more understanding and mitigation of RTAs. Despite the diverse methods, the fundamental characteristics of RTA, the limitation of data and methodological results are insufficient and not fully understood (Savolainen et al., 2011). Overall, Mannering (2018) pointed out three overall objectives on the statistics of RTA data analysis. First, the analysis of data with the purpose of quantifying an effect of significant determinants (explanatory variables) on the severity of crashes or likelihood. Second, analysing data with a plan to use the outcoming parameter estimates of the statistical models to forecast crashes severities or likelihoods. Third, studying before and after the specific day to evaluate the efficiency of identified safety countermeasure or changing a factor could influence the likelihood or severity of RTAs, RTIs or RTDs. However, Mannering and Bhat (2014) mentioned other advanced models that attempted to treat some subtle issues in traffic data like the impact of unobserved factors in accidents frequencies, spatio-temporal correlations, and so on that had been addressed with the stable progression of the advanced methodological techniques in this discipline [Figure(2.3)]. To date, many statistical models have been implemented and/or developed to perform RTA, RTI and RTD analysis and prediction.

Hughes et al. (2015) reviewed safety models from a variety of areas such as health, food, industry transport, construction, education and occupational safety. However, this study has categorised seven types of RTA models: sequence models, mathematical models, component models, intervention models, process models, system models and safety management models. Mathematical models are used for quantitative analysis of the data and their relations and used to analyse the impacts of precise strategy or policy (Hughes et al., 2015). Generally, in the RTA research field, it appears that two types of mathematical models were addressed: count data models or severity models of RTA. This study attempted for modelling RTA under consideration of both types, as shown in Figure (2.3).

Lord and Mannering (2010) surveyed several methodologies which have been utilised by researchers for decades. The most basic model is the Poisson regression model, where both mean and variance are assumed to be equal. As we know, the classical Poisson model gives the number of occurrences in the particular time period of an event (such as RTA, RTI, RTD, etc.), that occurs randomly at a constant average rate. Although this model is easy for estimation and fitting, it is affected negatively by small sample size bias and low sample means and cannot handle it over and under dispersion (Lord and Mannering, 2010). Jovanis and Chang (1986) applied Poisson model with six predictor variables and data were collected from Indiana toll road where a vehicle can solely enter and leave the tollway. Likewise, Joshua and Garber (1990) conducted linear and Poisson regressions to determine the influence of traffic and roadway geometric on trucks accidents rate annually. Although mathematical models have been progressed, Poisson variants dominate the methodological approach in the RTA discipline (Mannering and Bhat, 2014). Ye et al. (2013) used a method of maximum simulated likelihood estimation to apply a joint Poisson regression model with multivariate normal heterogeneities. The study found that the joint Poisson regression model improves the efficiency of coefficients as their standard deviations are decreased.

The amplified versions of Poisson model are the negative binomial or Poisson-gamma models that control possible overdispersions. Negative Binomial (Poisson-gamma) is a generalisation of the waiting time for the numbers of success in a sequence of independent experiments. The gamma probability distribution of the Poisson parameter is assumed in the negative binomial (Poisson-gamma) model (Lord and Mannering, 2010). However, modelling RTA by using Poisson-gamma model was done by some researchers over the decades such as Maher and Summersgill (1996); Amoros et al. (2003); Abbas (2004) etc.

Several investigators have offered using the Poisson lognormal model as a better option than the Poisson-gamma model for modelling traffic data (Miaou et al., 2005; Lord and Miranda-Moreno, 2008; Aguero-Valverde and Jovanis, 2008). The difference between the Poisson lognormal model and the Poisson-gamma model is that researchers use lognormal distribution to compute the Poisson parameter instead of the gamma distribution. Aguero-Valverde (2013) compared Poisson lognormal, Poisson gamma and Zero-inflated random effects by using 865 rural two-lane segment dataset from 2003 to 2006 in Pennsylvania. It appeared that ranking of the fixed over-time random effects models are very consistent and with the random effects over time, the standard errors of the RTA frequency estimates are significantly decreased for the most segments on the top of the ranking (Aguero-Valverde, 2013).

Since 1962 the Conway and Maxwell have introduced the Conway-Maxwell-Poisson model as the extension of the Poisson distribution in order to model queues and services in that time (Conway and Maxwell, 1962; Lord et al., 2008; Lord and Mannering, 2010). Later, the Conway-Maxwell-Poisson model utilised in RTA studies and is analogous to the negative binomial model for data characterised by over-dispersion. The drawback of Conway-Maxwell-Poisson model is that if there is a low mean then a high bias with small sample size is encountered. The multivariate application of this modeis not available until now to the best knowledge of (Lord and Mannering, 2010), however, some applications have been found in RTA investigation Lord et al. (2008, 2010).

On the other hand, RTA data includes more than one level of response variables such as RTI severity or survival and death. Thus, it would be useful to implement models that are associated with severity. Yasmin and Eluru (2013) compared between the ordered response models and unordered response models in a condition of the occupant, the driver (in-vehicle) or pedestrian (out-vehicle) injury severity in RTA data. Ordered response models include ordered logit (OL), generalised ordered logit (GOL) and mixed generalised ordered logit (MGOL). In contrast, unordered response models were multinomial logit (MNL), nested logit (NL), ordered generalised extreme value logit (OGEV) and mixed multinomial logit (MMNL). The results appeared in the literature for generalised ordered logit (GOL) and generalised ordered logit (MGOL) models are found to be reliable and those models are strong candidate models against the mixed multinomial logit (MMNL) model for RTA injury severity modelling (Yasmin and Eluru, 2013). Ordering within the response variable is a distinguished feature in the ordered response models and are more efficient in general. On the contrary, there is a restriction on the outcome process with the impact of exogenous variables in ordered response models. The parameters are allowed to vary across alternatives in unordered response models (Yasmin and Eluru, 2013). Some studies examined the influence of factors by employing ordered response models (for example, (Shibata and Fukuda, 1994; Khattak et al., 1998). Several researchers applied ordered response models within Oman through ROP data (Al Reesi et al., 2013a; Al-Bulushi et al., 2015b; Al-Aamri et al., 2017).

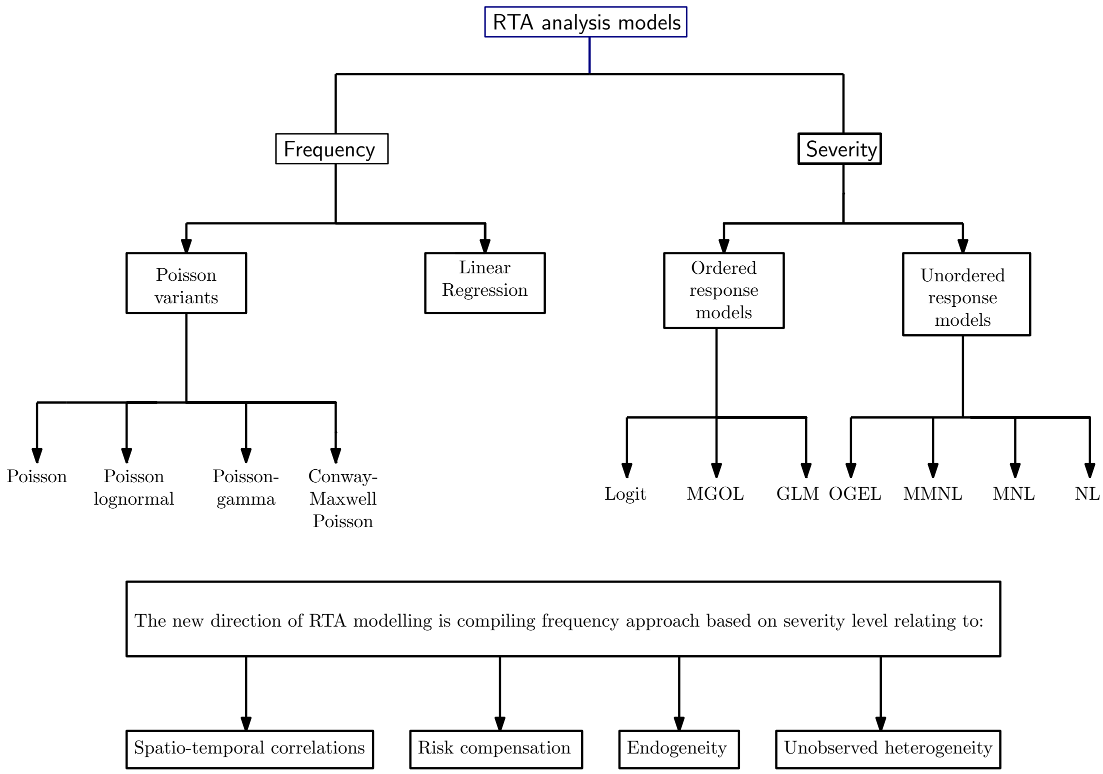


Figure 2.3: Categorisation of road traffic accident models

For trade-offs between ordered and unordered models or between frequency and severity models have revealed new paths to account for unobserved heterogeneity (Savolainen et al., 2011), thus indicated to address the correlation of spatial and temporal components of the RTA associated with injury severity models. Mannering and Bhat (2014) underlined that the next generation of methodological developments address complicated issues to spatio-temporal correlations, risk compensation, endogeneity and unobserved heterogeneity. Therefore, it seems that the future direction of mathematical models in RTA compiling frequency will be based on the severity, as shown in Figure (2.3).

2.5.1 Temporal models

Temporal models are based on time analysis that gives researchers an excellent way to identify RTA trend and forecasting. Time series data is a suitable data for temporal analysis and modelling in diverse disciplines. Time series is a sequence of values of a variable recorded over time, most often at a regular time interval. If observations are taken at discrete time points, then the time series is discrete and continuous if values taken in continuous time. Often continuous time series are usually discretised by sampling or by aggregation. Modelling and analysing time series data is a crucial way to reveal the trend, cycle and seasonality for predicting some future conditions which would, in turn, help with more useful and optimised planning for many organisations (Lavrenz et al., 2018). Time series are usually decomposed into four components: trend, cyclical pattern, seasonal variation and, random error. If the time-series data is available, it is possible for forecasting the future occurrence of RTAs, RTIs or, RTDs by using statistical methods and models under time series analysis strategies (Manikandan et al., 2018).

The prime argument of applying time series modelling strategies in RTA research is that RTA, RTI and RTD are all discrete variables. The broad application in RTA provides a different perspective on this during recent decades by Box and Jenkins (1970) approach (Quddus, 2008b; Liboschik et al., 2015; Narasimhan et al., 2017; Lavrenz et al., 2018). The application of some time series models such as ARMA, ARIMA, SARIMA, INAR, and GLARMA gives the real-valued time series data and strategies in this area (Quddus, 2008b). The major essential steps involved in time series modelling process are: model specification, fitting and diagnostics of the model (Cryer and Chan, 2008; Box et al., 2015). More details of time series strategies in the pieces of literature have been discussed in Section (4.2).

Several time series analysis of RTA and its impact have been published. Some studies conducted RTA time series (Sandt et al., 2016; Quddus, 2016; Narasimhan et al., 2017;

Jamal et al., 2020), some conducted RTI (Meyerhoff, 1978; Parvareh et al., 2018; Timmermans et al., 2019) and RTD (Quddus, 2008b; Manikandan et al., 2018; Al-Hasani et al., 2019b; Timmermans et al., 2019). Temporal unit which used to implement different time series models could be yearly (Quddus, 2008b; Narasimhan et al., 2017), quarterly (Sandt et al., 2016; Timmermans et al., 2019), monthly (Quddus, 2008b; Zhang et al., 2015; Narasimhan et al., 2017; Parvareh et al., 2018; Manikandan et al., 2018; Al-Hasani et al., 2019b; Jamal et al., 2020), weekly (Quddus, 2016; Timmermans et al., 2019), daily (Meyerhoff, 1978; Szeto et al., 2009) or hourly (Meyerhoff, 1978). Temporal scale relies on related areas that provide the data in a particular region or country. Often, studies used more than one temporal unit in the same study if the time series data are in years and months (Quddus, 2008b; Narasimhan et al., 2017) or quarters (season) and months (Timmermans et al., 2019).

Due to the government departments' serious concern in many developed and developing countries, there has been a significant shift in the trend of the number of accidents and injuries in recent years. Although there is a decline in the trend of the number of accidents in many Gulf countries, including Oman, the severity and number of injuries have been found yet to be significant. Recently a substantial decrease in the number of injuries has been observed for Oman. The number of accidents and injuries are highly volatile and there is also a shift of trends in accidents and injuries over time. They possess a high seasonality, which makes the analysis, model selection and forecasting complex tasks.

2.5.2 Spatial models

With sophisticated analysis tools, statistical modelling is becoming a key instrument in various disciplines. Generalised Linear Models (GLM) have been used as the common technique for the county-level modelling (Li et al., 2013). Within GLM models, fixed coefficient estimates explain the associations between the dependents variables and predictors in individual spatial units. Conversely, the GLM has a strong drawback as the model cannot capture spatial correlations existing in the RTA data (Li et al., 2013; Wang and Li, 2017). Moreover, some studies pointed to suspect outcomes predicting from global models due to the impact of predictors in individual spatial units either larger or smaller units (Nakaya et al., 2005; Li et al., 2013; Lu et al., 2014).

The spatial models provide a systematic procedure to scrutinise datasets that contain observations on the spatial or geographical units. The prime component of the spatial models is the weight matrix that gives the modeller a chance to explore the impact of neighbouring units on a particular location directly and indirectly (Anselin, 1988b; Azimian, 2018). Therefore, spatial models can generate consistent outcomes and can fit the location-specific data better than classical count data models (Aguero-Valverde and Jovanis, 2006; Quddus, 2008a). Hence, many methodological advances on spatial analysis have been developed to treat the subtle issues in traffic data, for instance, the impact of unobserved factors in spatial correlations, accident frequencies, unobserved heterogeneity, endogeneity, etc. (Mannering and Bhat, 2014). In addition, some empirical studies in the RTA field concluded that spatial models perform significantly better than classical models with RTA data (Aguero-Valverde and Jovanis, 2008; Quddus, 2008a; Barua et al., 2015; Rhee et al., 2016).

Spatial models are popular methodological techniques that have recently gained much attention for empirical research, including RTA, RTI or RTD analysis. The spatial methods have been applied to capture spatial effects and influencing factors for many RTA research. In the spatial investigation of RTA, the impact of population, density, number of vehicles, etc. on the number of accidents have been performed taking regions as the spatial units. However, identification or the justification of the factors are critical for spatial models. Rhee et al. (2016) indicated that improved prediction could be achieved by examining and control of the multiple spatial factors in road safety studies. Although there are many studies in accident research and many models have been suggested for each type of dataset, it is crucial to choose an appropriate model due to insufficient guidelines regarding spatial model selection. Equally, Quddus (2008a) argued that the estimation of spatial model parameters could be insignificant based on spatial model residuals according to spatial units levels in that model. However, the parameter estimation, hypothesis testing, model diagnostics, etc., are equally crucial for model fitting with spatial data. Moreover, model misspecification with the potential spatial correlation for the spatial models increases model residuals (Aguero-Valverde and Jovanis, 2008). Therefore, it is crucial to identify the most suitable model from the set of spatial models that are adequate to capture associated spatial factors and variations.

Investigators carried out a number of studies on RTA; however, they investigated different aspects for different countries. As such, Rhee et al. (2016) found that spatial error model (SEM) performs better than spatial lag model (SLM) and Ordinary least square model (OLS) for an accidents dataset for Seoul, Korea. In a similar study, Quddus (2008a) applied both models (SLM and SEM) to analyse data for 633 census wards in London metropolitan regions. However, due to the different choice of high-level spatial units such as regions or governorates, the model performance varies significantly. In Oman, as mentioned earlier, roads are the only possible option for transport both in terms of human occupants or goods to be transported instead of other transports like railways in many other countries. As a result, RTA has been a prominent public health problem in the country (Al-Lamki, 2010; Islam and Al Hadhrami, 2012; Al-Maniri et al., 2013; Al-Aamri et al., 2017). The spatial units of this analysis are the 11 governorates in the Sultanate of Oman. Although some researchers have carried out studies on the road safety of Oman, to the best of our knowledge, no study has been found that applies spatial models with RTA data in Oman. Therefore, the purpose of this study is to compare the most popular spatial models for the road traffic accidents and find the one that suits the most for the RTA data in Oman.

2.5.3 Spatio-temporal models

As discussed earlier, some issues are related to classical models, as they cannot account for the regression-to-the-mean bias and spatial effects, which can generate inadequate RTA evaluation, estimation or forecasting. Numerous research efforts have been made to explore RTA within global and local levels. However, there are some limitations such as temporal effects that have not been appropriately addressed, which can also cause flawed RTA frequency analysis and forecasting (Wang and Kockelman, 2013; Azimian, 2018). Besides, the nature of the relationship between RTAs in a particular area and the average time to the closest hospital has not been investigated in terms of approachability (Wang and Kockelman, 2013). Furthermore, local-level data have not been combined with area-based data to classify high-risk places, which may cause the incorrect identification of high-risk and low-risk spots. Thus, unobserved factors are likely to be correlated over time and space and if we do not take into account the temporal and spatial correlation of RTA data, then we may get inconsistent results (Mannering and Bhat, 2014). Nowadays, the new modelling frontier in RTA research refers to spatio-temporal correlations, risk compensation, endogeneity and unobserved heterogeneity (Savolainen et al., 2011; Mannering and Bhat, 2014; Mannering, 2018). The RTA frequency analysis should take spatio-temporal unobserved heterogeneity into account (Liu and Sharma, 2018; Cheng et al., 2020). Therefore, it is crucial to extend the investigation to include spatial and temporal effects jointly of the RTA.

Many researchers have recast analysis models for examining factors which caused the numbers of RTAs occurring in several geographical spaces during a particular period relying on available data (Lord and Mannering, 2010). Ma et al. (2017) investigated the application of multivariate space and time models to jointly analyse RTA frequency by the severity level of injuries with a temporal scale. They found that the model enabled to borrow strength across spatial units and over accident types due to capturing the underlying unobserved heterogeneity. Overall, since RTA data is aggregated over space and time, the spatio-temporal correlations are accountable for heterogeneity factors due to some unobserved characteristics (Liu and Sharma, 2018). Ma et al. (2017) found that the space-time model outperforms other alternatives, even spatial models or random parameter models in this discipline. However, Miaou et al. (2003) attempted to capture spatio-temporal factors to analysed RTAs in Texas counties for the data from 1992 to 1999. It is considered as the first study of space and time for modelling RTA data. Therefore, this study will attempt to develop and apply a spatio-temporal model that offers greater insights into RTA data analysis. However, the more sophisticated modelling approach is Bayesian hierarchical technique, which will also be investigated together with the spatial and temporal random effect components.

Several studies were conducted to model spatio-temporal unobserved heterogeneities among factors in the RTA field. There are diverse spatial units which have been implemented that consider the correlations for the neighbouring units such as segments (Aguero-Valverde, 2013; Wang et al., 2013b; Wen et al., 2019), intersections (Castro et al., 2012; Ma et al., 2017; Al-Aamri et al., 2020), traffic analysis zones (Dong et al., 2016; Meng et al., 2017; Cheng et al., 2020), corridors (Wang and Abdel-Aty, 2006; Guo et al., 2010), census tracts (MacNab, 2004; Wang and Kockelman, 2013), ward (Boulieri et al., 2017), counties (Liu and Sharma, 2018; Li et al., 2019), regions/governorates (Fu, 2016; Truong et al., 2016; Rahman et al., 2018). Similarly, various temporal scales were used to measure temporal correlation such as hour (Kilamanu et al., 2011; Meng et al., 2017; Al-Aamri et al., 2020), day (Fu, 2016; Ma et al., 2017; Rahman et al., 2018), week (Kilamanu et al., 2011; Liu et al., 2015), month (Quddus, 2008b; Liboschik et al., 2015; Rahman et al., 2018; Wen et al., 2019), and year (Castro et al., 2012; Wang et al., 2013b; Truong et al., 2016; Boulieri et al., 2017; Liu and Sharma, 2018; Li et al., 2019; Cheng et al., 2020). However, most of the further mentioned studies attempted to formulate space-time models, but some have been developed within a classical space-time framework with their data investigations. In fact, some researches are conducted on the impact of the space and the time by formulating in separate models for space effect and time effect (Miaou et al., 2003; Wang and Abdel-Aty, 2006). On the other hand, several researchers have developed their spatio-temporal models within Bayesian hierarchical framework, considered spatio-temporal unobserved factors jointly. Despite these spatio-temporal studies based on Bayesian hierarchical framework, they ignored the space-time interaction (Castro et al., 2012; Barua et al., 2015; Boulieri et al., 2017; Saha et al., 2018). There is the dearth of studies that took into account the space-time interaction components in the RTA modelling. As a result, analysis did not have solid meaningful interpretations due to not taking space-time interaction into models account (Abd Naeeim et al., 2020). In order to fill the gap in the literature, this study aimed to expand a hierarchical Bayesian space-time model that could have the best performance with Oman RTA data.

2.6 Chapter summary

In this chapter, we have introduced some background knowledge of RTA in the literature, including RTA situation globally with a focus on EMR, GCC and Oman. Worldwide, reports point out that the RTA are the ninth reason for injuries and deaths around while being the first reason for death among younger people aged 5-29 in 2016. We found that the RTA situation is the best in the Europe region because the death (RTD) rate is lower than in other world regions. The current situation in Americas and western Pacific regions is better than the remaining regions (Eastern Mediterranean, Africa and, South-East Asia). Eastern Mediterranean (EMR) region faces severe RTA burden and the number of deaths higher by 4.5 times than good road safety countries that requires more investigations. Moreover, the World Health Organisation (WHO) ranked EMR as the second costly region due to RTAs since EMR had 10% of RTDs among the world in 2010. State of arts on RTA field points out the depth of harms of RTA on the GCC states. This is because remarkable changes happened in those countries in term of modern lifestyle, high gross domestic product, increasing of population, enormous vehicles on roads, expanding roads network, etc. that play a negative role in terms of RTA.

We classified RTA factors into five domains, including human behaviour, socio-economic, environments and road defects, demographic and, spatial and spatio-temporal characteristics. Among GCC states, Oman has one of the high numbers of RTA worldwide, which is costing the country socio-economic burden through the last fifty years (Al-Lamki, 2010; Plankermann, 2014). In 2000, a national health survey in Oman demonstrated the first reason for morbidity and mortality is RTA that represented 61% (Islam and Al Hadhrami, 2012; Al Aamri, 2018). In Oman, several attempts have been made on different aspects of RTA, researchers underlined that more productive studies are required.

The existing state-of-the-art models for RTA analysis and modelling have also been discussed in this chapter. Overall, it has been found that the classical mathematical models into frequency and severity models that have been implemented with RTA data. We also reviewed temporal modelling in RTA discipline using time data to explore trend, cycle and seasonality. The state of arts pointed out the real value of time series data and strategies in RTA. They were implementing time series analysis and strategies support to forecast the future occurrence of RTAs. Spatial modelling has been reviewed with a comparison of Generalised linear models (GLM). Overall, since spatial models can capture unobserved factors in spatial correlation, can produce consistent outcomes, fit location-specific data, perform better than GLM. To best of our knowledge, no spatial study on RTA in Oman has previously been done. Finally, spatio-temporal models have been discussed in the chapter. Spatio-temporal modelling would capture the temporal and spatial effects simultaneously by combined space and time patterns. Recent references indicated to the new RTA modelling frontier associated with spatio-temporal correlation. Many researchers develop spatio-temporal models within Bayesian hierarchical framework. However, it is crucial to consider space-time interaction in spatio-temporal modelling.

Chapter 3

Data collection and computation packages

3.1 Introduction

Statistical modelling for road traffic accidents (RTA) can be used to investigate the temporal, spatial and spatio-temporal effects and variation by accident data. Therefore, it is a prime concern to decide how to select the explanatory variables into models. Selection of explanatory variables should be based on the RTA causal factors and the availability of the data are also crucial issues in model development. Often data are unavailable for all desired variables and may not be in good quality even if data are available. This chapter provides an overview of this study data characteristics and constraints of road traffic accidents and road traffic injuries as well as an overview of the statistical software packages. The chapter consists of four sections: introduction, data review, software and packages and a summary. Section (3.2) presents the source of three types of data for the purpose of this study. It is divided into three subsections conssisent with the study models: temporal data, spatial data and spatio-temporal data. Section (3.3) explains the tool used for fitting models and further analysis to conduct the study. At the end of this chapter, a summary (3.4) has been given in the fourth section.

3.2 Data review

In this study, three types of data are collected to perform modelling and analysis in different aspects. Data are the primary concern to achieve the study investigation and accomplish study objectives. Therefore, data for this study have been gathered from multiple sources. RTA data have been collected from the published reports by Royal Omani Police (ROP) (ROP, 2009b,a, 2014, 2015, 2017b,a). Other secondary data on different factors such as population size, area of Oman governorates, governorates density, etc. from the publication of the National Centre for Statistics & Information (NCSI) in Oman (NCSI, 2017a,b,c, 2018b,a,c,d, 2019a,b). Unemployment statistics were provided by the Public Authority of Manpower Register in Oman as requested for the purpose of this study.

It is worth mentioning that the ROP officers collect the data for all Oman traffic accidents at the scene for further investigation. Then, ROP officers complete the traffic accident report manually and send it to the Directorate General of Road Traffic in the capital city Muscat to be registered into the RTA database (Al-Maniri et al., 2013; Al-Reesi and Al-Maniri, 2014; Al-Bulushi et al., 2015b; Al-Aamri et al., 2017). However, the traffic accidents reports recording system in Oman taking into account RTA reports in three cases only (Al Bulushi, 2017). First, the ROP officers report the traffic accidents if there is a human injury RTI or death RTD (Al Bulushi, 2017; Al Aamri, 2018). The second case could be written the RTA report if public property damage is caused by that RTA even though there is no RTI or RTD (Al Bulushi, 2017; Al Aamri, 2018). Third, in case that there is no RTI or RTD and no public property damage, but the drivers of vehicles fail to determine the error and the person at fault (Al Bulushi, 2017; Al Aamri, 2018). Therefore, the RTA database in the ROP consists of serious RTAs only. Otherwise, all RTA cases are categorised as minor RTA that could be managed by insurance companies (Al Aamri, 2018). Moreover, some critical variable such as milage, accident coordinates, temperature could not take into account because data unavailable in the ROP data set. Consequently, the data provided by ROP is not accurate, which made limitation of the modelling and forecasting.

3.2.1 Temporal data

The data of road traffic accident and injuries (RTA and RTI) in Oman are maintained by the Royal Omani Police (ROP) and 'Statistical Summary Bulletins' are published annually by the Directorate of Road Traffic as part of the ROP. Summary data are issued by the National Centre for Statistics & Information (NCSI) in monthly reports, called, 'Monthly Statistical Bulletin' in Oman. Time series data is required for temporal modelling and investigation of RTA trend and forecast.

The RTA and RTI data for this part of the study have been collected from two sources: 'Statistical Summary Bulletins' from the Directorate of Road Traffic in the ROP and 'Monthly Statistical Bulletin' from the National Centre for Statistics & Information (NCSI). The ROP data cover all road accidents and injuries for the period of 2000 to 2016 published in (ROP, 2009a, 2014, 2017b). Additional data from 2016-19 have been collected from the 'Monthly Statistical Bulletin' of the National Centre for Statistics and Information (NCSI) NCSI (2017a,b,c, 2018b,a,c,d, 2019a,b). As the end, we consider monthly time series data of RTA and RTI from January 2000 to June 2019 in this study. Table (3.1) shows the temporal data used and implemented for time series models in Chapter (4).

Table 3.1: Temporal data of RTA and RTI in Oman 2000-2019

Time	RTA	RTI	Time	RTA	RTI	Time	RTA	RTI
January/2000	1122	883	February/2000	984	755	March/2000	1241	852
April/2000	1047	737	May/2000	1185	794	June/2000	1087	657
July/2000	1205	849	August/2000	1076	777	September/2000	1096	717
October/2000	1046	768	November/2000	987	691	December/2000	964	843
January/2001	1114	883	February/2001	999	689	March/2001	1055	856
April/2001	1010	716	May/2001	1081	767	June/2001	1028	819
July/2001	1014	827	August/2001	1043	774	September/2001	1160	761
October/2001	1283	857	November/2001	1203	869	December/2001	964	822
January/2002	1251	985	February/2002	1125	855	March/2002	1097	740
April/2002	696	563	May/2002	695	609	June/2002	635	641
July/2002	552	468	August/2002	662	686	September/2002	535	494
October/2002	587	696	November/2002	646	588	December/2002	626	629
January/2003	754	604	February/2003	941	642	March/2003	996	527
April/2003	950	546	May/2003	1026	567	June/2003	845	509
July/2003	838	608	August/2003	767	654	September/2003	711	444
October/2001	827	603	November/2003	758	551	December/2003	784	480
January/2004	876	560	February/2004	728	571	March/2004	735	477
April/2004	816	523	May/2004	846	537	June/2004	793	514
July/2004	804	567	August/2004	891	744	September/2004	743	503
October/2004	805	614	November/2004	688	563	December/2004	735	463
January/2005	696	462	February/2005	658	574	March/2005	758	497
April/2005	879	542	May/2005	842	542	June/2005	734	560

Continued on next page

Table 3.1 – continued from previous page

Time	RTA	RTI	Time	RTA	RTI	Time	RTA	RTI
July/2005	741	722	August/2005	798	666	September/2005	784	488
October/2005	857	595	November/2005	814	580	December/2005	686	487
January/2006	648	557	February/2006	578	415	March/2006	707	606
April/2006	702	586	May/2006	843	586	June/2006	928	590
July/2006	967	737	August/2006	818	578	September/2006	995	661
October/2006	855	779	November/2006	1052	827	December/2006	776	554
January/2007	896	661	February/2007	828	693	March/2007	862	617
April/2007	906	619	May/2007	802	619	June/2007	547	531
July/2007	617	768	August/2007	624	730	September/2007	618	734
October/2007	760	908	November/2007	672	740	December/2007	684	829
January/2008	712	775	February/2008	612	684	March/2008	601	768
April/2008	558	802	May/2008	579	802	June/2008	608	840
July/2008	801	1071	August/2008	673	971	September/2008	566	837
October/2008	771	1004	November/2008	787	1060	December/2008	714	923
January/2009	616	820	February/2009	577	793	March/2009	707	914
April/2009	663	880	May/2009	619	880	June/2009	650	821
July/2009	623	896	August/2009	568	824	September/2009	534	760
October/2009	623	771	November/2009	526	686	December/2009	547	741
January/2010	574	790	February/2010	545	694	March/2010	668	814
April/2010	658	984	May/2010	738	984	June/2010	580	786
July/2010	557	778	August/2010	652	780	September/2010	625	960
October/2010	652	844	November/2010	625	767	December/2010	697	972
January/2011	744	1089	February/2011	523	904	March/2011	707	1067

Continued on next page

Table 3.1 – continued from previous page

Time	RTA	RTI	Time	RTA	RTI	Time	RTA	RTI
April/2011	559	839	May/2011	598	838	June/2011	646	945
July/2011	816	1142	August/2011	671	1017	September/2011	620	844
October/2011	672	906	November/2011	515	856	December/2011	648	959
January/2012	646	967	February/2012	529	873	March/2012	749	1273
April/2012	771	1068	May/2012	771	1261	June/2012	771	983
July/2012	700	909	August/2012	628	988	September/2012	610	768
October/2012	598	834	November/2012	624	813	December/2012	680	881
January/2013	776	1048	February/2013	639	786	March/2013	714	962
April/2013	664	938	May/2013	808	997	June/2013	731	889
July/2013	649	874	August/2013	573	990	September/2013	584	830
October/2013	537	773	November/2013	553	831	December/2013	601	884
January/2014	641	379	February/2014	607	312	March/2014	591	337
April/2014	452	258	May/2014	591	356	June/2014	534	390
July/2014	584	448	August/2014	516	388	September/2014	501	264
October/2014	616	366	November/2014	538	212	December/2014	546	125
January/2015	568	339	February/2015	431	231	March/2015	516	300
April/2015	577	300	May/2015	611	323	June/2015	631	345
July/2015	626	408	August/2015	505	327	September/2015	542	308
October/2015	433	269	November/2015	449	240	December/2015	390	236
January/2016	443	280	February/2016	376	239	March/2016	438	307
April/2016	371	227	May/2016	405	278	June/2016	397	302
July/2016	388	252	August/2016	349	308	September/2016	352	222
October/2016	383	247	November/2016	393	273	December/2016	426	328

Continued on next page

Table 3.1 – continued from previous page

Time	RTA	RTI	Time	RTA	RTI	Time	RTA	RTI
January/2017	318	251	February/2017	287	202	March/2017	290	238
April/2017	355	247	May/2017	354	331	June/2017	385	320
July/2017	360	333	August/2017	232	243	September/2017	157	176
October/2017	354	236	November/2017	386	293	December/2017	296	228
January/2018	194	273	February/2018	224	191	March/2018	179	220
April/2018	200	242	May/2018	217	260	June/2018	222	255
July/2018	215	266	August/2018	215	238	September/2018	178	214
October/2018	202	229	November/2018	200	223	December/2018	207	225
January/2019	187	214	February/2019	156	177	March/2019	207	240
April/2019	163	194	May/2019	180	202	June/2019	210	240

3.2.2 Spatial data

The road traffic accidents (RTA) data of 2017 have been gathered for the Sultanate of Oman to study and model the spatial effects. The data includes the number of road traffic accidents in eleven governorates in Oman, which have been considered as spatial units in this study. The RTA data among Oman governorates (study spatial unites) has been requested from the Directorate of Road Traffic in the ROP at Muscat. Moreover, 'Statistical Summary Bulletins' which are published annually by the ROP include RTA data in Oman governorates (ROP, 2015, 2017a). This study also uses explanatory variables such as population size, population density, number of registered vehicles, number of unemployed persons, speed driving, season (February-June-July or otherwise). Population size and population density data are collected from the National Centre for Statistics & Information (NCSI) monthly reports, called 'Monthly Statistical Bulletin' (NCSI, 2017a,b,c, 2018b). The number of unemployed persons (jobs seekers) data in each governorate of

Oman in 2017 were collected from the Public Authority of Manpower register. Number of registered vehicles in each governorate, speed driving, season are collected from the RTA database. Table (3.2) presents a summary of the variables that will be used and implemented for spatial modelling in chapter (5).

Table 3.2: Descriptive statistics of the collected variables for spatial modeling

Variables	Mean	SD	Min	Median	Max
RTA	349.5	332.5	43	304	1,222
Population size	412,628	385,833.1	44,528	316,909	1,377,818
Jobs seekers	6,735	4,490.8	532	6,084	16,560
Registered vehicles	131,233	221,307.3	2,360	57,941	779,426
Population density	52.01	98.4	0.58	22.26	344.45
Speed driving	205.5	188.5	33	174	698
Season	99.64	92.4	14	79	348

3.2.3 Spatio-temporal data

The RTA data from 2013 to 2017 have been gathered and used to study the spatial and temporal effect simultaneously for the Sultanate of Oman. The study uses the dataset to identify spatio-temporal factors, their effect and variations in Chapter (6). However, this data is extended version of the data used in Chapter (5) used for spatial modelling. The data includes the number of road traffic accidents by eleven governorates in Oman considered as spatial units in spatio-temporal modelling. The temporal unit is the year, and the time scale is five years from 2013 to 2017. Figure (3.1) displays the visualisation of RTAs in Oman governorates between 2013 and 2017. Broadly, the highest values of all five years are clustered around the capital city Muscat governorate. The results are not surprising since this governorate has larger populations than other governorates, and therefore they are associated with more RTAs. The number of RTA data also revealed a cluster of low numbers of accidents in the country's southern parts. Table (3.3) presents a summary of the study variables. However, the visualisation maps of the explanatory variables are available in the Appendix (A-F).

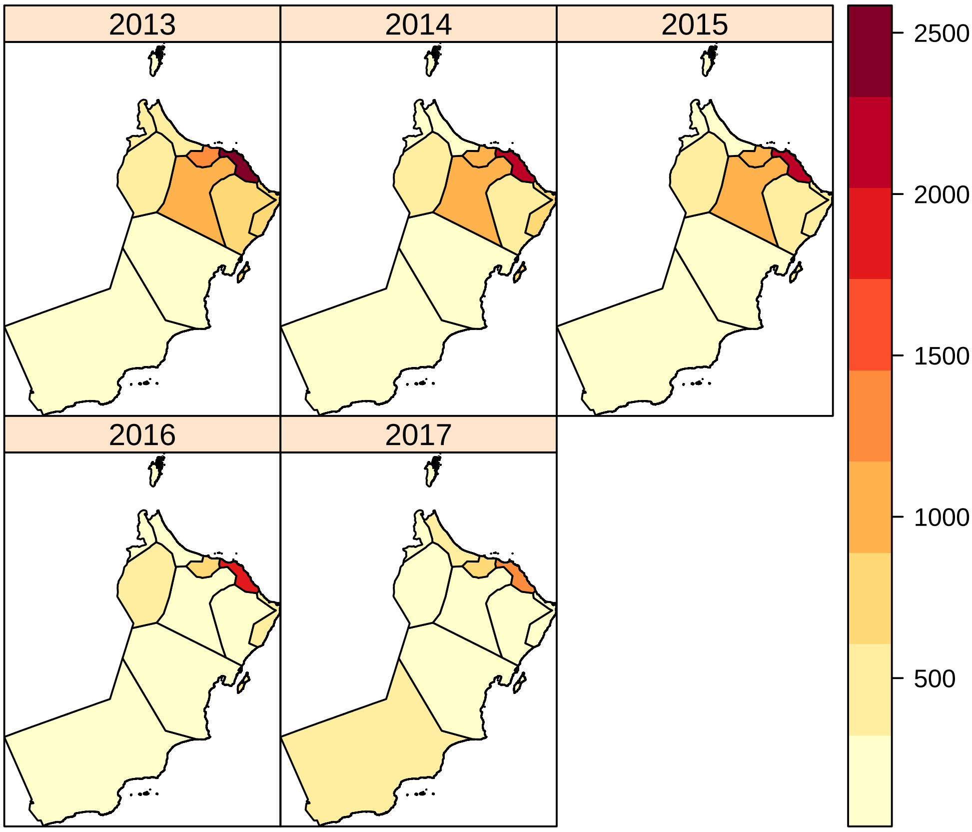


Figure 3.1: Frequencies of RTAs in Oman through governorates (2013-2017)

Table 3.3: Descriptive statistics of the collected variables for spatio-temporal modeling

Variables	Mean	SD	Min	Median	Max
RTA	534.4	567.74	43	325	2,582
Population size	376,488	354,022.66	10,148	325	1,441,622
Jobs seekers	8,238	7,453.02	346	5,359	34,799
Registered vehicles	115,831	191,666.51	1,511	52,037	779,426
Population density	48.31	90.47	0.497	20.43	360.40
Speed driving	286.1	276.74	25	181	1,184
Season	146.4	152.54	14	96	802

3.3 Software and packages

It is beyond doubt that information technology developments have promoted all other major natural, biological and physical science areas. Researchers have been constantly benefitted from the computer revolution to build up their research ideas and further enhance the scientific theory or empirical studies. There are many tools and softwares that can be implemented to model and analyse RTA data. However, it is crucial to identify the suitability if different tools for diverse study analysis.

Several papers discuss the statistical methods used in research and analysis (Ali and Bhaskar, 2016). Numerous statistical software systems are currently available to the public. In general, such software tools target three broader categories of audiences: humanities/social sciences, engineering/sciences, and medical/health sciences. Data analysis is at the heart of any quantitative research in natural or social sciences. The focus of such practice is to conclude collected data and design mathematical models (Ali and Bhaskar, 2016). The researchers in this discipline prefer to use the software tools like MS Excel, SPSS, SAS and, STATA along with the strong understanding of statistics to interpret results (Creswell, 1994; Albers, 2017; Bors, 2018). There are diverse statistical analysis applications in the areas ranging from artificial intelligence to automation and robotics in engineering and science fields. The programming environments needed for these areas are various in needs from low-level languages to high-level scripting languages. Hence, the engineering ecosystem promotes the use of analysis using the tools like MS Excel (Billo, 2007; Larsen, 2008), Python (Rogel-Salazar, 2018), and R programming (Patil, 2016; Bhanot et al., 2019; Díaz-Bejarano et al., 2019). Data and statistical analysis are used in medical and health sciences research for varying purposes from Epidemiology to Bioinformatics. Popular software tools for this discipline are R (Oster, 1998; Khan, 2013) and Python (Zhao et al., 2021).

Over the last two decades, the availability of temporal, spatial or spatio-temporal data has expanded virtually, and the prime reason is referred to the advances in computational tools that gives researchers opportunities to collect this kind of data (Blangiardo et al., 2013; Blangiardo and Cameletti, 2015; Wikle et al., 2019). Moreover, the investigator has the challenge to treat geo-referenced data consist of variables and information about time, space or both (space-time data) in many disciplines such as epidemiology, ecology, climatology and RTA (Blangiardo et al., 2013; Blangiardo and Cameletti, 2015).

Nowadays, considering the sophisticated software in data collection, leading to availability of big datasets, characterised by high spatial and temporal resolution as well as data from different sources is a critical issue (Blangiardo and Cameletti, 2015; Bakka et al., 2018; Wikle et al., 2019). However, R is a statistical computer program that provides a rich environment for data analysis and graphics of temporal, spatial or spatio-temporal data. Furthermore, this advanced tool could apply the theory of stochastic processes, has a computationally efficient implementation and has been widely used in practice (Bakka et al., 2018; Wikle et al., 2019). As a result, the computational efficiency of R software implementation, as well as the relative simplicity of the interface, has allowed applied researchers to fit a broad range of temporal, spatial and spatio-temporal models to a wide array of applications, particularly RTA field (Bakka et al., 2018; Wikle et al., 2019).

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. It is powerful in data manipulation, and it is found to be robust and versatile for data analysis. This open-source package has been seen as excellent in graph plotting. R is an implementation of the S programming language combined with lexical scoping semantics, inspired by Scheme. S was created by John Chambers in 1976, while at Bell Labs. R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently developed by the R Development Core Team (Chambers is a member). R is named partly after the first names of the first two R authors and partly as a follow-on on the name of S. The project was conceived in 1992, with an initial version released in 1995 and a stable beta version in 2000. The R Foundation has recently been awarded the Personality/Organization of the year 2018 award by the professional association of German market and social researchers. The open-source software R has been extensively used as a sophisticated tool to fit model and analyse different data Team (2000).

Several R-packages can be implemented to model and analyse road traffic accidents data. In this study, R-packages are used to model, analyse and visualise spatial or spatio-temporal data. Forecast package is implemented for temporal modelling and analysis with fitting time series models. Other packages are required for plotting such as ggplot2 and ggridges. On the other hand, R has extensive capabilities and flexibility to work with large amounts of spatial data and provides tools via multiple packages to model, analyse and visualise spatial data (Woods, 2017). In this study, as such R-packages is implemented for spatial modelling are rgdal, dplyr, spgwr and GWmodel. R software has also been implemented in our model framework to perform the spatio-temporal analysis and fitting Bayesian spatio-temporal models. Thus, some packages run for spatio-temporal modellings such as spacetime, INLA and spdep. Some more packages have also been used to visualise and estimate fixed and random parameters by maptools, broom, and colorspace packages.

3.4 Chapter summary

This chapter discussed data collection and software packages for the purpose to fulfill of the thesis aim and its objectives. Overall, multi-sources provided three types of Oman's RTA data. Monthly time series data have been gathered from ROP and NCSI Bulletins for temporal modelling. The total time series observation data is 234 months from January 2000 to June 2019. This study took into account eleven governorates in the Sultanate of Oman as eleven spatial units for spatial and spatio-temporal modelling. However, data has been requested from three related Authorities in Oman for spatial and spatio-temporal analysis. In this thesis, RTA data in 2017 is implemented for spatial modelling while RTA data from 2013 to 2017 is used for spatio-temporal modelling. Free R software packages have been implemented for all study models framework to fitting models and perform further analysis.

Chapter 4

Temporal modelling and forecasting of road traffic accidents data

4.1 Introduction

In recent years, detailed real-time data (collected over time) for road safety would be acquired at low cost and time series models can be well-suited to get perspective on road traffic accidents (RTA) research (Lavrenz et al., 2018). Although time series models for continuous variables have been well-studied, autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models by Box and Jenkins (1970) have also been used to model count data recently (Quddus, 2008b). In RTA research, several studies revealed that ARIMA is the best model performance due to the best-achieved values of diagnoses tests (Zhang, 2003; Raeside and White, 2004; Sheikh, 2009). However, finding an appropriate model for a time series data through suitable criteria and diagnostic checking are nontrivial tasks (Cryer and Chan, 2008). Along with the development of ARIMA in their milestone work, the authors Box and Jenkins also suggested a process for identifying, estimating, and checking models for a specific time-series dataset (Cryer and Chan, 2008). Referred to as the Box-Jenkins Method in the updated edition of the book (Box et al., 2015), the process of stochastic model building with an iterative approach consists of the following three essential steps:

1. Model Specification. Using the data and all related information to help select a sub-class of the model that may best summarise the data.

2. Model Fitting/Estimation. Use the data to train and estimate the parameters of the model (i.e. the coefficients).

3. Model Diagnostics/Checking. Evaluate the selected fitted model in the context of the available data and check for areas where the model may be improved (Cryer and Chan, 2008).

The iterative process continually loops through the cycle as new information is gained during diagnostics and incorporate that information into new model classes. The approach starts with the assumption that the process that generated the time series can be approximated using an ARMA model if it is stationary or an ARIMA model if it is non-stationary. Once a suitable model is fitted, diagnostic checking of the model is performed, which concerns evaluating the quality of the model. This ensures that the fitted model has reasonably well-satisfied the underlying assumptions. However, if there are no inadequacies found, the model fitting is assumed to be complete, and the model can then possibly be used to forecast future values. Otherwise, in the case of inadequacies, another model is searched, and thus, we return to the model specification step again (Cryer and Chan, 2008).

Diagnosis of time series models is a crucial step to examine the goodness of fit for the tentative model. Evaluation criteria such as Akaike Information Criterion (AIC) and Bayesian Information Criteria (BIC) have been used as a vital tool for selecting the best time series model from the group of models (Ozaki, 1977; Raeside and White, 2004; Ham et al., 2017). AIC and BIC found stable and reliable tools that can be utilised for the goodness of fit with temporal modelling (Ozaki, 1977; Raeside and White, 2004; Ham et al., 2017). Therefore, this thesis implemented AIC and BIC for temporal models fitting for RTA and RTI data from the Sultanate of Oman. To ensure that the selected model reasonably satisfies the underlying assumptions, it is necessary to apply diagnostics checking for evaluating the quality performance of the model. Consequently, the model can be used for forecasting if it scores fairly accurate or otherwise we return to the identification step again (Cryer and Chan, 2008). For checking the adequacy of the fitted time series model the residual diagnostics information is also quite useful. When the residuals behave like white noise, the model is adequate (Cryer and Chan, 2008; Lavrenz et al., 2018). There are several types of forecast-residuals to assess the accuracy of the time series (Hyndman et al., 2006). Therefore, our study evaluates time series models with the following very commonly used data analysis error metrics: root means square error (RMSE), mean absolute percentage error (MAPE) and mean absolute scaled error (MASE). At a more detailed level, using autocorrelation function (ACF) and partial autocorrelation function (PCAF) plots of the time series residuals may suggest whether the residuals are uncorrelated. Furthermore, there are several tests that are extremely used for diagnostic of the residuals' correlation such as: the Ljung-Box test (Zhang et al., 2015), Box-Pierce test (Szeto et al., 2009), portmanteau lack of fit test (Manikandan et al., 2018; Parvareh et al., 2018).

However, to the best of our knowledge, there are dearth of studies which have been conducted to model and analyse time series count data for accident research in Oman. In this research, the focus is on the times series model identification for the RTA and RTI data from a set of models, performing suitable diagnostic checks and forecasting accidents and injuries using the RTA and RTI data in Oman. The rest of the chapter is organised as follows: methodology in Section (4.2), results and discussion are given in the Section (4.3) and the chapter summary remarks in Section (4.4).

4.2 Modelling

Time series is a sequence of values of a variable recorded over time, most often at a regular time interval. If observations are taken at discrete time points, then the time series is discrete and continuous if values are taken in continuous time. Often continuous time series are usually discretised by sampling or by aggregation. Modelling and analysing time series data is a crucial way to reveal the trend, cycle and seasonality for predicting some future conditions which would, in turn, help with more effective and optimised planning for many organisations (Lavrenz et al., 2018). Time series are usually decomposed into four components: trend (T_t), cyclical pattern (C_t), seasonal variation (S_t) and random error (I_t). A time series can be expressed by an additive model defined as

$$X_t = T_t + C_t + S_t + I_t,$$

which can be used when the variation around the trend does not vary with the series. The multiplicative model defined as

$$X_t = T_t \times C_t \times S_t \times I_t,$$

is appropriate when the trend is proportional to the series. Often graphical approach (plot of a series) is used to identify whether a time series is additive or multiplicative. However, to find appropriate models for time series data, Box and Jenkins approach has been playing a significant role (Box et al., 2015). There are three main steps in the process, each of which may be used several times: (1) model identification (or specification), (2) model fitting, and (3) model diagnostics.

The primary tools used for model identification phase are plots of the series, correlograms - the plot of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) versus the lags. Stationarity is another crucial property of a time series modelling. A prime example of a stationary process is the so-called white noise process. However, we can frequently transform nonstationary series into stationary series by simple techniques such as differencing. There are a number of popular test procedures used by the researchers for checking if a series is stationary or not, e.g. Dickey-Fuller t test, the Said-Dickey t test, the Phillips-Perron t test, and the Elliott et al. ARIMA$(p, 1, q)$ test, etc. (Shively, 2004; Quddus, 2008b; Srivastava, 2015).

Appearing the count time series in the diversity of fields whenever nonnegative account observed events over a period (Liboschik et al., 2015). The introduction of correlation as a phenomenon that may be generated through lagged linear relations leads to proposing the autoregressive (AR) and autoregressive moving average (ARMA) models. Adding nonstationary component leads to the autoregressive integrated moving average (ARIMA) model popularised in the landmark work by Box and Jenkins (Box et al., 2015). These modelling approaches have widely been used to model count data as well (Quddus, 2008b; Zhang et al., 2015; Lavrenz et al., 2018).

Three parts constituted the simple ARIMA model: AR is the autoregressive part; I is the differencing part; and MA is the moving average part.

An ARIMA (p, d, q) model for a time series sequence $\{X_t,\, t = 1, 2, \ldots, n\}$ can be written as

$$\phi(B)(1 - B)^d X_t = \theta(B) A_t, \tag{4.1}$$

where p is the order of the AR process, d is an order of differences, q is the order of the MA process, A_t is the white noise sequence, ϕ is a polynomial of degree p, B is a backshift operator and θ is a polynomial of degree q.

However, an ARIMA model could not analyse time series with seasonal characteristics due to the three components of this model could not capture seasonal trend (Zhang et al., 2015). Therefore, seasonal autoregressive integrated moving average (SARIMA) models have been developed (Zhang et al., 2015). SARIMA models perform better than the historical average, linear regression, and simple ARIMA models for data with seasonal variations. In fact, SARIMA models are capable of taking into account the trend and seasonality. Moreover, in the RTA field, several studies revealed that SARIMA fitted data more accurate than ARIMA (Zhang et al., 2015; Manikandan et al., 2018; Halim et al., 2021). A SARIMA$(p, d, q)(P, D, Q)_s$ model can be expressed by the following equation

$$\phi(B)\Phi(B^s)(1 - B^s)^d X_t = \theta(B)\Theta(B^s)A_t, \tag{4.2}$$

where Φ, Θ, P, D and Q are seasonal counterparts of ϕ, θ, p, d and q, respectively, and s is the seasonality.

Fitting the best model of the time series data depends on finding the superior possible estimates of these unkown parameters for a given model. While there are essential approaches such as the least square method, method of moment and maximum likelihood method for model fitting, obtaining a suitable model can be cumbersome. However, several criteria such as Akaike information criterion (AIC) and Bayesian information criteria (BIC) have been used as a significant tool to choose the best model from a set of models (Ozaki, 1977; Raeside and White, 2004; Ham et al., 2017). Akaike information criterion (AIC) defined as

$$\text{AIC} = -2\log(L) + 2K \tag{4.3}$$

where L is the maximised likelihood, and K is the number of parameters, is used to obtain the order of the times series models (p, d, q, P, Q, D) which are the coefficients for

ARIMA$(p, d, q)(P, D, Q)_s$ model. Bayesian Information Criteria (BIC) is defined as

$$\text{BIC} = -2\log(L) + K\log(n), \tag{4.4}$$

where L is the maximised likelihood, K is the number of parameters, and n is the number of data points in the time series.

Model diagnostic checking is required when a suitable model was fitted. Diagnostic testing concerns evaluating the quality of the tentative model performance. This confirms that the fitted model has reasonably well-satisfied the underlying assumptions. Consequently, if the diagnostic test confirms no inadequacies, this is a good indicator for applying the same model to forecast future values. However, it could return to the identification step if the diagnostic test shows inadequacies outcomes (Cryer and Chan, 2008). There are several model diagnostic approaches used for examining the goodness of fit of the tentative model with a time series data. Residual analysis is a useful technique in checking whether a model has adequately captured the information in the data. Residual analysis is a useful technique in checking whether a model has adequately captured the information in the data. A good forecasting model yields residuals with the following properties: (i) the residuals are uncorrelated and (ii) the residuals have zero mean. If there are correlations between residuals, then there is information left in the residuals which should be used in computing forecasts. If the residuals have a mean other than zero, then the forecasts are biased. Any forecasting method that does not satisfy these properties can be improved. ACF and PCAF plots of residual series can suggest whether the residuals are uncorrelated. The Ljung-Box test (Zhang et al., 2015), Box-Pierce test (Szeto et al., 2009), portmanteau lack of fit test (Manikandan et al., 2018; Parvareh et al., 2018) are frequently used to identify the correlation of residuals. In our study, Ljung-Box test is used, which is defined as

$$Q = n(n+2)\sum_{k=1}^{h} \frac{\rho_k^2}{n-k} \tag{4.5}$$

where n is the sample size, ρ is the autocorrelation, K is the lags, and h is the lags to be tested.

For checking the adequacy of fitted model, the residual diagnostics can be used. When the residuals behave as the white noise, the model is adequate (Cryer and Chan, 2008; Lavrenz et al., 2018). There are several types of forecast-residuals to assess the accuracy of the time series (Hyndman et al., 2006). However, this study assesses time series models through three residuals: root mean square error (RMSE), mean absolute percentage error (MAPE) and mean absolute scaled error (MASE), which are frequently used in time series analysis. RMSE is the standard deviation of the residuals, defined as

$$\text{RMSE} = \left[\frac{\sum_{i=1}^{n}(x_{f_i} - x_{o_i})^2}{n}\right]^{1/2}, \tag{4.6}$$

where n is a simple size, x_{f_i} are the forecasted values, x_{o_i} are the observed values. The mean absolute percent error (MAPE) measures the size of the error in percentage terms. It is calculated as the average of the unsigned percentage error, as defined in the equation below

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_{f_i} - x_{o_i}|}{|x_{f_i}|} \times 100. \tag{4.7}$$

The mean absolute scale error (MASE) is used to compare models of a time series through scale-free for assessing forecast accuracy across series (Hyndman et al., 2006). MASE is defined as

$$\text{MASE} = \frac{1}{n}\sum_{i=1}^{n}\left(\left|\frac{e_t}{\frac{1}{n-1}\sum_{i=2}^{n}|x_{o_i} - x_{o_{i-1}}|}\right|\right), \tag{4.8}$$

where $e_t = x_{o_i} - x_{f_i}$ and the outcome values are independent of the data scale. However, if the outcome value is less than one, this indicates better forecasting. Alternatively, when the MASE value is greater than one, this means, the forecast is worse for the data.

4.3 Results and discussion

The time series data in this study represents the number of monthly road traffic accidents (RTA) and injuries (RTI) in Oman from January 2000 to June 2019. The resulting data consists of a total of 234 observations for both RTAs and RTIs.

The results show that over the past two decades, the incidence of RTAs in Oman has fallen from a high of 1,283 RTAs in October 2001 to a low 156 in February 2019 as shown in Figure (4.1). The mean is 660 for RTAs with a standard deviation 247.5. Similarly, RTIs varied from a high of 1,273 in March 2012 to a low of 125 in December 2014 with a mean of 620 RTI and standard deviation 265, as shown in Figure (4.2). Figure (4.2) shows that the monthly RTAs in Oman has a downward trend and non-stationary, which is confirmed by the Augmented Dickey-Fuller (ADF) test. The RTI time series data also have a non-stationary and downward trend as shown in Figure (4.2) and also confirmed through the ADF test.

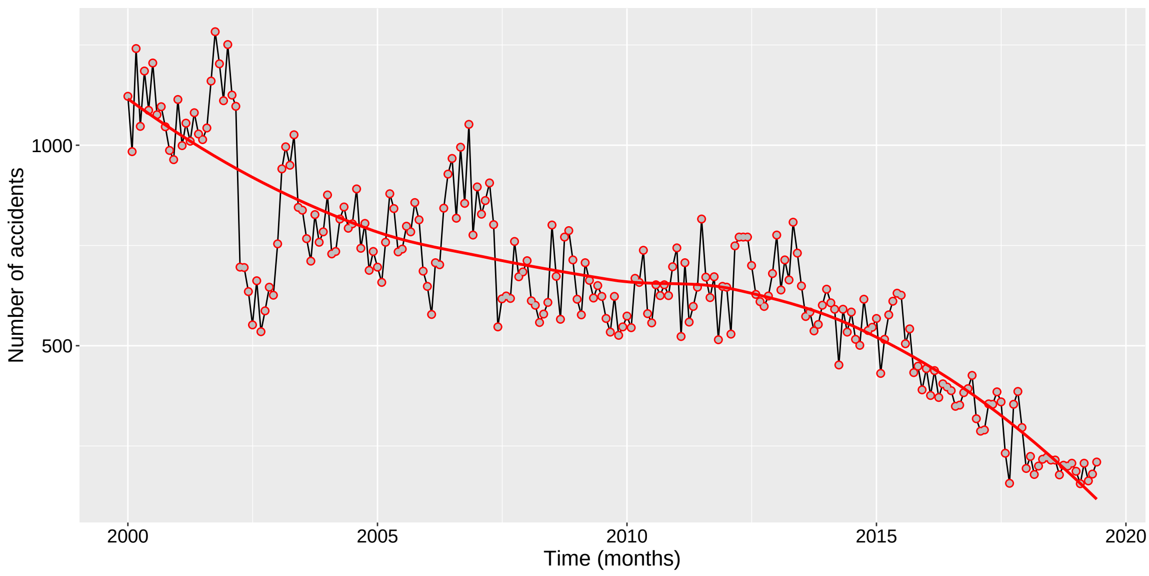


Figure 4.1: Road traffic accidents and injuries in Oman from January 2000 to June 2019

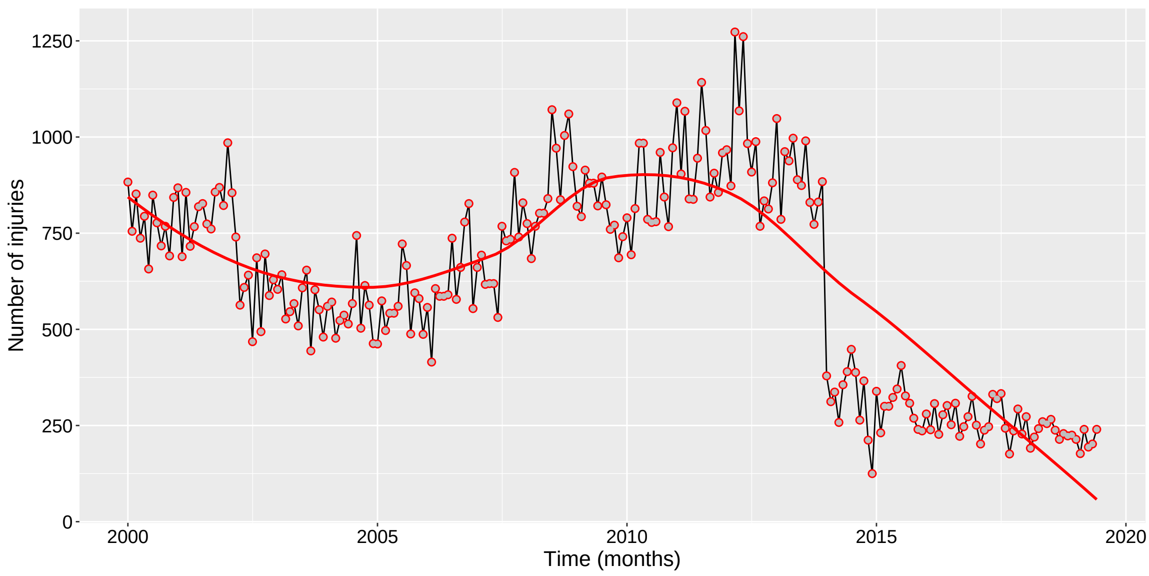


Figure 4.2: Road traffic injuries in Oman from January 2000 to June 2019

Data analysis and model fitting were conducted in R version 3.6.2. The components of the time series of RTA and RTI were decomposed and illustrated by Figure (4.3) and Figure (4.4), respectively, showing the observation, trend, seasonal and random components of both time series. Both figures suggest that the data are non-stationary and there are significant seasonalities including autoregressive and moving average components. Therefore, our primary focus is to develop suitable seasonal ARIMA models for both the series.

Different SARIMA models have been fitted in R and compared using the values of AIC, BIC, RMSE, MAPE and MASE. The values of AIC, BIC, RMSE, MAPE and MASE for different SARIMA models for RTA given in Table (4.1). While developing different models for RTA data, models with first-order difference ($d = 1$) are only considered as the data found to be stationary at lag 1. The analyses indicate that the best model for the RTA data in Oman is SARIMA $(3, 1, 1)(2, 0, 0)_{12}$ as the model has the lowest AIC (2744.69) and BIC (2768.84) values. For the RTI data, a number of models have been compared as shown in Table (4.2) and the model SARIMA$(0, 1, 1)(1, 0, 2)_{12}$ is found to be the best. Although BIC value (2821.58) is not the lowest for the SARIMA $(0, 1, 1)(1, 0, 2)_{12}$ model due to has more parameters than other models but the AIC value

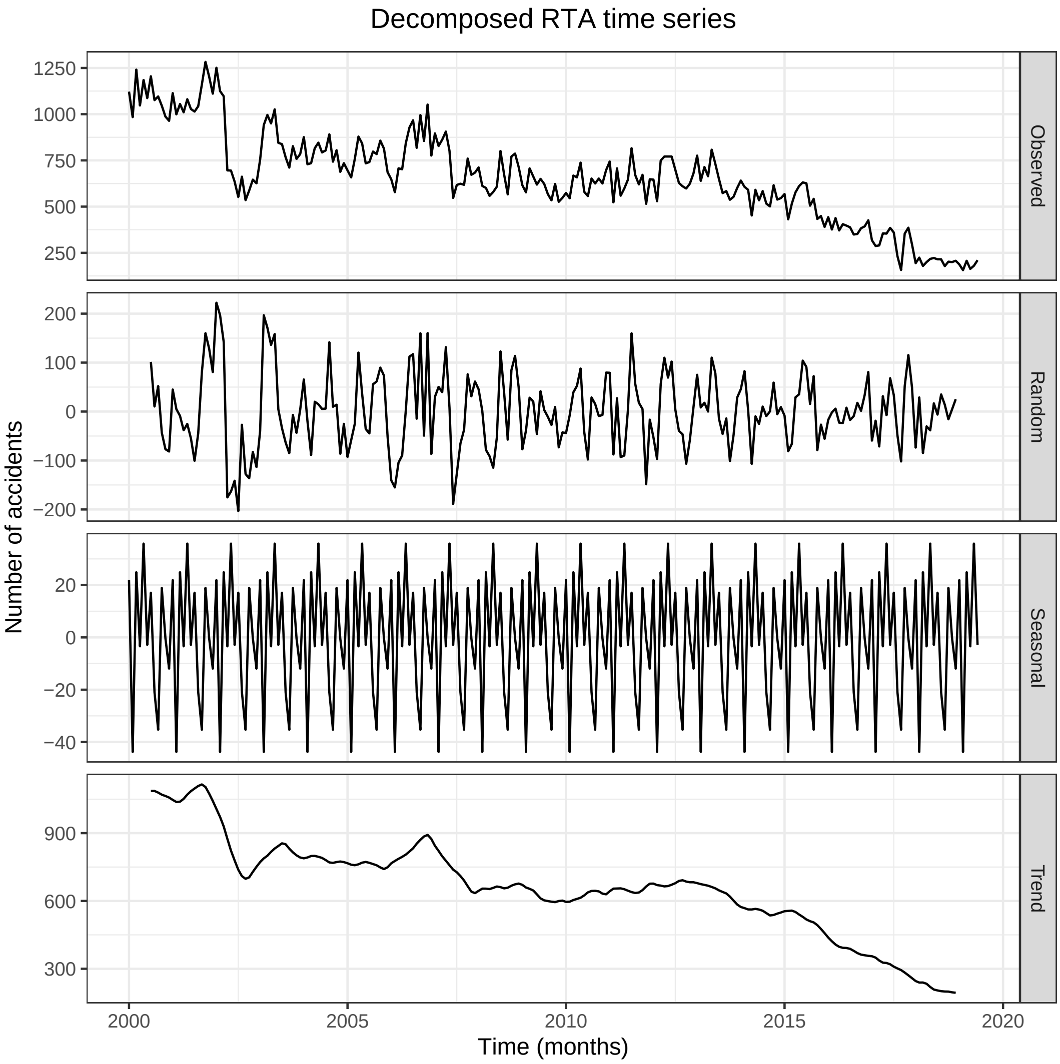


Figure 4.3: RTA time series decomposition

(2804.33) is the lowest. Additionally, this model has the lowest value in RMSE (96.73), MAPE (13.16) and MASE (0.83), which suggest that the model is better than the other models for the RTI time series in Oman.

Table 4.1: Assessment of different models for RTA data in Oman

Model	AIC	BIC	RMSE	MAPE	MASE
$(4,1,1)(2,0,0)_{12}$	2746.56	2774.17	84.47	10.93	0.89
$(4,1,1)(1,0,0)_{12}$	2748.10	2772.25	85.19	11.12	0.91
$(3,1,1)(2,0,0)_{12}$	2744.69	2768.84	84.49	10.95	0.89
$(5,1,1)(2,0,0)_{12}$	2747.42	2778.21	84.17	10.91	0.89
$(4,1,0)(2,0,0)_{12}$	2754.42	2778.58	86.35	10.85	0.91

The residuals of the final model were also evaluated by the root mean squared error (RMSE), mean absolute percentage error (MAPE) and mean absolute scale error (MASE)

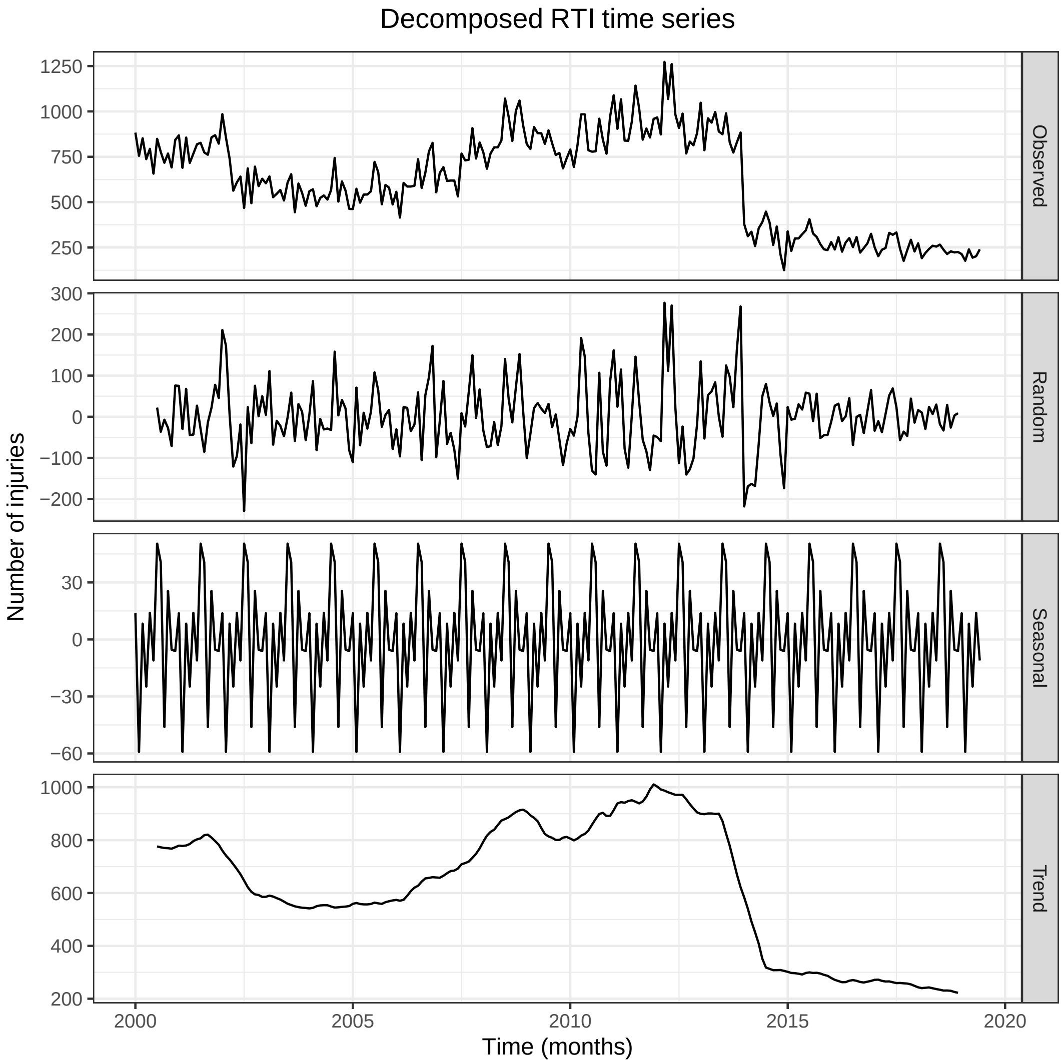


Figure 4.4: RTI time series components

Table 4.2: Assessment of different models for RTI data in Oman

Model	**AIC**	**BIC**	**RMSE**	**MAPE**	**MASE**
$(1, 1, 2)(0, 0, 2)_{12}$	2810.48	2831.19	97.75	13.66	0.85
$(0, 1, 1)(0, 0, 2)_{12}$	2807.22	2821.03	97.92	13.65	0.85
$(0, 1, 1)(1, 0, 2)_{12}$	2804.33	2821.58	96.73	13.16	0.83
$(0, 1, 1)(0, 0, 1)_{12}$	2808.89	2819.24	98.76	13.53	0.85
$(1, 1, 1)(0, 0, 2)_{12}$	2808.72	2825.97	97.80	13.68	0.85
$(0, 1, 2)(0, 0, 2)_{12}$	2808.74	2825.99	97.80	13.68	0.85

as displayed in Table (4.1). These suggest that although the model $(3, 1, 1)(2, 0, 0)_{12}$ for RTAs has slightly higher RMSE and MAPE errors, it is yet adequate and better than the other models considering AIC and BIC values. Moreover, the residual diagnostic, more specifically, the autocorrelation of the residuals were checked by the Ljung-Box test ($Q = 22.5$, p-value$= 0.21$), which shows that the test is insignificant. Figure (4.5) shows

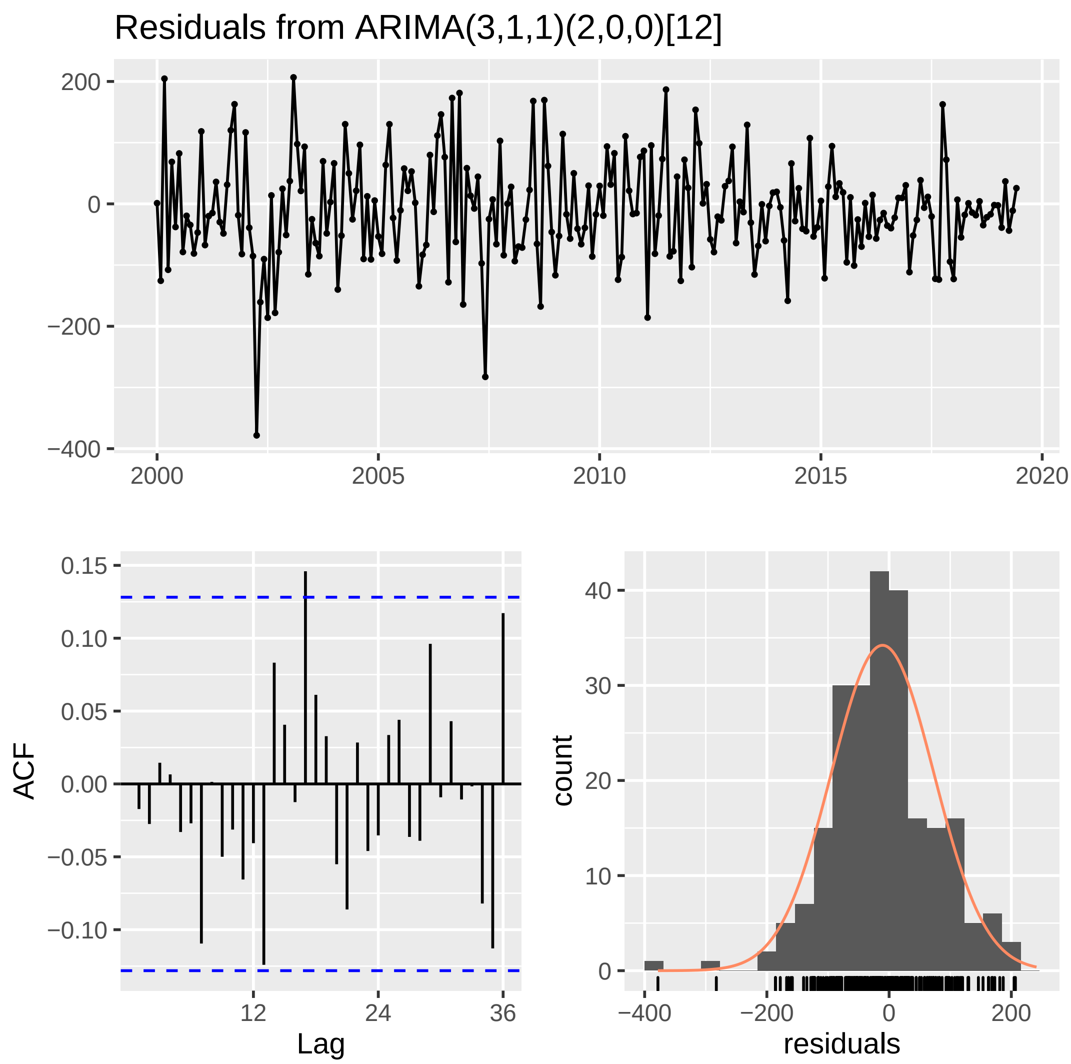


Figure 4.5: Residuals of the fitted model in RTA

Table 4.3: Coefficients and standard errors of parameters in the selected time series models

		RTA		RTI	
	Parameter	coefficient	standard error	coefficient	standard error
1	AR1	0.5565	0.0804	-	-
2	AR2	0.2440	0.0782	-	-
3	AR3	-0.0988	0.0697	-	-
4	MA1	-0.9190	0.0417	-0.4714	0.0643
5	SAR1	0.1103	0.0717	0.7401	0.3124
6	SAR2	0.1260	0.0694	-	-
7	SMA1	-	-	-0.6246	0.3154
8	SMA2	-	-	-0.0561	0.1222

that the model $(3,1,1)(2,0,0)_{12}$ residuals behaved as the white noise and ACF residuals fall near to the zero. It can be deduced from further goodness of fit analysis that the SARIMA$(3,1,1)(2,0,0)_{12}$ model fitted the data reasonably well.

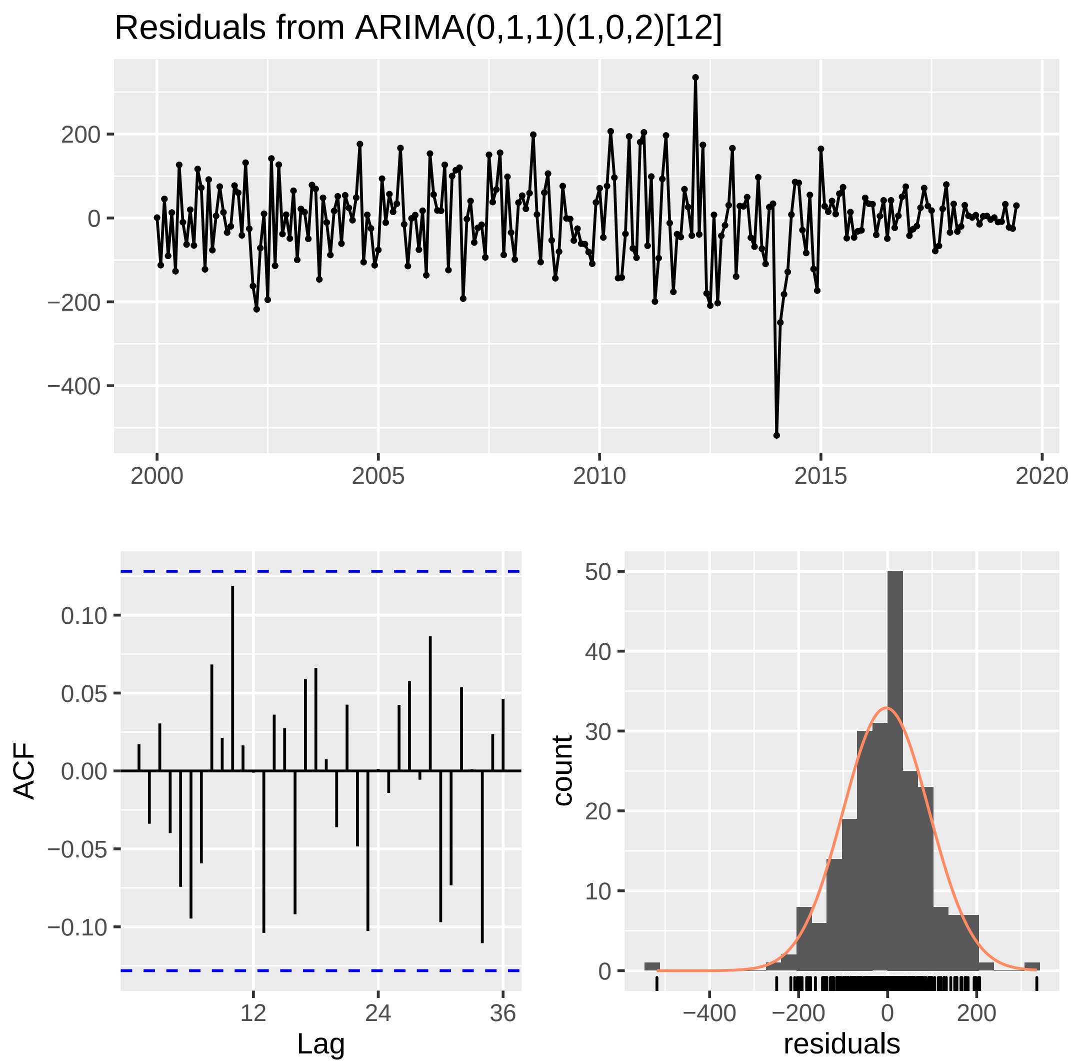


Figure 4.6: Residuals of the fitted model in RTI

Diagnostic checking and model validation were also performed as the procedures as mentioned earlier for the RTI data. The model SARIMA $(0,1,1)(1,0,2)_{12}$ has shown the highest adequacy than other models considering RMSE, MAPE and MASE, as shown in Table (4.2). Results of the Ljung-Box test ($Q = 21.6$, p-value$= 0.4136$) suggest that autocorrelation coefficients are not significantly different from zero. The ACF residuals indicate that autocorrelation is near to zero and do not deviate from a zero mean, i.e. the residuals follow a white noise process, as displying in Figure (4.6). These suggest that the SARIMA $(0,1,1)(1,0,2)_{12}$ model fitted injuries data in Oman well. Tables (4.3) shows the coefficients and standard errors of the selected models for RTA and RTI, respectively.

Now based on the final models for RTA and RTI, we have forecasted the number

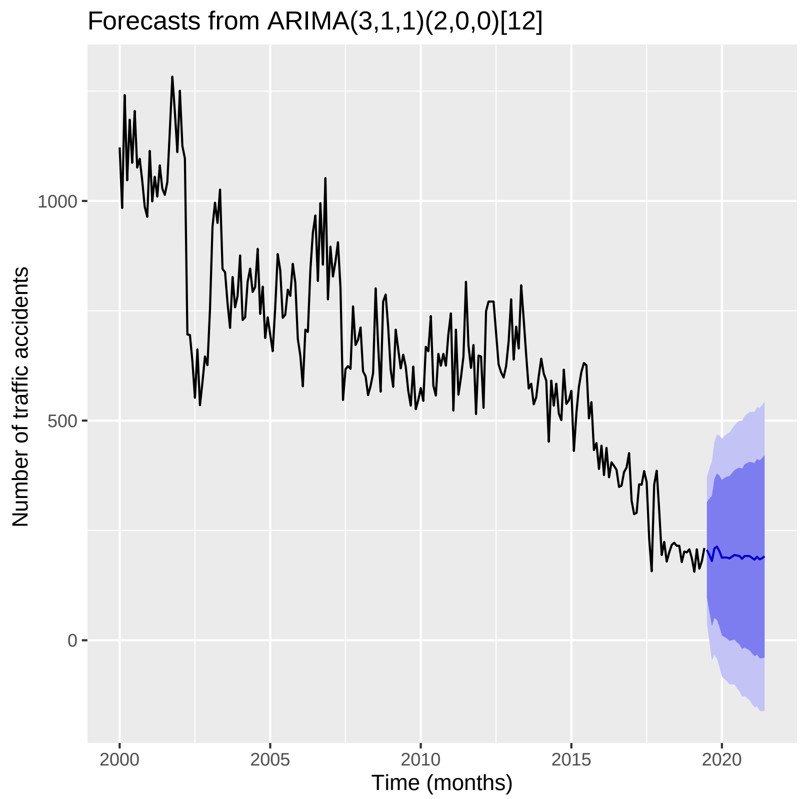


Figure 4.7: Observed (black) and forecasted values (blue) of traffic accidents in Oman

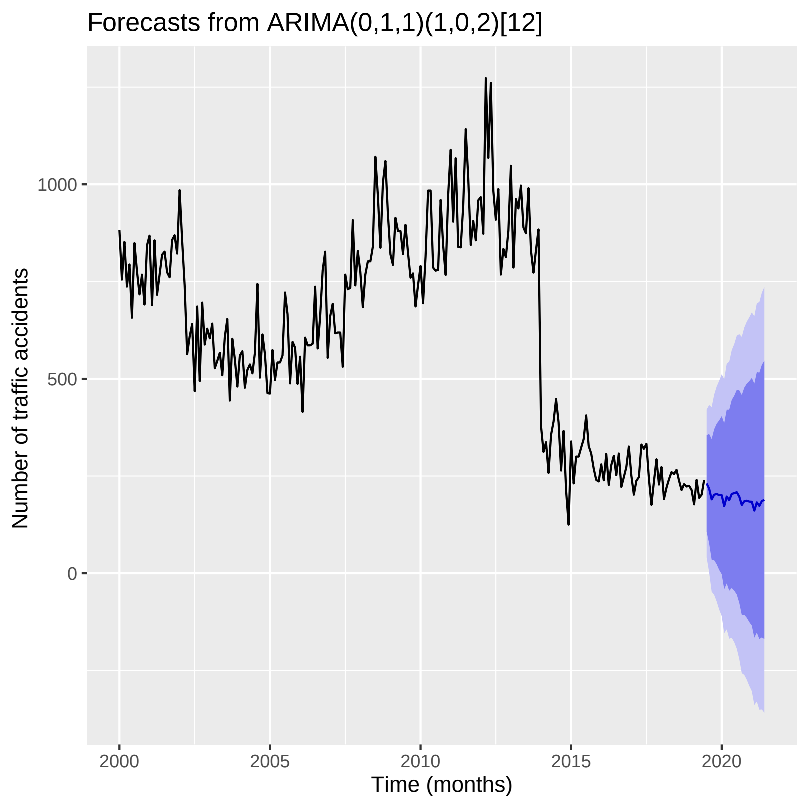


Figure 4.8: Observed (black) and forecasted values (blue) of traffic injuries in Oman

of accidents and injuries for the next 24 months. The SARIMA $(3,1,1)(2,0,0)_{12}$ model for RTA predicts the high occurrence of crashes in June, July and August in the next two years as shown in Figure (4.7). The higher number of RTAs occurred in June-July period, and the peak is in June during the summer holiday. This period is the fall season in the Dhofar region that is south monsoon rainfall time also, and the number of visitors from Arab Gulf countries increases during this period to enjoy of the fall season and visit Dhofar festival. Since people from Oman and other Gulf region are travelling between cities by highways during this period, the number of vehicles on the road rises and the chance of accidents increases. Similarly, May, June and July months are expected to have the highest number of traffic injuries in Oman forecasted by the SARIMA$(0,1,1)(1,0,2)_{12}$ model as shown in Figure (4.8). Therefore, summertime is the most accident-prone period and has a higher chance of having RTA and RTI in Oman.

4.4 Chapter summary

Road traffic accidents (RTA) and injuries (RTI) are global public health concerns to the societies both in developed and developing world resulting in loss of millions of lives, disabilities and a high volume of cost. Due to heavily stochastic pattern, seasonality, the shift of trends, forecasting road traffic accidents and injuries are highly complex. This part of the study aims to provide insights of choosing suitable time series models and analysing road traffic accidents and injuries taking RTA and RTI data in Oman as a case study as the country faces one of the highest numbers of road accidents per year. Data from January 2000 to June 2019 from several secondary sources were gathered. Time series decomposition, stationarity and seasonality were checked to identify the appropriate models for RTA and RTI. SARIMA (3, 1, 1)(2, 0, 0)(12) and SARIMA (0, 1, 1)(1, 0, 2)(12) models were found to be the best for the road traffic accident and injury data, respectively, comparing many different models. AIC, BIC and other error values were used to choose the best model. Model diagnostics were also performed to confirm the statistical assumptions and two-year forecasting was performed. The analyses in this study would help many Government Departments, academic researchers and decision-makers to generate policies to reduce accidents and injuries.

Chapter 5

Spatial modelling of road traffic accidents data

5.1 Introduction

Spatial models are found to be an elegant technique to capture localisation effects and influence of various factors in many different areas. As the effect of factors may vary with locations, spatial modelling techniques provides significantly improved estimates and better prediction in comparison to non-spatial models (Li et al., 2013). Ignoring the spatial effect in regression modelling may cause bias and a higher standard error in the estimation of model parameters and a lower amount of regression variation explained by the model (Aguero-Valverde and Jovanis, 2008; Barua et al., 2015). An enormous amount of studies have been carried out, in which spatial models capture spatial effects in a variety of ways based on the different forms of spatial distributions. Some of those spatial models are: spatial autoregressive (SAR) models (Rhee et al., 2016; Anselin, 1988a), spatial error models (SEM) (Al-Hasani et al., 2019a; Quddus, 2008a), multiple memberships models (MMM) (Cerasa and Cerioli, 2017; El-Basyouny and Sayed, 2009), extended multiple membership models (EMMM) (Browne et al., 2001; Shoukri et al., 2005; Spiess, 1998). Although these models are spatial models, the parameter estimates remain fixed for all locations. The spatial variations in these models are taken into account only through the spatial error structure (Pirdavani et al., 2014). However, there is another type of spatial modelling approach that provides a set of local models obtained by the calibration of multiple geographical entities. Two widely used forms of such models are– geographically weighted regression (GWR) models when the response variable is measured in continuous scale (Wang and Li, 2017; Fotheringham et al., 2002; Brunsdon et al., 1996) and geographically weighted Poisson regression (GWPR) models when the response variable is a count (Nakaya et al., 2005; Li et al., 2013; Pirdavani et al., 2014). These geographically weighted models allow the parameters to vary through the spatial units of a study area to reflect the local characteristics. Besides, these models focus on the geographic difference of factors affecting the outcome variable (Hezaveh et al., 2019). Despite of these promising features of a geographically weighted model, there are two main issues involved in fitting the geographically weighted regression models, which are the selection of bandwidth and kernel weighting function.

During the calibration of a geographically weighted model, the bandwidth, also referred to as window size, controls the fit of a localised model. Small bandwidths results in high spatial variation while the large bandwidth leads to the estimates close to the global models (Lu et al., 2014). Generally, there are two suggested methods in the literature for the selection of bandwidth– classical fixed bandwidth approach or adaptive bandwidth approach (Davies and Lawson, 2019). Several studies found that adaptive (optimal) bandwidth performs better than fixed bandwidth (John Braun and Rousson, 2000; Wang and Li, 2017; Davies and Lawson, 2019). Recently, Davies and Lawson (2019) compared the fixed and adaptive (optimal) bandwidth and recommended that adaptive bandwidth is optimal due to its performance, which is estimated through the integrated squared error between a given density estimate and the true density under a given scenario.

The main challenge of a GWR and GWPR model is to find the most suitable kernel weighting function which gives weights for the neighbouring observations during model calibration. If the generated weights through a kernel function are not accurate then the parameter estimates have large standard errors and less reliable predictions (Li et al., 2013). There are five different kernel functions used for GWR or GWPR modelling including box-car, bi-square, tri-cube, exponential and Gaussian weighted function (Li et al., 2013). These weighting functions generate weights for each observation and use them to calibrate the parameter estimates according to the spatial position of data points (Li et al., 2013). However, the potential problem for a modeller is to select the best possible kernel weighting function to generate the weights (Brunsdon et al., 1996).

Several studies applied different kernel functions for the geographically weighted models in many areas including road traffic accidents. Li et al. (2013) chose an adaptive bi-square kernel for weighting schemes in a GWPR model. Pirdavani et al. (2014) found that Gaussian weighted functions with adaptive bandwidth is the most suitable in GWRP models while comparing the bi-square and the Gaussian kernel weighting functions. Although some studies (Nakaya et al. (2005); Hadayeghi et al. (2010); Xu and Huang (2015); Hezaveh et al. (2019)) compared the bi-square and Gaussian kernel functions with both fixed and adaptive bandwidths. Broadly speaking, these empirical studies concentrated only on two kernel functions, bi-square and Gaussian weighted functions, in order to inspect spatial factors in GWPR models. Despite these efforts, it remains unclear which kernel function would provide the best fit for a geographically weighted Poisson regression (GWPR) model.

In this study, we compared the suitability of a GWPR model for five different kernel weighting functions: box-car, bi-square, tri-cube, exponential and Gaussian weighted function. The model formulation has been developed including derivation of the likelihood function, estimation of parameter through calibration of models by using different kernel weighting functions. The model framework has been applied to the road traffic accident (RTA) data in Oman to explore different factors associated with RTA and give insights of geographical variations of such factors. Several diagnostic tools, namely, geographically weighted deviance (GWD), Akaike information criterion (AIC) and the corrected Akaike Information Criterion (AICc) have been used to find the most suitable model with the associated kernel weighting function.

The rest of the chapter is structured as: the modelling includes the spatial models, spatial count model, kernel function, likelihood estimation and scoring and model diagnostic tools in Section (5.2), results and discussion in Section (5.3) and summary of this chapter in Section (5.4).

5.2 Modelling

5.2.1 Spatial model

In this study, we initially compare the spatial lag model (SLM) and the spatial error model (SEM) with road traffic accidents (RTA) data. Three models: Ordinary least square model (OLS), spatial lag model (SLM) and spatial error model (SEM) are applied in this study with the spatial data to compare their performance and accuracy. The ordinary least square model can be written as

$$Y_i = \beta_i X_i + \xi, \tag{5.1}$$

where Y_i is the number of RTAs as the dependent variable for $i = 1, 2, \ldots, n$, β_i is a vector of parameters, X_i is a matrix of independent variables and, ξ is a vector of unobserved error terms that assumed to be distributed normally $N(0, \sigma^2)$. The estimates of parameters β, in matrix notation, Brunsdon et al. (1996) can be given as

$$\hat{\beta_i} = (X^T X)^{-1} X^T Y, \tag{5.2}$$

where X^t is the transpose of data matrix, X, and Y is the observed vector of the dependent variable.

In this study, we evaluate two spatial models SLM and SEM, where the neighbouring (spatial) effect for a region i is considered to be affected by the other neighbouring regions $j, j = 1, 2, \ldots, n$ (Anselin, 1988b). Let us define the spatial dependence as

$$Y_i = f(Y_j), \quad i, j = 1, 2, \ldots, n; i \neq j, \tag{5.3}$$

where Y_i is the natural logarithm of the number of RTAs in i spatial units. However, this study applied two spatial models (SLM and SEM) compared with the OLS model as given in Equation (5.1).

The spatial lag model (SLM) can be written as

$$Y_i = \rho W_y + \beta_i X_i + \xi, \tag{5.4}$$

where Y_i is the vector of natural logarithm of RTAs for ith spatial unit (region), ρ is the spatial lag coefficient, W_y is the spatial weight matrix, X_i is the matrix of independent variables, β_i is the vector of parameters and ξ is the unobserved error terms vector.

The spatial error model (SEM) can be written as

$$Y_i = \beta_i X_i + \mu, \quad \text{with} \quad \mu = \lambda W_\mu + \xi, \tag{5.5}$$

where μ is the function of unobserved error terms, λ is a spatial error coefficient and W_μ is the spatial error weight matrix. The spatial weights are distinct elements in any cross-sectional analysis of spatial dependence. They are necessary components in the setting up of spatial autocorrelation statistics and provide calibration between units. For both SLM and SEM, the spatial weight matrix W can be defined as

$$W = \begin{bmatrix} W_{11} & W_{12} & \dots & W_{1n} \\ W_{21} & W_{22} & \dots & W_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{n1} & W_{n2} & \dots & W_{nn} \end{bmatrix}. \tag{5.6}$$

5.2.2 Spatial count model

Essentially, to model and analyse count data such as road traffic accident (RTA) data, Poisson regression model is the most basic model (Lord and Mannering, 2010). Aguero-Valverde (2013) pointed to the first stage of the RTA counts to model RTA using a Poisson process defined as:

$$Y \sim Poisson(\lambda). \tag{5.7}$$

Poisson regression model that gives the probability of (Y_i) the number of RTA in roadway entity (segment, intersection, region etc.) i for some time period (where $Y_i \geq 0$) is given by:

$$P\left(Y_i\right) = \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} \tag{5.8}$$

where Y_i is the number of RTA as the response variable and λ is the Poisson parameter of the road entity i, which is the expected number of accidents every year (mean, $E[y_i]$) (Lord and Mannering, 2010).

A Poisson regression model is fitted by estimating the Poisson parameters λ_i (expected number of accidents) as a function of explanatory variables, the most common functional form being (Lord and Mannering, 2010)

$$\lambda_i = e^{\beta X_i}, \tag{5.9}$$

where X_i is the vector of explanatory variables and, β is the vector of estimable parameters (Lord and Mannering, 2010). The Poisson mean is modelled by the following the log-link function (Aguero-Valverde, 2013)

$$\log(\lambda_i) = \beta_0 + \sum \beta_i X_i + \epsilon_i, \tag{5.10}$$

where β_0 is the intercept, β_i is the coefficient for covariate X_i, and ϵ_i is the error term.

Spatial count models, a special form of Poisson count model, allow parameter values to vary with spatial units $\mathbf{u}_i$ which is a vector describing the location i. A Poisson spatial model can be written as

$$Y_i \sim \text{Poisson}\left[\exp\left(\sum_k \beta_k(\mathbf{u}_i)x_{ik}\right)\right], \tag{5.11}$$

where x_{ik} is kth explanatory variable in location i, β_k is the parameter and $\mathbf{u}_i = (u_{l_i}, u_{h_i})$ is the vector describing the latitude and longitude at location i. This model clearly has a geographically varying coefficient and known as geographically weighted Poisson regression (GWPR). The covariate form of the model GWPR with a group of predictors, in which the parameters are allowed to vary over space can be written as

$$\ln(\mathbf{Y}) = \ln(\beta_0(\mathbf{u}_i)) + \beta_1(\mathbf{u}_i)\mathbf{X}_1 + \beta_2(\mathbf{u}_i)\mathbf{X}_2 + \cdots + \beta_K(\mathbf{u}_i)\mathbf{X}_K + \epsilon_i \tag{5.12}$$

where β_k is the function of the location $\mathbf{u}_i = (u_{l_i}, u_{h_i})$ denotes the two dimensional coordinates of the ith point in space. The parameters in the model $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_K)$ are allowed to be different between locations. Thus, the GWPR modelling framework addresses the spatial heterogeneity.

To estimate the parameters of a geographically weighted Poisson regression, a form of maximum likelihood principal is developed. The method is analogous to a locally maximum likelihood principle and can be referred to as geographically weighted likelihood method. The geographically weighted log-likelihood at location $\mathbf{u}_i$ can be given by

$$\max L(\mathbf{u}_i) = \sum_{j=1}^{N} \left(-\widehat{Y}_j \left(\boldsymbol{\beta}(\mathbf{u}_i) \right) + Y_j \log \widehat{Y}_j \left(\boldsymbol{\beta}(\mathbf{u}_i) \right) \right) \cdot \omega_{ij} \left(\|\mathbf{u}_i - \mathbf{u}_j\| \right), \tag{5.13}$$

where $\widehat{Y}_j \left(\beta(\mathbf{u}_i) \right)$ is the predicted number of events at location $\mathbf{u}_i = (u_{l_i}, u_{h_i})$ with parameters at regression point i is

$$\widehat{Y}_j \left(\beta(\mathbf{u}_i) \right) = \exp \left(\sum_k \widehat{\beta}_k(\mathbf{u}_i) x_{jk} \right), \tag{5.14}$$

and w_{ij} is the geographical weight of the jth observation at ith regression point. The weights ω_{ij} of the observations decrease gradually as the distance between the regression point i and the observation at location j gets larger.

The model needs to be calibrated as used in geographically weighted regression and smoothed with a spatial weighting function. The spatial weighting function is, in general, the spatial weighting kernel. These local statistics usually play a role in deploying the variation of variables of interest at the geographical location to display. In particular, the calibration of the GWPR model is localised by weighting each observation in the data set according to a proximity point or centred point. Based on a kernel weighting scheme, observations which are close to the proximity points are highly weighted than further points (Fotheringham et al., 2015). However, finding the appropriate kernel is crucial as the weights of the neighbouring observations contribute to the estimates through a distance decay function.

5.2.3 Kernel functions

There are several kernel functions used for weighting in GWPR model including box-car, bi-square, tri-cube, exponential and Gaussian. The simple form of a kernel function is the box-car kernel function. It is a discontinuous function and uses only the observations that are within a distance, say b, from the GWPR model calibration point (Gollini et al., 2013). However, the box-car function is computationally efficient as it uses a smaller subset of observations to fit the local model at each GWPR model calibration point. The box-car weighted kernel function can be given as

$$\omega_{ij} = \begin{cases} 1 & \text{if } |d_{ij}| < b, \\ \\ 0 & \text{otherwise.} \end{cases} \tag{5.15}$$

where ω_{ij} is the jth element of the diagnoal matrix when calibrating the model for i calibration point, d_{ij} is the distance between i and jth points and b is the bandwidth.

The bi-square is also a discontinuous function and gives null weights to spatial units observation with a distance greater than the bandwidth (Gollini et al., 2013). The bi-square kernel function is written as

$$\omega_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2 & \text{if } |d_{ij}| < b, \\ \\ 0 & \text{otherwise.} \end{cases} \tag{5.16}$$

Similar to bi-square, the tri-cube kernel gives a cubic weight and can be given as

$$\omega_{ij} = \begin{cases} \left(1 - \left(\frac{|d_{ij}|}{b}\right)^3\right)^3 & \text{if } |d_{ij}| < b, \\ \\ 0 & \text{otherwise.} \end{cases} \tag{5.17}$$

The exponential kernel is a continuous function of the distance between two spatial observation points or a calibration and an observation points. The weights can be a maximum of equal to 1 for an inspection at the GWPR model calibration point and will have an exponential decrease with the increase of the distance between calibration points (Gollini et al., 2013). The exponential kernel function is given by

$$\omega_{ij} = \exp\left(-\frac{|\, d_{ij} |}{b}\right) \tag{5.18}$$

The classical weighted kernel function is the Gaussian kernel function defined as

$$\omega_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right). \tag{5.19}$$

5.2.4 Likelihood estimation and scoring method for GWPR

Consider the GWPR model

$$Y_i \sim \text{Poisson}\left[\exp\left(\sum_k \beta_k(\mathbf{u}_i)x_{ik}\right)\right], \tag{5.20}$$

with the likelihood function

$$\max L(\mathbf{u}_i) = \sum_{j=1}^{N}\left(-\widehat{Y}_j\left(\boldsymbol{\beta}(\mathbf{u}_i)\right) + Y_j \log \widehat{Y}_j\left(\boldsymbol{\beta}(\mathbf{u}_i)\right)\right) \cdot \omega_{ij}\left(\|\mathbf{u}_i - \mathbf{u}_j\|\right). \tag{5.21}$$

The parameter vector $\boldsymbol{\beta}$ is estimated by solving

$$\frac{\partial L(\mathbf{u}_i)}{\partial \widehat{\boldsymbol{\beta}}(\mathbf{u}_i)} = 0. \tag{5.22}$$

Equation (5.22) gives

$$\frac{\partial}{\partial\widehat{\boldsymbol{\beta}}}\left[\sum_{j=1}^{N}\left(-\widehat{Y}_j\left(\boldsymbol{\beta}(\mathbf{u}_i)\right)+Y_j\log\widehat{Y}_j\left(\boldsymbol{\beta}(\mathbf{u}_i)\right)\right)\cdot\omega_{ij}\left(\|\mathbf{u}_i-\mathbf{u}_j\|\right)\right]=0 \tag{5.23}$$

Maximum likelihood estimates are obtained by using the iterative re-weighted least square method using (5.23) which can also be referred as local Fisher's scoring method and given by

$$\boldsymbol{\beta}^{(l+1)}(\mathbf{u}_i)=\left(\mathbf{X}^T\boldsymbol{\Psi}(\mathbf{u}_i)\boldsymbol{\Gamma}^{(l)}(\mathbf{u}_i)\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{\Psi}(\mathbf{u}_i)\boldsymbol{\Gamma}^{(l)}(\mathbf{u}_i)\mathbf{y}^{*(l)}(\mathbf{u}_i) \tag{5.24}$$

where $\boldsymbol{\beta}^{(l+1)}(\mathbf{u}_i)$ is a vector of local parameter estimates specific to location i while superscript $(l+1)$ gives the number of iterations. At lth stage, the parameter vector

$$\boldsymbol{\beta}^{(l)}(\mathbf{u}_i)=(\beta_0^{(l)}(\mathbf{u}_i),\beta_1^{(l)}(\mathbf{u}_i),...,\beta_K^{(l)}(\mathbf{u}_i))^T. \tag{5.25}$$

where X^T denotes the transpose of the design matrix X and gives as

$$\mathbf{X}=\begin{bmatrix}1 & x_{11} & \dots & x_{K1}\\ 1 & x_{12} & \dots & x_{K2}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & x_{1N} & \dots & X_{KN}\end{bmatrix}. \tag{5.26}$$

The spatial weights matrix $\boldsymbol{\Psi}(\mathbf{u}_i)$ for location i is a diagonal matrix and given as

$$\boldsymbol{\Psi}(\mathbf{u}_i)=\begin{bmatrix}\omega_{i1} & 0 & \dots & 0\\ 0 & \omega_{i2} & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \omega_{iN}\end{bmatrix} \tag{5.27}$$

The variance weights matrix $\mathbf{\Gamma}^{(l)}(\mathbf{u}_i)$ for the location i is

$$\mathbf{\Gamma}^{(l)}(\mathbf{u}_i) = \begin{bmatrix} \widehat{Y}_1(\boldsymbol{\beta}^{(l)}(\mathbf{u}_i)) & \dots & 0 & \\ \dots & \widehat{Y_2}(\boldsymbol{\beta}^{(l)}(\mathbf{u}_i)) & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \widehat{Y}_N(\boldsymbol{\beta}^{(l)}(\mathbf{u}_i)) & \end{bmatrix} \tag{5.28}$$

The vector of adjusted dependent variables $\mathbf{y}^{*(l)}(\mathbf{u}_i)$ is given by

$$\mathbf{y}^{*(l)}(\mathbf{u}_i) = \left(y_1^{*(l)}(\mathbf{u}_i), y_2^{*(l)}(\mathbf{u}_i), \cdots, y_N^{*(l)}(\mathbf{u}_i)\right)^T \tag{5.29}$$

where

$$\begin{aligned} y_j^{*(l)}(\mathbf{u}_i) &= \left(\beta_0^{(l)}(\mathbf{u}_i) + \sum_k^K \beta_k^{(l)}(\mathbf{u}_i)x_{jk}\right) + \frac{Y_j - \widehat{Y}_j(\boldsymbol{\beta}^{(l)}(\mathbf{u}_i))}{\widehat{Y}_j(\boldsymbol{\beta}^{(l)}(\mathbf{u}_i))} \\ &= \eta_j(\boldsymbol{\beta}^{(l)}(\mathbf{u}_i)) + \frac{Y_j - \widehat{Y}_j(\boldsymbol{\beta}^{(l)}(\mathbf{u}_i))}{\widehat{Y}_j(\boldsymbol{\beta}^{(l)}(\mathbf{u}_i))} \end{aligned} \tag{5.30}$$

where η_j is a linear predictor of jth observation. The sets of the local parameter can be estimated by repeating the iterative procedure for each location i. The final estimate can be written as

$$\boldsymbol{\beta}(\mathbf{u}_i) = \left(\mathbf{X}^T\mathbf{\Psi}(\mathbf{u}_i)\mathbf{\Gamma}(\mathbf{u}_i)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{\Psi}(\mathbf{u}_i)\mathbf{\Gamma}(\mathbf{u}_i)\mathbf{y}^*(\mathbf{u}_i) \tag{5.31}$$

The prediction of each observation j at each regression points i would be

$$\widehat{Y}_{ij} = \begin{bmatrix} \widehat{\eta}_1(u_1) & \widehat{\eta}_1(u_2) & \dots & \widehat{\eta}_1(u_N) \\ \widehat{\eta}_2(u_1) & \widehat{\eta}_2(u_2) & \dots & \widehat{\eta}_2(u_N) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\eta}_N(u_1) & \widehat{\eta}_N(u_2) & \dots & \widehat{\eta}_N(u_N) \end{bmatrix}$$

$$= \mathbf{C} \begin{bmatrix} y_1^*(\mathbf{u}_1) & y_1^*(\mathbf{u}_2) & \dots & y_1^*(\mathbf{u}_N) \\ y_2^*(\mathbf{u}_1) & y_2^*(\mathbf{u}_2) & \dots & y_2^*(\mathbf{u}_N) \\ \vdots & \vdots & \ddots & \vdots \\ y_N^*(\mathbf{u}_1) & y_N^*(\mathbf{u}_2) & \dots & y_N^*(\mathbf{u}_N) \end{bmatrix} \tag{5.32}$$

where the ith row of the matrix $\mathbf{C}$, $\mathbf{c}_i$, is

$$\mathbf{c}_i = \mathbf{x}_i(\mathbf{X}^T\mathbf{\Psi}(\mathbf{u}_i)\mathbf{\Gamma}(\mathbf{u}_i)X)^{-1}\mathbf{X}^T\mathbf{\Psi}(\mathbf{u}_i)\mathbf{\Gamma}(\mathbf{u}_i) \tag{5.33}$$

To predict the number of events at a location, the diagonal elements of the matrix on the lefthand side in Equation (5.24) are utilised. The mapping from adjusted dependent variable to linear predictor is represented as given below

$$\widehat{\eta} = \mathbf{S}y^* \tag{5.34}$$

where

$$\widehat{\eta} = (\widehat{\eta}_1(u_1), \widehat{\eta}_2(u_2), \dots, \widehat{\eta}_N(u_N))^T \text{ and } y^* = (y_1^*(\mathbf{u}_1), y_2^*(\mathbf{u}_2), \dots, y_N^*(\mathbf{u}_N))^T.$$

Analogous to the hat matrix, $\mathbf{S}$ is the matrix with the (ij)th element S_{ij} being

$$S_{ij} = C_{ij}\frac{y_i^*(\mathbf{u}_j)}{y_j^*(\mathbf{u}_j)}, \tag{5.35}$$

where C_{ij} is the (ij)th element of the matrix $\mathbf{C}$.

5.2.5 Model diagnostic tools

To compare and evaluate the GWPR model performance with five the aforementioned kernel weight functions, we have used three indicators: Akaike information criterion (AIC), corrected Akaike information criterion (AICc) and geographically weighted deviance (GW). The AIC value of a model with bandwidth b is defined as

$$\text{AIC} = D(b) + 2k(b), \tag{5.36}$$

where D is the deviance of the model, k is the number of parameters in the model and b is bandwidth used in the kernel weighting for the model. A model with minimum AIC value has the higher goodness of fit (Nakaya et al., 2005). However, in case of local spatial regression model with a small degrees of freedom, a bias adjustment in the AIC is introduced (Pirdavani et al., 2014). The Akaike information criterion with bias correction (AICc) can then be defined as

$$\text{AICc} = D(b) + 2k(b) + 2\frac{k(b)(k(b) + 1}{n - k(b) - 1} \tag{5.37}$$

where n is the number of spatial observations. We can rewrite (5.37) as

$$\text{AICc} = \text{AIC} + 2\frac{k(b)(k(b) + 1}{n - k(b) - 1}. \tag{5.38}$$

For a geographically weighted Poisson regression model the deviance statistic, referred to as geographically weighted deviance (GWD), can be given by

$$\text{GWD}(b) = \left[\sum_{i=1}^{N} Y_i \log\Big(Y_i/\widehat{Y}_i\left(\boldsymbol{\beta}(\mathbf{u}_i), b\right)\Big) - \sum_{i=1}^{N}\Big(Y_i - \widehat{Y}_i\left(\boldsymbol{\beta}(\mathbf{u}_i), b\right)\Big)\right] \tag{5.39}$$

where Y_i is observed number of events in the ith location, $\widehat{Y_i}$ is the predicted value for the ith location and $\boldsymbol{\beta}(\mathbf{u}_i)$ is the parameter vector for the set of k explanatory variables. The most suitable would be the one with the lowest values of the Akaike information criterion (AIC), corrected Akaike information criterion (AICc) and geographically weighted deviance (GWD).

5.3 Results and discussion

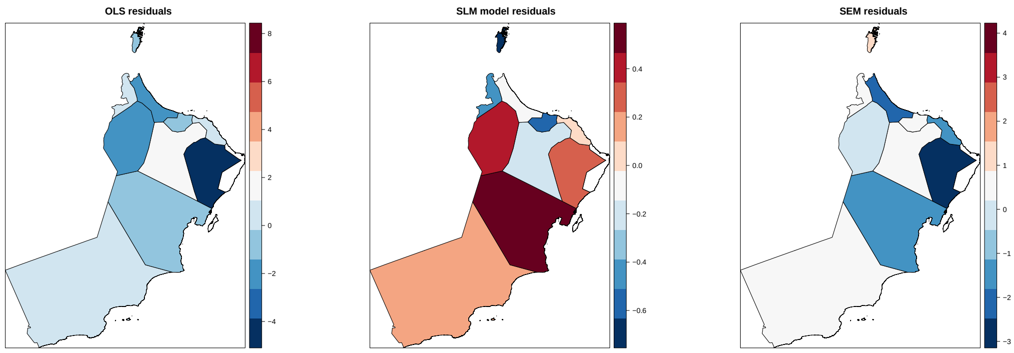
5.3.1 Spatial model

In this study, three models have been applied which are ordinary least square (OLS), spatial lag model (SLM) and spatial error model (SEM) to the RTA data of Oman for the year 2017. In the regression models, the number of road traffic accidents (RTA) in different governorates (regions) in Oman is considered as the dependent variable. The independent variables in this study are population density, population size, number of vehicles, number of job seekers (unemployment), number of accidents caused by speed in 2017 and number of accidents occurring during seasonal months in 2017 at the Sultanate of Oman. There are eleven governorates in Oman representing the spatial units in this study. The results are shown in Table (5.1) that are obtained from the fitting of the three models– OLS, SLM and SEM.

Table 5.1: Results of parameter estimates in different spatial models for RTA data.

Dependent variable	OLS	SLM	SEM
Spatial lag coefficient,ρ	-	0.061498	-
Spatial error coefficient,λ	-	-	-1.5208
Constant	-1.96e+01**	-3.76e+01***	-2.06e+01***
population density	4.90e-01*	2.39e-01***	5.11e-01***
Population size	2.85e-05	-3.76e-05***	4.40e-05***
No. of vehicles	-3.99e-04**	-2.20e-04***	-4.04e-04***
No. of job seekers	-1.10e-03	-1.07e-04	-1.20e-03*
Speed driving	6.64e-01**	6.00e-01***	6.53e-01***
Season	2.56e+00***	2.76e+00***	2.53e+00***

Level of signficance: $* : P < 0.05$, $** : P < 0.001$, $*** : P < 0.0001$



(a) Residuals for OLS (b) Residuals for SLM (c) Residuals for SEM

Figure 5.1: Models residuals through Oman's Governorates.

The results shown in the Table (5.1) give the comparison of the estimated parameters and their significance levels ($* : P < 0.05$, $** : P < 0.001$, $*** : P < 0.0001$). The effects of two variables (number of speed driving and season accidents) obtained are similar in all three models while there is a clear difference in population density, population size and number of vehicles. Spatial error model gives very close coefficient to ordinary least square to the number of job seekers variable.

For checking the adequacy of models, the residual plots can be used. The residuals of the three models were also evaluated as displayed in Figure (5.1). This suggests that the SLM for RTA data has higher accuracy than others. To select the most suitable model, Akaike information criteria (AIC) and Bayesian information criteria (DIC) are the two major criteria widely used including spatial data analysis (Anselin, 1988b).

Table 5.2: Evaluation of different spatial models for RTA data.

Diagnostic Criteria	OLS	SLM	SEM
log-likelihood	-28.18	-6.34	-25.78
AIC	72.37	30.69	69.55
BIC	75.54	34.27	73.13

As showing in Table (5.2), the values of the log-likelihood, AIC and BIC of OLS, SLM and SEM models were compared in order to determine the best model. The most interesting finding is that the SLM outperforms the SEM due to best values of log-likelihood = −6.34, AIC = 30.69 and BIC = 34.27 while the corresponding values for SEM are −25.78, 69.55 and 73.13, respectively. Therefore, the SLM model is found to be the best to identify associated factors for the road traffic accidents in Oman.

5.3.2 Spatial count model

In this study, five geographically weighted Poisson regression (GWPR) models are evaluated and compared with different kernel weighting functions including box-car, bi-square, tri-cube, exponential and Gaussian kernel weighting function. Each model used different kernel weighting function to evaluate the model performance and find out the most suitable one. There are eleven spatial units in this study, namely, governorates (regions) in Oman. One of the secondary aspects of spatial modelling is to choose the bandwidth size to give weight for the nearest spatial units. The adaptive bandwidth was employed for each kernel weighting functions in our work. Three diagnostic tools were used to evaluate the performance of the GWPR models: geographically weighted deviance (GWD), Akaike information criterion (AIC) and the corrected Akaike Information Criterion (AICc).

The secondary data consist of seven variables: number of RTA as the dependent variable and six variables: population density, population size, number of registered vehicles, number of unemployed persons, speed driving and seasons as explanatory variables for the Sultanate of Oman in 2017. There are eleven governorates considered as the eleven spatial units for the GWPR model. A primary analysis has been conducted by fitting a Poisson generalised linear model to learn about the significant factors. Table (5.3) provides the fitting of the global Poisson regression model in generalised linear modelling (GLM) framework. However, the Poisson GLM model shows that population size, number of registered vehicles, number of unemployed persons and season are significant.

The mean and variance of the number of accidents in Oman are found to be 350

Table 5.3: Poisson GLM fitting

Coefficients	Estimate	Standard Error	Z-value	p-value
Intercept	$4.40e{+}00$	$5.19e{-}02$	84.82	$< 2e{-}16^{***}$
Population size	$8.42e{-}07$	$4.15e{-}07$	2.03	0.042^{*}
Population density	$8.28e{-}04$	$1.46e{-}03$	0.57	0.570
No. of registered vehicles	$-3.16e{-}06$	$7.75e{-}07$	−4.08	$4.54e{-}05^{***}$
No. of unemployed persons	$2.42e{-}05$	$1.46e{-}05$	1.66	0.097
Speed driving	$1.48e{-}03$	$1.31e{-}03$	1.13	0.260
Season	$7.08e{-}03$	$2.96e{-}03$	2.39	0.017^{*}

Level of significance: $^{*}p < 0.05$, $^{**}p < 0.001$, $^{***}p < 0.0001$

and 110,577, respectively per governorate per year. This clearly indicates the over-dispersion in the data and the inadequacy of the fitted global Poisson GLM model in Table (5.3). Therefore, we fitted a generalised Poisson GLM (Consul and Famoye, 1992) and a negative binomial GLM (Hilbe, 2011) and the evaluation of model fitting is shown in Table (5.4) in the next section. However, the high variation of the number of accidents at different governorate level provides an indication that different factors may have impacted the RTA differently at each governorate level and local level models may explain the variation more clearly. This motivates applying the geographically weighted Poisson regression.

5.3.3 GWPR model evaluation

One of the main aims of this study is to evaluate different kernels for the GWPR models with traffic accidents data and suggest the most suitable one. The results shown in Table (5.4) are obtained from the evaluation of the five GWPR models with five different kernel weighting functions. As mentioned earlier, the most suitable model is the one which has the lowest values of GWD, AIC and AICc. Although bi-square and Gaussian kernel functions have been used for the GWPR models as found in the literature, in particular, for modelling and analysing road traffic accident data. However, the other kernels method are also found to be equally potential from the analytical point of view. Box-car kernel weighting function is the simplest form of the kernel functions and GWPR model with box-car function ($GWPR_1$) gives the highest values for all three indicators: GWD equals to 107.76, AIC 121.76 and AICc 159.09. Gaussian kernel function seems to be the most popular kernel method for spatial analysis and also for road traffic accident data. Interestingly, the GWPR model with Gaussian kernel ($GWPR_5$)is found to be the second least preferred method for our data as the model gives second largest values of GWD equals to 98.95, AIC 113.22 and AICc 153.74. The GWPR model with bi-square kernel function gives substantially better result than the model with Gaussian kernel function as shown in Table (5.4). Comparison between the GWPR models with bi-square and tri-cube kernel function is difficult as the GWD and AIC values are lower for the model with the tri-cube kernel while the AICc value is higher ($GWPR_2$ vs $GWPR_3$).

Table 5.4: Evaluation of different GWPR models for different kernel weighting functions

Model	Kernel weighting function	Deviance* or GWD	AIC	AICc
GLM (Poisson)	—	107.76*	202.29	239.62
GLM (Generalised Poisson)	—	107.76*	108.14	145.48
GLM (Negative binomial)	—	11.60*	137.18	209.18
$GWPR_1$	Box-car	107.76	121.76	159.09
$GWPR_2$	Bi-square	84.14	99.00	147.75
$GWPR_3$	Tri-cube	80.56	95.79	150.61
$GWPR_4$	Exponential	75.37	90.46	143.09
$GWPR_5$	Gaussian	98.95	113.22	153.74

Deviance* is used for GLMs (Poisson, generalised Poisson and negative binomial models)

Clearly, GWPR model with exponential kernel weighting function ($GWPR_4$) has the lowest values of GWD (75.36), AIC (90.46) and AICc (143.09) in comparison to those obtained from the other models. It can be inferred that the exponential kernel function fits the study covariates through local areas better than other kernel functions. Thus it provides a perfect calibrations and the model outperforms. Here all three indicators give the lowest value for the model and can be considered as the most suitable one. Note that we used adaptive kernel method for bandwidth selection as it has been suggested as the optimal bandwidth for spatial analysis.

5.3.4 GWPR model estimates

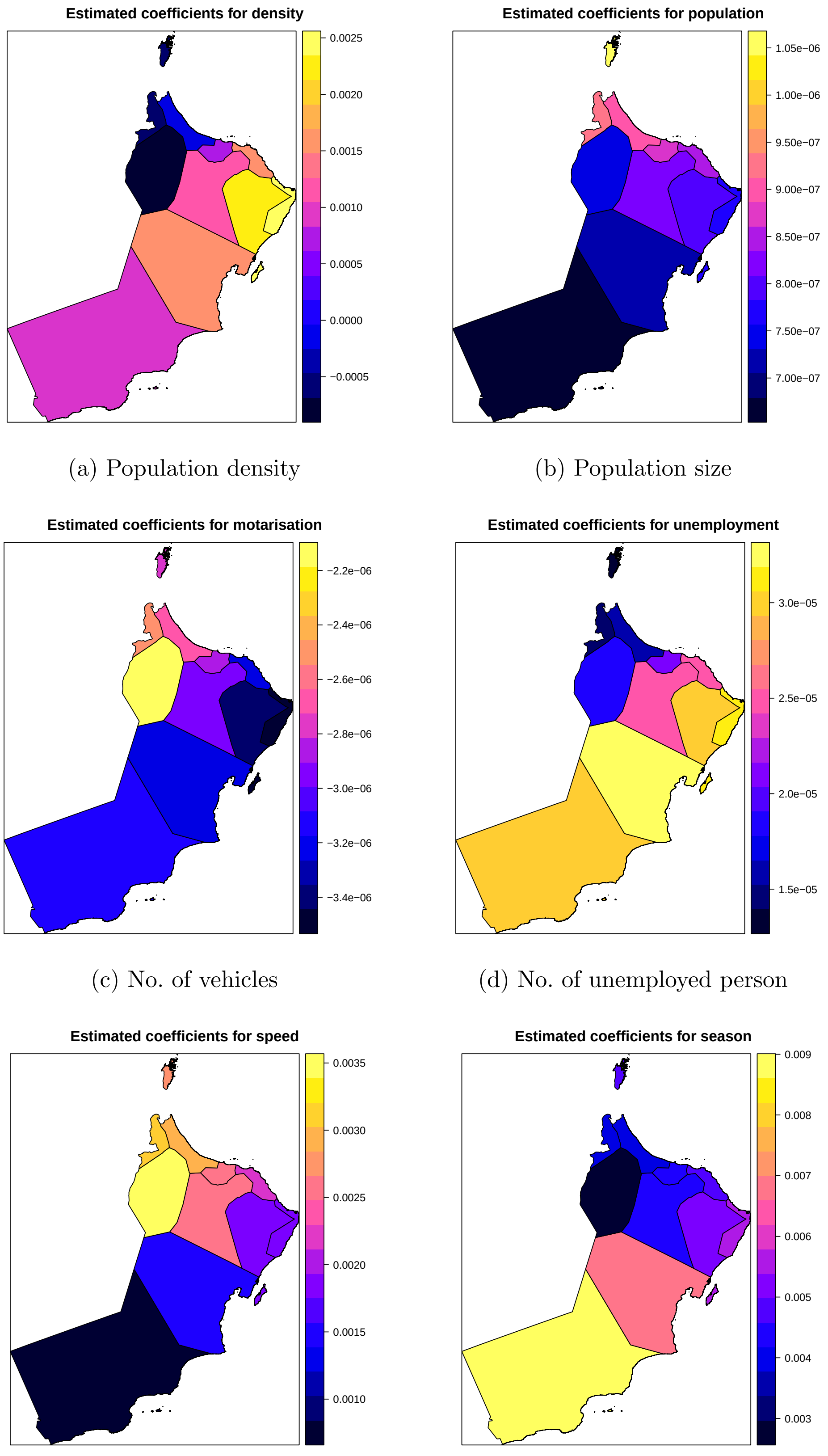
In our analysis, GWPR model (with exponential kernel weighting function and adaptive bandwidth) ($GWPR_4$) is found to be the most suitable one for road traffic accident data in Oman. Therefore, we produce the local parameter estimates from the fitted model and also the maps for the parameter estimates for the variables: population density, population size, number of vehicles, number of unemployed person, speed driving and season.

Table (5.5) provides the summary statistics of the local parameter estimates obtained from the $GWPR_4$ model. The model estimates for the local coefficients are plotted by regions (governorates) in maps. The maps demonstrate the spread of local coefficients values over the geographical regions in the country. Figure (5.2) presents the extent of local parameter values over eleven governorates (the spatial units) in the study area. The estimated coefficients show that population size, density, number of unemployed person, speed driving and season have positive impact on the number of accidents in the governorates in Oman. Interestingly, the number of vehicles found to have negative impact on accidents in Oman.

Table 5.5: Summary of local parameter estimates by GWPR model with exponential kernel weighting function

Variable	Min.	1st Quartile	Median	3rd Quartile	Max.
Intercept	$4.37e{+}00$	$4.41e{+}00$	$4.44e{+}00$	$4.45e{+}00$	4.450
Population density	$-6.90e{-}04$	$-2.97e{-}04$	$9.08e{-}04$	$1.51e{-}03$	0.002
Population size	$6.78e{-}07$	$7.62e{-}07$	$8.11e{-}07$	$8.87e{-}07$	0.000
No. of vehicles	$-3.45e{-}06$	$-3.20e{-}06$	$-2.94e{-}06$	$-2.74e{-}06$	0.000
No. of unemployed persons	$1.39e{-}05$	$1.75e{-}05$	$2.43e{-}05$	$2.95e{-}05$	0.000
Speed driving	$8.38e{-}04$	$1.82e{-}03$	$2.49e{-}03$	$2.81e{-}03$	0.003
Season	$2.96e{-}03$	$4.11e{-}03$	$4.65e{-}03$	$5.46e{-}03$	0.009

The distribution of the $GWPR_4$ model parameter estimates over the governorates in Oman is shown in Figure (5.2). The parameter estimates of population density ranges



(a) Population density (b) Population size

(c) No. of vehicles (d) No. of unemployed person

(e) Speed driving (f) Season

Figure 5.2: The estimates of local coefficients from the GWPR model with exponential kernel.

from 0.0005 to 0.0025 and the number of accidents in east governorates (both North and South Sharqiyah) are affected the most compared to other regions while the governorates in the north-west are less affected as displayed in Figure (5.2a). This is because the density of rular areas in North Sharqiyah and South Sharqiyah governorates are more than other governorates. Many inhabitants like to stay near their farms, camels and livestock. People who live in these governorates move to visit each other by car on any occasion and sometimes without occasions. Also, people from urban areas use vehicles to visit their relatives in the rular or bedouin areas and vice versa. Moreover, Bedouin people in these governorates like to work in logistic transports since they carry fish and livestock across the country and other Arabian Gulf regions. The estimates of the parameter for population size range between $7.00e-7$ and $1.05e-06$ and Musandam (in the most north of the country) is the highest affected governorate by population size [see Figure (5.2b)]. This is unexpected because this governorate has the lowest number of the population, but the nature of the terrain is challenging. Most of Musandam governorate land consists of high mountains which are very challenging for human living and constructed proper roads while most of the people live in limited areas or islands. Figure (5.2d and 5.2f), shows that Dhofar is the governorate in the south, where the number of RTA is affected by the unemployment and season. Moreover, the region is less affected by population size, number of vehicles and speed driving than other governorates in Oman as showing in Figure (5.2b, 5.2c and 5.2e). The variable season found to have the highest impact to the region is Dhofar because Dhofar celebrates the 'Salalah Tourism Festival' in June-August period and the number of accidents is significantly higher during this period.

Our model showed the spatial variation of the estimates of the parameters for the data and considered the effect of neighbouring regions. The map gives a clear indication to the policy-makers, relevant government departments and researchers to identify the associated factors related to the number of accidents in Oman and take suitable actions and measures to prevent accidents in Oman.

5.4 Chapter summary

The chapter compared several spatial models: spatial lag model, spatial error model with RTA data and developed a framework for geographically weighted Poisson regression models (GWPR). A geographically weighted Poisson regression models (GWPR) is a class of spatial count regression model that captures the localisation effect and influencing factors. One major challenge with the GWPR models is to set appropriate kernel function to give weights for each neighbouring point during the model calibration. In this study, a GWPR model formulation has been developed for many different kernel functions, including box-car, bi-square, tri-cube, exponential and Gaussian function. Derivation of the likelihood function, parameter estimation and model selection criteria have been discussed. We applied the model formulation to the road traffic accident data in Oman. Akaike information criterion (AIC), corrected Akaike information criterion (AICc) and geographically weighted deviance (GWD) have been used to assess the model fitting. The model with the exponential kernel weighted function provides the best fitting for the data and captures the spatial heterogeneity and factors better with the exponential kernel weighted function. This suggests GWPR captures the spatial heterogeneity factors better with the exponential kernel weighted function.

Chapter 6

Spatio-temporal modelling of road traffic accident data

6.1 Introduction

Although previous chapters have included either temporal or spatial effects for the data separately, it is crucial to measure the space-time unobserved effects jointly. Often some unobserved factors are likely to be correlated over time and space and if we do not take into account the temporal and spatial correlation of the data, undoubtedly a consequence will result in inefficient or inconsistent coefficient estimates (Mannering and Bhat, 2014). Moreover, spatio-temporal models accurately take account of both spatial dependence and uncorrelated heterogeneity (Quddus, 2008a). Therefore, spatio-temporal models are more suitable to examine the relationships and impactive factors over the spatial units.

The spatio-temporal models mainly formulated in Bayesian hierarchical framework, particularly with count data. Bayesian hierarchical framework is typically based on GLM (Tonui et al., 2018). However, the prime idea of the Bayesian hierarchical framework is to describe the lowest level of the model units and then build up into a hierarchy of higher-level units successively (Azimian, 2018). Moreover, Bayesian hierarchical models can handle data with low counts and are more flexible to take into account the spatial correlation (Boulieri et al., 2017). Furthermore, Bayesian hierarchical methods produce stable estimation by taking the information from the neighbouring units even with low counts (Ghosh et al., 1998; Maiti, 1998). Therefore, this method could reduce the high variability of parameter estimate and bias compared to the classical techniques that are generally used in count data models (Ghosh et al., 1994; Tonui et al., 2018). However, space-time models within the Bayesian framework requires new and more flexible statistical methods and faster computational ways to fit the model (Tonui et al., 2018). In addition, the model can be estimated and fitted using two techniques: empirical Bayes (EB) (Clayton and Kaldor, 1987; Marshall, 1991; Zou et al., 2018; Li et al., 2019) and, fully Bayes (FB) approach (Barua et al., 2015; Liu and Sharma, 2017; Abd Naeeim et al., 2020; Cheng et al., 2020). Persaud et al. (2010) concluded that FB technique with use prior distributions into account is better on the flexibility than EB technique for uncertainty in the sample data. Besides, FB approach provides whole pictures about model parameters because they can estimate posterior by marginal distribution as an alternative option of estimate by a single point only (Tonui et al., 2018). Nevertheless, Ghosh et al. (1994) pointed to the complex calculation, interact computation and time-consuming in using FB approach for estimation or fitting hierarchical models. Nowadays, with the vast amplification of sophisticated software and packages as advanced statistical tools, they can be used for fitting or estimating posterior parameters in complicated models such as fully Bayes (Blangiardo and Cameletti, 2015; Dong et al., 2016; Abd Naeeim et al., 2020). Because of these issues, several studies applied FB technique such as Markov chain Monte Carlo (MCMC) algorithm computation (Fu, 2016; Dong et al., 2016; Cheng et al., 2020) and integrated nested Laplace approximation (INLA) (Altieri et al., 2016; Liu and Sharma, 2017; Saha et al., 2018). However, the main challenge for implementation of MCMC when the model is very complex included time and space effects, need to evaluate the convergence of posterior sampling, which could be very time-consuming (Schrödle et al., 2011; Liu and Sharma, 2017; Tonui et al., 2018). A different perspective on this is provided by INLA (Rue et al., 2009), that numerically approximate and fit latent Gaussian models are faster than MCMC and more accurate deriving the posterior densities (Schrödle et al., 2011; Altieri et al., 2016; Liu and Sharma, 2017).

In the context of spatio-temporal Bayesian hierarchical model structure, the model consists of three main components which are spatial, temporal and the interaction between space and time. Both spatial components and temporal components could take into account the structured and unstructured heterogeneity effects. Essentially, the intrinsic conditional autoregressive (ICAR) is suitable to model for the spatial structure to the count data modelling (Boulieri et al., 2017; Liu and Sharma, 2017; Cheng et al., 2020). However, Besag–York–Molli (BYM) model proposed by Besag (1974); Besag et al. (1991) covers both spatially structured by implementing ICAR, and unstructured random effects by assuming Gaussian exchangeable accounting for heterogeneity as well as for the spatial correlation based on neighbourhoods (Boulieri et al., 2017; Liu and Sharma, 2018; Abd Naeeim et al., 2020). The BYM model has recently been applied in RTA space-time modelling in several studies (Boulieri et al., 2017; Liu and Sharma, 2018; Li et al., 2019). On the downside, there is inconvenient large-scale characteristic in the ICAR, which could result in negative pairwise correlation through regions located further apart (MacNab, 2011; Botella-Rocamora et al., 2013; Tonui et al., 2018). Moreover, variance components in the BYM convolution model could not be determined (MacNab, 2014; Rampaso et al., 2016). Unlike this, Saha et al. (2018) found that ICAR model performed better than the LCAR model when the hierarchical Bayesian model was implemented on Florida RTA data. However, this study did not take into account temporal effects or space-time interaction effects with Bayesian hierarchical formulation. There is no such study in this field that used LCAR within the spatio-temporal Bayesian hierarchical model to the best of our knowledge. On the other hand, temporal structure part can be modelled as first-order random walk (RW1) (Boulieri et al., 2017; Ma et al., 2017; Liu and Sharma, 2018; Li et al., 2019), or the first-order autoregressive (AR1) model (Liu and Sharma, 2017; Cheng et al., 2020), and linear temporal model (Liu and Sharma, 2017; Cheng et al., 2018, 2020). The unstructured temporal part would also account by assuming following the normal distribution as did with unstructured spatial parts (Liu and Sharma, 2017; Ma et al., 2017).

For the spatio-temporal interaction part, some studies considered spatio-temporal interactions but they are assumed to follow a normal distribution (Jiang et al., 2014; Liu and Sharma, 2017; Ma et al., 2017; Liu and Sharma, 2018). However, some of those studies did not show any data-driven evidence to corroborate their assumptions (Jiang et al., 2014). As mentioned above, the space-time interaction part in the spatio-temporal Bayesian hierarchical models instead of assume normality solely. Although Boulieri et al. (2017) suggested considering space-time interaction to model RTA data, limited number of studies have been found in the literature. Also, Liu and Sharma (2018) argued that the spatial correlations should have been considered which would evolve dynamically over time. Furthermore, however, space-time interaction dynamically could be extended into considering four interaction types of prior distributions within Bayesian hierarchical (Knorr-Held, 2000).

In order to bridge these gaps in previous studies, this study aimed to develop a hierarchical Bayesian spatio-temporal model that has the best performance with RTAs data. The rest of this chapter is divided into four subsections in the modeling Section (6.2) which consist of details of the model framework, INLA, Random effects analysis and models checking criteria. The results and discussion in Section (6.3) and summary of this chapter in Section (6.4).

6.2 Modelling

In this section we propose the model formulation and derivation of spatio-temporal models and apply the most suitable models with RTAs data in Oman.

6.2.1 Statistical framework

The statistical framework of Bayesian spatio-temporal hierarchical architecture includes three parts; the spatial, temporal and spatio-temporal effect components. The Bayesian spatio-temporal hierarchical model can be written as follow:

$$\log(\lambda_{it}) = \alpha + \beta \mathbf{X}_{it} + \mu_i + \nu_i + \gamma_t + \phi_t + \eta_{it}, \tag{6.1}$$

where i is a spatial unit (governorates in Oman), t is a temporal unit (index of year), λ_{it} is the average of accidents frequency in governorate i and year t, α is an intercept term, β is a vector of regression coefficients, $\mathbf{X}_{it}$ is a vector of covariate of governorate i in year t, μ_i is a structured spatial random effect of governorate i, ν_i is an unstructured spatial random effect of governorate i, γ_t is a structured temporal random effect in year t, ϕ_t is an unstructured temporal random effects and η_{it} is the spatio-temporal interaction effect.

However, it is often difficult to implement full Bayesian hierarchical model equation (6.1) for different reasons that leads to neglecting some parts in the full Bayesian spatio-temporal hierarchical model (6.1) and we assume to zero.

Spatial components

In the spatial component $(\mu_i + \nu_i)$ of the Bayesian spatio-temporal model, considering $\gamma_t = \phi_t = \eta_{it} = 0$. In that case, can be modelled either parametic spatial Bayesian equation (6.2), or non-parametic spatial Bayesian Equation (6.3).

$$\log(\lambda_{it}) = \alpha + \beta \mathbf{X}_{it} + \mu_i + \nu_i \tag{6.2}$$

$$\log(\lambda_{it}) = \alpha + \mu_i + \nu_i \tag{6.3}$$

There are numerous studies which described taking the spatial dependence into account for Bayesian hierarchical formulation (Besag, 1974; Besag et al., 1991; Stern and Cressie, 2000; MacNab and Dean, 2000; Leroux et al., 2000). Yet the BYM model is the most implemented form in the spatial components (structured and unstructured) in RTA discipline (Boulieri et al., 2017; Liu and Sharma, 2017; Li et al., 2019). BYM model has been formulated μ_i the structured spatial effect using an intrinsic conditional autoregressive (ICAR) structure as follows

$$\mu_i|\mu_{j\neq i} \sim N\left(\frac{\sum_{j\in N_(i)} \mu_j}{\#N(i)}, \frac{\tau_\mu^{-1}}{\#N(i)}\right). \tag{6.4}$$

While ν_i the unstructured spatial effect is given by the following normal distribution

$$\nu_i \overset{iid}{\sim} N(0, {\tau_\nu}^{-1}) \tag{6.5}$$

where τ_μ and τ_ν are precisions and $\#N(i)$ are number of neighbours for the ith governorate.

However, as mentioned above in the Section (6.1), ICAR model may lead to negative pairwise correlation through regions located further apart (MacNab, 2011; Botella-Rocamora et al., 2013; Tonui et al., 2018). Moreover, the variance components is unkown in the BYM convolution model (MacNab, 2014; Rampaso et al., 2016). As an alternative, we considered the Leroux conditional autoregressive (LCAR) model with spatial components that takes into accounts the limitation arose in Leroux et al. (2000). These prior distributions take into account both parts: the structured and the unstructured spatial variability (Leroux et al., 2000). Thus, the distribution of LCAR can be expressed as

$$\mu_i|\mu_{j\neq i} \sim N\left(\frac{\rho\sum_{j\in N_(i)}\mu_j}{\#N(i)}, \frac{\tau_\mu^{-1}}{\#N(i)}\right). \tag{6.6}$$

As mentioned earlier, Mannering (2018) indicated researchers have most often implicitly assumed that the effects of the statistically identified determinants are constant over time (temporally stable). He suggested that temporal instability is likely to exist for a number of fundamental behavioural reasons and this temporal instability is supported by the findings of several recent accident data analyses. Ma et al. (2017) investigated the application of multivariate space and time models. The joint analysis of RTA frequency by the severity level of injuries with temporal scale found that the model enables to capture underlying unobserved heterogeneity.

Temporal components

Like spatial components, temporal components ($\gamma_t + \phi_t$) of the Bayesian spatio-temporal model consider $\mu_i = \nu_i = \eta_{it} = 0$. Then, equation (6.1) reduces to

$$\log(\lambda_{it}) = \alpha + \beta\mathbf{X}_{it} + \gamma_t + \phi_t \tag{6.7}$$

and

$$\log(\lambda_{it}) = \alpha + \gamma_t + \phi_t. \tag{6.8}$$

In the temporal component, four models can be considered: linear temporal model, the 1st order autoregressive (AR1) model, 1st random walk (RW1) model and 2nd order random walk (RW2). Equations (6.9) and (6.10) show the linear temporal components.

$$\gamma_t = (\beta_2 + \delta_i) * t \tag{6.9}$$

and

$$\delta_i \overset{iid}{\sim} N(0, {\tau_\delta}^{-1}) \tag{6.10}$$

where τ_δ is the precision, δ_i is the interaction between time and spatial unit i and β_2 is a global time trend.

The AR1 model can be given by

$$\gamma_t \sim \begin{cases} N(0, (\tau_\gamma(1-\rho^2))^{-1}) & \text{for } t = 1, \\ \rho\gamma_{t-1} + \epsilon_t & \text{for } t = 2, 3, ...10 \end{cases} \tag{6.11}$$

with

$$|\rho| < 1, \tag{6.12}$$

and

$$\varepsilon_t \overset{iid}{\sim} N(0, {\tau_\varepsilon}^{-1}), \tag{6.13}$$

where τ is the precision, ρ is the correlation coefficient, and ε_t is the white noise.

The RW1 model can be written as

$$\gamma_{t+1} = \gamma_t + \epsilon_t \tag{6.14}$$

and

$$\epsilon \overset{iid}{\sim} N(0, \tau_\epsilon{-1}) \tag{6.15}$$

where τ_ϵ is the precision and ϵ_t is the white noise.

Similarly, the RW2 model can be defined as

$$\gamma_{t+2} = \gamma_{t+1} + \gamma_t + \epsilon_t \tag{6.16}$$

where

$$\epsilon \overset{iid}{\sim} N(0, \tau_{\epsilon}{-}1). \tag{6.17}$$

Spatio-Temproal interaction components

In the spatio temporal interaction component, η_{it} is assumed to follow the zero-mean normal distribution

$$\eta_{it} \overset{iid}{\sim} N(0, {\tau_{\eta}}^{-1}), \tag{6.18}$$

where τ_{η} is a precision. Clearly, if all $\eta_{it} = 0$ the model in equation (6.1) will be changed to the main effect model that can be written as follows

$$\log(\lambda_{it}) = \alpha + \beta \mathbf{X}_{it} + \mu_i + \nu_i + \gamma_t + \phi_t. \tag{6.19}$$

However, to produce consistent and better fit by spatio-temporal interaction models compared with classical count models underlined by (Wang et al., 2013b; Aguero-Valverde and Jovanis, 2006). It can be inferred that η_{it} can capture the variations that can not be explained with the main effect model (Li et al., 2019; Abd Naeeim et al., 2020). Knorr-Held Knorr-Held (2000) proposed four types to identify the possible combinations of space-time interactions based on Kronecker product of the structure matrices (Clayton, 1996). Since there are two types of temporal effects that could interact with two types of spatial effects four possibilities of spatio-temporal interaction are obtained. Table (6.1) gives identification and comparison of four types of possibilities for the inseparable spatio-temporal interaction effects.

As shown in the Table (6.1), type$I$ interactions have been expected to interact between unstructured spatial ν_i and unstructured temporal ϕ_t while there is no structure factors in this type (Blangiardo and Cameletti, 2015; Li et al., 2019). It has meant that any model, in this case, are considering unobserved covariates they capture any effect in

Table 6.1: Identification of four types of spatio-temporal interaction effects

	Type$I$	TypeII	TypeIII	TypeIV
Prior dependence	Unstructured spatial, unstructured temporal	Unstructured spatial, structured temporal	Structured spatial, unstructured temporal	Structured spatial, structured temporal
Interacting parameters	ν_i and ϕ_t	ν_i and γ_t	μ_i and ϕ_t	μ_i and γ_t
Description	There are no spatial or temporal patterns in interactions	Interactions vary in time with trends differing by neighbourhood	Interactions have a trend in space. neighbouring governorates have similar differences from the overall trend in time	Full spatio-temporal interactions. neighbouring governorates have similar to the average trend but different temporal trends
Prior distribution used	Normal distribution	Random walk (RW1 or RW2)	ICAR or LCAR	The Kronecker product of random walk in time (RW1 or RW2) and CAR in space (ICAR or LCAR)
Structure matrix $\mathbf{R}$	$\mathbf{I}$	$\mathbf{R}_{\gamma_t}$	$\mathbf{R}_{\mu_i}$	$\mathbf{R}_{\mu_i} \otimes \mathbf{R}_{\gamma_t}$

the structured space-time factors. Consequently, Kronecker product will be as follows (Clayton, 1996)

$$\mathbf{R}_{\eta_{it}} = \mathbf{R}_{\nu_\mathbf{i}} \otimes \mathbf{R}_{\phi_t} = \mathbf{I} \otimes \mathbf{I} = \mathbf{I}. \tag{6.20}$$

In typeII the interactions are expected between unstructured spatial ν_i with structured temporal γ_t. Thus, models with typeII interactions can not take into account structured spatial, but it considers for temporal trends vary through governorates or any spatial unit. Therefore, according to Clayton's rule, the product in this type is given by

$$\mathbf{R}_{\eta_{\mathbf{it}}} = \mathbf{R}_{\nu_\mathbf{i}} \otimes \mathbf{R}_{\gamma_\mathbf{t}} = \mathbf{I} \otimes \mathbf{R}_{\gamma_\mathbf{t}} = \mathbf{R}_{\gamma_\mathbf{t}}, \tag{6.21}$$

then each η_i, $i = 1, 2, ..., n$ follows a random walk (RW1 or RW2) independent of all other governorates. In typeIII interaction structured spatial effects μ_i and unstructured temporal effects ϕ_t are expected to interact. Here, models take into account spatial trends through different years or (time points) but do not consider structured temporal part. Thus, according to Clayton's rule, the Kronecker product in this type is

$$\mathbf{R}_{\eta_{\mathbf{it}}} = \mathbf{R}_{\mu_{\mathbf{i}}} \otimes \mathbf{R}_{\phi_{\mathbf{t}}} = \mathbf{R}_{\mu_{\mathbf{i}}} \otimes \mathbf{I} = \mathbf{R}_{\mu_{\mathbf{i}}}. \tag{6.22}$$

Table 6.2: Identification of models

Full Bayesian model	Spatial effects	Temporal effects	Spatio-temporal effects	Interaction type
$S_{ICAR}T_{RW1}ST_0$	$\mu_i \sim ICAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW1$, $\phi_t \sim N(\cdot)$	- -	Main
$S_{ICAR}T_{RW1}ST_I$	$\mu_i \sim ICAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW1$, $\phi_t \sim N(\cdot)$	$\nu_i \sim N(\cdot)$ $\phi_t \sim N(\cdot)$	type$I$
$S_{ICAR}T_{RW1}ST_{II}$	$\mu_i \sim ICAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW1$, $\phi_t \sim N(\cdot)$	$\nu_i \sim N(\cdot)$, $\gamma_t \sim RW1$	typeII
$S_{ICAR}T_{RW1}ST_{III}$	$\mu_i \sim ICAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW1$, $\phi_t \sim N(\cdot)$	$\mu_i \sim ICAR$, $\phi_t \sim N(\cdot)$	typeIII
$S_{ICAR}T_{RW1}ST_{IV}$	$\mu_i \sim ICAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW1$, $\phi_t \sim N(\cdot)$	$\mu_i \sim ICAR$, $\gamma_t \sim RW1$	typeIV
$S_{LCAR}T_{RW1}ST_0$	$\mu_i \sim LCAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW1$, $\phi_t \sim N(\cdot)$	- -	Main
$S_{LCAR}T_{RW1}ST_I$	$\mu_i \sim LCAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW1$, $\phi_t \sim N(\cdot)$	$\nu_i \sim N(\cdot)$, $\phi_t \sim N(\cdot)$	type$I$
$S_{LCAR}T_{RW1}ST_{II}$	$\mu_i \sim LCAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW1$, $\phi_t \sim N(\cdot)$	$\nu_i \sim N(\cdot)$, $\gamma_t \sim RW1$	typeII
$S_{LCAR}T_{RW1}ST_{III}$	$\mu_i \sim LCAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW1$, $\phi_t \sim N(\cdot)$	$\mu_i \sim LCAR$, $\phi_t \sim N(\cdot)$	typeIII
$S_{LCAR}T_{RW1}ST_{IV}$	$\mu_i \sim LCAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW1$, $\phi_t \sim N(\cdot)$	$\mu_i \sim LCAR$, $\gamma_t \sim RW1$	typeIV
$S_{ICAR}T_{RW2}ST_0$	$\mu_i \sim ICAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW2$, $\phi_t \sim N(\cdot)$	- -	Main
$S_{ICAR}T_{RW2}ST_I$	$\mu_i \sim ICAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW2$, $\phi_t \sim N(\cdot)$	$\nu_i \sim N(\cdot)$, $\phi_t \sim N(\cdot)$	type$I$
$S_{ICAR}T_{RW2}ST_{II}$	$\mu_i \sim ICAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW2$, $\phi_t \sim N(\cdot)$	$\nu_i \sim N(\cdot)$, $\gamma_t \sim RW2$	typeII
$S_{ICAR}T_{RW2}ST_{III}$	$\mu_i \sim ICAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW2$, $\phi_t \sim N(\cdot)$	$\mu_i \sim ICAR$, $\phi_t \sim N(\cdot)$	typeIII
$S_{ICAR}T_{RW2}ST_{IV}$	$\mu_i \sim ICAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW2$, $\phi_t \sim N(\cdot)$	$\mu_i \sim ICAR$, $\gamma_t \sim RW2$	typeIV
$S_{LCAR}T_{RW2}ST_0$	$\mu_i \sim LCAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW2$, $\phi_t \sim N(\cdot)$	- -	Main
$S_{LCAR}T_{RW2}ST_I$	$\mu_i \sim LCAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW2$, $\phi_t \sim N(\cdot)$	$\nu_i \sim N(\cdot)$, $\phi_t \sim N(\cdot)$	type$I$
$S_{LCAR}T_{RW2}ST_{II}$	$\mu_i \sim LCAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW2$, $\phi_t \sim N(\cdot)$	$\nu_i \sim N(\cdot)$, $\gamma_t \sim RW2$	typeII
$S_{LCAR}T_{RW2}ST_{III}$	$\mu_i \sim LCAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW2$, $\phi_t \sim N(\cdot)$	$\mu_i \sim LCAR$, $\phi_t \sim N(\cdot)$	typeIII
$S_{LCAR}T_{RW2}ST_{IV}$	$\mu_i \sim LCAR$, $\nu_i \sim N(\cdot)$	$\gamma_t \sim RW2$, $\phi_t \sim N(\cdot)$	$\mu_i \sim LCAR$, $\gamma_t \sim RW2$	typeIV

Finally, typeIV interaction includes two dependent and structured spatial μ_i and temporal γ_t to be interacted (Blangiardo and Cameletti, 2015; Li et al., 2019). Models under typeIV take into account both structured spatial effects and structured temporal effects. Clearly, this type of interaction is the most complex type of all the spatio-temporal interactions (Tonui et al., 2018). In typeIV interaction, models are suitable if the temporal trends vary from governorates to governorates; however, they are more likely to be similar for adjacent governorates (Tonui et al., 2018). The structure matrix $R_{\eta_{it}}$ rely on characteristics of both conditional autoregression (ICAR or LCAR) and random walk (RW1 or RW2). Therefore, based on Clayton's rule, the Kronecker product matrices in this type is defined as

$$\mathbf{R}_{\eta_{\mathbf{it}}} = \mathbf{R}_{\mu_{\mathbf{i}}} \otimes \mathbf{R}_{\gamma_{\mathbf{t}}}. \tag{6.23}$$

The analysis of the data in this study is carried out within a Bayesian hierarchical framework. However, our model formulation consists of spatial components, temporal components and, spatio-temporal components. All combinations of spatial, temporal and spatio-temporal proposed models identification in this study are listed in Table (6.2). Table (6.2) gives a summary for all models being used and the interaction type for each model.

6.2.2 Integrated Nested Laplace Approximation (INLA)

As mentioned in the literature section, it would be better to use integrated nested Laplace approximation (INLA) to solve spatio-temporal Bayesian models. INLA is a numerical technique to estimate the full Bayesian proposed by Rue et al. (2009), that has been found to be faster and more accurate than MCMC technique (Martins et al., 2013).

Let us assume that, y be the vector of response variables, β the vector of the target parameter, and ω the vector of the hyper-parameter. The posterior probability densities of parameter elements in the hierarchical Bayesian models (Blangiardo et al., 2013;

Blangiardo and Cameletti, 2015) can be given by

$$p(\beta_i|y) = \int p(\beta_i, \omega|y)d\omega \tag{6.24}$$
$$= \int p(\beta_i|\omega, y)p(\omega|y)d\omega,$$

where β_i is the ith parameter. The posterior probability for each element of the hyper-parameter vector is

$$p(\omega_k|y) \int p(\omega|y)d\omega_{-k}, \tag{6.25}$$

where ω_k is the kth hyper parameter, and ω_{-k} is the complement hyper parameter set to ω_k. Hence, it should take two steps for performing, first compute $p(\omega|y)$ of the approximation to the joint posterior of the hyperparameters as (Blangiardo et al., 2013; Blangiardo and Cameletti, 2015)

$$p(\omega|y) = \frac{p(\beta, \omega|y)}{p(\beta|\omega, y)} \tag{6.26}$$
$$= \frac{p(y|\beta, \omega)p(\beta, \omega)}{p(y)} \cdot \frac{1}{p(\beta|\omega, y)}$$
$$\propto \frac{p(y|\beta, \omega)p(\beta|\omega\omega)p(\omega)}{p(\beta|\omega, y)}$$
$$\approx \frac{p(y|\beta, \omega)p(\beta|\omega)p(\omega)}{\widetilde{p}(\beta|\omega, y)}|_{\beta=\overset{*}{\beta}(\omega)} =: \widetilde{p}(\omega|y), \tag{6.27}$$

where $\widetilde{p}(\omega|y)$ is the Gaussian approximation – given by the Laplace method – of $p(\beta|\omega, y)$ and $\overset{*}{\beta}(\omega)$ is the mode for a given ω.

The second step is compute $p(\beta_i|\omega, y)$, which is required for computing the parameter marginal posteriors $p(\beta_i|y)$. However, this step will be more complex, more cost of computation due to more elements in β than ω (Rue et al., 2009). Thus, we need to rewrite the parameters vector β as $\beta = (\beta_i, |\beta_{-i})$ and then we use a second time Laplace approximation to obtain

$$
\begin{aligned}
p(\beta|\omega, y) &= \frac{p(\beta_i, \beta_{-i}|\omega, y)}{p(\beta_{-i}|\beta_i, \omega, y)} &(6.28)\\
&= \frac{p(\beta, \omega|y)}{p(\omega|y)p(\beta_{-i}|\beta_i, \omega, y)}\\
&\propto \frac{p(\beta, \omega|y)}{p(\beta_{-i}|\beta_i, \omega, y)}\\
&\approx \frac{p(\beta, \omega|y)}{\widetilde{p}(\beta_{-i}|\beta_i, \omega, y)}\Big|_{\beta_{-i}=\overset{*}{\beta}_{-i}(\beta_i,\omega)} =: \widetilde{p}(\beta_i|\omega, y), &(6.29)
\end{aligned}
$$

where $\widetilde{p}(\beta_{-i}|\beta_i, \omega, y)$ is the Laplace Gaussian approximation to $p(\beta_{-i}|\beta_i, \omega, y)$ and $\overset{*}{\beta}_{-i}$ (β_i, ω) is its mode. Because the random variables $\beta_{-i}|\beta_i, \omega, y$ are in general reasonably normal, the approximation provided by (6.29) typically works very well. Owing to locating the mode by grid search, INLA can obtain the marginal posterior $p(\beta_i|y)$ (6.27) by an approximate as follow

$$
\widetilde{p}(\beta_i|y) \approx \int \widetilde{p}(\beta_i|\omega, y)\widetilde{p}(\omega|y)d\omega. \tag{6.30}
$$

For each ω^* in the set of points Θ with corresponding weight Υ_{ω^*}, a conditional posteriors $\widetilde{p}(\beta_i|\omega, y)$ are obtained by grid search also. Numerical integration can be used to obtain the marginal posterior $\widetilde{p}(\beta_i|y)$

$$
\widetilde{p}(\beta_i|y) \approx \sum_{\omega^*\in\Theta} \widetilde{p}(\beta_i|\omega^*, y)\widetilde{p}(\omega^*|y)\Upsilon_{\omega^*}. \tag{6.31}
$$

6.2.3 Random effects analysis

For spatio-temporal models, researchers are interested to estimate the relative contribution of the structured and unstructured effects over the marginal totals (Boulieri et al., 2017; Liu and Sharma, 2017; Li et al., 2019). Since the spatial effects are divided into a structured block and instructured block, it would be possible to determine which one has more effects by using the fraction of the smoothing precisions for both blocks. If the structured effect exceeds, the spatial dependency is greater and vice versa. The spatial fraction of smoothing precisions is given by

$$f_i = \frac{m(\tau_\mu)}{m(\tau_\mu) + m(\tau_\nu)}, \tag{6.32}$$

where f_i is the criteria for measuring the spatial dependency, $m(\tau_\mu)$ is the mean values of the smoothing precisions of the structured spatial effects and τ_ν is the mean values of the smoothing precisions of the unstructured spatial effects. If the spatial fraction f_i is close to one, then the strucuctured spatial impact is smoother than unstructured spatial impact. Thus, in this case the spatial dependency is unobvious (Li et al., 2019). Otherwise, in case of f_i is close to zero this means that the structured impact dominates, and the spatial dependency is more significant.

Similary, the temporal fraction of smoothing precisions is given by

$$f_t = \frac{m(\tau_\gamma)}{m(\tau_\gamma) + m(\tau_\Phi)}, \tag{6.33}$$

where f_t is the criteria measuring temporal dependency, $m(\tau_\gamma)$ is the mean values of the smoothing precisions of the structured temporal effects and τ_Φ is the mean values of the smoothing precisions of the unstructured temporal effects. If f_t is close to one, then the tempral dependency is negligible. Otherwise, the temporal dependency is larger.

6.2.4 Models checking and comparison

The deviance information criterion (DIC) is more appropriate for Bayesian hierarchical models than Akaike Information Criteria AIC (equation 5.36) and BIC (equation 4.4). The DIC is a generalisation of AIC, which trades-off model fit against a measure of model complexity (Spiegelhalter et al., 2014; Ma et al., 2017; Saha et al., 2018). Spiegelhalter et al. (2002) proposed the DIC based on the principle of gather the goodness of fit and the complexity to obatain DIC value. Therefore, the DIC can be given as follows

$$DIC = D(\bar{\beta}) + 2P_D = \bar{D} + P_D, \tag{6.34}$$

where $D(\bar{\beta})$ is the deviance using the posterior mean values of the parameters of interst β, $\bar{\beta}$ is the posterior mean of deviance $D(\beta)$ and P_D is the number of the effective parameters. Similar to AIC and BIC, the best performace model is the one which has the smallest DIC value. However, there are three possible differences of the DIC value which could result: greater than 10 rules out the model, differences between 5 and 10 are substantial, and differences less than 5 mean that the models are not significantly different (Liu and Sharma, 2017; Ma et al., 2017). However, it is possible that the complex CAR model random-effects DIC value is under-penalize (Plummer, 2008; Keller, 2013). Therefore, more evaluation criteria should be used such as the conditional predictive ordinate (CPO)(Pettit, 1990) and the cross-validated probability integral transform (PIT) (Dawid, 1984; Liu and Sharma, 2017)

$$CPO_i = \Pi(x_i|x_{-i}) \tag{6.35}$$

and

$$PIT = p(x_i \leqslant x_i|x_{-i}) \tag{6.36}$$

where x_i is the ith observation and X_{-i} acts all observation except the ith one.

The negative mean logarithmic $C\bar{P}O$ utilising as a magnitude of the predictive quality of the model (Stone, 1977; Schrödle et al., 2011; Liu and Sharma, 2017). It has been proved that the negative mean logarithmic $C\bar{P}O$ [equation (6.37)] is asymptotically equivalent to AIC (Stone, 1977). Thus, $C\bar{P}O$ will be used as the goodness of fit in this study also and the lower value indicates the best model. The $C\bar{P}O$ can be given as follows

$$C\bar{P}O = -\frac{1}{n}\sum_{n=1}^{n} log(CPO_i). \tag{6.37}$$

6.3 Results and discussion

The proposed full Bayesian spatio-temporal models listed in the Table (6.2) are implemented in the R platform through INLA (Rue et al., 2009; Blangiardo and Cameletti, 2015). The value of effective parameters (p_D), deviance information criterion (DIC) and negative mean logarithmic conditional predictive ordinate $C\bar{P}O$, which indicate the model performance are used. The goodness of fit measures for all the models listed in Table (6.2) are presented in Table (6.3).

Table 6.3: Goodness of fit for the proposed models

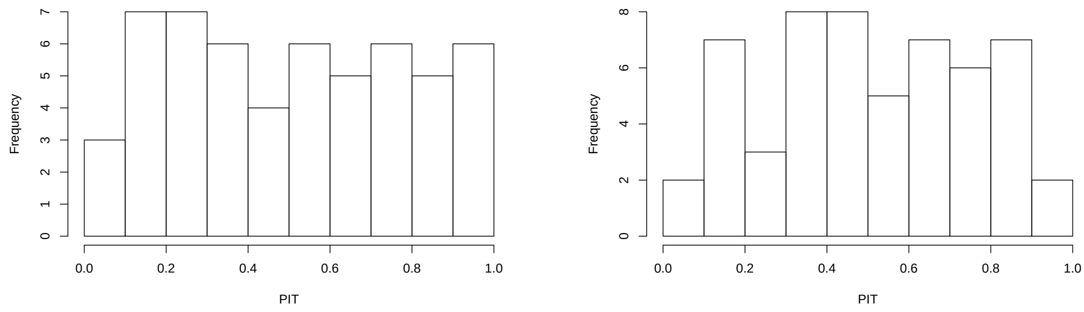
Model	pD	DIC	$C\bar{P}O$	Interaction type
$S_{ICAR}T_{RW1}ST_0$	20.94	694.58	7.53	main
$S_{ICAR}T_{RW1}ST_I$	51.59	526.80	5.57	type$I$
$S_{ICAR}T_{RW1}ST_{II}$	46.48	519.91	5.24	typeII
$S_{ICAR}T_{RW1}ST_{III}$	51.34	526.78	5.58	typeIII
$S_{ICAR}T_{RW1}ST_{IV}$	46.97	522.65	5.41	typeIV
$S_{LCAR}T_{RW1}ST_0$	20.92	694.63	7.53	main
$S_{LCAR}T_{RW1}ST_I$	51.05	525.91	5.52	type$I$
$S_{LCAR}T_{RW1}ST_{II}$	48.14	520.86	5.25	typeII
$S_{LCAR}T_{RW1}ST_{III}$	50.20	526.63	5.56	typeIII
$S_{LCAR}T_{RW1}ST_{IV}$	49.25	523.72	5.45	typeIV
$S_{ICAR}T_{RW2}ST_0$	21.03	695.19	7.57	main
$S_{ICAR}T_{RW2}ST_I$	50.77	525.88	5.53	type$I$
$S_{ICAR}T_{RW2}ST_{II}$	43.63	525.88	5.53	typeII
$S_{ICAR}T_{RW2}ST_{III}$	50.60	525.41	5.54	typeIII
$S_{ICAR}T_{RW2}ST_{IV}$	46.24	532.19	5.70	typeIV
$S_{LCAR}T_{RW2}ST_0$	20.99	695.23	7.56	main
$S_{LCAR}T_{RW2}ST_I$	50.46	527.07	5.61	type$I$
$S_{LCAR}T_{RW2}ST_{II}$	45.35	522.67	5.37	typeII
$S_{LCAR}T_{RW2}ST_{III}$	50.29	526.69	5.58	typeIII
$S_{LCAR}T_{RW2}ST_{IV}$	46.47	533.33	5.76	typeIV

As shown in the Table (6.3), it is clear that spatio-temporal interactions models perform better than spatio-temporal main models ($S_{ICAR}T_{RW1}ST_0$, $S_{LCAR}T_{RW1}ST_0$, $S_{ICAR}T_{RW2}ST_0$ and, $S_{LCAR}T_{RW2}ST_0$). In spatio-temporal main models we considered spatial factors and temporal factors separately with-out taking into account the spatio-temporal interactions (Castro et al., 2012; Barua et al., 2015). Although spatio-temporal main models are parsimonious with small values of the effective number of parameters (p_D = 20.94, 20.92, 21.03 and, 20.99) for the ($S_{ICAR}T_{RW1}ST_0$, $S_{LCAR}T_{RW1}ST_0$, $S_{ICAR}T_{RW2}ST_0$ and, $S_{LCAR}T_{RW2}ST_0$) models respectively but the values of DIC, $C\bar{P}O$ are very high and the difference is very large. Overall, among spatio-temporal main models, both (ICAR or LCAR) in spatial part and (RW1 or RW2) in temporal part did not affect differently in these measures. Consequently, unseparated spatio-temporal Bayesian models are playing essential roles to capture unobserved heterogeneity factors. This implies that one should model and analyse RTA data for Oman using unseparated spatio-temporal Bayesian models for suitable forecasting.

It is interesting to note that in all four groups listed in the Table (6.3), typeII spatio-temporal interaction models outperform in each group in terms of DIC and $C\bar{P}O$ measures. DIC values for each spatio-temporal interaction typeII models are $S_{ICAR}T_{RW1}ST_{II} = 519.91$, $S_{LCAR}T_{RW1}ST_{II} = 520.86$, $S_{ICAR}T_{RW2}ST_{II} = 525.88$ and $S_{LCAR}T_{RW2}ST_{II} = 522.67$, which represented the lowest values in each group. Equally, those models have obtained the best values of $C\bar{P}O$ in their groups, $S_{ICAR}T_{RW1}ST_{II} = 5.24$, $S_{LCAR}T_{RW1}ST_{II} = 5.25$, $S_{ICAR}T_{RW2}ST_{II} = 5.53$ and $S_{LCAR}T_{RW2}ST_{II} = 5.37$. In terms of the trade-off between model fitting and complexity, p_D values are better for typeIII than other types (type$I$, typeII and, typeIV) of spatio-temporal interaction models in the groups. It can be inferred that spatio-temporal interaction typeII models which include interaction between unstructured spatial effects and structured temporal (either $RW1$ or $RW2$) effects are the best models to capture endogeneity factors in RTA data. Likewise, the implication mentioned earlier in this study, is confirming the findings observed in (Li et al., 2019) for RTA data. This study produced results which corroborate the findings of a great deal of the previous work in this field. However, interested researchers and concerned authorities in Oman and the Arabian Gulf should consider the interaction of the RTA temporal trends through the spatial unit.

Another important finding was that we no longer need to consider $RW2$ in the full model for the study data. This is because values of DIC and $C\bar{P}O$ in $S_{ICAR}T_{RW1}ST_{II}$ ($DIC = 519.91$ and $C\bar{P}O = 5.24$) are lower than $S_{ICAR}T_{RW2}ST_{II}$ ($DIC = 525.88$ and $C\bar{P}O = 5.53$) which show significant difference and rule out the model $S_{ICAR}T_{RW1}ST_{II}$. Also, when Leurex (LCAR) technique uses with structured spatial in models $S_{LCAR}T_{RW1}ST_{II}$ ($DIC = 520.86$ and $C\bar{P}O = 5.25$) and $S_{LCAR}T_{RW2}ST_{II}$ ($DIC = 522.67$ and $C\bar{P}O = 5.37$). However, the data set consists of five years temporal scale that still sought longer periods. It is suggested that the second random walk ($RW2$) in the full Bayesian spatio-temporal models can be fitted with dataset for more than ten years. In Oman, it is crucial to produce RTA, RTI and RTD data in more structured spatio-temporal dataset format for each governorate and city for every year.

On the contrary to expectations, this study did not find a significant difference of DIC and $C\bar{P}O$ values between $S_{ICAR}T_{RW1}ST_{II}$ model ($DIC = 519.91$ and $C\bar{P}O = 5.24$) and $S_{LCAR}T_{RW1}T_{II}$ model ($DIC = 520.86$ and $C\bar{P}O = 5.25$). Even though the study by Saha et al. (2018) found a substantial difference between ICAR and LCAR models. However, the models were fitted for only spatial factors with no consideration of temporal or spatio-temporal unobserved heterogeneity effects. Therefore, the Leroux car (LCAR) model performance was much better as a part of the full Bayesian spatio-temporal interaction model than separate spatial models that were the most interesting finding in this study. It suggests that one should consider the Leroux car (LCAR) model in the full Bayesian spatio-temporal framework by implementing with multivariate severity levels. However, since the DIC and $C\bar{P}O$ values are not significantly different, the trade-off between fitting model and complexity have been required to choose the final model. Also, it can be seen from the Table (6.3) that $S_{ICAR}T_{RW1}T_{II}$ ($p_D = 46.48$) has less number of effective parameters than $S_{LCAR}T_{RW1}T_{II}$ ($p_D = 48.14$) but the difference



(a) PIT values of the $S_{ICAR}T_{RW1}ST_{II}$ model

(b) PIT values of the $S_{LCAR}T_{RW1}ST_{II}$ model

Figure 6.1: Histogram of the PIT values

is solely 1.66. Overall, the model $S_{ICAR}T_{RW1}T_{II}$ is slightly better as pointed out in the table but still the difference is insignificant. Thus, the model $S_{ICAR}T_{RW1}T_{II}$ and model $S_{LCAR}T_{RW1}T_{II}$ are preferred and they are selected for further analysis in our study.

Figure (6.1) displays PIT histogram of the $S_{ICAR}T_{RW1}ST_{II}$ model (left) and PIT histogram of the $S_{LCAR}T_{RW1}ST_{II}$ model (right). The empirical distribution of the PIT is used to assess the predictive performance of the model (Gneiting et al., 2007; Blangiardo and Cameletti, 2015). Both models have obtained a fair uniform distribution of the PIT values that also indicated a good model performance. It means that both models have the same predictive distribution for the data. However, as shown in Figure (6.1) PIT histogram of the $S_{ICAR}T_{RW1}ST_{II}$ model tends to have more uniformity. On the otherhand, both models fit the data reasonably well because the area of p-values are in the middle as the histogram of the posterior predictive p-value in Figure (6.2). However, the study found that the posterior predictive p-values of the $S_{LCAR}T_{RW1}ST_{II}$ model is slightly better than the $S_{ICAR}T_{RW1}ST_{II}$ model as shown in Figure (6.2 b). Broadly speaking, these two models ($S_{ICAR}T_{RW1}ST_{II}$ and, $S_{LCAR}T_{RW1}ST_{II}$) are found to be the best for estimation and prediction for our data.

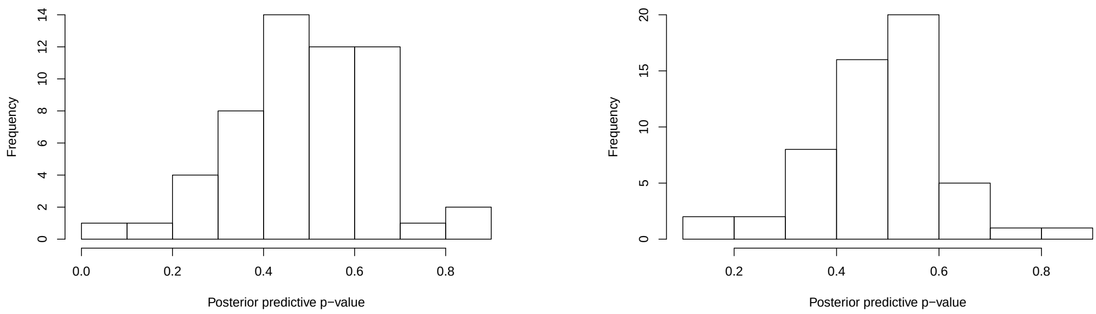
Table (6.4) summarises the parameters estimates for both models ($S_{ICAR}T_{RW1}ST_{II}$ and $S_{LCAR}T_{RW1}ST_{II}$). Table (6.4) shows the means of the estimated parameters of the

Table 6.4: Bayesian estimation of parameters for RTA

Variables	$S_{ICAR}T_{RW1}ST_{II}$ model				$S_{LCAR}T_{RW1}ST_{II}$ model			
	Mean	SD	BCI 2.5%	BCI 97.5%	Mean	SD	BCI 2.5%	BCI 97.5%
Intercept	$2e+00$	$7.75e-01$	$4.83e-01$	$3.52e+00$	$4.68e+00$	$1.77e-01$	$4.32e+00$	$5.026e+00$
Population size	$3.85e-07$	$1.14e-06$	$-1.84e-06$	$2.65e-06$	$1.24e-06$	$7.68e-07$	$-3.08e-07$	$2.75e-06$
Jobs seekers	$4.87e-06$	$4.30e-06$	$-3.62e-06$	$1.33e-05$	$-1.73e-06$	$6.11e-06$	$-1.42e-05$	$9.78e-06$
Registered vehicles	$3.33e-06$	$1.96e-06$	$-5.56e-07$	$7.16e-06$	$-8.024e-07$	$1.94e-06$	$-4.57e-06$	$3.082e-06$
Population density	$3.17e-03$	$5.59e-03$	$-7.89e-03$	$1.41e-02$	$-3.54e-03$	$3.98e-03$	$-1.16e-02$	$4.10e-03$
Speed driving	$3.02e-03$	$4.30e-04$	$2.17e-03$	$3.86e-03$	$3.26e-03$	$5.09e-04$	$2.27e-03$	$4.28e-03$
Season	$7.13e-04$	$6.10e-04$	$-4.87e-04$	$1.92e-03$	$-1.09e-04$	$6.73e-04$	$-1.44e-03$	$1.21e-03$

Table 6.5: Bayesian estimation with random effect for RTA

Variables	$S_{ICAR}T_{RW1}ST_{II}$ model				$S_{LCAR}T_{RW1}ST_{II}$ model			
	Mean	SD	BCI		Mean	SD	BCI	
			2.5%	97.5%			2.5%	97.5%
Intercept	2.4	$7.75e-01$	0.92	3.89	4.9	0.08	4.74	5.06
Speed driving	$3.01e-03$	$2.86e-03$	$2.58e-03$	$3.70e-03$	$3.09e-03$	$2.8e-04$	$2.54e-03$	$3.64e-03$
τ_μ	18776.82	18410.43	1268.43	67338.55	3.15	1.88	8.49	7.96
τ_υ	18711.95	18440.29	1282.90	67436.23	18404	18238	1259	66710
τ_γ	18585.42	18375.42	1248.70	67208.04	2062.76	12261.72	20.26	13780
τ_ϕ	18985.28	18557.76	1377.6	67828.95	19374.96	18646.69	1430.81	68471.55
τ_η	57.49	14.55	33.70	90.55	32.62	8.62	18.62	52.21
f_i	0.50		–	–	0.0001	–	–	–
f_t	0.49	–	–	–	0.96	–	–	–



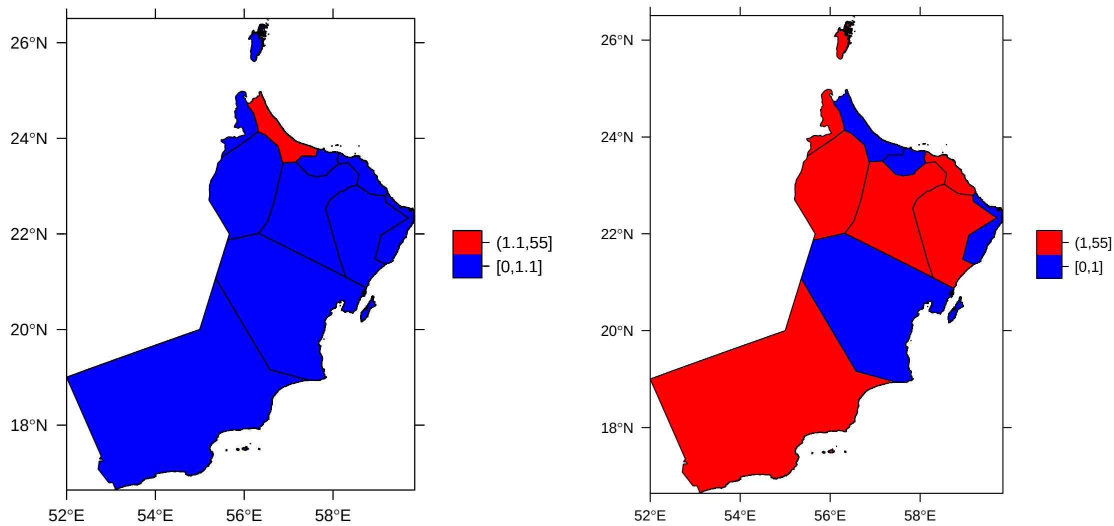
(a) Predicted p-values for the $S_{ICAR}T_{RW1}ST_{II}$ model

(b) Predicted p-values for the $S_{LCAR}T_{RW1}ST_{II}$ model

Figure 6.2: Predicted P.values for the both models.

posterior distribution, their standard deviation and, 2.5 and 97.5 the percentile values of the Bayesian credible interval (BCI). Bayesian inference relay of the posterior density interval, which is broadly known as Bayesian credible interval (BCI) (Rue et al., 2009; Saha et al., 2018). Therefore, 95% BCI is utilised to locate the credibility of the study variables. The 95% BCI includes values between 2.5% and 97.5% of the posterior probability distribution such as presenting in the Table (6.4). However, the parameters will be credible (significant) if the 95% BCI does not contain zero in their intervals. Moreover, the credible parameter has positive (increasing) effects of the number of RTA if two extremes of coefficient in the 95% BCI is greater than zero. Similarly, the credible parameter has negative (decreasing) effects of the number of RTA if two extremes of coefficient in the 95% BCI is less than zero.

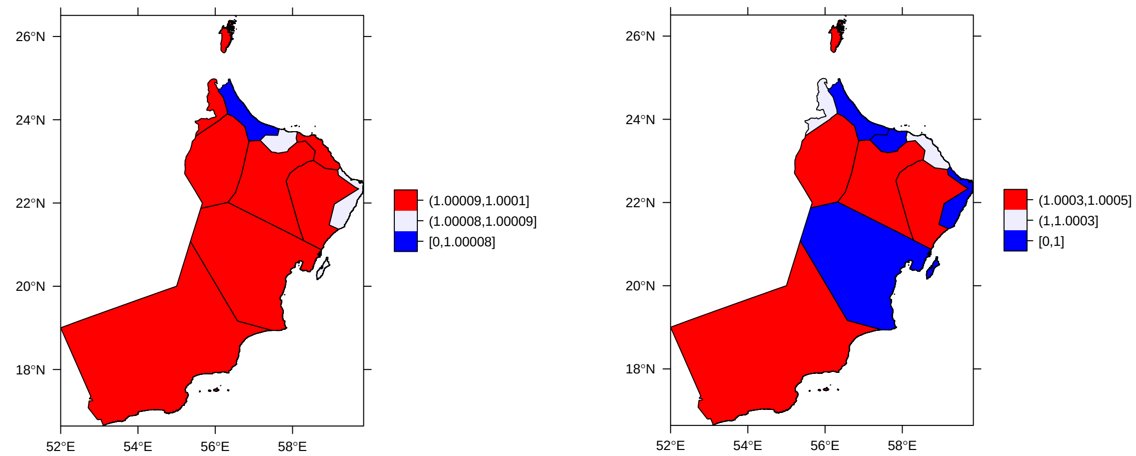
The intercept and speed driving parameters had credible positive impacts as have been obtained from both models as shown in Table (6.4). These imply that speed driving factor plays a significant role which superfacts the frequency of road traffic accidents across Oman governorates. This finding is in line with numerous previous studies (Al-Reesi and Al-Maniri, 2014; Al Reesi et al., 2016; Al-Aamri et al., 2017; Al Aamri, 2018). However, both models obtained opposite signs for other variables except the population size was positive for both models. However, both models agreed for non-credibility of all other variables. These findings are consistent with a previous study (Liu and Sharma, 2017, 2018). This is because those factors may affect in the short term with the whole area as proved in Section (5.3) and study in (Al-Hasani et al., 2019a), but their influence is less with a long term on individual small areas. However, to estimate random effects we should rebuild the models by considering only credible variables (Liu and Sharma, 2017; Li et al., 2019).



(a) $S_{ICAR}T_{RW1}ST_{II}$ model (b) $S_{LCAR}T_{RW1}ST_{II}$ model

Figure 6.3: Exponential posterior means of the structured spatial effect $[exp(\mu_i)]$.

As displayed in Table (6.5), the mean of smoothing precisions for the structured spatial effects τ_{μ} is positive and significant from both models. The mean values of structured spatial precisions are 18776.82 and 3.15, which are obtained from $S_{ICAR}T_{RW1}ST_{II}$ model and $S_{LCAR}T_{RW1}ST_{II}$ model respectively. Similarly, the mean of smoothing precisions for the unstructured spatial effects is positive and significant (18711.95 and 18404 from $S_{ICAR}T_{RW1}ST_{II}$ model and $S_{LCAR}T_{RW1}ST_{II}$ model respectively). However, for the model $S_{LCAR}T_{RW1}ST_{II}$, the smoothing precision's spatial fraction is 0.50. As a consequence, outcomes of $S_{ICAR}T_{RW1}ST_{II}$ model pointed out that the structured and unstructured spatial impacts are playing the same role in RTA frequency in Oman. Unlike the outcome of $S_{LCAR}T_{RW1}ST_{II}$ model, the result indicated that the structured spatial h

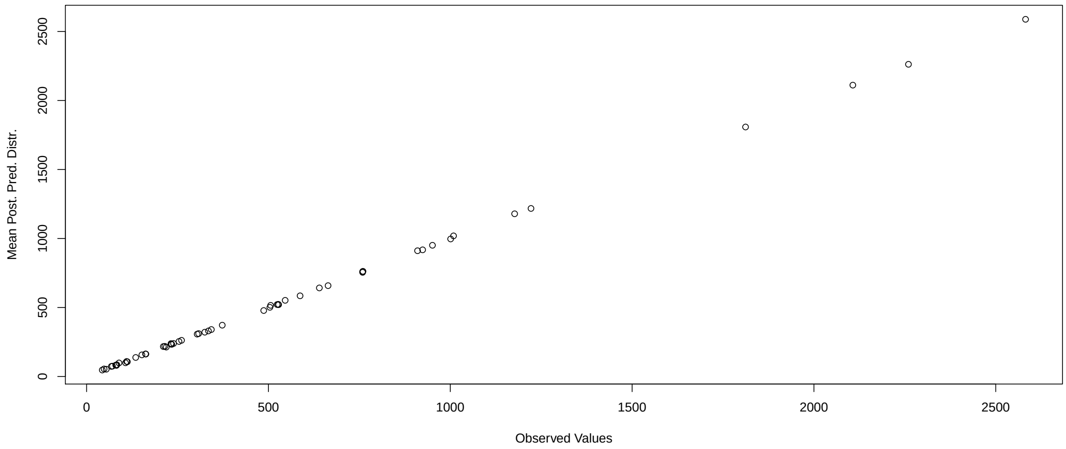


(a) $S_{ICAR}T_{RW1}ST_{II}$ model (b) $S_{LCAR}T_{RW1}ST_{II}$ model

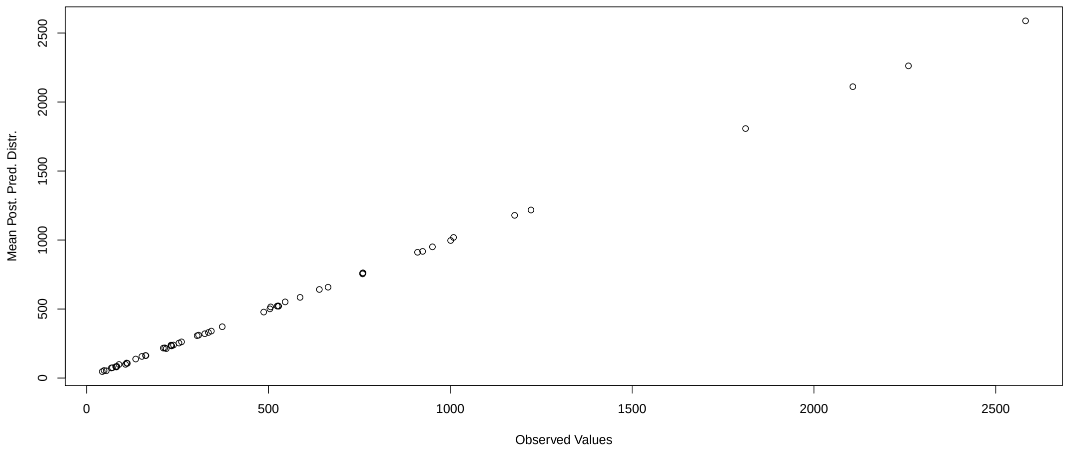
Figure 6.4: Exponential posterior means of the unstructured spatial effect $[exp(\nu_i)]$.

effect is stronger because the $f_i = 0.0001$. However, the exponential posterior means of the structured spatial effect is demonstrated in Figure (6.3) while the exponential posterior means of the unstructured spatial effect in Figure (6.4). Clearly, for $S_{LCAR}T_{RW1}ST_{II}$ model, North Al-Batinha governorate only tends to have higher RTAs [Figure (6.3, a) with exponential posterior means of the structured spatial effect while other governorates have lower. In the meantime, the same governorate (North Al-Batinah) tends to have lower RTAs [(Figure (6.4), a] with exponential posterior means of the unstructured spatial effect than other governorates.

For the temporal side, temporal fraction of smoothing precisions f_t are 0.49 and, 0.96 for $S_{ICAR}T_{RW1}ST_{II}$ model and $S_{LCAR}T_{RW1}ST_{II}$ model respectively. It means that $S_{ICAR}T_{RW1}ST_{II}$ model outcomes have been indicated that the same amount of temporal effects for structured γ and unstructured ϕ components on RTA in the Sultanate of Oman. However, $S_{ICAR}T_{RW1}ST_{II}$ points out that unstructured temporal effects are exceeded on RTA frequency at Oman. Both models underline that temporal structured and temporal unstructured effects have significantly increased effects. Both models agreed that there are spatio-temporal interaction effects on RTA frequency in Oman. As shown in Table



(a) Forecasted mean values vs observed values for the $S_{ICAR}T_{RW1}ST_{II}$ model



(b) Forecasted mean values vs observed values for the $S_{LCAR}T_{RW1}ST_{II}$ model

Figure 6.5: Relation between forecasted and observed values for the both models.

(6.5), RTA in Oman have been influenced by interactions varying in time with trends differing by neighbourhood governorates. Therefore, various levels of countermeasures policies through decision-makers, investigators, researchers, etc. should be considered for the country.

Figure (6.5) shows the relation between the mean of forecasting values and observed values obtained from both models. Clearly, both models could be used for the high-efficiency forecasting, as shown in Figure (6.5). Thus, both models have obtained strong positive relation for forecasting RTA in the Sultanate of Oman. However, more recent data are required to generate such forecasting.

6.4 Chapter summary

This chapter showed a detailed analysis of the spatio-temporal unobserved effects jointly of the road traffic accident (RTA) data in Oman. Spatio-temporal models accurately took account of both spatial dependence and uncorrelated heterogeneity. However, it is crucial to consider the interaction between spatial effects and temporal effects. The spatio-temporal models have been formulated into Bayesian hierarchical methods that produce stable estimation due to capturing the information from the neighbouring units, which consider an excellent facility to benefits even with low counts. The challenge was to build the spatio-temporal model's three components by selecting the perfect techniques to parts for space effects, time effects and the interaction between space and time effects.

This study attempted to develop a spatio-temporal model for RTA data. Integrated nested Laplace approximation (INLA) which is fully Bayes tools have been used for fitting models, estimating fixed, random or posterior parameters. It is interesting to confirm that unseparated spatio-temporal Bayesian models are playing essential roles to capture unobserved heterogeneity for different factors. Another important finding is that spatio-temporal interaction typeII models that include interacting between unstructured spatial effects and structured temporal (either RW1 or RW2) effects are the best models to capture the endogeneity of factors RTA data. However, the main finding is that the Leroux car (LCAR) model performance is much higher for the spatial part of full Bayesian spatio-temporal interaction models than separate spatial models. Although the evaluation criteria values of $S_{ICAR}T_{RW1}ST_{II}$ ($DIC = 519.91, C\bar{P}O = 5.24$ and $P_D = 46.48$) are slightly lower than $S_{LCAR}T_{RW1}ST_{II}$ ($DIC = 520.86, C\bar{P}O = 5.25$ and $P_D = 48.14$), the study has not found a significant difference. It suggests examining the second-order random walk (RW2) model in the full Bayesian spatio-temporal framework by implementing with data for more than ten years. It is worth considering the Leroux car (LCAR) model in the full Bayesian spatio-temporal framework by implementing with multivariate severity levels.

Chapter 7

Conclusion and future work

The study attempted to characterise the temporal and spatial unobserved endogeneity effects of road traffic accidents (RTAs) by implementing spatio-temporal models for Oman RTAs data. A number of temporal, spatial and spatio-temporal models have been investigated to fit RTA data and select the most suitable model. Forecasting results have also been provided with the suitable models. The results of the study have provided significant insights on the nature, extent and factors of road traffic accidents in Oman and would help the policy-makers to measure appropriate interventions to reduce the accidents in Oman.

7.1 Temporal modelling

In Chapter (4), a study was conducted to develop a suitable temporal model, perform analysis and forecast RTA for Oman. Thesis studied the time series model identification for the road traffic accident (RTA) and road traffic injuries (RTI) data from a set of models, performing suitable diagnostic checks and forecasting accidents and injuries using the RTA and RTI monthly time series data in Oman. The time series modelling and analysis showned that there was a downward trend in RTAs in the period between January 2000 to June 2019 with a mean 660 and standard deviation of 247.5 RTAs.

However, in October 2001, a peak occurred with $1,283$ RTAs which then declined to the lowest point of 156 RTAs in February 2019. Based on the Box and Jenkins approach, SARIMA$(3,1,1)(2,0,0)_{12}$ model was selected to forecast RTAs for the next 24 months in Oman. This model forecasted the high occurrence of RTA in June, July and August in the following years. The policymakers in Oman should keep under their consideration the results of this study. Both models in RTA and RTI expected the high chance for accidents and injuries during the summer season. Therefore, summer periods are more accident-prone, and the chances of RTA and RTI are higher in Oman. The study recommended considering the season as exposure factor that influences the RTA frequency in Oman and should be taken into account when models are built with multi studies in this discipline.

7.2 Spatial modelling

Considering spatial variation on RTA in the Chapter (5), the thesis also investigated some spatial models. Different spatial models were fitted and investigated: the spatial lag model (SLM), spatial error model (SEM), geographically weighted regression (GWR) model and, geographically weighted Poisson regression (GWPR) model. The spatial variations in the SLM and SEM models took the errors into account only through the spatial error structure (Pirdavani et al., 2014). However, the geographically weighted models (GWR and GWPR) allowed the parameters to vary through the spatial units in a study region to capture the local factors. The main challenge of a GWR and GWPR model was to find the most suitable kernel weighting function which gives weights for the neighbouring observations during model calibration. The study found that GWPR models can substantially capture the heterogeneity of the spatial factors over the regions or spatial units in RTA field. The study compared the suitability of a GWPR model for five different kernel weighting functions: box-car, bi-square, tri-cube, exponential and Gaussian weighted function. The model formulation was shown in details including the likelihood function, estimation of parameter through calibration of models for different kernel weighting functions. The framework was applied to the road traffic accident (RTA) data in Oman to explore different factors associated with RTA and give insights into geographical variations of such factors. The crucial finding to emerge from this study was that GWPR model with exponential kernel function and adaptive bandwidth is the most suitable for modelling, fitting and analysing road accident data in Oman. Although several studies underlined that GWPR model with bi-square kernel function with adaptive bandwidth is more suitable (Nakaya et al., 2005; Hadayeghi et al., 2010; Xu and Huang, 2015; Hezaveh et al., 2019), but our results [Chapter (5)] suggested that the GWPR model with exponential kernel weighting function outperformed over other weighting functions. Furthermore, previous studies only compared the bi-square and Gaussian kernel functions with both fixed and adaptive bandwidths with ignoring other kernel functions. I collected the road traffic accident data for Oman in 2017 along with some explanatory variables which are population size, population density, number of registered vehicles, number of unemployed persons, speed driving and season from the secondary sources. The parameter estimates are produced and shown on the maps. Several interesting insights are demonstrated so that policy-makers can take appropriate actions to reduce the number of accidents in the Sultanate of Oman. As such that, the number of accidents in east governorates (both North and South Sharqiyah) are affected the most by population density in comparison to other regions while the governorates in the north-west are less affected. Also, the study found that Dhofar is the governorate in the south, where the number of RTA is affected by the unemployment and season. Moreover, the region is less affected by population size, the number of registered vehicles and speed driving compared by other governorates in Oman. The variable season is found to have the highest impact on the region of Dhofar because Dhofar celebrates the ‘Salalah

Tourism Festival' in the June-August period and the number of accidents is significantly higher during that period.

7.3 Spatio-temporal modelling

This thesis also developed Bayesian hierarchical spatio-temporal interaction models with RTA as account data as presented in Chapter (6). I developed four groups of Bayesian hierarchical spatio-temporal models and compared the performance of the models by using the number of the effective parameters (p_D), deviance information criterion (DIC) and, the negative mean logarithmic ($C\bar{P}O$), which gave indication of better models. This study found that generally unseparated spatio-temporal interaction models are playing essential roles to capture unobserved heterogeneity factors impacts. It implies that Bayesian hierarchical spatio-temporal models perform better and take into account the interaction between spatial and temporal effects. There are four types of spatio-temporal interaction could be occur between spatial (structured or unstructured) effects and temporal (structured or unstructured) effects. One of the significant findings to emerge from this study is that spatio-temporal interaction typeII models which include interacting between unstructured spatial effects and structured temporal (either $RW1$ or $RW2$) effects are the best models to capture endogeneity factors in RTA data. This study found that the interactions of RTA are varying in time with trends differing by neighbourhoods which should be taken under consideration of modellers in the RTA research field. The present finding seems to be consistent with Li et al. (2019) which found that type II of spatio-temporal models outperformed with Idaho dataset. However, another important finding was that we no longer need to consider $RW2$ in the full model for the study data. The Leroux car (LCAR) model performed much better as a part of full Bayesian spatio-temporal interaction models than separate spatial models, which were considered the most interesting finding in this study as discussed in Chapter (6).

This investigation found that spatio-temporal unobserved heterogeneity effects exist for RTA data in the Sultanate of Oman. By implementing two Bayesian hierarchical spatio-temporal interaction models, ($S_{ICAR}T_{RW1}ST_{II}$ and $S_{LCAR}T_{RW1}ST_{II}$) this study explored those spatio-temporal heterogneity. However,$S_{ICAR}T_{RW1}ST_{II}$ consists of car techniques in spatial part while $S_{LCAR}T_{RW1}ST_{II}$ model has the Leroux car (LCAR) in the spatial part. This happened because this study did not find a significant difference of p_D, DIC and $C\bar{P}O$ values between $S_{ICAR}T_{RW1}T_{II}$ model ($p_D = 46.48$, $DIC = 519.91$ and $C\bar{P}O = 5.24$) and $S_{LCAR}T_{RW1}T_{II}$ model ($p_D = 48.14$, $DIC = 520.86$ and $C\bar{P}O = 5.25$). Both models proved that speed driving factor plays a significant role which increases the frequency of road traffic accidents across Oman governorates. The mean of smoothing precisions for the structured spatial effects and the mean of smoothing precisions for the unstructured spatial effects are positive and significant for both models. However, outcomes of $S_{ICAR}T_{RW1}ST_{II}$ model pointed that the structured and unstructured spatial impacts are playing the same role of RTA frequency in Oman, while the outcome of $S_{LCAR}T_{RW1}ST_{II}$ model indicated that the structured spatial effect is stronger. On the other hand, $S_{ICAR}T_{RW1}ST_{II}$ model outcomes indicated that same portion of temporal effects (structured and unstructured) impacted on RTA in Oman while $S_{LCAR}T_{RW1}ST_{II}$ pointed out that the unstructured temporal effects are exceeded on RTA frequency. The most crucial finding to emerge from this study is that RTA in Oman have been influenced by interactions vary in time with trends differing by neighbourhood governorates. Thus, various levels of countermeasures policies by the decision-makers, investigators and researchers should be considered.

Finally, there is no work without limitation. We believe that some important limitations need to be mentioned. First, models build-up to the explanatory variables were limited to the six available variables only. However, primary variables like weather condition, road infrustructure, gender, age, etc. were not considered. Second, there are manually recorded traffic accidents in Oman ignoring some necessary information in this era such as accidents coordinates and driver's mileage. Third, there was no linkage of related data in the country to demonstrate the RTA burden among all Oman governorates. Fourth, the unprecedented circumstances represented by Covid-19 in whole the world affected the forecasting of the RTA frequency.

7.4 Future work

This research contributes to the topic of spatio-temporal modelling of RTA data by considering space and/or time endogeneity and unobserved effects. This work is represented as an empirical study which implemented formulation models with Oman RTA data. There were four directions of modelling that have been conducted which are: temporal modelling, spatial modelling, spatio-temporal separated modelling and spatio-temporal interaction modelling. The findings of this study have several important implications for future practice. This section discusses possible directions for future research. There are two main aspects suggested by the work in this thesis which are methodological and computational aspect of RTA modelling and Oman RTA aspect.

First, according to a methodological and computational aspect of RTA modelling, there are several natural directions suggested by the work in thesis chapters. From Chapter (4), it can be attempted to assess optimal (automatic) time series models, generated by auto.arima( ) function from the forecast R package. It would be interesting to consider some machine learning models with RTA time series data as well, such as artificial neural network, support vector machine, random forest regression, etc. if sufficient data are available. In Chapter (6), it is showed that there is no need to consider $RW2$ in the full spatio-temporal model because $RW1$ as the prior distribution of a structured temporal part in the full model was better. However, this may be the case for the current study data. The natural extension needs to be done to establish whether prior temporal distribution of second-order random walk $RW2$ in the full Bayesian spatio-temporal models is outperforming with data set for longer time scale and/or with different temporal units. Besides, for the prior spatial distribution account Leroux car (LCAR) in the full Bayesian spatio-temporal models by multivariate severity levels and/or with different spatial units is essential to evaluate the full model performance. It is possible to compare proposed models with different RTA data sets or with different count data from other disciplines. Broadly, any other spatial and prior temporal distribution can be accounted in full Bayesian spatio-temporal model framework. In addition, it is recommended to compare and present MCMC and INLA outcomes to fit and solve Bayesian hierarchical spatio-temporal models in RTA research. Thus, future research may be conducted to expand the methodological and computational aspect of RTA data.

Second, according to our Oman RTA analysis, there are several future possibilities for the policy-makers and researchers. The model results can be used for forecasting RTA in the Sultanate of Oman with more, recent and advanced data for each governorates. For deeper analysis, the $GWPR$ model formulation in Chapter (5) would be implemented for specific regions in the country. Furthermore, investigation of the socio-economic impact of the RTAs in Oman would be considered. In Chapter (6), findings of this Chapter have established a baseline for Oman spatio-temporal effects of RTA while the natural progression of this work is to concentrate on specific governorates or areas or regions. A greater focus on a small domain in time and space could produce more interesting findings that would account more of increased risk and deeper analysis. Aiming at identifying clusters of the certain governorate, smaller spatial units could be used. On the other hand, policy-makers can check the accident prone hot spots and could generate their road safety countermeasures to reduce RTAs in different governorates in Oman . However, the developed models in this study can be used not only for the GCC countries but also for other countries considering similar factors and more other factors such as weather, lifestyle, environment, etc.

References

Abbas, K. A. (2004). Traffic safety assessment and development of predictive models for accidents on rural roads in egypt, *Accident Analysis & Prevention* **36**(2): 149–163.

Abd Naeeim, N. S., Abdul Rahman, N. and Muhammad Fahimi, F. A. (2020). A spatial–temporal study of dengue in peninsular malaysia for the year 2017 in two different space–time model, *Journal of Applied Statistics* **47**(4): 739–756.

Abdo, A. A. and Al-Ojaili, A. (2015). Assessment of awareness of livestock-vehicle traffic accidents in Dhofar region, Oman, *International Journal of Applied Engineering Research* **10**(18): 38955–38959.

Abdo, A. M. A. (2006). Sensitivity analysis of traffic accidents causes in sultanate of oman.

Abdo, A. M. A. and Al-Ibrahim, Y. A. (2016). Assessment of traffic safety and awareness among road users in salalah, sultanate of oman, *Research Journal of Applied Sciences, Engineering and Technology* **12**(5): 574–581.

Aguero-Valverde, J. (2013). Full bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: comparing the precision of crash frequency estimates, *Accident Analysis & Prevention* **50**: 289–297.

Aguero-Valverde, J. and Jovanis, P. (2008). Analysis of road crash frequency with spatial models, *Transportation Research Record: Journal of the Transportation Research Board* (2061): 55–63.

Aguero-Valverde, J. and Jovanis, P. P. (2006). Spatial analysis of fatal and injury crashes in pennsylvania, *Accident Analysis & Prevention* **38**(3): 618–625.

Ahmed, H. and El-Sadig, M. (2002). *Epidemiology and economic impact of road traffic accidents in the United Arab Emirates*, PhD thesis, University of Abertay Dundee.

Al Aamri, A. K. (2018). *Quantifying the patterns of road traffic crashes in the Sultanate of Oman: statistical evaluation of aggregate data from police records*, PhD thesis, University of Southampton.

Al-Aamri, A. K., Hornby, G., Zhang, L.-C., Al-Maniri, A. A. and Padmadas, S. S. (2020). Mapping road traffic crash hotspots using gis-based methods: A case study of muscat governorate in the sultanate of oman, *Spatial Statistics* p. 100458.

Al-Aamri, A., Padmadas, S., Zhang, L.-C. and Al-Maniri, A. (2017). Disentangling age–gender interactions associated with risks of fatal and non-fatal road traffic injuries in the Sultanate of Oman, *BMJ Global Health* **2**(3): e000394.

Al Bulushi, I. A. (2017). *Heavy vehicle safety in Oman: a situational analysis*, PhD thesis, Queensland University of Technology.

Al-Bulushi, I., Edwards, J., Davey, J., Armstrong, K., Al-Maniri, A. and Al-Shamsi, K. (2015a). Socio-cultural influences on vehicle defects in the omani heavy vehicle industry.

Al-Bulushi, I., Edwards, J., Davey, J., Armstrong, K., Al-Reesi, H. and Al-Shamsi, K. (2015b). Heavy vehicle crash characteristics in Oman 2009–2011, *Sultan Qaboos University Medical Journal* **15**(2): e191.

Al-Hasani, G., Asaduzzaman, M. and Soliman, A.-H. (2019a). Comparison of spatial regression models with road traffic accidents data, *ICSTA'19 International Conference*, AVESTIA, pp. 31:1–31:4.

Al-Hasani, G., Khan, A. M., Al-Reesi, H. and Al-Maniri, A. (2019b). Diagnostic time series models for road traffic accidents data, *International Journal of Applied Statistics and Econometrics, 2: 19*.

Al-Lamki, L. (2010). Life loss and disability from traffic accidents: It is imperative we all act now, *Sultan Qaboos University Medical Journal* **10**(1): 1.

Al-Madani, H. M. and Al-Janahi, A. (2018). Socio-economic models for over speeding voilaters in Qatar and Oman, *Population* **1000**: 593.

Al-Maimani, E. Y. and Yahia, H. A. (2019). An investigation of driver attitudes towards road safety in oman, *Journal of Student Research*.

Al-Maniri, A. A. N., Al-Reesi, H., Al-Zakwani, I. and Nasrullah, M. (2013). Road traffic fatalities in Oman from 1995 to 2009: evidence from police reports, *International Journal of Preventive Medicine* **4**(6): 656.

Al Reesi, H., Al Maniri, A., Adawi, S. A., Davey, J., Armstrong, K. and Edwards, J. (2016). Prevalence and characteristics of road traffic injuries among young drivers in Oman, 2009–2011, *Traffic Injury Prevention* **17**(5): 480–487.

Al Reesi, H., Al Maniri, A., Plankermann, K., Al Hinai, M., Al Adawi, S., Davey, J. and Freeman, J. (2013a). Risky driving behavior among university students and staff in the Sultanate of Oman, *Accident Analysis & Prevention* **58**: 1–9.

Al-Reesi, H. and Al-Maniri, A. (2014). Road traffic injuries among young drivers in oman, *Oman medical journal* **29**(5): 313.

Al Reesi, H., Freeman, J., Davey, J., Al Adawi, S. and Al Maniri, A. (2018). Measuring risky driving behaviours among young drivers: Development of a scale for the Oman setting, *Transportation Research Part F: Traffic Psychology and Behaviour* **55**: 78–89.

Al-Reesi, H., Ganguly, S. S., Al-Adawi, S., Laflamme, L., Hasselberg, M. and Al-Maniri, A. (2013b). Economic growth, motorization, and road traffic injuries in the Sultanate of Oman, 1985–2009, *Traffic Injury Prevention* **14**(3): 322–328.

Al-Risi, A. (2014). *Characteristics of road traffic injuries and potential risk factors in Oman*, PhD thesis, University of Otago.

Al-Sinawi, H., Al-Alawi, M., Al-Lawati, R., Al-Harrasi, A., Al-Shafaee, M. and Al-Adawi, S. (2012). Emerging burden of frail young and elderly persons in oman: For whom the bell tolls?, *Sultan Qaboos University Medical Journal* **12**(2): 169.

Albers, M. J. (2017). *Introduction to quantitative data analysis in the behavioral and social sciences*, John Wiley & Sons.

Ali, Z. and Bhaskar, S. B. (2016). Basic statistical tools in research and data analysis, *Indian journal of anaesthesia* **60**(9): 662.

Altieri, L., Cocchi, D., Greco, F., Illian, J. B. and Scott, E. (2016). Bayesian p-splines and advanced computing in r for a changepoint analysis on spatio-temporal point processes, *Journal of Statistical Computation and Simulation* **86**(13): 2531–2545.

Amoros, E., Martin, J. L. and Laumon, B. (2003). Comparison of road crashes incidence and severity between some french counties, *Accident Analysis & Prevention* **35**(4): 537–547.

Andreassen, D. C. (1985). Linking deaths with vehicles and population, *Traffic Engineering and Control* **26**(11): 547–549.

Ansari, S., Akhdar, F., Mandoorah, M. and Moutaery, K. (2000). Causes and effects of road traffic accidents in saudi arabia, *Public health* **114**(1): 37–39.

Anselin, L. (1988a). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity, *Geographical analysis* **20**(1): 1–17.

Anselin, L. (1988b). Model validation in spatial econometrics: a review and evaluation of alternative approaches, *International Regional Science Review* **11**(3): 279–316.

Avery, J. (1995). Accident prevention–injury control–injury prevention–or whatever?, *Injury prevention* **1**(1): 10.

Azimian, A. (2018). Bayesian spatio-temporal analysis of road traffic crashes.

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D. and Lindgren, F. (2018). Spatial modeling with r-inla: A review, *Wiley Interdisciplinary Reviews: Computational Statistics* **10**(6): e1443.

Barua, S., El-Basyouny, K. and Islam, M. T. (2015). Effects of spatial correlation in random parameters collision count-data models, *Analytic Methods in Accident Research* **5**: 28–42.

Bener, A., Achan, N., Sankaran-Kutty, M., Ware, J., Cheema, M. and Al-Shamsi, M. A. (1994). Casualty risk reduction from safety seat belts in a desert country, *Journal of the royal society of health* **114**(6): 297–299.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society: Series B (Methodological)* **36**(2): 192–225.

Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics, *Annals of the institute of statistical mathematics* **43**(1): 1–20.

Bhanot, N., Singh, H., Sharma, D., Jain, H. and Jain, S. (2019). Python vs. r: A text mining approach for analyzing the research trends in scopus database, *arXiv preprint arXiv:1911.08271*.

Bijur, P. E. (1995). What's in a name? comments on the use of the terms' accident'and'injury'., *Injury Prevention* **1**(1): 9.

Billo, E. J. (2007). *Excel for scientists and engineers: numerical methods*, John Wiley & Sons.

Blangiardo, M. and Cameletti, M. (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*, John Wiley & Sons.

Blangiardo, M., Cameletti, M., Baio, G. and Rue, H. (2013). Spatial and spatio-temporal models with r-inla, *Spatial and spatio-temporal epidemiology* **4**: 33–49.

Bors, D. (2018). *Data analysis for the social sciences: Integrating theory and practice*, Sage.

Botella-Rocamora, P., López-Quílez, A. and Martinez-Beneito, M. (2013). Spatial moving average risk smoothing, *Statistics in medicine* **32**(15): 2595–2612.

Boulieri, A., Liverani, S., de Hoogh, K. and Blangiardo, M. (2017). A space–time multivariate Bayesian model to analyse road traffic accidents by severity, *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **180**(1): 119–139.

Box, G. E., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M. (2015). *Time series analysis: forecasting and control*, John Wiley & Sons.

Browne, W. J., Goldstein, H. and Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models, *Statistical Modelling* **1**(2): 103–124.

Brunsdon, C., Fotheringham, A. S. and Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity, *Geographical analysis* **28**(4): 281–298.

Butt, F. M., Ashiq, M., Rehman, S. U., Minhas, K. S. and Khan, M. A. (2020). Bibliometric analysis of road traffic injuries research in the gulf cooperation council region, *F1000Research*.

Castro, M., Paleti, R. and Bhat, C. R. (2012). A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections, *Transportation research part B: methodological* **46**(1): 253–272.

Cerasa, A. and Cerioli, A. (2017). Outlier-free merging of homogeneous groups of pre-classified observations under contamination, *Journal of Statistical Computation and Simulation* **87**(15): 2997–3020.

Chekijian, S., Paul, M., Kohl, V. P., Walker, D. M., Tomassoni, A. J., Cone, D. C. and Vaca, F. E. (2014). The global burden of road injury: its relevance to the emergency physician, *Emergency medicine international*.

Cheng, W., Gill, G. S., Zhang, Y. and Cao, Z. (2018). Bayesian spatiotemporal crash frequency models with mixture components for space-time interactions, *Accident Analysis & Prevention* **112**: 84–93.

Cheng, W., Gill, G. S., Zhang, Y., Vo, T., Wen, F. and Li, Y. (2020). Exploring the modeling and site-ranking performance of bayesian spatiotemporal crash frequency models with mixture components, *Accident Analysis & Prevention* **135**: 105357.

Chitme, H. R., Al-Kashmiri, A., Al-Thehli, H. M., Al-Qanoobi, M. J., Al-Mushefri, M. M. and Venuvgopal, J. (2018). Impact of medical conditions and medications on road traffic safety, *Oman medical journal* **33**(4): 316.

Clayton, D. and Kaldor, J. (1987). Empirical bayes estimates of age-standardized relative risks for use in disease mapping, *Biometrics* pp. 671–681.

Clayton, D. G. (1996). Generalized linear mixed models, *Markov chain Monte Carlo in practice* **1**: 275–302.

Consul, P. and Famoye, F. (1992). Generalized Poisson regression model, *Communications in Statistics - Theory and Methods* **21**(1): 89–109.

Conway, R. W. and Maxwell, W. L. (1962). Network dispatching by the shortest-operation discipline, *Operations Research* **10**(1): 51–73.

Creswell, J. W. (1994). *Research design: Qualitative & quantitative approaches.*, Sage Publications, Inc.

Cryer, J. D. and Chan, K.-S. (2008). *Time series analysis with application in R*, Springer.

Dahdah, S. and Bose, D. (2013). Road traffic injuries: a public health crisis in the middle east and north africa, *Transport Notes TRN-4*.

Davies, T. M. and Lawson, A. B. (2019). An evaluation of likelihood-based bandwidth selectors for spatial and spatiotemporal kernel estimates, *Journal of Statistical Computation and Simulation* pp. 1–22.

Dawid, A. P. (1984). Present position and potential developments: Some personal views statistical theory the prequential approach, *Journal of the Royal Statistical Society: Series A (General)* **147**(2): 278–290.

Díaz-Bejarano, E., Francia, V. and Coletti, F. (2019). 7 data analysis for chemical engineers: Introduction to r, *Introduction to Software for Chemical Engineers, Second Edition*, CRC Press, pp. 221–250.

Dong, N., Huang, H., Lee, J., Gao, M. and Abdel-Aty, M. (2016). Macroscopic hotspots identification: a bayesian spatio-temporal interaction approach, *Accident Analysis & Prevention* **92**: 256–264.

El-Basyouny, K. and Sayed, T. (2009). Urban arterial accident prediction models with spatial effects, *Transportation Research Record* **2102**(1): 27–33.

El-Sadig, M., Norman, J. N., Lloyd, O. L., Romilly, P. and Bener, A. (2002). Road traffic accidents in the united arab emirates: trends of morbidity and mortality during 1977–1998, *Accident Analysis & Prevention* **34**(4): 465–476.

Farag, S. G. and Hashim, I. H. (2017). Safety performance appraisal at roundabouts: Case study of salalah city in oman, *Journal of Transportation Safety & Security* **9**(sup1): 67–82.

Farag, S. G., Hashim, I. H. and El-Hamrawy, S. A. (2014). Analysis and assessment of accident characteristics: Case study of dhofar governorate, sultanate of oman, *Int J Traffic Trans Eng* **3**: 189–98.

Fotheringham, A., Brunsdon, C. and Charlton, M. (2002). Geographically weighted regression: the analysis of spatially varying relationships.

Fotheringham, A. S., Crespo, R. and Yao, J. (2015). Geographical and temporal weighted regression (GTWR), *Geographical Analysis* **47**(4): 431–452.

Fu, S. (2016). Hierarchical bayesian lasso for a negative binomial regression, *Journal of Statistical Computation and Simulation* **86**(11): 2182–2203.

Ghosh, M., Natarajan, K., Stroud, T. and Carlin, B. P. (1998). Generalized linear models for small-area estimation, *Journal of the American Statistical Association* **93**(441): 273–282.

Ghosh, M., Rao, J. et al. (1994). Small area estimation: an appraisal, *Statistical science* **9**(1): 55–76.

Gneiting, T., Balabdaoui, F. and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**(2): 243–268.

Gollini, I., Lu, B., Charlton, M., Brunsdon, C. and Harris, P. (2013). GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models, *arXiv preprint arXiv:1306.0413*.

Guo, F., Wang, X. and Abdel-Aty, M. A. (2010). Modeling signalized intersection safety with corridor-level spatial correlations, *Accident Analysis & Prevention* **42**(1): 84–92.

Hadayeghi, A., Shalaby, A. S. and Persaud, B. N. (2010). Development of planning level transportation safety tools using geographically weighted Poisson regression, *Accident Analysis & Prevention* **42**(2): 676–688.

Halim, H., Bustam, B. and Saing, Z. (2021). The forecasting of a traffic accident in the pandemic of covid-19, *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* **12**(6): 3664–3669.

Ham, S., Kim, S., Lee, N., Kim, P., Eom, I., Lee, B., Tsai, P.-J., Lee, K. and Yoon, C. (2017). Comparison of data analysis procedures for real-time nanoparticle sampling data using classical regression and arima models, *Journal of Applied Statistics* **44**(4): 685–699.

Hamed, A.-r. (2016). *The Psychosocial Influences on Risky Driving Behaviors among Young Drivers in Oman: A Social Learning Approach*, PhD thesis, Sultan Qaboos University.

Hamid, S. and Davoud, K.-Z. (2019). Road traffic injuries measures in the eastern mediterranean region: findings from the global status report on road safety–2015, *Journal of injury and violence research* **11**(2): 149.

Henrich, H. (1959). Industrial accident prevention: A scientific approach, *Edition New York Wiley.*

Hezaveh, A. M., Arvin, R. and Cherry, C. R. (2019). A geographically weighted regression to estimate the comprehensive cost of traffic crashes at a zonal level, *Accident Analysis & Prevention* **131**: 15–24.

Hilbe, J. M. (2011). *Negative binomial regression*, Cambridge University Press.

Huang, H. and Abdel-Aty, M. (2010). Multilevel data and bayesian analysis in traffic safety, *Accident Analysis & Prevention* **42**(6): 1556–1565.

Hughes, B., Newstead, S., Anund, A., Shu, C. and Falkmer, T. (2015). A review of models relevant to road safety, *Accident Analysis & Prevention* **74**: 250–270.

Hyndman, R. J. et al. (2006). Another look at forecast-accuracy metrics for intermittent demand, *Foresight: The International Journal of Applied Forecasting* **4**(4): 43–46.

Islam, M. M. and Al Hadhrami, A. Y. (2012). Increased motorization and road traffic accidents in Oman, *Journal of Emerging Trends in Economics and Management Sciences* **3**(6): 907.

Jackisch, J., Sethi, D., Mitis, F., Szymañski, T. and Arra, I. (2015). *European facts and the global status report on road safety 2015*, World Health Organization. Regional Office for Europe.

Jamal, A., Rahman, M. T., Al-Ahmadi, H. M. and Mansoor, U. (2020). The dilemma of road safety in the eastern province of saudi arabia: consequences and prevention strategies, *International journal of environmental research and public health* **17**(1): 157.

Javid, A. M. and Al-Neama, Y. M. (2018). Identification of factors causing driver's distraction in oman, *Journal of Applied Engineering Science* **16**(2): 153–160.

Jiang, X., Abdel-Aty, M. and Alamili, S. (2014). Application of poisson random effect models for highway network screening, *Accident Analysis & Prevention* **63**: 74–82.

John Braun, W. and Rousson, V. (2000). An autocorrelation criterion for bandwidth selection in nonparametric regression, *Journal of Statistical Computation and Simulation* **68**(1): 89–101.

Joshua, S. C. and Garber, N. J. (1990). Estimating truck accident rate and involvements using linear and Poisson regression models, *Transportation planning and Technology* **15**(1): 41–58.

Jovanis, P. P. and Chang, H.-L. (1986). Modeling the relationship of accidents to miles traveled, *Transportation Research Record* **1068**: 42–51.

Kaplan, S., Guvensan, M. A., Yavuz, A. G. and Karalurt, Y. (2015). Driver behavior analysis for safe driving: A survey, *IEEE Transactions on Intelligent Transportation Systems* **16**(6): 3017–3032.

Karvonen, M., Mikheev, M., Organization, W. H. et al. (1986). *Epidemiology of occupational health*, World Health Organization. Regional Office for Europe.

Keller, J. (2013). Penalized loss functions for bayesian model comparison.

Khan, A. M. (2013). R-software: A newer tool in epidemiological data analysis, *Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine* **38**(1): 56.

Khattak, A., Kantor, P. and Council, F. (1998). Role of adverse weather in key crash types on limited-access: roadways implications for advanced weather systems, *Transportation Research Record: Journal of the Transportation Research Board* (1621): 10–19.

Kilamanu, W., Xia, J. C. and Caulfield, C. (2011). Analysis of spatial and temporal distribution of single and multiple vehicle crash in western australia: a comparison study, *Proceedings of International Congress on Modelling and Simulation*, Modelling and Simulation Society of Australia and New Zealand, pp. 683–689.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk, *Statistics in medicine* **19**(17-18): 2555–2567.

Larsen, G. (2008). Excel for scientists and engineers: Numerical methods, *Technometrics* **50**(2): 236.

Lavrenz, S. M., Vlahogianni, E. I., Gkritza, K. and Ke, Y. (2018). Time series modeling in traffic safety research, *Accident Analysis & Prevention* **117**: 368–380.

Leroux, B. G., Lei, X. and Breslow, N. (2000). Estimation of disease rates in small areas: a new mixed model for spatial dependence, *Statistical models in epidemiology, the environment, and clinical trials*, Springer, pp. 179–191.

Li, Z., Chen, X., Ci, Y., Chen, C. and Zhang, G. (2019). A hierarchical bayesian spatiotemporal random parameters approach for alcohol/drug impaired-driving crash frequency analysis, *Analytic methods in accident research* **21**: 44–61.

Li, Z., Wang, W., Liu, P., Bigham, J. M. and Ragland, D. R. (2013). Using geographically weighted Poisson regression for county-level crash modeling in California, *Safety Science* **58**: 89–97.

Liboschik, T., Fokianos, K. and Fried, R. (2015). *tscount: An R package for analysis of count time series following generalized linear models*, Universitätsbibliothek Dortmund.

Liu, C. and Sharma, A. (2017). Exploring spatio-temporal effects in traffic crash trend analysis, *Analytic methods in accident research* **16**: 104–116.

Liu, C. and Sharma, A. (2018). Using the multivariate spatio-temporal bayesian model to analyze traffic crashes by severity, *Analytic methods in accident research* **17**: 14–31.

Liu, C., Gyawali, S., Sharma, A. and Smaglik, E. (2015). A methodological approach for spatial and temporal analysis of red light running citations and crashes: a case-study in lincoln, nebraska, *Technical report*.

Lord, D. and Mannering, F. (2010). The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives, *Transportation Research Part A: Policy and Practice* **44**(5): 291–305.

Lord, D. and Miranda-Moreno, L. F. (2008). Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a Bayesian perspective, *Safety Science* **46**(5): 751–770.

Lord, D., Geedipally, S. R. and Guikema, S. D. (2010). Extension of the application of conway-maxwell-poisson models: Analyzing traffic crash data exhibiting underdispersion, *Risk Analysis* **30**(8): 1268–1276.

Lord, D., Guikema, S. D. and Geedipally, S. R. (2008). Application of the Conway–Maxwell–Poisson generalized linear model for analyzing motor vehicle crashes, *Accident Analysis & Prevention* **40**(3): 1123–1134.

Lu, B., Harris, P., Charlton, M. and Brunsdon, C. (2014). The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models, *Geo-spatial Information Science* **17**(2): 85–101.

Ma, X., Chen, S. and Chen, F. (2017). Multivariate space-time modeling of crash frequencies by injury severity levels, *Analytic Methods in Accident Research* **15**: 29–40.

Machado-León, J. L., de Oña, J., de Oña, R., Eboli, L. and Mazzulla, G. (2016). Socio-economic and driving experience factors affecting drivers? perceptions of traffic crash risk, *Transportation research part F: traffic psychology and behaviour* **37**: 41–51.

MacNab, Y. and Dean, C. (2000). Parametric bootstrap and penalized quasi-likelihood inference in conditional autoregressive models, *Statistics in medicine* **19**(17-18): 2421–2435.

MacNab, Y. C. (2004). Bayesian spatial and ecological models for small-area accident and injury analysis, *Accident Analysis & Prevention* **36**(6): 1019–1028.

MacNab, Y. C. (2011). On gaussian markov random fields and bayesian disease mapping, *Statistical Methods in Medical Research* **20**(1): 49–68.

MacNab, Y. C. (2014). On identification in bayesian disease mapping and ecological–spatial regression models, *Statistical methods in medical research* **23**(2): 134–155.

Maher, M. J. and Summersgill, I. (1996). A comprehensive methodology for the fitting of predictive accident models, *Accident Analysis & Prevention* **28**(3): 281–296.

Maiti, T. (1998). Hierarchical bayes estimation of mortality rates for disease mapping, *Journal of Statistical Planning and Inference* **69**(2): 339–348.

Manikandan, M., Prasad, V., Mishra, A. K., Konduru, R. K. and Newtonraj, A. (2018). Forecasting road traffic accident deaths in india using seasonal autoregressive integrated moving average model, *International Journal of Community Medicine and Public Health* **5**(9): 3962–3968.

Mannering, F. (2018). Temporal instability and the analysis of highway accident data, *Analytic Methods in Accident Research* **17**: 1–13.

Mannering, F. and Bhat, C. (2014). Analytic methods in accident research: methodological frontier and future directions, *Analytic Methods in Accident Research* **1**: 1–22.

Mannering, F. L., Shankar, V. and Bhat, C. R. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data, *Analytic methods in accident research* **11**: 1–16.

Marshall, R. J. (1991). Mapping disease and mortality rates using empirical bayes estimators, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **40**(2): 283–294.

Martins, T. G., Simpson, D., Lindgren, F. and Rue, H. (2013). Bayesian computing with inla: new features, *Computational Statistics & Data Analysis* **67**: 68–83.

Mastinu, G. and Plöchl, M. (2014). *Road and off-road vehicle system dynamics handbook*, CRC press.

McIlvenny, S. (2006). Road traffic accidents-a challenging epidemic, *Sultan Qaboos University Medical Journal* **6**(1): 3.

Meng, F., Xu, P., Wong, S., Huang, H. and Li, Y. (2017). Occupant-level injury severity analyses for taxis in hong kong: A bayesian space-time logistic model, *Accident Analysis & Prevention* **108**: 297–307.

Meyerhoff, N. J. (1978). The influence of daylight saving time on motor vehicle fatal traffic accidents, *Accident Analysis & Prevention* **10**(3): 207–221.

Miaou, S.-P., Bligh, R. and Lord, D. (2005). Part 1: Roadside safety design: Developing guidelines for median barrier installation: Benefit-cost analysis with texas data, *Transportation Research Record: Journal of the Transportation Research Board* (1904): 2–19.

Miaou, S.-P., Song, J. J. and Mallick, B. K. (2003). Roadway traffic crash mapping: a space-time modeling approach, *Journal of transportation and Statistics* **6**: 33–58.

Nakaya, T., Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2005). Geographically weighted Poisson regression for disease association mapping, *Statistics in Medicine* **24**(17): 2695–2717.

Narasimhan, G., Ephrem, B. G., Cheriyan, S. and Balasupramanian, N. (2017). Predictive analytics of road accidents in oman using machine learning approach, *2017 Interna-*

tional Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), IEEE, pp. 1058–1065.

NCSI (2017a). *Monthly Statistical Bulletin April 2017*, National Centre for Statistics & Information, Sultanate of Oman.

NCSI (2017b). *Monthly Statistical Bulletin July 2017*, National Centre for Statistics & Information, Sultanate of Oman.

NCSI (2017c). *Monthly Statistical Bulletin September 2017*, National Centre for Statistics & Information, Sultanate of Oman.

NCSI (2018a). *Monthly Statistical Bulletin April 2018*, National Centre for Statistics & Information, Sultanate of Oman.

NCSI (2018b). *Monthly Statistical Bulletin January 2018*, National Centre for Statistics & Information, Sultanate of Oman.

NCSI (2018c). *Monthly Statistical Bulletin June 2018*, National Centre for Statistics & Information, Sultanate of Oman.

NCSI (2018d). *Monthly Statistical Bulletin November 2018*, National Centre for Statistics & Information, Sultanate of Oman.

NCSI (2019a). *Monthly Statistical Bulletin April 2019*, National Centre for Statistics & Information, Sultanate of Oman.

NCSI (2019b). *Monthly Statistical Bulletin November 2019*, National Centre for Statistics & Information, Sultanate of Oman.

Nofal, F. H., Saeed, A. A. and Anokute, C. C. (1996). Aetiological factors contributing to road traffic accidents in riyadh city, saudi arabia, *Journal of the Royal Society of Health* **116**(5): 304–311.

Oster, R. A. (1998). An examination of five statistical software packages for epidemiology, *The American Statistician* **52**(3): 267–280.

Ozaki, T. (1977). On the order determination of arima models, *Applied Statistics* pp. 290–301.

Parvareh, M., Karimi, A., Rezaei, S., Woldemichael, A., Nili, S., Nouri, B. and Nasab, N. E. (2018). Assessment and prediction of road accident injuries trend using time-series models in Kurdistan, *Burns & Trauma* **6**(1): 9.

Patil, S. (2016). Big data analytics using r, *International Research Journal of Engineering and Technology (IRJET) Volume* **3**: 78–81.

Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A. A., Jarawan, E., Mathers, C. D. et al. (2004). World report on road traffic injury prevention.

Persaud, B., Lan, B., Lyon, C. and Bhim, R. (2010). Comparison of empirical bayes and full bayes approaches for before–after road safety evaluations, *Accident Analysis & Prevention* **42**(1): 38–43.

Pettit, L. (1990). The conditional predictive ordinate for the normal distribution, *Journal of the Royal Statistical Society: Series B (Methodological)* **52**(1): 175–184.

Pirdavani, A., Bellemans, T., Brijs, T. and Wets, G. (2014). Application of geographically weighted regression technique in spatial analysis of fatal and injury crashes, *Journal of Transportation Engineering* **140**(8): 04014032.

Plankermann, K. (2014). *Human factors as causes for road traffic accidents in the Sultanate of Oman under consideration of road construction designs*, PhD thesis.

Plummer, M. (2008). Penalized loss functions for bayesian model comparison, *Biostatistics* **9**(3): 523–539.

Quddus, M. A. (2008a). Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data, *Accident Analysis & Prevention* **40**(4): 1486–1497.

Quddus, M. A. (2008b). Time series count data models: an empirical application to traffic accidents, *Accident Analysis & Prevention* **40**(5): 1732–1741.

Quddus, M. A. (2016). Non-gaussian interrupted time series regression analysis for evaluating the effect of smart motorways on road traffic accidents, *Technical report*.

Raeside, R. and White, D. (2004). Predicting casualty numbers in Great Britain, *Transportation Research Record: Journal of the Transportation Research Board* (1897): 142–147.

Rahman, M. K., Crawford, T. and Schmidlin, T. W. (2018). Spatio-temporal analysis of road traffic accident fatality in bangladesh integrating newspaper accounts and gridded population data, *GeoJournal* **83**(4): 645–661.

Ramana, G. V., Reddy, N. S. and Praneeth, D. (2018). Analysis on traffic safety and road accidents: a case study on selected roads in muscat, oman, *Int. J. Civ. Eng. Technol* **9**(2): 417–427.

Rampaso, R. C., de Souza, A. D. P. and Flores, E. F. (2016). Bayesian analysis of spatial data using different variance and neighbourhood structures, *Journal of Statistical Computation and Simulation* **86**(3): 535–552.

Rhee, K.-A., Kim, J.-K., Lee, Y.-I. and Ulfarsson, G. F. (2016). Spatial regression analysis of traffic crashes in Seoul, *Accident Analysis & Prevention* **91**: 190–199.

Rogel-Salazar, J. (2018). *Data science and analytics with Python*, CRC Press.

ROP (2009a). *Traffic Statistic 2013*, Director General of Traffic, Royal Oman Police.

ROP (2009b). *Traffic Statistics Report, Muscat, Sultanate of Oman, Directorate General of Traffic*, Muscat: Royal Oman Police.

ROP (2014). *Traffic Statistic 2013*, Director General of Traffic, Royal Oman Police.

ROP (2015). *Facts and Figures.*, Director General of Traffic, Royal Oman Police.

ROP (2017a). *Facts and Figures*, Director General of Traffic, Royal Oman Police.

ROP (2017b). *Traffic Statistic annualy book 2016*, Director General of Traffic, Royal Oman Police.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations, *Journal of the Royal Statistical Society: Series b (statistical methodology)* **71**(2): 319–392.

Ruikar, M. et al. (2013). National statistics of road traffic accidents in india, *Journal of Orthopedics, Traumatology and Rehabilitation* **6**(1): 1.

Saha, D., Alluri, P., Gan, A. and Wu, W. (2018). Spatial analysis of macro-level bicycle crashes using the class of conditional autoregressive models, *Accident Analysis & Prevention* **118**: 166–177.

Sandt, L. S., Marshall, S. W., Rodriguez, D. A., Evenson, K. R., Ennett, S. T. and Robinson, W. R. (2016). Effect of a community-based pedestrian injury prevention program on driver yielding behavior at marked crosswalks, *Accident Analysis & Prevention* **93**: 169–178.

Savolainen, P. T., Mannering, F. L., Lord, D. and Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives, *Accident Analysis & Prevention* **43**(5): 1666–1676.

Schrödle, B., Held, L., Riebler, A. and Danuser, J. (2011). Using integrated nested laplace approximations for the evaluation of veterinary surveillance data from switzerland:

a case-study, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **60**(2): 261–279.

Sengoelge, M., Laflamme, L. and El-Khatib, Z. (2018). Ecological study of road traffic injuries in the eastern mediterranean region: country economic level, road user category and gender perspectives, *BMC public health* **18**(1): 236.

Sheikh, M. M. R. (2009). *A statistical analysis of road traffic accidents and casualties in Bangladesh*, PhD thesis, Edinburgh Napier University.

Shibata, A. and Fukuda, K. (1994). Risk factors of fatality in motor vehicle traffic accidents, *Accident Analysis & Prevention* **26**(3): 391–397.

Shively, P. (2004). Testing for a unit root in ARIMA processes, *Journal of Applied Statistics* **31**(7): 785–798.

Shoukri, M., Chaudhary, M. and Al-Halees, A. (2005). Estimating $P(Y < X)$ when X and Y are paired exponential variables, *Journal of Statistical Computation and Simulation* **75**(1): 25–38.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of the royal statistical society: Series b (statistical methodology)* **64**(4): 583–639.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van der Linde, A. (2014). The deviance information criterion: 12 years on, *Journal of the Royal Statistical Society: Series B: Statistical Methodology* pp. 485–493.

Spiess, M. (1998). A mixed approach for the estimation of probit models with correlated responses: some finite sample results, *Journal of Statistical Computation and Simulation* **61**(1-2): 39–59.

Srivastava, T. (2015). A complete tutorial on time series modeling in R, *Analytics Vid.*

Staton, C., Vissoci, J., Gong, E., Toomey, N., Wafula, R., Abdelgadir, J., Zhao, Y., Liu, C., Pei, F., Zick, B. et al. (2016). Correction: Road traffic injury prevention initiatives: a systematic review and metasummary of effectiveness in low and middle income countries, *PLoS one* **11**(2): e0150150.

Stern, H. S. and Cressie, N. (2000). Posterior predictive model checks for disease mapping models, *Statistics in medicine* **19**(17-18): 2377–2397.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and akaike's criterion, *Journal of the Royal Statistical Society: Series B (Methodological)* **39**(1): 44–47.

Szeto, W., Ghosh, B., Basu, B. and O'Mahony, M. (2009). Multivariate traffic forecasting technique using cell transmission model and SARIMA model, *Journal of Transportation Engineering* **135**(9): 658–667.

Team, R. C. (2000). R language definition, *Vienna, Austria: R foundation for statistical computing*.

Timmermans, C., Alhajyaseen, W., Al Mamun, A., Wakjira, T., Qasem, M., Almallah, M. and Younis, H. (2019). Analysis of road traffic crashes in the state of qatar, *International journal of injury control and safety promotion* **26**(3): 242–250.

Tonui, B., Mwalili, S. and Wanjoya, A. (2018). Spatio-temporal variation of hiv infection in kenya, *Open Journal of Statistics* **8**(5): 811–830.

Truong, L. T., Kieu, L.-M. and Vu, T. A. (2016). Spatiotemporal and random parameter panel data models of traffic crash fatalities in vietnam, *Accident Analysis & Prevention* **94**: 153–161.

Wang, C., Quddus, M. A. and Ison, S. G. (2013a). The effect of traffic and road characteristics on road safety: A review and future research direction, *Safety Science* **57**: 264–275.

Wang, C., Quddus, M. and Ison, S. (2013b). A spatio-temporal analysis of the impact of congestion on traffic safety on major roads in the UK, *Transportmetrica A: Transport Science* **9**(2): 124–148.

Wang, W. and Li, D. (2017). Structure identification and variable selection in geographically weighted regression models, *Journal of Statistical Computation and Simulation* **87**(10): 2050–2068.

Wang, X. and Abdel-Aty, M. (2006). Temporal and spatial analyses of rear-end crashes at signalized intersections, *Accident Analysis & Prevention* **38**(6): 1137–1150.

Wang, Y. and Kockelman, K. M. (2013). A poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods, *Accident Analysis & Prevention* **60**: 71–84.

Wen, H., Zhang, X., Zeng, Q. and Sze, N. (2019). Bayesian spatial-temporal model for the main and interaction effects of roadway and weather characteristics on freeway crash incidence, *Accident Analysis & Prevention* **132**: 105249.

WHO (1989). Manifesto for safe communities adopted at the first world conference on accident and injury prevention 1989, *Karolinska Institute: Stockholm, Sweden.*

WHO (2010). Eastern mediterranean status report on road safety: call for action, *Technical report*, World Health Organization.

WHO (2013). Global Status Report on Road Safety 2013: supporting a decade of action: summary, *Technical report*, World Health Organization.

WHO (2015). *Global Status Report on Road Safety 2015*, World Health Organization.

WHO (2018). Global status report on road safety 2018: Summary, *Technical report*, World Health Organization.

WHO (2019). *Global status report on alcohol and health 2018*, World Health Organization.

Wikle, C. K., Zammit-Mangion, A. and Cressie, N. (2019). *Spatio-temporal Statistics with R*, CRC Press.

Woods, L. (2017). *A statistical and geographical analysis of workplace accidents in England and Wales*, PhD thesis, University of Liverpool.

Xu, P. and Huang, H. (2015). Modeling crash spatial heterogeneity: random parameter versus geographically weighting, *Accident Analysis & Prevention* **75**: 16–25.

Yasmin, S. and Eluru, N. (2013). Evaluating alternate discrete outcome frameworks for modeling crash injury severity, *Accident Analysis & Prevention* **59**: 506–521.

Ye, X., Pendyala, R. M., Shankar, V. and Konduri, K. C. (2013). A simultaneous equations model of crash frequency by severity level for freeway sections, *Accident Analysis & Prevention* **57**: 140–149.

Zhang, G. P. (2003). Time series forecasting using a hybrid arima and neural network model, *Neurocomputing* **50**: 159–175.

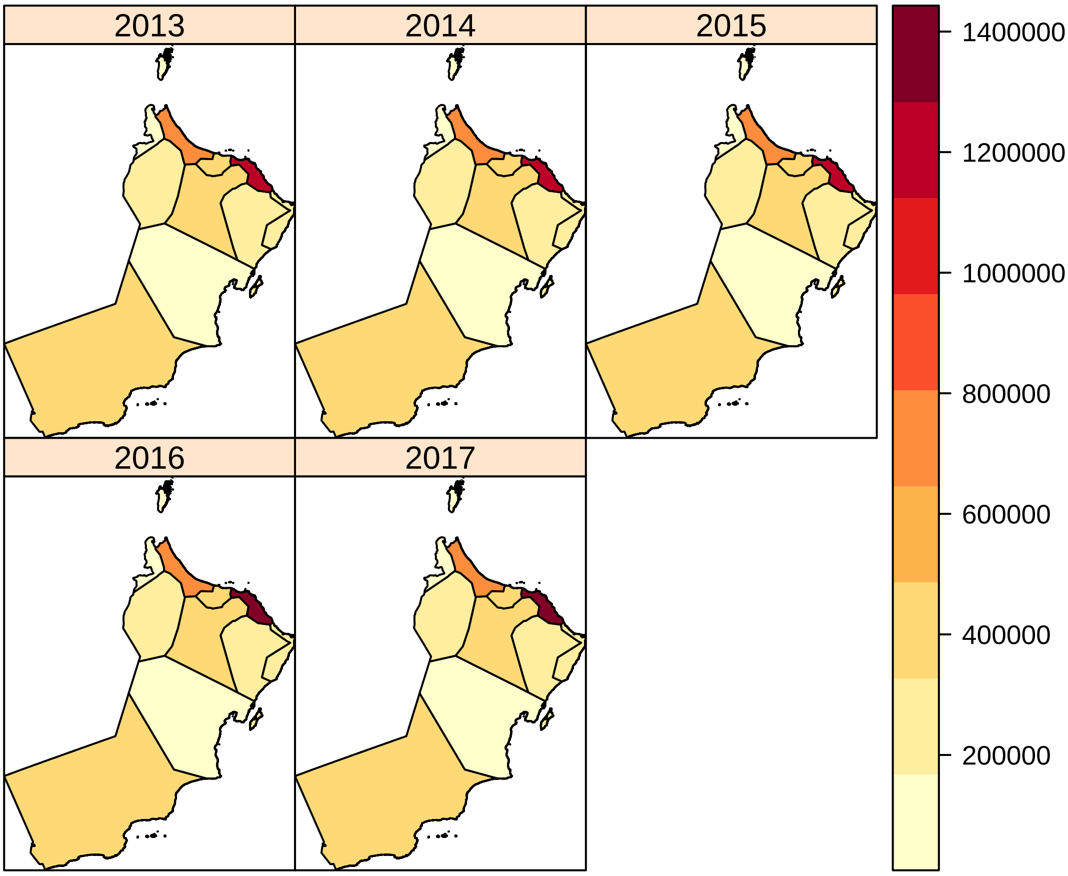
Zhang, X., Pang, Y., Cui, M., Stallones, L. and Xiang, H. (2015). Forecasting mortality of road traffic injuries in China using seasonal autoregressive integrated moving average model, *Annals of Epidemiology* **25**(2): 101–106.

Zhao, Y., Qiao, Z., Xiao, C., Glass, L. and Sun, J. (2021). Pyhealth: A python library for health predictive models, *arXiv preprint arXiv:2101.04209*.

Zou, Y., Ash, J. E., Park, B.-J., Lord, D. and Wu, L. (2018). Empirical bayes estimates of finite mixture of negative binomial regression models and its application to highway safety, *Journal of Applied Statistics* **45**(9): 1652–1669.

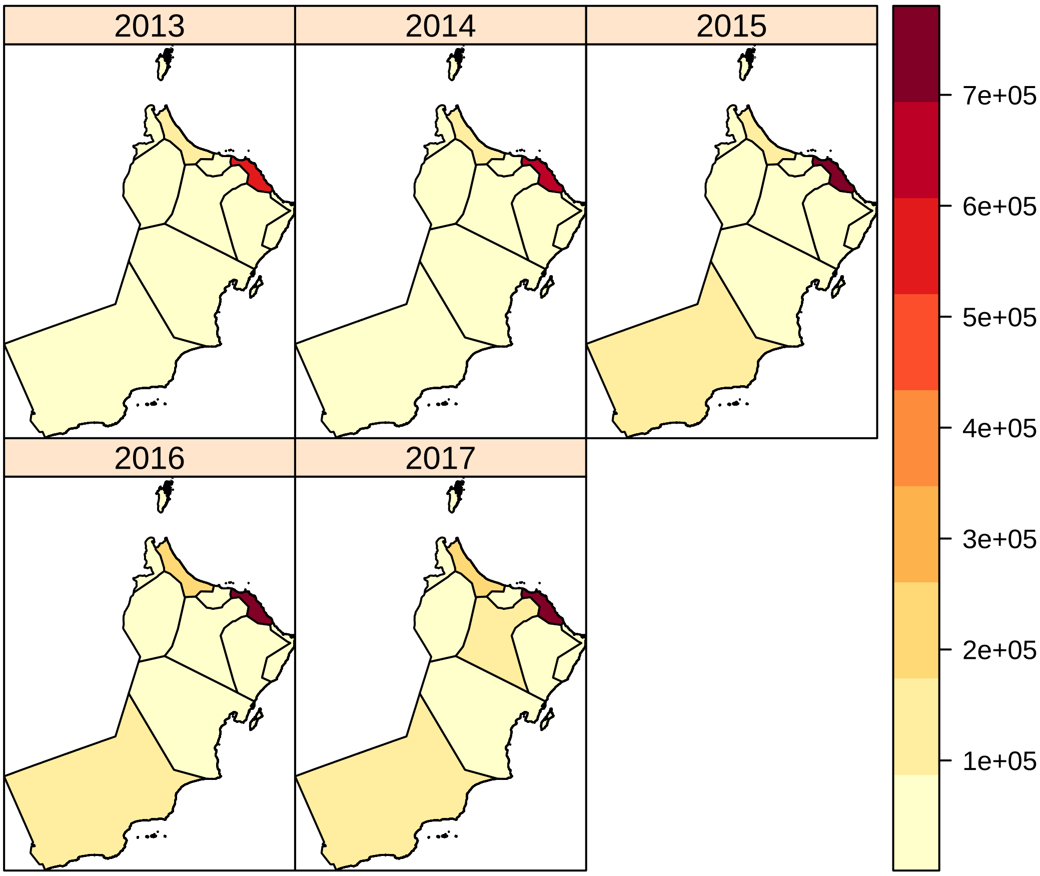
Appendix

Appendix A: Spatio-temporal pattern of population among Oman governorates



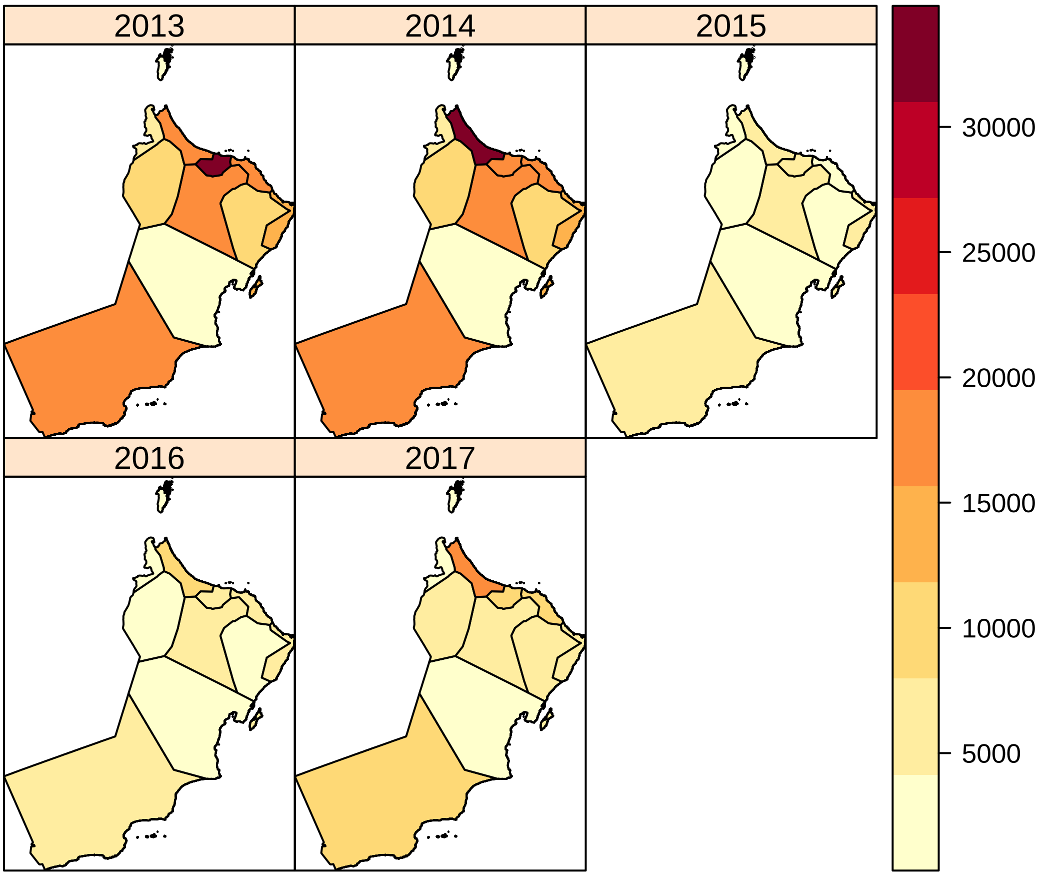
Popoulation size among Oman governorates (2013-2017)

Appendix B: Spatio-temporal pattern of registered vehicles among Oman governorates



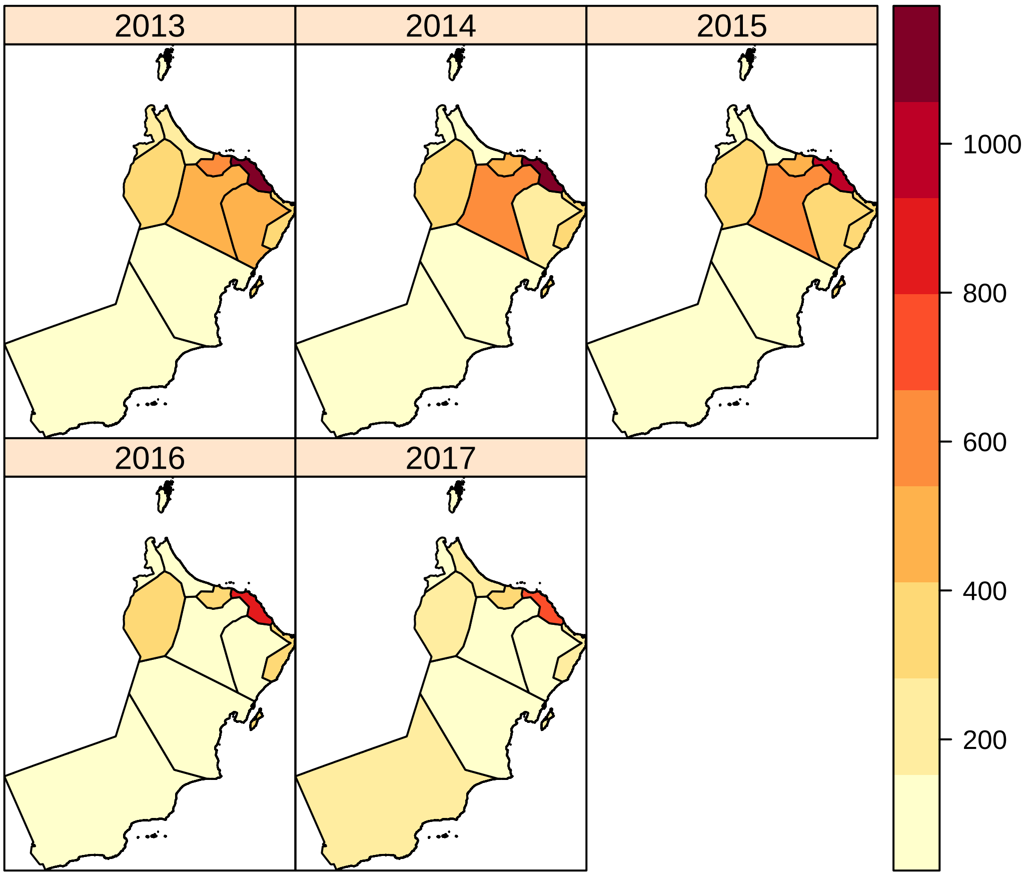
Registered vehicles among Oman governorates (2013-2017)

Appendix C: Spatio-temporal pattern of jobs seekers among Oman governorates



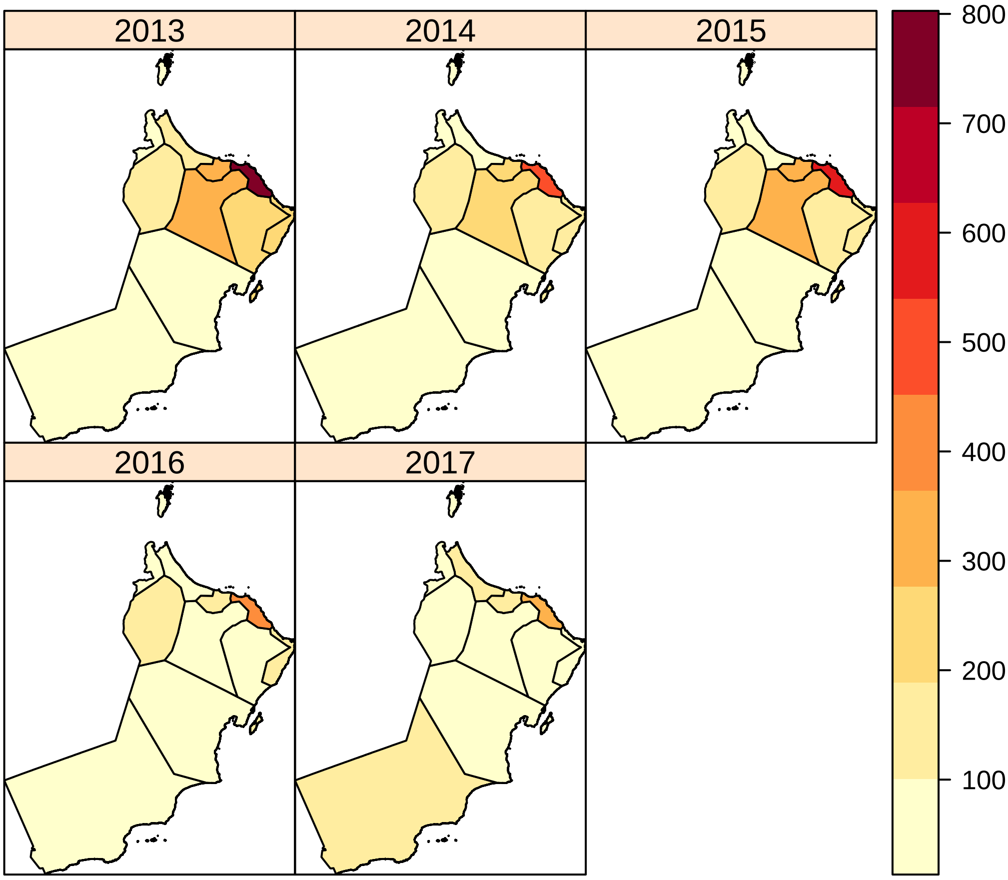
Unemployed persons among Oman governorates (2013-2017)

Appendix D: Spatio-temporal pattern of speed driving RTAs among Oman governorates



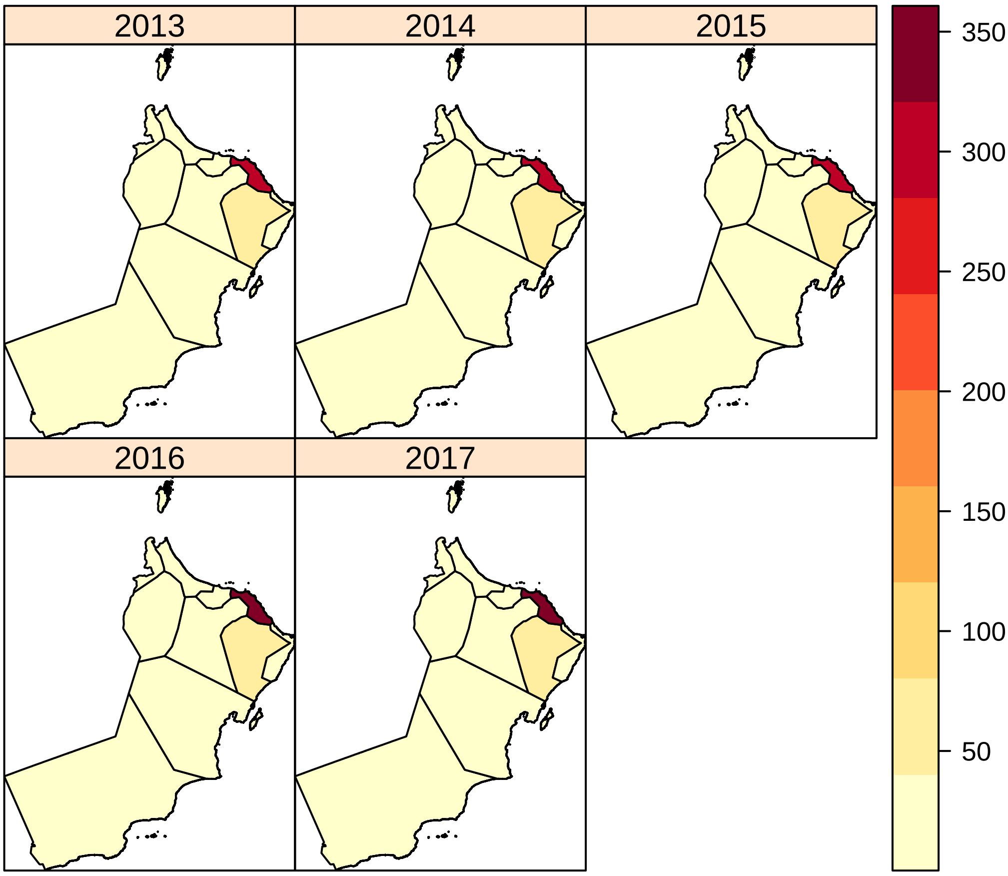
Speed driving RTA among Oman governorates (2013-2017)

Appendix E: Spatio-temporal pattern of seasonal RTAs among Oman governorates



Seasonal RTAs among Oman governorates (2013-2017)

Appendix F: Spatio-temporal pattern of population density in each governorate for Oman



Population density among Oman governorates (2013-2017)