Modelling and predicting collision risks between wildlife and moving vehicles across space and time

By

Casey C. Visintin
ORCID: 0000-0003-2245-8998

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy

March 2018

School of BioSciences
Faculty of Science
The University of Melbourne
Abstract

In brief: Wildlife-vehicle collisions pose considerable economic and environmental costs. Drawing on concepts from risk theory, ecology, and transportation planning, I investigate collision risk at varying spatial and temporal scales, for seven mammal species, and on two distinct vehicular networks (road and rail transport). I develop methods to quantify, analyse, and predict collision risk with a conceptual model framework that utilises publicly-accessible sources of data. The methods can be adapted to desktop-based tools and software applications to assist wildlife and road managers, as well as to GPS and smart-phone technologies to communicate risk to the general public.

As humans distribute across and modify landscapes, they increase their proximity and exposure to other biotic organisms. The global extent of roads pose several ecological threats, including destruction of habitat through fragmentation, to wildlife. Perhaps the most directly visible effect of roads is wildlife mortality from collisions with moving vehicles. From an anthropocentric perspective, animal-vehicle collisions with larger species cause injuries, property damage, and occasional loss of human life which compels managers to find solutions. Worldwide, many studies have been undertaken to understand and prevent wildlife-vehicle collisions; some studies related collisions to animal biology, behaviour and characteristics, whilst others assessed the effectiveness of on-road mitigation strategies. Road ecology currently lacks a generalised conceptual framework to organise all of the disparate factors relating to wildlife-vehicle collisions and support managers to make better decisions. The broad aim of this thesis is to model and predict where and when wildlife will be at high risk from collisions with moving vehicles across large transportation networks, and is intended to support decision-making by managers working in the wildlife and transportation fields. I address this broad aim in six chapters by developing and testing a conceptual risk modelling framework and highlighting its potential extension as a management and public outreach tool. Chapter 1 provides the general context for the work.

In Chapter 2, I introduce a novel method to predict collision risk by modelling hazard (presence and movement of vehicles) and exposure (animal presence) across a large transportation network. To estimate the hazard, I predict relative traffic volume and speed along road segments in a case study area (Victoria, south-east Australia) using regression models based on human demographic variables. I model exposure by predicting suitable habitat for a case study species (Eastern Grey kangaroo) based
on existing fauna survey records and geographic and climatic variables. The collision model is trained with records of reported kangaroo-vehicle collisions to investigate how these factors collectively contribute to collision risk. This framework is useful because it disentangles natural and anthropogenic effects on the likelihood of wildlife-vehicle collisions by representing hazard and exposure with separate, tunable sub-models.

I extend the conceptual risk framework to five additional native mammal species, using the same spatial extent, in Chapter 3. For each species, I construct independent statistical models that use wildlife atlas data to predict occurrence across geographic space. Traffic volume and speed on road segments, predicted from the work in Chapter 2, characterise the magnitude of threatening processes. I model collision risk for each species using the conceptual framework, compare results, and combine predictions to represent aggregated risk across the Victorian road network. The results indicate that the model framework generalises well to multiple species and provides useful outputs for supporting road management.

In Chapter 4, I apply the risk framework to a different species (Mule deer) and geographical location (central California, USA). This represents the first work to apply a generalised model framework to analyse and predict WVC in two continents. I use Eastern Grey kangaroos in Victoria (Chapter 2) as a comparative study. Despite being different species occurring in different geographic areas, I examine whether collision risk correlates similarly with animal behaviour (occurrence of the species) and human behaviour (traffic volume and speed). To assess the performance of the collision models, I use different sets of data, police records of road accidents, to evaluate each model’s predictions. The results demonstrate that the conceptual risk framework generalises well to different species and different geographical areas. My analysis also suggests appropriate mitigation for both species may include reduced speeds or fencing on road segments where species occurrence is predicted to be high.

Chapter 5 demonstrates an application of the conceptual risk framework to a different mode of transport (trains). Further, I add an additional feature to the collision model to account for temporal variation. To assess the risk of wildlife-train collisions, I quantify regional train movements in space and time, and determine the likelihood of occurrence of Eastern Grey kangaroos in Victoria, Australia. I model collisions between trains and kangaroos, accounting for time of day, train occurrence and speed, and kangaroo occurrence, and predict collision rates on a passenger railway network based on three management scenarios that influence train speed and occurrence of kangaroos near the railway lines. The results indicate that train speed is the most influential variable followed by presence of kangaroos. Reducing speeds in areas of high predicted kangaroo occurrence, during periods of peak animal activity within a twenty-four hour cycle, reduces collision rate the most. This work further supports the generality of the conceptual risk framework, and more specifically, that predictions from the model can help managers decide where, when and how best to mitigate collisions between animals and trains.

To further examine the performance of the work in Chapter 2, I compare the collision model framework predictions using several independent datasets in Chapter 6. The results demonstrate that whilst the model framework is sensitive to the choice of
training data, combining otherwise disparate sources of information can improve modelling performance and subsequent predictions of collision risk. This result supports the need for a global, publicly-accessible, unified system for storing and disseminating data on wildlife-vehicle collisions.

Lastly, in Chapter 7 I discuss the main findings of my thesis, caveats and limitations of the work, and future research opportunities. Further, I highlight key applications of the work to support management as a decision-making tool and educate the public as a mobile device-based information system.
Declaration

This is to certify that:

i. the thesis comprises only my original work towards the PhD except where indicated in the Preface,

ii. due acknowledgement has been made in the text to all other material used,

iii. the thesis is fewer than 100 000 words in length, exclusive of tables, maps, bibliographies and appendices.

______________________________

Casey Visintin
March, 2018
Preface

This thesis has been supported through collaborations with: my two primary supervisors, Michael A. McCarthy (MMC) and Rodney van der Ree (RvdR); a University of Melbourne collaborator, Graeme Coulson (GC); an overseas collaborator in California, Fraser Shilling (FS); and my PhD mentor, Nick Golding (NG). To facilitate journal publications throughout and subsequent to my PhD candidature, Chapters 2 to 6 are written as independent studies and, therefore, include detailed introductions and discussions to adequately frame each body of work. As such, there will be some unavoidable repetition in the chapters. I have written these chapters using the collective first-person plural “we”, consistent with academic publishing practice; collaborator contributions were:

Chapter 2. This chapter is reproduced, in its entirety, from the paper *A simple framework for a complex problem? Predicting wildlife-vehicle collisions*, published during my PhD candidature. The content in the chapter is nearly identical to the journal article with the following exceptions: (a) supporting information listed in the publication has been included within the thesis chapter for completeness; (b) figure and table numbers have changed to be in accordance with the thesis structure; and (c) some minor edits to text and figures have been made. MMC and I conceived the modelling ideas. I developed the methods, completed all of the analyses, and wrote the first draft of the manuscript. MMC and RvdR provided editorial comments on the chapter.

Chapter 3. This chapter is reproduced, in its entirety, from the paper *Consistent patterns of vehicle collision risk for six terrestrial mammal species*, published during my PhD candidature. I conceived the original ideas, developed the methods, completed all of the analyses, and wrote the first draft of the manuscript. MMC and RvdR provided editorial comments on the chapter.

Chapter 4. This chapter represents the submitted paper *A generalised predictive model for wildlife-vehicle collision risk with large mammals*. I conceived the original ideas, developed the methods, completed all of the analyses, and wrote the first draft of the manuscript. FS provided the data on deer collisions in California used in the chapter. MMC, RvdR, and FS provided editorial comments on the chapter.

Chapter 5. This chapter represents the paper *Managing the timing and speed of vehicles reduces wildlife-transport collision risk*, published during my PhD candidature. The original idea was developed through conversations with GC. I developed the methods, completed all of the analyses, and wrote the first draft of the manuscript. NG created the temporal function used in the analysis. MMC and RvdR provided
editorial comments on the chapter.

Chapter 6. I conceived the original idea through conversations with RvdR. I developed the methods, completed all of the analyses, and wrote the chapter. MMC and RvdR provided editorial comments on the chapter.

In this thesis, I also train and test statistical models with existing data on species occurrence and collision events. All of these data were originally collected by several different organisations (whom also currently maintain these data). More information on these organisations is available in Appendix C.

My PhD candidature and associated work was supported by a University of Melbourne International Research Scholarship and the Australian Research Council Centre of Excellence for Environmental Decisions.
First authored publications


4. Casey Visintin, Fraser Shilling, Rodney van der Ree, Michael A. McCarthy (in review) A generalised predictive model for wildlife-vehicle collision risk with large mammals. *Biological Conservation*
Transitioning from industry to academia can be a somewhat daunting experience. But for those who persevere, the journey can be quite rewarding. With generous support along the way, I have made that journey and there are many people I have to thank for such a memorable experience.

I would like to begin by thanking my two PhD supervisors, Michael A. McCarthy and Rodney van der Ree. Thanks for everything. Mick McCarthy taught me a huge amount about ecological modelling and environmental decision-making. He also improved my skills as an academic researcher. The Quantitative and Applied Ecology (QAECO) Lab is absolutely stellar and I owe Mick some serious gratitude for getting me in the door. Rod has not only taught me a great deal about road and urban ecology, but also significantly enriched my PhD experience. Thanks for bridging all the government collaborations and looking after me. And thanks for the extra-curricular opportunities; teaching the public about microbats in the Royal Botanic Gardens was a highlight. I truly envy Mick and Rod’s enthusiasm, optimism, and wisdom and could not have asked for better supervisors.

To Peter Vesk, my PhD committee chair, and Yung en Chee, PhD committee member, thanks for providing really useful guidance and support throughout my candidature. I would also like to thank everyone who attended my pre-completion talk – Graeme Coulson, Kath Handasyde, Andrew Robinson, Mark Burgman, Jane Elith, and Sam Parsons – the feedback was invaluable. And thanks for humouring my idea of a ‘mock defence’ session.

Amy Hidge, Karen Masson, and James Johnson at Wildlife Victoria deserve special recognition. Their support (and data) have really made my research possible. And a big thanks to others whom have generously shared their data; VicRoads, V/Line, the City of Bendigo, and Insurance Australia Group. To Fraser Shilling, my collaborator in California, thanks for the great data and providing engaging discussions on road ecology and academia – and keep up the good fight against the assault on science in the USA.

To my two wonderful mentors, Nick Golding and Kylie Soanes, thanks for all the support. Nick Golding is highly accomplished yet refreshingly humble and Kylie Soanes is an all-around inspiration. Both of them have generously counselled me on the ins and outs of academia. I look forward to possible collaborations with them in the future.

QAECO is a large group and several members have really made my tenure a special experience. Els van Burm, my long-lost Belgian twin and academic comrade, put in...
the long library hours with me. Michaela Plein read my drafts and taught me other important life lessons like making granola and growing vegetables. Darren Southwell was kind enough to share his long list of potential interview questions and look over some of my architect-turned-ecologist job applications. It was pleasure catching small mammals and frogs with Hugh Davies and Matt West, respectively. I shared an office with an amazing crew – a big thanks to Esti, Michelle, Saras and Holly for the delightful afternoon walks and morning coffee runs. To Freya, Clair, Hannah, Liz, Natasha, Gerry and Chris, thanks for being just plain awesome. I am humbled by the talents of Jian, José, Guru, Heini, Reid, Pia, Cindy, and Nat – all of whom I give thanks to for helping me along in this PhD journey. It has been a pleasure to know Pauline Byron and Dolla Boutros who have been super friendly and helpful and always looked after me. Thanks to everyone else for the great laughs and conversation during sunny lunches in the systems garden.

A huge thanks goes to Dave Duncan and Belinda Ashe for hosting me in their lovely home whilst I finished the thesis. Dave provided great advice, having completed his own PhD – not to mention prepared Melbourne’s finest tacos. Belinda generously proofread my drafts and kept me laughing when PhD stresses loomed about. To my family, of course, I give thanks for the support and understanding that a thesis makes one antisocial at times. And special thanks to my sister, Faren, for coming along to hear my completion talk.

Finally, thank you Australian government. My PhD candidature was generously supported by a University of Melbourne International Research Scholarship and by the Australian Research Council Centre of Excellence for Environmental Decisions. I hope to give back to Australia in the near future; what an amazing land.
## Contents

Abstract iii  
Declaration vii  
Preface ix  
Acknowledgements xiii  
List of Figures xix  
List of Tables xxi  

1 General Introduction 1  
1.1 The wildlife and roads conundrum 2  
1.1.1 Human-wildlife interactions 2  
1.1.2 Impacts of linear infrastructure on wildlife 4  
1.1.3 Wildlife mortality on roads 6  
1.2 Improving management and decision-making 6  
1.2.1 Understanding wildlife-vehicle collisions 6  
1.2.2 Incorporating risk theory & modelling 8  
1.2.3 The challenges of disparate data 9  
1.3 Thesis aims and scope 10  

2 A conceptual risk framework for modelling and predicting wildlife-vehicle collisions 13  
2.1 Introduction 15  
2.2 Materials and Methods 16  
2.2.1 Conceptual Model Framework 16  
2.2.2 Study Area & Species 18  
2.2.3 Species Occurrence Sub-Model 19  
2.2.4 Traffic Volume & Speed Sub-Models 21  
2.2.5 Collision Modelling 22  
2.3 Results 23  
2.4 Discussion 29  
2.4.1 Applied Management Implications 31
## Contents

3 Consistent patterns of vehicle collision risk for six terrestrial mammal species

3.1 Introduction .................................................. 35
3.2 Materials & Methods ........................................ 36
  3.2.1 Study Area .............................................. 36
  3.2.2 Study Species .......................................... 36
  3.2.3 Conceptual Model Framework ......................... 38
  3.2.4 Species Occurrence Models ......................... 38
  3.2.5 Collision Model ........................................ 42
3.3 Results ...................................................... 44
3.4 Discussion .................................................. 55
  3.4.1 Conclusions & Implications for Management ....... 57

4 A generalised predictive model for wildlife-vehicle collision risk with large mammals

4.1 Introduction .................................................. 61
4.2 Materials & Methods ........................................ 62
  4.2.1 Overview of Workflow ................................. 62
  4.2.2 Study Areas ........................................... 62
  4.2.3 Data ..................................................... 64
  4.2.4 Collision Modelling & Validation .................... 65
4.3 Results ...................................................... 69
4.4 Discussion .................................................. 73
  4.4.1 Management Implications ............................. 75
4.5 Supplemental Information ................................ 75
  4.5.1 Species Distribution Modelling ..................... 75

5 Predicting wildlife-train collisions across space and time to inform railway operations

5.1 Introduction .................................................. 83
5.2 Materials & Methods ........................................ 84
  5.2.1 Study Area .............................................. 84
  5.2.2 Data Preparation ....................................... 85
  5.2.3 Species Occurrence .................................... 86
  5.2.4 Characteristics of Railway Network ................. 86
  5.2.5 Statistical Modelling .................................. 87
5.3 Results ...................................................... 90
5.4 Discussion .................................................. 95

6 Validating collision model predictions with disparate datasets: a case for a unified system

6.1 Introduction .................................................. 101
6.2 Materials & Methods ........................................ 103
6.3 Results ...................................................... 106
6.4 Discussion .................................................. 111
7 General Discussion
7.1 Synthesis of work ........................................ 116
7.2 Benefits of spatial modelling .......................... 120
7.3 Limitations of models and caveats .................. 121
7.4 Novel future research ................................. 122
7.5 Generalising to other threats to wildlife .......... 123
7.6 Potential applications of research ................. 125
7.6.1 Centralised data collection and reporting system . 125
7.6.2 Integration of thesis research with existing and new technology 127
7.7 Concluding remarks ................................... 128

References .................................................. 129

Appendix A: Computer code ............................... 149
A.1 Links to R code used in Chapter 2 ............... 150
A.2 Links to R code used in Chapter 3 ............... 150
A.3 Links to R code used in Chapter 4 ............... 151
A.4 Links to R code used in Chapter 5 ............... 151
A.5 Links to R code used in Chapter 6 ............... 151

Appendix B: Species used in research ............... 153
B.1 Mule deer ............................................. 154
B.2 Eastern Grey kangaroos .............................. 154
B.3 Swamp wallabies .................................... 155
B.4 Wombats ............................................. 155
B.5 Koalas ............................................... 155
B.6 Brushtail possums .................................. 156
B.7 Ringtail possums .................................... 156

Appendix C: Data sources used for research .......... 157
C.1 Wildlife Victoria ..................................... 158
C.2 California Roadkill Observation System .......... 158
C.3 VicRoads ............................................. 158
C.4 V/Line .............................................. 158
C.5 City of Bendigo ..................................... 159
C.6 Crashstats .......................................... 159
C.7 Insurance Australia Group .......................... 159
C.8 Victorian Biodiversity Atlas ....................... 159
C.9 Global Biodiversity Information Facility .......... 160
C.10 WorldClim ......................................... 160
C.11 Moderate Resolution Imaging Spectroradiometer .... 160
C.12 Defense Meteorological Satellite Program ....... 161
C.13 Australian Grassland and Rangeland Assessment by Spatial Simulation 161
C.14 Geoscience Australia ................................ 161
C.15 United States Geologic Survey .................... 161
C.16 Global Land Cover Facility ................................. 162
C.17 VicMap ...................................................... 162
C.18 SoilGrids .................................................... 162
C.19 Australian Bureau of Statistics ......................... 162
C.20 United States Census Bureau ............................. 163

Appendix D: List of studies used for comparing modelling variables 165
List of Figures

1.1 Conceptual diagram for framing human-wildlife interactions .................. 4
1.2 Global distribution of known roads ........................................ 5
1.3 Variables used in wildlife-vehicle collision studies ......................... 7

2.1 Diagram of the modelling framework ........................................ 18
2.2 Predicted relative likelihood of grey kangaroo occurrence in Victoria .... 20
2.3 Predicted relative traffic volume in Victoria ................................ 21
2.4 Predicted relative traffic speed in Victoria ................................ 22
2.5 Map of collision risk per road segment in Victoria ......................... 23
2.6 Effects of predictor variables on relative likelihood of grey kangaroo
    occurrence ................................................................. 24
2.7 Predicted relative likelihood of grey kangaroo occurrence across Australia 26
2.8 Effects of predictor variables on traffic volume and speed .................. 27
2.9 Effects of predictor variables on relative likelihood of collision .......... 28

3.1 Wildlife-vehicle collisions in Victoria ..................................... 37
3.2 Predicted relative likelihood of six mammal species occurrence in Victoria 40
3.3 Road segments with predictions of high collision likelihood for six mam-
    mal species ................................................................. 43
3.4 Most significant predictor variables on relative likelihood of occurrence
    for six mammal species .................................................. 46
3.5 Spatial autocorrelation in occupancy models residuals for six mammal
    species ................................................................. 48
3.6 Marginal effect of species occurrence on relative likelihood of collision
    for six mammal species ................................................ 49
3.7 Marginal effect of traffic volume on relative likelihood of collision for
    six mammal species ..................................................... 50
3.8 Marginal effect of traffic speed on relative likelihood of collision for six
    mammal species ......................................................... 51
3.9 Marginal effect of species occurrence on relative likelihood of collision
    for six mammal species by season ..................................... 52
3.10 Marginal effect of traffic volume on relative likelihood of collision for
    six mammal species by season ......................................... 53
3.11 Marginal effect of traffic speed on relative likelihood of collision for six
    mammal species by season ............................................. 54
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Locations of WVC for Eastern Grey kangaroos in Victoria and Mule deer in central California</td>
<td>63</td>
</tr>
<tr>
<td>4.2</td>
<td>Spatial autocorrelation in randomised quantile model residuals for kangaroos and deer</td>
<td>66</td>
</tr>
<tr>
<td>4.3</td>
<td>Predicted relative collision risk for all sealed roads in Victoria and central California</td>
<td>68</td>
</tr>
<tr>
<td>4.4</td>
<td>Marginal effects of predictor variables on relative likelihood of collision for kangaroos and deer</td>
<td>72</td>
</tr>
<tr>
<td>4.5</td>
<td>Spatial autocorrelation in occupancy model residuals for kangaroos and deer</td>
<td>77</td>
</tr>
<tr>
<td>4.6</td>
<td>Most significant predictor variables on relative likelihood of occurrence for kangaroos and deer</td>
<td>78</td>
</tr>
<tr>
<td>4.7</td>
<td>Predicted relative likelihood of occurrence for kangaroos in Victoria and deer in central California</td>
<td>79</td>
</tr>
<tr>
<td>5.1</td>
<td>Regional passenger train network in Victoria</td>
<td>85</td>
</tr>
<tr>
<td>5.2</td>
<td>Spatial grid used to organise modelling data</td>
<td>86</td>
</tr>
<tr>
<td>5.3</td>
<td>Components used to build crepuscular term for kangaroo-train collision model</td>
<td>88</td>
</tr>
<tr>
<td>5.4</td>
<td>Binned randomised quantile residuals plots for kangaroo-train collision model</td>
<td>92</td>
</tr>
<tr>
<td>5.5</td>
<td>Effects of predictor variables on relative likelihood of kangaroo-train collisions</td>
<td>93</td>
</tr>
<tr>
<td>5.6</td>
<td>Calibration and ROC plots for kangaroo-train collision model</td>
<td>94</td>
</tr>
<tr>
<td>6.1</td>
<td>Study area for model validation</td>
<td>104</td>
</tr>
<tr>
<td>6.2</td>
<td>Marginal effects of predictor variables on relative likelihood of collision using independent datasets to train models</td>
<td>107</td>
</tr>
<tr>
<td>6.3</td>
<td>Model calibration for all combinations of original and independent data</td>
<td>109</td>
</tr>
<tr>
<td>6.4</td>
<td>Comparisons of observations versus model predictions for all combinations of original and independent data</td>
<td>109</td>
</tr>
<tr>
<td>6.5</td>
<td>Model calibration for all combinations of training data using the town-level independent data for validation</td>
<td>110</td>
</tr>
<tr>
<td>6.6</td>
<td>Model performance for all combinations of training data using the town-level independent data for validation</td>
<td>110</td>
</tr>
<tr>
<td>7.1</td>
<td>Total collisions by hour for six mammal species</td>
<td>119</td>
</tr>
<tr>
<td>7.2</td>
<td>Conceptual risk model framework for other threats</td>
<td>124</td>
</tr>
<tr>
<td>7.3</td>
<td>Centralised data collection and reporting system</td>
<td>126</td>
</tr>
<tr>
<td>7.4</td>
<td>Decision support tool for wildlife-vehicle collisions</td>
<td>128</td>
</tr>
</tbody>
</table>
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Variables used in statistical models</td>
<td>20</td>
</tr>
<tr>
<td>2.2</td>
<td>Statistical models used in the conceptual framework</td>
<td>25</td>
</tr>
<tr>
<td>2.3</td>
<td>Summary of collision and alternative collision models</td>
<td>28</td>
</tr>
<tr>
<td>3.1</td>
<td>Six mammal species most frequently reported in wildlife-vehicle collisions</td>
<td>38</td>
</tr>
<tr>
<td>3.2</td>
<td>Predictor variables used in species occurrence models and collision risk models</td>
<td>39</td>
</tr>
<tr>
<td>3.3</td>
<td>Statistical model performance for six mammal species</td>
<td>41</td>
</tr>
<tr>
<td>3.4</td>
<td>Summary of collision models for six mammal species</td>
<td>45</td>
</tr>
<tr>
<td>3.5</td>
<td>Predictor variables used in species occurrence models and collision risk models</td>
<td>45</td>
</tr>
<tr>
<td>4.1</td>
<td>Performance of traffic models</td>
<td>65</td>
</tr>
<tr>
<td>4.2</td>
<td>Summary of collision models for kangaroos and deer</td>
<td>69</td>
</tr>
<tr>
<td>4.3</td>
<td>Statistical model performance for kangaroos and deer</td>
<td>70</td>
</tr>
<tr>
<td>4.4</td>
<td>Ten highest ranked road segments for predicted collision risk for kangaroos and deer</td>
<td>71</td>
</tr>
<tr>
<td>4.5</td>
<td>Predictor variables used in species occurrence models for kangaroos and deer</td>
<td>76</td>
</tr>
<tr>
<td>5.1</td>
<td>Predictor variables used in kangaroo-train collision model</td>
<td>89</td>
</tr>
<tr>
<td>5.2</td>
<td>Summary of kangaroo-train collision model</td>
<td>90</td>
</tr>
<tr>
<td>5.3</td>
<td>Predicted kangaroo-train collisions based on different management scenarios</td>
<td>91</td>
</tr>
<tr>
<td>6.1</td>
<td>Datasets used to fit collision models and validate predictions</td>
<td>106</td>
</tr>
<tr>
<td>6.2</td>
<td>Summary of model fits using original and validation data</td>
<td>107</td>
</tr>
<tr>
<td>6.3</td>
<td>Discrimination ability of models using all combinations of independent data</td>
<td>108</td>
</tr>
<tr>
<td>7.1</td>
<td>Examples of human-wildlife conflicts and conceptualised parameters</td>
<td>125</td>
</tr>
</tbody>
</table>
1

General Introduction
1.1 The wildlife and roads conundrum

1.1.1 Human-wildlife interactions

We are now in what some refer to as the Anthropocene, where *Homo sapiens* exert strong influences on climatic, geologic and biotic systems (Crutzen, 2006). Studies estimate that humans have transformed more than seventy-five percent of the earth’s ice-free terrestrial surface (e.g. Ellis & Ramankutty, 2008). These modifications result from large-scale activities; perhaps the most influential being intensified agricultural production, provision of linear infrastructure (e.g. roads, rail, and utility easements), and urban development (Vitousek et al., 1997; Sanderson et al., 2002; Foley et al., 2005). The human population is forecast to continue to grow in the future, intensifying existing activities in modified landscapes and increasing pressure on unmodified landscapes. For example, recent projections indicate a human population of between 9.6 and 12.3 billion in the year 2100 (Gerland et al., 2014). The largest rates of increase are projected for developing countries and these will have environmental consequences such as land transformation. Further, these issues are compounded by the high amount of biodiversity and low amount of conservation resources and legislative protection in developing countries.

Increases in human settlements, industry, and related infrastructure advance the proximity between people and wildlife. The ensuing ‘human-wildlife interactions’ (see Manfredo, 2008) may be positive (e.g. psychological well-being, see Dallimer et al., 2012), negative (e.g. swooping birds, see Jones & Thomas, 1999) or neutral. Herein, I define wildlife as any non-domesticated organism in the biological kingdom of Animalia.

Human-wildlife interactions are multifaceted because the same organism may be associated with positive interactions for one person or part of society, and negative interactions for another. For example, some people may appreciate the existence of pigeons in urban environments whilst others may view the birds as a nuisance or invasive species (Harris, de Crom & Wilson, 2016). Cultural values have a large influence on these views; some groups celebrate wildlife species that are believed to have spiritual kinship (e.g. totemism) or persecute wildlife due to religious beliefs (e.g. killing of Aye-ayes on sight in Madagascar, see Hill & Webber, 2010). Further, the human dimension of wildlife interactions may involve humans in the capacity of actors, interacting directly with wildlife, or observers who witness how others do or have interacted with wildlife. In wildlife-vehicle collisions, for example, people are both actors as vehicle occupants and, subsequently, observers as motorists passing road-killed wildlife.

Interactions can be frequent or infrequent, fleeting or long-run, and have large or small impact either individually or in aggregate. These impacts can be considered from the perspective of humans, wildlife, or both. The outcomes or consequences for wildlife can be tied to vital rate parameters, such as mating and breeding success, birth and death rates, and be applicable to individual animals, populations, or even evolutionary significant units or species. Positive outcomes of interactions
with wildlife for humans are diverse and can be grouped as consumptive and non-
consumptive (Chardonnet et al., 2002). Negative consequences for humans include
nuisance, damage to property, or risk to health or life and, like positive outcomes,
can be contemplated at the level of individual actors/observers, or the level of soci-
ety. These impacts can be conceptually diagrammed as a product of the frequency,
or duration, and consequence of the interaction (Figure 1.1). Although many interac-
tions usually have opposite effects on humans and wildlife (i.e. good for humans, bad
for wildlife, and vice versa), some human-wildlife interactions may have benefits (i.e.
‘mutualism’) or negative outcomes (e.g. wildlife-vehicle collisions) for both humans
and wildlife. Each situation requires unique management strategies; for instance, the
severity and nature of elephant raids on crops of subsistence farmers in a developing
country will warrant different mitigation than coyote attacks on domesticated pets in
affluent neighbourhoods.

In urban areas, exposure to wildlife is due to tolerant species of wildlife remaining or
re-establishing in human-dominated landscapes (Soulsbury & White, 2016). Although
the suite of organisms that humans have contact with is quite varied – the most
frequent contact possibly being with invertebrates – medium to large species (i.e.,
>10 kg) tend to get more attention (Seoraj-Pillai & Pillay, 2016). Certain kinds of
interactions, usually where wildlife and humans compete for resources, or intersect in
a way that causes nuisance or harm to human activities are referred to as ‘human-
wildlife conflicts’ (but see Peterson et al., 2010). Studies have demonstrated notable
impacts on society (e.g. human injuries, illness, and economic losses) from wildlife (e.g.
Conover et al., 1995). Larger species are more likely to cause significant harm to person
or damage to property (Conover, 2001) and are less easily managed by individuals
that lack proper training. Therefore, the public expects government intervention to
manage human-wildlife conflicts (Reiter, Brunson & Schmidt, 1999). Further, the
general public often views the management of human-wildlife conflicts to reside in
the scope of wildlife-focussed disciplines. However, issues arising from human-wildlife
interactions are complex and may require knowledge from several non-ecological fields
(Decker & Chase, 1997; Madden, 2004). For example, specialist knowledge on both
animal behaviour (zoology) and human behaviour (sociology) may be combined to
arrive at effective solutions (Dickman, 2010).
Figure 1.1: Conceptual diagram for framing human-wildlife interactions from both a human perspective (upper panels) and wildlife perspective (lower panels). The consequence (perceived or actual) of an interaction can range from positive (blue shading in right-hand panels) to negative (red shading in left-hand panels), or be all negative or all positive. For example, bird feeding by residents in backyards (🔗) is all positive (ignoring possible spread of disease for illustration purposes) and a wildlife-vehicle collision with a large-bodied animal (.cloud) is only negative (hence only a single representation of those icons in each panel). Crop-raiding by elephants (🐘) has only negative consequences for humans but the consequences for elephants may be negative, due to dispersal or control measures to prevent crop raiding, or positive if the elephant successfully obtains food. Note, perceived consequences can also vary by individual or group. Impacts are amplified by the frequency or duration of an interaction. Rarer events with higher negative consequences may be equivalent to frequent events with lower negative consequences such as raccoons feeding from rubbish bins ( gölge ). The concentric zones and numbers are based on a relative Likert-scale and may be used to sum interaction scores between the two perspectives to determine priorities for mitigation.

1.1.2 Impacts of linear infrastructure on wildlife

Linear infrastructure can traverse vast areas between human settlements, and penetrate remaining natural and semi-natural areas. In some cases, these easements may be the only remaining semi-natural refuges for wildlife (Bennett, 1991). Linear infrastructure is most commonly thought of as roads, however, can also include railways, utility easements, and other continuous human-engineered resources in the landscape.
Impacts of linear infrastructure on wildlife can be positive, neutral, or negative. For example, disused road reserves lead to retention of habitat that may benefit wildlife as they often contain old and diverse remnant vegetation (Lentini et al., 2011). Further, linear infrastructure without moving vehicles may provide connection corridors and assist migration for some species (van der Ree, Smith & Grilo, 2015). Effects may be neutral when suitable habitat is provided on road reserves with very little traffic volume or occupied by species with low mobility. Nonetheless, negative effects of linear infrastructure on wildlife persistence may certainly outweigh the benefits.

There are few places left on Earth that are not accessible by some form of road (Figure 1.2) and, consequently, the vast sphere of ecological influence makes roads a global concern (Laurance et al., 2014). The field of ‘road ecology’ evolved to study problems and solutions associated with the construction and use of roads across the natural landscape (see Forman et al., 2003). It is well documented that roads have negative impacts on wildlife (Forman & Alexander, 1998; Spellerberg, 1998; van der Ree, Smith & Grilo, 2015), both indirectly and directly. The construction and use of road networks contributes to fragmentation, modification and loss of habitat, generates environmental pollution and facilitates human disturbance to natural landscapes (e.g. clearing, logging, mining, hunting). Roads also have direct impacts on wildlife populations through mortality resulting from wildlife-vehicle collisions (Fahrig & Rytwinski, 2009).

Figure 1.2: Global distribution of known roads. Actual distribution may be under-represented in some developing countries due to data deficiencies. Data used to create the map was sourced from the Center for International Earth Science Information Network (www.ciesin.org); accessed 3 February, 2017.
1.1.3 Wildlife mortality on roads

Human-wildlife conflicts occur where moving vehicles and animals co-occur in space and time, and this is facilitated by transportation networks. Private transport is increasingly common; people travel further and faster. Growth of roads is particularly large in developing nations, which will increase as developing nations become more affluent (van der Ree, Smith & Grilo, 2015). The rate of mortality of wildlife will also increase due to collisions with moving vehicles. It is estimated that billions of vertebrate fauna die annually due to collisions with vehicles on transportation networks (Seiler & Helldin, 2006). Animals that are struck and die on roadsides or railway corridors attract predators and may result in secondary collisions which further increase mortality rates (Spellerberg, 1998). From a conservation perspective, these mortalities can result in population effects and disruption to food-chain networks (Ramp & Ben-Ami, 2006; Chambers & Bencini, 2010; Polak et al., 2014). Although collisions with wildlife bring about matters of animal welfare, biodiversity, and wildlife persistence, perhaps the largest concern to road managers is human safety.

From an anthropocentric perspective, wildlife-vehicle collisions with larger species cause injuries, property damage, and occasional loss of human life which compels road and wildlife managers to find solutions. Both the visibility of the problem – resulting in negative public reactions – and potential threats to human safety have caused some road authorities to regularly patrol problematic roads and collect carcasses. In some cases the data are recorded (Huijser et al., 2007b). This has also initiated partnerships between academics and road authorities to conduct scientific research on animal-vehicle collisions. These endeavours have created a wealth of research and literature on larger common species (e.g. Bissonette, Kassar & Cook, 2008; Clevenger et al., 2002; Romin & Bissonette, 1996; Sudharsan, Riley & Campa, 2009), but proportionally lower amounts on smaller (and perhaps more threatened) vertebrate fauna (e.g. Clevenger, Chruszcz & Gunson, 2003).

1.2 Improving management and decision-making

1.2.1 Understanding wildlife-vehicle collisions

The factors influencing the abundance of collision between wildlife and moving vehicles (wildlife-vehicle collisions) are quite broad and can be grouped into four major categories: animal behaviour, human behaviour, the road environment, and the landscape environment (Forman et al., 2003). For example, increases in collisions may be due to the following: animal behaviours, such as migration, flight response, or feeding patterns, thereby increasing species presence on roads; the operation of motor vehicles by humans at high volumes, high speeds, and at dangerous times; landscape features attracting animals to, and influencing movement across, roads and; road features hindering safe vehicle operation. In seventy-one studies that analysed wildlife-vehicle collision risk, traffic related variables (volume and speed) – which may be considered human behaviour – featured in the top four most commonly used predictor variables
1.2 Improving management and decision-making

See Appendix D for all variables and studies used in the analysis.

(Figure 1.3). Collision occurrence is related to many combinations of environmental and anthropogenic variables (Barnum, Rinehart & Elbroch, 2007). Some of these factors are more easily managed than others and that presents an optimisation problem for managers; where to have the largest effect for a given expenditure depends on knowing the relative influence of each factor on collision risk. In some cases the conflicts can be mitigated with engineering solutions, such as tunnels and land bridges (Bond & Jones, 2008), or rope ladders (Soanes et al., 2013) which are more suited to known species movements or behaviours. Fencing, also an engineering solution – and often used in concert with the aforementioned mitigation to direct animal movements – is primarily focussed on human safety through reducing the rate of wildlife-vehicle collisions (Langley, Higgins & Herrin, 2006). Where mobile wildlife are in contact diffusely over a road network it is a more difficult problem to manage. In these cases, mitigation can either focus on excluding or discouraging wildlife from roads or changing human behaviour.

Mitigation strategies to reduce wildlife-vehicle collisions have varying costs and effectiveness (Huijser et al., 2009) and costs can be compounded by additional collisions occurring from ineffective mitigation. Fencing is very effective at excluding animals from roads and reducing WVC, however, can result in other negative impacts to wildlife such as limitation of dispersal (i.e. barrier effects, see Jaeger & Fahrig, 2004). Fences are often used in conjunction with crossing structures to facilitate
connectivity whilst reducing collisions but these installations can be expensive. Dis-
couragement devices to scare animals off of roads are much less expensive but have
marginal effectiveness (Bender, 2003; Scheifele, Browning & Collins-Scheifele, 2003).
Influencing human behaviour is perhaps the least expensive form of mitigation yet
difficult to implement effectively. Road signage is the most common form, however, is
predominantly ineffective as humans become easily conditioned to warning messages
(Bond & Jones, 2013). For costly mitigation, it is therefore important to know exactly
where to mitigate along transportation networks.

In addition to knowing how and where to mitigate, an important objective in
managing wildlife-vehicle collisions is also knowing when to mitigate. This applies
to non-permanent forms of mitigation such as changes in vehicular movements (e.g.
frequency and speed), seasonal road closures (see Hilty, Lidicker Jr & Merenlender,
2012), or warning systems (but see Huijser & McGowen, 2003, for review of auto-
mated signs). Temporal variation occurs in both natural and anthropogenic systems.
Wildlife species have different patterns of activities (e.g. diurnal, nocturnal, crepus-
cular) that make them more or less susceptible to wildlife-vehicle collisions during the
day. When predictions of collision risk are made using temporal information, they
are referred to as ‘hot moments’. In the few studies that represent temporal variation
at the hourly scale (e.g. Neumann et al., 2012), the models do not consider changing
patterns of daylight throughout the year – which often influences wildlife activity. For
example, using hour as a predictor will have considerable error when species have
consistent daily patterns of movement, relative to dawn/dusk, throughout the year.
Further, seasonal factors such as migration and mating periods may influence species
movements and are therefore useful to include in the modelling.

1.2.2 Incorporating risk theory & modelling

Preventing wildlife-vehicle collision may be viewed as a risk analysis problem. Risks
are chances of adverse events, that have defined consequences, happening in specified
time frames (Burgman, 2005). Managing risks involves a) assessing hazards (i.e. po-
tentially harmful situations), exposure to hazards, and probabilities of hazards occur-
ring, b) simulating outcomes under different scenarios, and c) making decisions based
on acceptable thresholds. To assess risk, it is useful to employ quantitative methods
either independently or in combination with qualitative methods. Quantitative anal-
ysis is used to understand, communicate, and manage uncertainty about risk in many
fields such as engineering (e.g. Apostolakis, 2004), biological sciences (e.g. Suter II,
2016), and finance (e.g. McNeil, Frey & Embrechts, 2015). All of these fields define
risk differently and the parameterisations used in their analyses vary accordingly.

Many studies have utilised environmental modelling to analyse and predict risks
to species (fauna and flora) and biotic systems resulting from human activity. The
outputs of these models have many benefits including informing management opera-
tions, supplementing additional models, facilitating scientific research, and optimising
mitigation strategies. The use of modelling in conservation efforts is well established
(Starfield & Bleloch, 1986), however, predicting risks to wildlife by vehicles is a relatively new practice. The most common use of modelling has been to identify collision ‘hotspots’ in geographic space in an effort to direct management efforts. For example, Malo, Suárez & Díez (2004) predicted collision risk along a 4-lane highway in Spain for multiple species at two spatial scales using logistic regression. Similar methods have been used to predict collision risk with wombats (Roger & Ramp, 2009), and deer (Sudharsan, Riley & Campa, 2009).

Whilst the predictions from such models may be useful to managers, a comparative framework that describes simple relationships between variables, with tunable parts, and allows simulation of outcomes is more useful. First, it can identify simple predictors for collision risk that are able to be mitigated, such as traffic volume. This also allows sensitivity analyses on proposed management scenarios to be conducted (i.e. varying predictor values and analysing resulting changes in collision occurrence). Further, although management implications are highlighted in many studies, it is not always clear how the results of the analyses could be easily applied due to unknown interactions between variables and confounding effects in the model predictors (Gunson, Moutrakis & Quackenbush, 2011). Some studies explicitly suggest mitigation that relates directly to environmental variables based on model results (e.g. Grilo, Bissonette & Santos-Reis, 2009). Although these recommendations are useful in the specific contexts of the studies, they may not generalise to varying spatial scales – specifically large areas. In response to these shortcomings, road ecology needs a general framework to combine the disparate factors relating to wildlife-vehicle collisions (Clevenger et al., 2015) and better support management objectives, globally.

For the purpose of this work, I use ‘collision risk’ throughout to refer to ‘relative abundance of collisions between wildlife and moving vehicles’. In contrast, many road agencies view collision risk in the context of individual drivers. In this respect, the implications of risk are predominately viewed as human injury or death resulting from vehicle collisions with other objects - which in some cases include large animals. The modelling framework described herein operates irrespective of this conceptualisation of risk when used with large species that are considered road hazards and included in the suite of threats that road agencies protect drivers against. However, in this thesis, ‘collision risk’ encapsulates all the problems faced when wildlife are struck by moving vehicles from both animal (e.g. population effects, welfare) and human (e.g. injury, property damage) perspectives.

### 1.2.3 The challenges of disparate data

The contemporary modeller has considerable access to publicly-available data. Anthropogenic data, such as demography, exist for most developed parts of the world. Environmental data, particularly satellite-based and remotely-sensed, have global coverage. Collecting data can be quite expensive, and this information provides low-cost sources of information and has the potential to significantly expand the scope of modelling wildlife-vehicle collisions. Further, citizen-science data can supplement data-deficient studies, although the use of this data in collision modelling is quite low
relative to the amount in existence (but see Paul et al., 2014). Worldwide, collision data is collected by non-government organisations, private-sector entities (e.g. insurance companies), academic institutions, government agencies (e.g. road authorities and environmental managers), community groups, and individual citizen scientists but these data are not always easily attainable due to privacy restrictions or data formats.

Consequently, many studies that use predictive modelling to determine where and when WVC will occur are limited to single sources of data – either field collected based on predetermined project requirements (e.g. Langen, Ogden & Schwarting, 2009; Roger & Ramp, 2009) or pre-existing from independent surveys or routine collections (e.g. Hothorn, Brandl & Müller, 2012; Malo, Suárez & Díez, 2004). Bias and error may arise from the data sources and collection methods, however, these are not often explicitly reported or accounted for in the modelling methods. Further, few studies utilise independently-obtained data to validate model performance. Although cataloguing and reporting systems are used to store data on WVC, few are open access or incorporate data from other institutions or previous surveying events (but see Shilling, Perkins & Collinson, 2015). As disparate sources of collision data are combined, management will require an analytic modelling tool that is flexible to these data and can explicitly handle associated biases and errors. Methods such as multilevel modelling using Bayesian inference may be used to express explanatory variables as distributions with associated uncertainty rather than point estimates. This is especially useful when models are fitted to data in which the predictor variables are outputs (e.g. predictions) generated by other models. Further, these methods also accommodate measurement error (e.g. survey bias for collision reporting) by representing the dependent variable as a distribution with associated uncertainty.

1.3 Thesis aims and scope

It is clear that wildlife-vehicle collisions are a complex problem that have global implications. This problem will continue into the future as more developing nations expand their linear transportation networks thereby bringing humans and wildlife into closer proximity. Effective management will require knowing where, when, and how to mitigate collision risks, but this will require a clear understanding of the factors, and their interactions, relating to wildlife-vehicle collisions. Previous work on modelling and predicting collision risk have been contextually useful but their generality or relationship to management objectives were not explored. There is a critical need for a conceptual risk framework for modelling and predicting wildlife-vehicle collisions that is flexible to different parameterisations, informative for management objectives, generalisable to different spatial scales and transportation networks, and responsive to temporal information. These features will support global efforts to reduce wildlife-vehicle collisions.

My aim is to introduce a conceptual model framework to analyse and predict where and when wildlife will be at high risk of collisions with moving vehicles across
large transportation networks, irrespective of species or geographic area. As I intend the work to support managers working in the wildlife and transportation fields, I develop analyses that are generalisable, transferable, reproducible, and based on open-source data and methods. In each chapter, I test different model applications and demonstrate management value. I introduce a general conceptual framework to model wildlife-vehicle collision risk using tunable sub-models and publicly-accessible data in Chapter 2. Using the same geographic area, in Chapter 3 I test the ability of the conceptual model framework to generalise to six mammal species commonly involved in wildlife-vehicle collisions. In Chapter 4, I apply the conceptual framework to a different geographic area (California, USA) and a different species (Mule deer) to test its flexibility to varying locations and spatial scales. In addition to road vehicles, trains also hit wildlife but collision risk on railways is seldom studied. I test the conceptual model’s ability to generalise to other transportation networks and analyse and predict wildlife-train collisions in both space and time in Chapter 5. To assess the performance of the model introduced in Chapter 2, and demonstrate the importance of unified and standardised data repositories of wildlife-vehicle collisions, I compare model predictions to additional datasets collected by various organisations in Chapter 6. In Chapter 7, I summarise a selection of key findings within the chapters whilst suggesting novel areas for further research and potential applications of the work.
A conceptual risk framework for modelling and predicting wildlife-vehicle collisions
Abstract

Collisions of vehicles with wildlife kill and injure animals, and are also a risk to vehicle occupants, but preventing these collisions is challenging. Surveys to identify problem areas are expensive and logistically difficult. Computer modelling has identified correlates of collisions yet these can be difficult for managers to interpret in a way that will help them reduce collision risk.

We introduce a novel method to predict collision risk by modelling hazard (presence and movement of vehicles) and exposure (animal presence) across geographic space. To estimate the hazard, we predict relative traffic volume and speed along road segments across south-east Australia using regression models based on human demographic variables. We model exposure by predicting suitable habitat for our case study species (Eastern Grey Kangaroo *Macropus giganteus*) based on existing fauna survey records and geographic and climatic variables. Records of reported kangaroo-vehicle collisions are used to investigate how these factors collectively contribute to collision risk.

The species occurrence (exposure) model generated plausible predictions across the study area, reducing the null deviance by 30.4%. The vehicle (hazard) models explained 54.7% variance in the traffic volume data and 58.7% in the traffic speed data. Using these as predictors of collision risk explained 23.7% of the deviance in incidence of collisions. Discrimination ability of the model was good when predicting to an independent dataset.

The research demonstrates that collision risks can be modelled across geographic space with a conceptual analytical framework using existing sources of data, reducing the need for expensive or time-consuming field data collection. The framework is novel because it disentangles natural and anthropogenic effects on the likelihood of wildlife-vehicle collisions by representing hazard and exposure with separate, tunable sub-models.
2.1 Introduction

Roads have well-documented negative ecological impacts (Forman & Alexander, 1998; Spellerberg, 1998; van der Ree, Smith & Grilo, 2015) including effects on terrestrial fauna. Road construction and use fragments and destroys habitat, causes pollution (e.g. noise, light and chemical run-off), and kills and injures animals. Perhaps the most visible impact is direct mortality through wildlife-vehicle collisions (WVC) – billions of vertebrate fauna are killed annually around the world (Seiler & Helldin, 2006). Such an issue has prompted many road management authorities to routinely collect animal carcasses struck and killed by moving vehicles to reduce visual impacts for road travellers (Huijser et al., 2007a) and avoid secondary collisions with scavenging wildlife species. In addition, many governments around the world incur significant costs installing wildlife-proof fencing and under- and over-passes to reduce the rate of wildlife-vehicle collisions and improve landscape connectivity (van der Ree, Smith & Grilo, 2015).

Worldwide, the frequency, magnitude and distribution of WVC have been widely studied. Many such studies relate rate of collisions to environmental conditions, anthropogenic variables, and animal biology, behaviour and characteristics. While many studies have used statistical modelling to determine hotspots for WVC, most are limited to a single stretch of road or small collections of roads (e.g. Gundersen & Andreassen, 1998; Clevenger et al., 2002; Clevenger, Chruszczy & Gunson, 2003; Ramp et al., 2005; Ramp & Ben-Ami, 2006; Gomes et al., 2008; Langen, Ogden & Schwarting, 2009; Hurley, Rapaport & Johnson, 2009; Roger & Ramp, 2009; Hothorn, Brandl & Muller, 2012; Markolt et al., 2012; Santos et al., 2013; Seo et al., 2015). These models perform relatively well at local scales, and some have been employed to communicate areas of high-risk to road managers, but many cannot extrapolate to other sections of road or entire networks. We extend this work by developing and testing a framework that may be applied at much larger scales; consistent with the boundaries of road authority jurisdictions (i.e. state/provincial). Clevenger et al. (2015) asserts that the variability of significant predictors among regions and geographic scales highlights a critical need for a useful broad-scale conceptual framework to analyse/predict WVC.

Managers often have limited time and budgets to survey WVC across large areas or road networks. Many existing studies utilise data collected at fine spatial scales to model WVC, which is only feasible for limited areas because it is prohibitively expensive to collect data at the required regional scale. Methods to incorporate publicly-accessible, existing sources of data or model data-deficient parameters are useful to reduce costs. For example, remotely-sensed data is useful to determine environmental influences, while GIS-based census data may be used to characterise road conditions. Our framework is a desktop-based exercise for managers to determine risk across a large network with the ability to fine-tune based on available data. Moreover, it can be adapted with additional or modified data without changing the underlying methodology.

When modelling is used to determine hotspots for WVC, predictions are made using single models that combine both environmental and anthropogenic variables
A conceptual risk framework for modeling and predicting wildlife-vehicle collisions (e.g. Malo, Suárez & Díez, 2004; Ramp et al., 2005; Gomes et al., 2008; Roger & Ramp, 2009; Hothorn, Brandl & Müller, 2012; Barthelmess, 2014; Meisingset et al., 2014; Snow, Williams & Porter, 2014). These studies are valuable for managers to identify locations for mitigation or further study, however, they are often limited in their ability to extrapolate beyond the study area and do not clearly indicate potential confounding effects or suggest what mitigation to use (e.g. on the road environment or on the species). For example, vegetation on the road verge is commonly used to model collisions, and has been shown to be an effective predictor in previous work. But as a single covariate among others in a regression model, it is difficult to determine if vegetation is related to collisions based on its affect on visibility for drivers, or its attraction for animals. We extend the utility of this research by disentangling the effects of human activity and wildlife behaviour, which has not been achieved before. We aim to improve the accessibility of collision analysis and predictions by relating risk to two components, exposure and hazard. This hierarchical classification and structure of the predictors enables a more straightforward calibration and interpretation of the models – a useful feature for managers – without compromising predictive performance.

2.2 Materials and Methods

2.2.1 Conceptual Model Framework

Risk (as a rate) of animal collisions can be expressed as a function of exposure and hazard:

\[ R_i = a E_i H_i \]  

(2.1)

where \( R_i \) is the risk, \( E_i \) is the exposure, \( H_i \) is the hazard, \( a \) is a constant of proportionality and \( i \) represents a modelling unit (e.g. site, road, road segment). This equation is multiplicative and to enable linear analysis, we logarithmically transform the variables:

\[ \log(R_i) = \log(a) + \log(E_i) + \log(H_i) \]  

(2.2)

This suggests risk is perfectly related to both exposure and hazard and we allow modified responses by expressing the constant of proportionality as an intercept and introducing regression coefficients such that risk is modelled as:

\[ \log(R_i) = \beta_0 + \beta_1 \log(E_i) + \beta_2 \log(H_i) \]  

(2.3)

–or equivalently as–

\[ R_i = e^{\beta_0 + \beta_1 \log(E_i) + \beta_2 \log(H_i)} \]  

(2.4)

Here, \( \beta_1 \) and \( \beta_2 \) indicate the relative influence of each predictor on risk. This makes inferences on the model fit more tractable and clearly identifies areas for management focus (e.g. the exposure or the hazard). If risk is exactly related to exposure and hazard as in equation 2.2, then both regression coefficients \( \beta_1 \) and \( \beta_2 \) will equal one.
For this study, we represent exposure with animal presence, and hazard with both traffic volume and speed as neither one in isolation would be a realistic threat. We envisage risk as being measured by the rate of collisions. Using this configuration, the rate of collisions ($C_i$) is modelled as a function of three predictors of (1) species occurrence, (2) traffic volume and (3) traffic speed:

$$C_i = e^{\beta_0 + \beta_1 \log(O_i) + \beta_2 \log(V_i) + \beta_3 \log(S_i)}$$

(2.5)

where $O_i$ is species occurrence, $V_i$ is traffic volume, $S_i$ is traffic speed, in a given place $i$.

While we consider risk as being measured by the rate of collisions (e.g. collisions per month), our data consist of binary events (observations of individual collisions and no-collisions, coded as ones or zeroes, respectively). If collisions are treated as events in a Poisson encounter model, then the probability of no collisions occurring equals $e^{-C_i}$. Let $Y_i = 1$ indicate a collision occurred at site $i$, and $Y_i = 0$ indicate no collision, with $p_i = \Pr(Y_i = 1)$. Since the probability of no-collision is also equal to one minus the probability of collision:

$$\log(1 - p_i) = -e^{\beta_0 + \beta_1 \log(O_i) + \beta_2 \log(V_i) + \beta_3 \log(S_i)}$$

(2.6)

–or equivalently as–

$$\log(-\log(1 - p_i)) = \beta_0 + \beta_1 \log(O_i) + \beta_2 \log(V_i) + \beta_3 \log(S_i)$$

(2.7)

The left-hand side of equation 2.7 is the complementary log-log link function, which has similar properties to the more common logit function of logistic regression. However, we use the complementary log-log function here, because it relates more clearly to the rate of collisions (2.5), and to our conceptual risk model (2.1). Thus:

$$\text{cloglog}(p_i) = \beta_0 + \beta_1 \log(O_i) + \beta_2 \log(V_i) + \beta_3 \log(S_i)$$

(2.8)

Information on species occurrence, traffic volume and traffic speed are unlikely to be available for every place $i$ and we propose that this can be modelled. Several transportation modelling methods exist for estimating traffic patterns: generalised linear modelling (e.g. Seaver, Chatterjee & Seaver, 2000; Zhao & Chung, 2001), neural network analysis (e.g. Duddu & Pulugurtha, 2013), empirical Bayes estimation (e.g. Yang & Davis, 2002), universal kriging (e.g. Eom et al., 2006; Selby & Kockelman, 2013), and support vector machines (e.g. Castro-Neto et al., 2009). Modelling approaches to estimate species occurrence include correlative (e.g. Guisan & Thuiller, 2005; Elith & Leathwick, 2009) and mechanistic (e.g. Kearney & Porter, 2009) species distribution modelling (SDM) and extensions of population viability analysis (PVA) (e.g. Gilpin & Soule, 1986; Shaffer, 1990).

We apply our framework to study vehicle collisions with kangaroos in Australia using species distribution modelling to estimate kangaroo occurrence and linear regression to estimate traffic volume and speed (Figure 2.1).
2.2.2 Study Area & Species

We selected the State of Victoria in south-east Australia as a study area as its diversity across its area of 227,819 square kilometres provides a good platform to illustrate the framework. Our study combines all sealed roads within the state (≈150,000 kilometres) and predicts collision risk across six motorway class types. To organise our spatial data, we overlaid a spatial grid of one square kilometre resolution on the study area. To produce modelling units for the collision model, we further segmented the roads by intersecting all roads and the spatial grid. The open-source software package R (R Core Team, 2016) was used to perform all spatial and statistical analyses.

We used the native species Eastern Grey Kangaroo (*Macropus giganteus*, Shaw, hereafter referred to as “grey kangaroo”) as the case study species (see Appendix B for more information). In Victoria, the grey kangaroo is the most abundant of the macropod family and frequently involved in wildlife-vehicle collisions. Between 2005 and 2013, over 600 incidents were reported to the Victorian Police and documented in the VicRoads Crashstats database (see Appendix C). Actual incident rates, however, are much higher as the largest Victorian wildlife organisation, Wildlife Victoria (see Appendix C), received over 5000 reports of grey kangaroo-vehicle collisions over the
2.2 Materials and Methods

same period. Grey kangaroos are the second largest native terrestrial mammal in Australia and share many similar characteristics and management issues with ungulates found in North America and Europe (Croft, 2004; Coulson & Eldridge, 2010). Moreover, there is a research gap involving collision modelling of single macropod species in Australia (Bond & Jones, 2014).

2.2.3 Species Occurrence Sub-Model

We downloaded survey records of grey kangaroos from the Victorian Biodiversity Atlas (VBA, see Appendix C) occurring in the period 2000–2014, and spatial accuracy within 500 metres. The presences were both systematic targeted surveys and incidental sightings from data maintained by the Arthur Rylah Institute, a division of the Victorian Department of Environment, Land, Water and Planning. Kangaroos are generalists and widely distributed; it was therefore assumed that although abundance may fluctuate based on drought conditions, distribution would not change significantly from the year 2000. We used data from this period because it improved the sample size for occurrence modelling. We derived background data by randomly sampling 10,000 points across the entire study area. After sampling values from the covariate rasters and eliminating null values, the final dataset included 901 species presence and 9957 background observations.

To estimate occurrence of grey kangaroos, we used Boosted Regression Trees (BRT, see Friedman, 2002). BRT modelling offers advantages of handling different types of predictor variables, accommodating missing data and outliers, fitting complex non-linear relationships, and incorporating interaction effects between predictors (Elith, Leathwick & Hastie, 2008). We selected a tree complexity of five (limit on number of terminal nodes per tree used to include potential interactions) and a learning rate of .005 (contribution of each tree to the model). Classification methods have an established use in studies of species distributions (see Walker, 1990; Skidmore, Gauld & Walker, 1996), however, our framework is not limited to any particular modelling method. We selected seven predictors (Table 2.1) based on the biology and behaviour of grey kangaroos. All of the species occurrence sub-model predictor variables were below a pairwise correlation threshold of 0.75 to reduce potential effects of multicollinearity. We predicted relative likelihood of grey kangaroo occurrence from the model fit at a one square kilometre resolution across Victoria (Figure 2.2).
A CONCEPTUAL RISK FRAMEWORK FOR MODELLING AND PREDICTING WILDLIFE-VEHICLE COLLISIONS

Table 2.1: Variables used in statistical models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species Occurrence</td>
<td>KANG</td>
<td>presences and pseudo-absences of Eastern Grey Kangaroos</td>
</tr>
<tr>
<td></td>
<td>ELEV</td>
<td>elevation of terrain in meters above sea level</td>
</tr>
<tr>
<td></td>
<td>GREEN</td>
<td>remote-sensed mean seasonal change in greenness (2003–2013) in vegetation</td>
</tr>
<tr>
<td></td>
<td>LIGHT</td>
<td>remote-sensed relative artificial light intensity</td>
</tr>
<tr>
<td></td>
<td>MNTEMPWQ</td>
<td>mean temperature of wettest quarter in °C</td>
</tr>
<tr>
<td></td>
<td>PRECDM</td>
<td>precipitation of driest month in millimetres</td>
</tr>
<tr>
<td></td>
<td>SLOPE</td>
<td>slope of terrain in decimal percent rise</td>
</tr>
<tr>
<td></td>
<td>TREEDENS</td>
<td>tree canopy coverage within 1 square kilometre in decimal percentage</td>
</tr>
<tr>
<td>Traffic Volume</td>
<td>AADT</td>
<td>average annual daily traffic counts per road segment</td>
</tr>
<tr>
<td></td>
<td>KMTODEV</td>
<td>distance in kilometres to commercial planning zones</td>
</tr>
<tr>
<td></td>
<td>KMTOWHY</td>
<td>distance in kilometres to major road segments (freeways and highways)</td>
</tr>
<tr>
<td></td>
<td>POPDENS</td>
<td>2011 population divided by area in square kilometres</td>
</tr>
<tr>
<td></td>
<td>RDCLASS</td>
<td>road class category (‘freeway’, ‘highway’, ‘arterial’, ‘sub-arterial’, ‘collector’ or ‘local’)</td>
</tr>
<tr>
<td></td>
<td>RDDENS</td>
<td>total lengths in kilometres of road segments within 1 square kilometre</td>
</tr>
<tr>
<td>Traffic Speed</td>
<td>SPEEDLMT</td>
<td>posted speed limit per road segment</td>
</tr>
<tr>
<td></td>
<td>RDCLASS</td>
<td>road class category (see above)</td>
</tr>
<tr>
<td></td>
<td>RDDENS</td>
<td>total lengths in kilometres of road segments within 1 square kilometre</td>
</tr>
<tr>
<td>Collision</td>
<td>COLL</td>
<td>presences and pseudo-absences of grey kangaroo-vehicle collisions</td>
</tr>
<tr>
<td></td>
<td>EGK</td>
<td>predicted relative likelihood of kangaroo presence</td>
</tr>
<tr>
<td></td>
<td>TVOL</td>
<td>predicted traffic volume (number of vehicles per day) per road segment</td>
</tr>
<tr>
<td></td>
<td>TSPD</td>
<td>predicted posted traffic speed (kilometres per hour) per road segment</td>
</tr>
</tbody>
</table>

Figure 2.2: Predicted relative likelihood of grey kangaroo presence in study area. Darker shades indicate higher relative probabilities of occurrence (mean: 0.057; range: 0.002–0.986).
2.2 Materials and Methods

2.2.4 Traffic Volume & Speed Sub-Models

Average Annual Daily Traffic (AADT) represents the sum of traffic travelling in both directions which pass a roadside observation point during a full year divided by 365 days for a given road segment. AADT volume is usually only available for major road segments and we did not have data for most local, collector and sub-arterial roads under municipal district control. We predicted volume estimates for all road segments in the study area with Random Forest regression (see Breiman, 2001). The dependent variable was 2013 AADT recorded by VicRoads on 3174 road segments. We included five predictor variables (Table 2.1) that related to processes in traditional four-step traffic demand modelling (trip generation, trip distribution, mode choice and route assignment). All of the traffic model predictor variables were below a pairwise correlation threshold of 0.7 to reduce potential effects of multicollinearity. The traffic volume sub-model used the log-link function on the dependent variable (Table 2.2) due to the approximate log-normal distribution of AADT. We predicted AADT to all road segments using the model fit (Figure 2.3).

Figure 2.3: Predicted relative traffic volume in number of vehicles day\(^{-1}\) per road segment in study area. Darker shades indicate higher predicted traffic volumes (mean: 4481; range: 274–60850).

Traffic speed was modelled and predicted using a similar methodology to the traffic volume model. As with the traffic volume data, we did not have access to municipal records, however, we required speed values for all road segments across Victoria. We obtained posted speed limit data, for the year 2014, for all major road segments (n=42,439) and used road density and road class predictors (Table 2.1) in a linear regression model (Table 2.2) to predict speed for all road segments (Figure 2.4).
A conceptual risk framework for modelling and predicting wildlif...h hour$^{-1}$ per road segment in study area. Darker shades indicate higher predicted traffic speeds (mean: 62; range: 42–106).

### 2.2.5 Collision Modelling

To produce a kangaroo-vehicle collision dataset, we obtained records from the Wildlife Victoria database for a three year period between 1 January, 2010 and 1 January, 2013. The data for collisions were selected from a period where consistent techniques were used to collect and record the data. Pre-2010 records exist, however, are sparse and more prone to error. We first verified all automatically geocoded coordinates for accuracy using an online latitude/longitude mapping system (www.gpsvisualizer.com) and excluded all records with a geographical accuracy exceeding 300 metres. We selected road segments that were closest to the collision records in space and coded them with ones. To produce background points, we randomly selected approximately twice the number of collision-coded road segments and coded them with zeros. We combined the collision and background segments to produce a final dataset of 2264 collision and 4489 background segments. Using the same methodology, we developed an additional dataset of road segments for a period of one year between 1 January, 2013 and 1 January, 2014 for model validation (2125 collision and 4212 background records). Each segment contained predicted values for both traffic speed and volume from the previously described sub-models. As species occurrence predictions were expressed across a one square kilometre raster grid, we used the mid-point of each road segment for sampling the species occurrence sub-model predictions.

Our collision model fitted predicted values from the sub-models to collision or background occurrences at each road segment (Table 2.2). We used the fitted collision model to predict relative probabilities of collision on all road segments in the study area.
and, by using symbol classification (colour and line thickness) in QGIS (www.qgis.org), produced a map (Figure 2.5).

![Map of collision risk per road segment. Darker shades indicate higher relative risk of collisions with kangaroos (mean: 0.24; range: 0.01–0.99).](image)

To determine the effectiveness of partitioning the variables into logical sub-models, we also modelled collision risk with a combined set of all variables (Table 2.2). That is, we analysed a logistic regression model that related collisions directly to the original variables used to model kangaroo occurrence, traffic volume and traffic speed. This emulated past work for purposes of comparison. We also compiled a list of variables used in seventy-one collision modelling papers (see Appendix D) and verified that all of our predictors (with the exception of artificial light) were used in at least three or more published studies. The remaining highly-used variables were either irrelevant to our scope or difficult to obtain.

### 2.3 Results

The species occurrence model predicted the relative likelihood of kangaroo occurrence to vary (.002 to .986) across the study area using 5400 regression trees. The deviance explained by the model was approximately 30.4% (null deviance = 0.572, estimated cross-validation (CV) deviance = 0.398 ±0.008, CV AUC = 0.878 ±0.006; see Table 2.2). The three most influential variables were artificial light (19.6% contribution to model), elevation (18.4%), and precipitation of the driest month (14.8%). Partial dependence plots demonstrated plausible relationships between the predictors and grey kangaroo presence (Figure 2.6). We extended our model predictions to continental Australia (including areas well beyond Victoria) which aligned well with known grey kangaroo range (Figure 2.7). Approximately fifty percent of the grey kangaroo
A conceptual risk framework for modelling and predicting wildlife-vehicle collisions were within areas where the predicted probability of grey kangaroo occurrence was above 0.2. Collision records in areas of lower predicted occurrence areas may be due to misclassification errors (i.e. misidentification of the species of kangaroo), reporting bias/spatial error in the collision data, or sampling bias in the data used to train the kangaroo occurrence model.

![Figure 2.6: Effects of three most influential predictors on relative likelihood of grey kangaroo occurrence. Grey kangaroo occurrence is expressed on a relative probability scale. Artificial Light (LIGHT) is a surrogate for urban development where higher positive values represent intensely urbanised and populated areas. Elevation (ELEV) is measured in metres above sea level. Precipitation of the driest month (PRECDM) is the mean amount of rainfall in the summer expressed in millimetres.](image)

The traffic model volume explained 54.4% of the variation in the AADT data and the traffic speed model explained 58.7% of the variation in the posted speed limit data (Table 2.2). Each model used the default value of 500 trees to fit the data. All predictor variables used in the traffic models demonstrated plausible relationships to both AADT and speed (Figure 2.8). Traffic volume showed a generally decreasing trend with increasing distance to development and major thoroughfares and a generally increasing trend with increasing population and road density. Traffic speed showed a generally decreasing trend with increasing road density. Both traffic speed and volume decreased as road class changed from ‘freeway’ to ‘local’. The most influential variable for traffic volume was road class (33.8% relative contribution to reduction of variance), followed by distance to urban development (19.7%) and distance to freeways and highways (16.4%). For traffic speed, the most influential variable was road class, however, road density was nearly as influential.

The collision model explained 23.7% of the deviance (Table 2.2). Using the independent dataset to verify the predictive accuracy of the model resulted in an ROC score of 0.81. All three variables were highly significant (p < .0001) with traffic speed and grey kangaroo occurrence contributing the most to overall reduction in deviance; 27.3% and 72.7%, respectively (Table 2.3). The Akaike Information Criterion (AIC)
2.3 Results

Table 2.2: Statistical models used in the conceptual framework. Deviance explained is the percentage of variance in the data explained by the model. The receiver operator characteristic (ROC) is the ability of the model to correctly identify collisions in independent datasets (1 being perfect, and 0.5 no better than random).

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Model</th>
<th>Deviance Explained</th>
<th>ROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species Occurrence</td>
<td>Pr(KANG = 1) ≈ logit⁻¹(β₀ + β₁ELEV + β₂GREEN + β₃LIGHT + β₄MNTEMPWQ + β₅PRECDM + β₆SLOPE + β₇TREEDENS)</td>
<td>30.4%</td>
<td>0.88</td>
</tr>
<tr>
<td>Traffic Volume</td>
<td>log(AADT) ≈ β₀ + β₁KMTODEV + β₂KMTOWHY + β₃POPdens + β₄RDDENS + β₅RDCLASS</td>
<td>54.4%</td>
<td>—</td>
</tr>
<tr>
<td>Traffic Speed</td>
<td>SPEEDLMT ≈ β₀ + β₁RDCLASS + β₂RDDENS</td>
<td>58.7%</td>
<td>—</td>
</tr>
<tr>
<td>Collision</td>
<td>cloglog(Pr(COLL = 1)) ≈ β₀ + β₁ln(EGK) + β₂ln(TVOL) + β₃ln(TSPD)</td>
<td>23.7%</td>
<td>0.81</td>
</tr>
<tr>
<td>Alternative Collision</td>
<td>cloglog(Pr(COLL = 1)) ≈ β₀ + β₁ELEV + β₂GREEN + β₃KMTODEV + β₄KMTOWHY + β₅LIGHT + β₆MNTEMPWQ + β₇POPdens + β₈PRECDM + β₉RDCLASS + β₁₀RDDENS + β₁₁SLOPE + β₁₂TREEDENS</td>
<td>24.9%</td>
<td>0.84</td>
</tr>
</tbody>
</table>

score – measuring the quality of the model given the data and the parameters – was 6,579. All predictor variables demonstrated logical relationships to collision likelihood in the partial dependency plots (Figure 2.9). The rapid ascent and gradual levelling off of collision risk to increasing traffic volume suggests a threshold of around 2,000 vehicles day⁻¹.

Our alternative collision model (all sub-model predictor variables combined in a single model) explained 24.9% of the deviance; less than 2% more than the four-model framework. The model AIC was 6,496.2. The most influential variables were road class (42.1% relative contribution to reduction of deviance), distance to urban development (20.8%), elevation (17.3%) and population density (9.2%). We used the same independent dataset to verify the predictive accuracy of the model, resulting in an ROC score of 0.84.
Figure 2.7: Predicted relative likelihood of grey kangaroo occurrence across Australia. Darker shades indicate higher relative likelihood of occurrence (mean: 0.023; range: 0.001–0.986). The dashed line represents the existing known range of grey kangaroos as reported by the International Union for the Conservation of Nature. Victoria is located in the south-east corner of the mainland.
2.3 Results

Figure 2.8: Effects of predictor variables on traffic volume and speed. AADT is the predicted mean average annual daily traffic in vehicles day\(^{-1}\). SPEED is the predicted traffic speed in km hour\(^{-1}\). POPDENS is population density in persons square kilometre\(^{-1}\). RDDENS is road density per road segment in kilometres square kilometre\(^{-1}\). KMTODEV is road segment distance to urban land use in kilometres. KMTOHWY is road segment distance to freeway/highway in kilometres. RDCLASS is designated road segment functional classification.
Table 2.3: Summary of collision and alternative collision model fits. Coefficients and significance of variables are shown with relative contribution to model fit. Highly significant variables are marked with an asterisk. Contribution of variables to reduction in deviance (via analysis of variance) are expressed as decimal percentages (ANOVA).

| Model Type       | Variable      | Coefficient | Std. Error | Wald-$\chi^2$ | Pr(>|Z|) | ANOVA     |
|------------------|---------------|-------------|------------|---------------|----------|-----------|
| Collision        | Intercept     | -12.82      | 0.6635     | -19.33        | <2e$^{-16}$ | —         |
|                  | EGK           | 0.6583      | 0.0206     | 32.01         | <2e$^{-16}$ | 0.7268    |
|                  | TVOL          | 0.2715      | 0.0252     | 10.77         | <2e$^{-16}$ | 0.0005    |
|                  | TSPD          | 2.694       | 0.1308     | 20.59         | <2e$^{-16}$ | 0.2726    |
| Alternative Collision | Intercept  | -2.567      | 0.2642     | -9.716        | <2e$^{-16}$ | —         |
|                  | ELEV          | 0.003177    | 0.0002233  | 14.23         | <2e$^{-16}$ | 0.1729    |
|                  | GREEN         | 1.405       | 0.344      | 4.085         | 4.42e$^{-5}$ | 0.0011    |
|                  | KMTODEV       | -0.02784    | 0.002097   | -13.28        | <2e$^{-16}$ | 0.2079    |
|                  | KMTOHWY       | 0.006001    | 0.004206   | 1.427         | 0.1537    | 0.0004    |
|                  | LIGHT         | 0.004142    | 0.002046   | 2.025         | 0.043     | 0.0119    |
|                  | MNTEMPWQ      | 0.1519      | 0.01673    | 9.077         | <2e$^{-16}$ | 0.0398    |
|                  | POPDENS       | -0.0006986  | 0.0000535  | -13.06        | <2e$^{-16}$ | 0.0922    |
|                  | PRECDM        | 0.02573     | 0.003404   | 7.559         | 4.05e$^{-14}$ | 0.0483    |
|                  | RDDCLASS      | -0.4166     | 0.01467    | -28.4         | <2e$^{-16}$ | 0.4205    |
|                  | RDDENS        | 0.02353     | 0.008762   | 2.686         | 0.0072    | 0.0038    |
|                  | SLOPE         | 0.009252    | 0.00739    | 1.117         | 0.2641    | 0.0004    |
|                  | TREEDENS      | -0.1761     | 0.1387     | -1.27         | 0.2041    | 0.0008    |

Figure 2.9: Effects of predictor variables on relative likelihood of collision. EGK is the relative likelihood of kangaroo occurrence. TVOL is the predicted daily traffic volume in vehicles day$^{-1}$. TSPD is the predicted traffic speed in kilometres hour$^{-1}$. 
2.4 Discussion

This research introduces a new conceptual framework for predicting risk of wildlife-vehicle collisions. It is distinct from other collision models in its treatment of the analytical modelling framework. Where other research has related multiple environmental and anthropogenic variables to collision occurrence in a single statistical model (e.g. Lee et al., 2004; Ramp et al., 2005; Klöcker, Croft & Ramp, 2006; Litvaitis & Tash, 2008; Gunson et al., 2009; Roger & Ramp, 2009) and predicted risk from model fits (e.g. Malo, Suárez & Díez, 2004; Sudharsan, Riley & Campa, 2009; Gunson, Ireland & Schueler, 2012), we separate and develop the predictors in sub-models that would most suitably inform management. Each sub-model may be independently scrutinised for bias, uncertainty, and spatial autocorrelation and tuned accordingly. In this case study, our approach identifies relationships between, and effects of, species presence and traffic volume/speed on collision risk to grey kangaroos. In a similar manner, Bauduin et al. (2013) developed an index of co-occurrence to assess collision risk between manatees and recreational watercraft, however, the management implications were not extensively discussed. Particular to our study, road managers and environmental managers may be interested in whether to reduce collisions by focusing on the road environment (e.g. Clevenger et al., 2002; Jaeger & Fahrig, 2004; Bond & Jones, 2008, 2013), traffic conditions, or species (e.g. Huijser & McGowen, 2003; Huijser et al., 2006). Both the collision and alternative collision models had similar fits and made similar predictions. However, we argue that cause and effect is easier to interpret when using our proposed framework. For example, the variable of population density (human) is shown to contribute to collisions, however, it is unclear to what degree it is correlated with grey kangaroo occurrence, traffic conditions, or both, in the alternative collision model.

It should be noted that whilst the model framework allows clearer inferences on major factors contributing to collision risk, managers are still required to determine if mitigating these factors are contrary to competing objectives. As an example, consider adjusting posted speed limits to affect collision risk. Although the model predictions demonstrate reduced risk at lower traffic speeds, this strategy may lead to “speed dispersion” – larger variation in vehicular movement which increases unsafe overtaking and causes additional safety risks to humans. Moreover, efficient transportation is also a goal of road managers and reducing speed may conflict with this objective. Nonetheless, the negative effects associated with these management strategies can be quantified and included in cost-benefit analyses, along with other effects, to assist decision-making about existing and future road planning. Knowledge about the effects of traffic speed and volume on wildlife-vehicle collision risk may assist planners to design roads in regional areas that are treated differently within and outside of wildlife habitat yet maintain overall network efficiencies.

The models used in the study were all correlative and assumed static equilibrium in the environment. Temporal patterns of wildlife-vehicle collisions exist and would be useful to incorporate into the collision model, but the variable of time was not easily integrated into this study. The original collision dataset indicated that the lowest
number of incidents occurred in summer (December–February) and the highest were reported in August. The highest reporting times were consistent with the crepuscular (active between dusk and dawn) nature of kangaroo movements (McCullough & McCullough, 2000) and peaked at approximately 7:00 and 17:00 hours. Comparing the model performance on data accounting for time of day and of year is an area for future research. Other modelling methods that explicitly address interactions between space and time exist, such as multi-dimensional Poisson process models, and would be useful to incorporate into the framework.

This study demonstrates the usefulness of existing sources of data for scientific research and environmental management. Data provided by non-government organisations such as Wildlife Victoria, are not only inexpensive to collate, but are also valuable ecological indicators and sources of information, in particular, on species distributions. However, it is cautioned that use of such data should be subject to rigorous quality control and verification. A large amount of entered records were not useful for the scale of this study due to incomplete data or geographical ambiguity. Fauna atlases have the same potential issues. Many of the records in the Victorian Biodiversity Atlas were subject to geographical bias (e.g. close to roads and towns) and some potential inaccuracies (e.g. misidentification of species or approximate location). Graham et al. (2004) elaborates on the use of such data in scientific studies. Moreover, testing the model with alternative sources of collision reports (e.g. insurance records) would help account for biases present in the data. Although under-reporting of collision data has limited effects on model robustness, spatial biases can adversely affect model performance (Snow, Porter & Williams, 2015). The Wildlife Victoria dataset is an example of opportunistic, citizen-science data. Whilst this information is useful for modelling collision risk to large common species, it does not often provide adequate information for modelling rare or smaller species that are less reported. These data are more difficult to obtain and usually require targeted surveys – often labour-intensive searching on foot (Santos et al., 2016) – or novel survey techniques such as specimen or environmental DNA analyses (Klippel et al., 2015). Data collected through the use of these techniques may be used in combination with citizen-science data to correct for detection bias; methods have been demonstrated in species distribution modelling (Fithian et al., 2015). In our modelling framework, incorporating correction factors for biases (e.g. detection) is possible through the use of an offset parameter to account for survey effort. In this context, the offset term is used to standardise reporting effort by incorporating spatial (i.e. length of road segment) and temporal (i.e. reporting period) information.

Road managers are primarily concerned with human safety and thus large wildlife species may be a focus – as is the case of kangaroos in Victoria where animals cause direct damage or injury. But these concerns may extend to smaller species if drivers swerve to avoid collisions and strike other objects. Road managers are also motivated by preserving positive public opinion and dead animals on the road are unfavourable in any situation. Environment managers also respond to public opinion, and although mortality of common species due to road use may not constitute a significant threat to biological conservation, wildlife-vehicle collisions are still perceived as negative events.
Animal welfare is one such negative outcome as animals injured by strikes from moving vehicles often suffer slow and agonising deaths. Further, as ecological systems are complex, mortality caused by road vehicles may also have unanticipated effects on other species by disrupting food chains (e.g. secondary collisions with scavengers).

Our statistical modelling methods were chosen to match the data and purpose (see Wintle, Elith & Potts, 2005; Guiller-Arroita et al., 2015), and are based on well-established analyses in the modelling literature. However, our framework is not restricted to the particular models that we built; any appropriate statistical methods may be used for each of the sub-models. For example, boosted regression trees are a well developed method for species distribution modelling (Elith, Leathwick & Hastie, 2008) but may also be suited to predict traffic volume and speed. Further exploration of modelling methods, including those based on machine learning, may result in better calibrated models with less uncertainty and more robust inferences.

The use of mechanistic models to explain population dynamics (requiring collection of additional information such as age, sex, and size of species involved in wildlife-vehicle collisions) may be beneficial as the integration of Population Viability Analysis into species distribution and collision models can be informative (see Tyre, Possingham & Lindenmayer, 2001; Elith, Kearney & Phillips, 2010; Polak et al., 2014).

All of the models with binary dependent variables were subject to spatial autocorrelation producing potentially biased standard errors and predictions of grey kangaroo presence and collision risk. Techniques to incorporate autocovariates in the models similar to Crase, Liedloff & Wintle (2012) might improve the predictive performance but may only be logical for particular species (e.g. ranging or nomadic species). Further, we chose to test the model framework using a binomial dependent variable, however, it can be adapted to analyse other data types such as non-negative integers (e.g. number of collisions) or categorical information (e.g. groupings of number of collisions). These alternative methods might usefully identify hotspots (i.e. sites with multiple collisions) at smaller scales, such as road segments, or simulate effects of planned road expansion projects. For example, population effects on target species can be quantified and estimated based on predicting the magnitudes of collision hotspots.

2.4.1 Applied Management Implications

To realise the full potential use by managers, a model framework must be conceptually simple, flexible and adaptable. All of the input data must be accessible and the framework must allow inferences which are relevant and draw conclusions which are tractable. Our study uses vehicle collisions with grey kangaroos in Australia to demonstrate this analytic framework, however, we envision extending the model to (i) explore and predict risk to other species (e.g. wombats or deer) arising from vehicle collisions by identifying and quantifying probabilities of occurrence using alternative species distribution models and (ii) quantify risks to wildlife arising from other anthropogenic threats such as linear infrastructure (e.g. electrocutions), pollution (e.g. entanglements) or introduction of domestic predators (e.g. dog and cat attacks). Managers can use our framework across several disciplines, and it would
A conceptual risk framework for modelling and predicting wildlife-vehicle collisions benefit from additional testing across several spatial and temporal scales to determine its full potential flexibility and generality.
Consistent patterns of vehicle collision risk for six terrestrial mammal species
Abstract

The occurrence and rate of wildlife-vehicle collisions is related to both anthropocentric and environmental variables, however, few studies compare collision risks for multiple species within a model framework that is adaptable and transferable. Our research compares collision risk for multiple species across a large geographic area using a conceptually simple risk framework.

We used six species of native terrestrial mammal often involved with wildlife-vehicle collisions in south-east Australia. We related collisions reported to a wildlife organisation to the co-occurrence of each species and a threatening process (presence and movement of road vehicles). For each species, we constructed statistical models from wildlife atlas data to predict occurrence across geographic space. Traffic volume and speed on road segments (also modelled) characterised the magnitude of threatening processes.

The species occurrence models made plausible spatial predictions. Each model reduced the unexplained variation in patterns and distributions of species between 29.5% (swamp wallaby) and 34.3% (koala). The collision models reduced the unexplained variation in collision event data between 9.1% (swamp wallaby) and 19.4% (common ringtail possum), predictor variables correlating similarly with collision risk across species.

Road authorities and environmental managers need simple and flexible tools to inform projects. Our model framework is useful for directing mitigation efforts (e.g. on road effects or species presence), predicting risk across differing spatial and temporal scales and target species, inferring patterns of threat, and identifying areas warranting additional data collection, analysis, and study.
3.1 Introduction

Roads have considerable negative ecological impacts, prompting numerous scientific studies and mitigation projects over the past five decades (Forman & Alexander, 1998; Spellerberg, 1998; van der Ree, Smith & Grilo, 2015). The mortality of wildlife from collisions with vehicles is a significant problem throughout most of the developed world, and will likely become more so in the next few decades in the developing world (van der Ree, Smith & Grilo, 2015). Wildlife-vehicle collisions (WVC) are estimated to kill billions of vertebrate fauna annually on transportation networks (Seiler & Helldin, 2006), fostering an interdisciplinary management problem. To address this problem, studies seek to understand the frequency, magnitude and distribution of WVC. Through this understanding, managers can identify and apply appropriate mitigation strategies.

WVC are financially costly (Bissonette, Kassar & Cook, 2008; Huijser et al., 2009; Rowden, Steinhardt & Sheehan, 2008) and thus knowing where and how to mitigate is important. If mitigation measures are not appropriately specified or located in the landscape, costs arise from both wasted installation labour, materials and time, and on-going collisions resulting from the omission. Moreover, some forms of mitigation, such as fencing, also create “barrier-effects” – a direct impediment to the movement of many species which affects dispersal and gene flow (Epps et al., 2005). Thus, strategic, informed use of mitigation is important for both conservation and road safety objectives.

The literature contains many studies on WVC with several papers utilising statistical modelling methods. Quantitative models are a useful way to support decision-making for managers by helping to clearly organise problems, test inputs, and make inferences (Anderson et al., 2015). By determining the factors that contribute to collisions with quantitative analysis, managers can begin to understand relationships and optimise mitigation strategies.

We view two challenges in current WVC modelling and management practices. First, methods that generalise to multiple species and across taxonomic groups are under-represented in scientific studies (see Farmer & Brooks, 2012). This is likely due to the complexity in resource requirements for different taxa. Another challenge is identifying relationships between predictors and collisions that are able to be mitigated. Although management implications are highlighted in many studies, it is not always clear how the results of the analyses could be easily applied due to potential confounding effects in the model predictors (Gunson, Mountrakis & Quackenbush, 2011). Some studies explicitly suggest mitigation that relates directly to environmental variables based on model results (see Grilo, Bissonette & Santos-Reis, 2009). Although these recommendations are useful in the specific contexts of the studies, they may not generalise to varying spatial scales – specifically large areas. We argue that generality in methods is useful and it is important to create analytical tools that help managers identify risk to both individual species and taxonomic groups across large areas, with predictors that are manageable, or suggest areas for more specific examination.
In this study we have two objectives: testing a conceptual risk model framework and analysing patterns of WVC for multiple species. We first model wildlife-vehicle collisions by expanding upon the conceptual analytical framework in Chapter 2 that relates risk to exposure (species occurrence) and hazard (traffic volume and speed). As this model framework is able to quantify risk over large spatial scales (demonstrated on a large region of south-east Australia as a case study), we apply it to six Australian terrestrial mammal species commonly involved in WVC to test its flexibility and potential to support management decision-making. We analyse model outputs to determine the relationships of common factors contributing to wildlife-vehicle collisions and make predictions of risk to identify/prioritise areas requiring mitigation. As the need for, and location of, WVC mitigation is often derived using species characteristics and movements (e.g. Clevenger et al., 2002), we use the model framework to determine the importance of species occurrence in collision risks on road segments. Ignoring variables that influence habitat preferences of species may lead to less robust inference by managers (Roger & Ramp, 2009). Where occurrence is influential for multiple species, managers may decide to control animal presence on or near the roads and this may involve a mixture of mitigation strategies (see Beckmann et al., 2010). Likewise, if traffic speed or volume is a significant driver of collisions for many species, managers may consider control mechanisms such as speed enforcement or alternative transportation planning.

3.2 Materials & Methods

3.2.1 Study Area

We used the 227,819 square kilometres state of Victoria in south-east Australia as a study area (Figure 3.1). The Victorian road authority (VicRoads) manages 25,256 kilometres of major roadway within the state. The remaining sealed roads (≈ 125,000 kilometres) are controlled by seventy-nine municipal districts. Our study analysed collision risk across 147,970 kilometres of sealed roads. To organise our spatial data and modelling, we overlaid a spatial grid of 1-km² resolution (extents: −58000,5661000 x 764000,6224000, projection: GDA94 MGA zone 55, number of cells: 462,786) on the study area. Each grid cell was the modelling unit for species occurrence. All roads in the study area were bisected by the grid resulting in road segments that were approximately one kilometre or less in length. We used these 612,791 segments as our modelling units for the collision model.

3.2.2 Study Species

Wildlife Victoria is a not-for-profit organisation that maintains a professional database of state-wide reported wildlife injuries and includes spatial and temporal data for more than 80,000 records (see Appendix C for more information). We selected the species of mammal most frequently recorded as ‘hit by vehicle’ in the Wildlife Victoria database
(Table 3.1). Only mammals with at least 80 reported collision events were selected for use in the study, resulting in six study species.

Eastern grey kangaroos (*Macropus giganteus*, Shaw) are the second largest mammal (up to 85 kilograms for males) in Australia and share many management issues (e.g. over-abundance) with ungulates found in North America and Europe (Croft, 2004; Coulson & Eldridge, 2010). They occur in groups and have home range sizes between 25 and 125 hectares (Dawson, 2012). Swamp wallabies (*Wallabia bicolor*, Desmarest) are medium-sized (13 to 17 kilograms), solitary mammals more often found at higher elevations and in areas of denser foliage (Van Dyck & Strahan, 2008). Common wombats (*Vombatus ursinus*, Shaw) are medium-sized (22 to 39 kilograms), burrowing mammals and one of three wombat species in Australia (Van Dyck & Strahan, 2008). Although wombats occasionally share burrows, they are typically solitary animals and occupy home range sizes of about 20 hectares. Common possums, brushtail (*Trichosurus vulpecula*, Kerr) and ringtail (*Pseudocheirus peregrinus*, Boddaert), are small arboreal mammals (one to four kilograms) that are often more abundant throughout Victorian urban and suburban environments than other arboreal mammals. Koalas (*Phascolarctos cinereus*, Goldfuss) are medium-sized (eight to 15 kilograms) arboreal mammals. They are mainly sedentary due to their exclusive diet of Eucalyptus leaves which have low caloric content and nutritional value and are toxic to many other species (Van Dyck & Strahan, 2008). More information on these species can be found in Appendix B.
Consistent patterns of vehicle collision risk for six terrestrial mammal species

Table 3.1: Top six species most frequently reported in wildlife-vehicle collisions in the Wildlife Victoria database (WVD) between the years of 2010 and 2014. Species observations from the Victorian Biodiversity Atlas were used to train the species occurrence models and recorded collisions from the WVD for training the collision risk models. The datasets used in the models are comprised of both presence (P) and background (B) observations. Note, some species occurrence modelling datasets incorporate less than 10,000 background observations due to elimination of spatial duplicates (i.e., presence and background grid cells sharing the same coordinates were reduced to single presence observations).

<table>
<thead>
<tr>
<th>Species Name</th>
<th>Species Code</th>
<th>Species Occurrence Model Data</th>
<th>Collision Model Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern Grey Kangaroo</td>
<td>EGK</td>
<td>10508 : 703(P) / 9805(B)</td>
<td>612791 : 1344(P) / 611447(B)</td>
</tr>
<tr>
<td>Common Brushtail Possum</td>
<td>BTP</td>
<td>10808 : 1021(P) / 9787(B)</td>
<td>612791 : 257(P) / 612534(B)</td>
</tr>
<tr>
<td>Common Ringtail Possum</td>
<td>RTP</td>
<td>10590 : 791(P) / 9799(B)</td>
<td>612791 : 212(P) / 612579(B)</td>
</tr>
<tr>
<td>Swamp Wallaby</td>
<td>BSW</td>
<td>10548 : 755(P) / 9793(B)</td>
<td>612791 : 108(P) / 612683(B)</td>
</tr>
<tr>
<td>Common Wombat</td>
<td>WOM</td>
<td>10330 : 515(P) / 9815(B)</td>
<td>612791 : 156(P) / 612635(B)</td>
</tr>
<tr>
<td>Koala</td>
<td>KOA</td>
<td>10304 : 489(P) / 9815(B)</td>
<td>612791 : 81(P) / 612710(B)</td>
</tr>
</tbody>
</table>

3.2.3 Conceptual Model Framework

We employed the quantitative risk framework in Chapter 2 to examine how collision risk of each species relates to occurrence of the species, traffic volume, and traffic speed on all road segments. Using the open-source software package ‘R’ version 3.3.0 (R Core Team, 2016) to perform all statistical analyses, we developed species distribution models (SDM) to predict occurrence across the study area for each of the six species, and linear regression models to predict traffic volume and speed. Predicted traffic volume and speed values for all road segments were modelled by regressing annual average daily traffic (AADT) counts and posted speed limit data on anthropogenic variables using random forests; detailed methods are provided in Chapter 2.

3.2.4 Species Occurrence Models

We obtained observation records of the six study species from the Victorian Biodiversity Atlas (VBA, see Appendix C) that satisfied the following criteria: survey date between 1 January 2000 and 31 December 2013 and spatial coordinate certainty of ≤500 metres. As our occurrence models are correlative, we grouped the records based on identical spatial coordinates regardless of observation dates; thus multiple observations were aggregated to single presence observations in space. For each of the six study species, we selected all 1-km² grid cells that contained at least one occurrence record to represent presences across the study area. This ensured that we maintained at least 1,000 meters between all observations. As we did not have access to recorded absence data, we randomly selected 10,000 1-km² grid cells (without replacement) within the study area as background data and combined them with presence data.
3.2 Materials & Methods

(Table 3.1). Note, any spatially duplicated presence and background grid cells were reduced to single presence observations; therefore, some datasets may incorporate less than 10,000 background observations.

We selected predictor variables that influenced the biology, behaviour, and characteristics of all species based on literature review, consultation with species experts, and ecological principles. We considered a range of variables that describe three strata that our six species occupy; the subsurface zone for burrowing animals, the ground plane for ranging species, and the canopy for tree-dwelling wildlife. Each predictor was represented as a 1-km\(^2\) raster grid and values were sampled using the respective grid centroid coordinates in the presence/background datasets for each species. To reduce effects of collinearity, we used a predictor set that exhibited Pearson correlation coefficients less than 0.70 for all pairs of explanatory variables. There were two exceptions; elevation and slope (0.71), which were expected to exhibit some collinearity, and tree density and precipitation of the driest month (0.72).

For each species, we fit Boosted Regression Trees (BRT, see Friedman, 2002) models to the occurrence data using tools and methods by Elith, Leathwick & Hastie (2008). These methods included the use of internal cross-validation to reduce overfitting. BRT fit complex non-linear relationships and automatically incorporate interaction effects between predictors (Elith & Leathwick, 2009). We selected a tree complexity of five (limit on number of terminal nodes per tree used to regulate interactions), a learning rate of .005 (contribution of each tree to the model), and bag fraction of 0.5 (decimal percent of data values used to cross validate the model predictions). For each model, we employed a selection mechanism – based on Elith, Leathwick & Hastie (2008) – to determine the most useful and parsimonious combination of predictors (Table 3.3) from a full set of 18 variables (Table 3.2).

Table 3.3: Final predictors selected for each species in species occurrence models. The relative importance of each variable in the models is expressed as percentage contribution to reduction in unexplained variation in the data. ‘—’ indicates variable was not selected for use in the model and bold text signifies most influential variable per species.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>EGK</th>
<th>BTP</th>
<th>RTP</th>
<th>BSW</th>
<th>WOM</th>
<th>KOA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELEV</td>
<td>9.73</td>
<td>7.67</td>
<td>4.72</td>
<td>7.52</td>
<td>7.25</td>
<td>6.86</td>
</tr>
<tr>
<td>GRASS</td>
<td>8.33</td>
<td>6.64</td>
<td>5.44</td>
<td>6.6</td>
<td>6.32</td>
<td>10.34</td>
</tr>
<tr>
<td>GREEN</td>
<td>6.71</td>
<td>7.27</td>
<td>5.68</td>
<td>8.54</td>
<td>7.05</td>
<td>7.98</td>
</tr>
<tr>
<td>HYDRODIST</td>
<td>2.31</td>
<td>2.88</td>
<td>2.07</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>HYDROFLOW</td>
<td>—</td>
<td>2.56</td>
<td>—</td>
<td>2.6</td>
<td>3.22</td>
<td>—</td>
</tr>
<tr>
<td>ISOTHERM</td>
<td>2.49</td>
<td>2.82</td>
<td>3.09</td>
<td>2.7</td>
<td>2.29</td>
<td>—</td>
</tr>
<tr>
<td>LIGHT</td>
<td><strong>14.85</strong></td>
<td>10.19</td>
<td><strong>13.44</strong></td>
<td>7.58</td>
<td>8.18</td>
<td>5.37</td>
</tr>
<tr>
<td>MNTEMPWQ</td>
<td>3.57</td>
<td>5.44</td>
<td>6.32</td>
<td>4.42</td>
<td>4.61</td>
<td>4.54</td>
</tr>
<tr>
<td>PRECDM</td>
<td>7.13</td>
<td>7.49</td>
<td>7.98</td>
<td>4.41</td>
<td>14.73</td>
<td>7.06</td>
</tr>
<tr>
<td>PRECSEAS</td>
<td>3.41</td>
<td>5.27</td>
<td>4.23</td>
<td>4.45</td>
<td>4.75</td>
<td>8.48</td>
</tr>
<tr>
<td>ROADDIST</td>
<td>4.66</td>
<td>7.03</td>
<td>8.08</td>
<td>9.07</td>
<td>6.44</td>
<td>8.9</td>
</tr>
<tr>
<td>SLOPE</td>
<td>7.41</td>
<td>5.04</td>
<td>5.55</td>
<td>6.33</td>
<td>6.9</td>
<td>6.37</td>
</tr>
<tr>
<td>SOILBULK</td>
<td>7.94</td>
<td>5.18</td>
<td>4.98</td>
<td>4.49</td>
<td>6.87</td>
<td>5.53</td>
</tr>
<tr>
<td>SOILEXCH</td>
<td>2.28</td>
<td>2.93</td>
<td>3.43</td>
<td>3.24</td>
<td>3.95</td>
<td>—</td>
</tr>
<tr>
<td>SOILSAND</td>
<td>3.73</td>
<td>2.91</td>
<td>2.94</td>
<td>3.4</td>
<td>3.15</td>
<td>4.46</td>
</tr>
<tr>
<td>TEMPANRANGE</td>
<td>5.22</td>
<td><strong>12.15</strong></td>
<td>10.63</td>
<td><strong>11.5</strong></td>
<td>6.79</td>
<td><strong>12.94</strong></td>
</tr>
<tr>
<td>TOWNDIST</td>
<td>2.73</td>
<td>2.54</td>
<td>2.83</td>
<td>2.9</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>TREEDENS</td>
<td>7.51</td>
<td>6.53</td>
<td>8.88</td>
<td>10.35</td>
<td>4.58</td>
<td>11.18</td>
</tr>
</tbody>
</table>
We refit the models using the best combination of predictors and predicted relative likelihood of occurrence to 1-km$^2$ spatial grids for six species (Figure 3.2). We examined the rasters for unrealistic predictions of distributions by plotting atlas records of occurrences on prediction grids and also cross-referencing with expert-derived species range maps. Spatial autocorrelation was also assessed for each species by calculating Moran’s $I$ from model residuals and spatial coordinates and plotting against distance in one kilometre bins.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.2}
\caption{Predicted relative likelihood of occurrence of each species across the State of Victoria. Darker shading indicates higher relative likelihood of occurrence.}
\end{figure}


3.2 Materials & Methods

Table 3.2: Predictor variables used in species occurrence (SO) models and collision risk (CR) models. The spatial coordinates of centroids for grids with species presences and 10,000 randomly selected background grids were used to sample from 1-km$^2$ resolution predictor variable grids for occurrence models. The mid-points of road segments were used to sample from 1-km$^2$ resolution occurrence model predictions. Note, reported means and ranges are for entire study area. More information on the sources of these data can be found in Appendix C.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Units</th>
<th>Range</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELEV (SO)</td>
<td>Elevation of Terrain Above Sea Level</td>
<td>m</td>
<td>−0.39–1890.66 : 252.83</td>
<td></td>
</tr>
<tr>
<td>GRASS (SO)</td>
<td>Mean Annual Grass Growth</td>
<td>$kg ha^{-1}$</td>
<td>0–15.74 : 5.45</td>
<td></td>
</tr>
<tr>
<td>GREEN (SO)</td>
<td>Mean Seasonal Change in Vegetation Greenness</td>
<td></td>
<td>−62–68 : 0.21</td>
<td></td>
</tr>
<tr>
<td>HYDRODIST (SO)</td>
<td>Distance from Major Waterway</td>
<td>$km$</td>
<td>0–95 : 10.09</td>
<td></td>
</tr>
<tr>
<td>HYDROFLOW (SO)</td>
<td>Hydrological Flow Accumulation</td>
<td></td>
<td>1–32767 : 197.95</td>
<td></td>
</tr>
<tr>
<td>ISOTHERM (SO)</td>
<td>Isothermality</td>
<td>$^\circ C * 10$</td>
<td>0–53 : 48.40</td>
<td></td>
</tr>
<tr>
<td>LIGHT (SO)</td>
<td>Remote-sensed Artificial Light Intensity</td>
<td></td>
<td>0–63 : 1.23</td>
<td></td>
</tr>
<tr>
<td>MNTEMPWQ (SO)</td>
<td>Mean Temperature of Wettest Quarter</td>
<td>$^\circ C * 10$</td>
<td>−8–162 : 102.18</td>
<td></td>
</tr>
<tr>
<td>PRECDM (SO)</td>
<td>Precipitation of Driest Month</td>
<td>$mm$</td>
<td>0–81 : 34.70</td>
<td></td>
</tr>
<tr>
<td>PRECSEAS (SO)</td>
<td>Precipitation Seasonality</td>
<td>$mm$</td>
<td>0–43 : 23.99</td>
<td></td>
</tr>
<tr>
<td>ROADDIST (SO)</td>
<td>Distance from Roadway</td>
<td>$km$</td>
<td>0–36.55 : 2.02</td>
<td></td>
</tr>
<tr>
<td>SLOPE (SO)</td>
<td>Slope of Terrain</td>
<td>%</td>
<td>0–41.61 : 2.69</td>
<td></td>
</tr>
<tr>
<td>SOILBULK (SO)</td>
<td>Mean Soil Bulk Density at 2.5cm Depth</td>
<td>$kg m^{-3}$</td>
<td>0–1365 : 1129.13</td>
<td></td>
</tr>
<tr>
<td>SOILEXCH (SO)</td>
<td>Mean Soil Cation Exchange Capacity at 2.5cm Depth</td>
<td>$cmolckg^{-1}$</td>
<td>10–102 : 24.84</td>
<td></td>
</tr>
<tr>
<td>SOILSAND (SO)</td>
<td>Mean Soil Texture Fraction Sand at 2.5cm Depth</td>
<td>%</td>
<td>23–100 : 55.11</td>
<td></td>
</tr>
<tr>
<td>TEMPANRANGE (SO)</td>
<td>Temperature Annual Range</td>
<td>$^\circ C * 10$</td>
<td>0–289 : 245.10</td>
<td></td>
</tr>
<tr>
<td>TOWNDIST (SO)</td>
<td>Distance from Town Center</td>
<td>$km$</td>
<td>0–46.10 : 5.76</td>
<td></td>
</tr>
<tr>
<td>TREEDENS (SO)</td>
<td>Tree Cover</td>
<td>decimal %</td>
<td>0–1 : 0.28</td>
<td></td>
</tr>
<tr>
<td>BSW (CR)</td>
<td>Relative Probability of Wallaby Occ. per km$^2$</td>
<td>—</td>
<td>0.00–0.96 : 0.18</td>
<td></td>
</tr>
<tr>
<td>BTP (CR)</td>
<td>Relative Probability of Brushtail Possum Occ. per km$^2$</td>
<td>—</td>
<td>0.00–0.99 : 0.20</td>
<td></td>
</tr>
<tr>
<td>EGK (CR)</td>
<td>Relative Probability of Kangaroo Occ. per km$^2$</td>
<td>—</td>
<td>0.00–0.99 : 0.13</td>
<td></td>
</tr>
<tr>
<td>KOA (CR)</td>
<td>Relative Probability of Koala Occ. per km$^2$</td>
<td>—</td>
<td>0.00–0.97 : 0.12</td>
<td></td>
</tr>
<tr>
<td>RTP (CR)</td>
<td>Relative Probability of Ringtail Possum Occ. per km$^2$</td>
<td>—</td>
<td>0.00–0.98 : 0.16</td>
<td></td>
</tr>
<tr>
<td>TSPD (CR)</td>
<td>Predicted Vehicle Speed per Road Segment</td>
<td>$kmh^{-1}$</td>
<td>20.36–107.55 : 61.41</td>
<td></td>
</tr>
<tr>
<td>TVOL (CR)</td>
<td>Predicted Vehicle Volume per Road Segment</td>
<td>$vehicles day^{-1}$</td>
<td>0.29–47639.76 : 4309.27</td>
<td></td>
</tr>
<tr>
<td>WOM (CR)</td>
<td>Relative Probability of Wombat Occ. per km$^2$</td>
<td>—</td>
<td>0.00–0.97 : 0.12</td>
<td></td>
</tr>
</tbody>
</table>
3.2.5 Collision Model

We obtained spatially unique collision record coordinates from the Wildlife Victoria database for each species spanning a four year period between 1 January, 2010 and 31 December, 2013 and with a spatial accuracy <300 metres. For each target species, we selected the road segments that intersected with the reported location of each collision and re-coded them with ones. We selected the remaining road segments in the study area, coded them with zeros to represent background data, and combined them with the ‘collision’ records (Table 3.1). We sampled the relative species occurrence predictions at mid-points of the road segments in the collision dataset by intersecting the road features with the respective occurrence spatial grids. Traffic modelling in Chapter 2 had already predicted traffic speed and volumes for each road segment. This resulted in six modelling datasets (one for each species), each with a binary dependent variable of collision (1) or background (0) and three continuous predictors of species occurrence, traffic volume, and traffic speed.

We regressed collisions on the predictor variables using a complementary log-log (cloglog) link. This link considers the collision data as resulting from a truncated count process which agrees with both our underlying theory about how collisions occur (as a rate following a Poisson distribution) and our data construction methodology. Further, this link can be used to specify a model that is invariant to the spatial scale of the observations allowing greater flexibility as a modelling tool. The cloglog link contrasts with the popular logit link, commonly used for binomial data, which does not have the same interpretation. We included two variations on the original model presented in Chapter 2. First, we added an offset term to account for variation in road lengths and the temporal span of the data. We intend our model outputs to be suitable for management and the offset allows us to represent risk across the network using annual collisions per kilometre. Second, we included a quadratic term for traffic volume to allow a unimodal response shape. Former studies indicate that some species are repelled by roads when traffic volume increases beyond a threshold value (e.g. Seiler, 2005; Gagnon et al., 2007). Other possible determinants of this response shape include changes in local wildlife abundance resulting from previous wildlife-vehicle collisions, absence of wildlife in highly urbanised areas (i.e., heavy traffic areas), and reporting bias due to stopping difficulty on roads with high traffic volumes. We tested quadratic terms for the other predictors, however, they were not informative in this work.

Our final model is expressed as:

\[
\text{cloglog}(p_i) = \beta_0 + \beta_1 \log(O_i) + \beta_2 \log(V_i) + \beta_3 (\log(V_i))^2 + \beta_4 \log(S_i) + \log(L_i \times T) \tag{3.1}
\]

where \( p_i = \Pr(Y_i = 1) \) is the relative likelihood of a collision occurring, \( O_i \) is species occurrence, \( V_i \) is traffic volume, \( S_i \) is traffic speed, and \( L_i \) is the length of a given road segment \( i \). \( T \) is a constant representing the total time span of the data in years and is used in the road length offset to scale the model to an annual baseline.

Using the model fits, we predicted relative collision risk to all roads segments for each of the six study species. To validate the predictions of the model, we created
independent datasets for each species using the same methods as the modelling data but for collisions recorded in the year 2014. We calculated receiver operator characteristic (ROC) scores (Metz, 1978), also known as area under the curve (AUC), to assess the ability of the models to correctly predict collision events. Although other metrics are used to measure model discrimination (e.g. true skill statistic (TSS), see Allouche, Tsoar & Kadmon, 2006), they are often only suited to presence-absence data where prevalence is known. ROC uses non-specific threshold values and is an appropriate measure for assessing the discriminative ability of presence-only models (Lawson et al., 2014).

To determine collective risk on each road segment, we summed the total relative collision rates for all six species on each segment. These results were mapped to highlight road segments with generally high risk regardless of target species (see Figure 3.3). For each species, we also identified road segments that ranked in the top 0.1% for predicted collision rates within each respective range; collision rates varied by species due to differences in reported collisions. We recorded the frequency that individual road segments appeared in this category, the values ranging from zero (no species) to six (all species).

Figure 3.3: Road segments with summed collision rates for all six species. Darker, thicker lines indicate segments with higher predicted annual collisions per kilometre. The entire road network is shown as faint gray lines for context.

All of the modelling in our study was correlative, however, we explored the influence
of seasonal phenomena by partitioning the collision data into four seasons (summer, autumn, winter, spring) and running the models for each subset. Note, due to the small sample bias through partitioning, we fit the data using penalised likelihood methods (see Firth, 1993). We plotted the marginal effects of the predictors on collisions to determine if any season deviated from expected patterns.

### 3.3 Results

The species occurrence models produced ecologically-plausible predictions about the relative likelihood of species presence across Victoria. The deviance (unexplained variation in the data) reduced by the models ranged from 29.5% to 34.3% and the mean cross-validated ROC scores varied between 0.87 and 0.92 (Table 3.4). The model that best fit the data was for the koala.

Five variables were not selected in every species model and the least selected variables were distance to major waterway and hydrological flow accumulation (Table 3.3). The three most influential predictors were annual range of temperature, artificial light, and precipitation in the driest month (Figure 3.4). There was a generally decreasing trend in occurrence likelihood as annual temperature ranges increased for all species except kangaroos and wombats. Brushtail possums and swamp wallabies both had the largest spikes in relative occurrence at values of approximately 18 degrees Celsius. Relative likelihood of occurrence increased at higher values of artificial light for all species except koalas and wombats. At high values of precipitation in the driest month, occurrence of the two possums deviated considerably from the trends observed in all other species.

Spatial autocorrelation patterning was evident in the species occurrence model residuals between two and approximately twelve kilometres (Figure 3.5). Correlation values (Moran’s I) exceeded 0.4 at distances of two kilometres or less for all species except the ringtail possum. The highest overall spatial autocorrelation was in the koala model residuals, however, the wombat, koala, and swamp wallaby models all demonstrated high levels of spatial autocorrelation at one kilometre distances.

All of the collision models fitted the data as expected. Reduction on the null deviance was between 7.4% and 19.4% and all of the collision models attained high ROC scores when predictions were compared to the independent validation data (Table 3.4). The swamp wallaby model had the lowest ability to discriminate between false positives and false negatives (0.80) and the koala model had the highest (0.89). All of the predictors were highly significant ($p < 0.001$) in all of the models except traffic speed which was not a highly significant variable for the three arboreal mammal species.

All of the predictor variables demonstrated plausible relationships to collision likelihood in the partial dependency plots (Figures 3.6 to 3.8) and the signs of the coefficients were as expected (Table 3.5). Further, these response shapes were similar across all seasons for each species except the two possums (Figures 3.9 to 3.11). Based on analysis of variation (ANOVA), species occurrence had the largest contribution to reduction in deviance for all of the models.

Among species, the response of collision risk to occurrence likelihood demonstrated
3.3 Results

Table 3.4: Statistical model performance as percent reduction of deviance on the null model and receiver operator characteristic (ROC) scores for species occurrence (SO) and collision risk (CR) models by species. Note, due to the nature of gradient boosting and internal cross-validation, errors on both deviance and ROC scores are reported as ± values for species occurrence models.

<table>
<thead>
<tr>
<th>Species Name</th>
<th>% Deviance (SO)</th>
<th>ROC (SO)</th>
<th>% Deviance (CR)</th>
<th>ROC (CR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern Grey Kangaroo</td>
<td>30.7±0.7</td>
<td>0.89±0.004</td>
<td>11.8</td>
<td>0.80</td>
</tr>
<tr>
<td>Common Brushtail Possum</td>
<td>32.6±1.2</td>
<td>0.89±0.005</td>
<td>16.1</td>
<td>0.83</td>
</tr>
<tr>
<td>Common Ringtail Possum</td>
<td>33.7±1.1</td>
<td>0.90±0.005</td>
<td>19.4</td>
<td>0.87</td>
</tr>
<tr>
<td>Swamp Wallaby</td>
<td>29.5±1.5</td>
<td>0.87±0.007</td>
<td>9.1</td>
<td>0.80</td>
</tr>
<tr>
<td>Common Wombat</td>
<td>33.7±1.9</td>
<td>0.91±0.006</td>
<td>15.7</td>
<td>0.89</td>
</tr>
<tr>
<td>Koala</td>
<td>34.3±1.9</td>
<td>0.92±0.007</td>
<td>7.4</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Table 3.5: Summary of collision model fit for each species. Coefficients and significance of variables are shown with relative percent contribution (ANOVA) to model fits for each species. Highly significant variables (p < .001) are marked with asterisks.

| Species Name          | Variable | Coefficient | Std. Error | Z-value | Pr(>|Z|) | ANOVA |
|-----------------------|----------|-------------|------------|---------|---------|-------|
| Eastern Grey Kangaroo | Intercept| −46.55      | 1.961      | −23.74  | <.001*  | —     |
|                       | EGK      | 0.666       | 0.02413    | 27.6    | <.001*  | 69.53 |
|                       | TVOL     | 7.981       | 0.4631     | 17.23   | <.001*  | 7.179 |
|                       | TVOL\(^2\) | −0.4688     | 0.02899    | −16.17  | <.001*  | 11.63 |
|                       | TSPD     | 2.187       | 0.1261     | 17.35   | <.001*  | 11.66 |
| Common Brushtail Possum| Intercept| −57.19      | 9.55       | −5.989  | <.001*  | —     |
|                       | BTP      | 0.8599      | 0.08795    | 9.777   | <.001*  | 73.29 |
|                       | TVOL     | 11          | 2.006      | 5.483   | <.001*  | 19.16 |
|                       | TVOL\(^2\) | −0.589      | 0.115      | −5.121  | <.001*  | 7.518 |
|                       | TSPD     | 0.1785      | 0.3644     | 0.4898  | 0.620   | 0.031 |
| Common Ringtail Possum| Intercept| −146.7      | 25.95      | −5.651  | <.001*  | —     |
|                       | RTP      | 1.079       | 0.1027     | 10.51   | <.001*  | 74.44 |
|                       | TVOL     | 30.47       | 5.675      | 5.37    | <.001*  | 13.93 |
|                       | TVOL\(^2\) | −1.681      | 0.319      | −5.271  | <.001*  | 11.04 |
|                       | TSPD     | 0.9182      | 0.4093     | 2.244   | 0.025   | 0.590 |
| Swamp Wallaby         | Intercept| −64.59      | 8.024      | −8.049  | <.001*  | —     |
|                       | BSW      | 0.729       | 0.08703    | 8.377   | <.001*  | 60.03 |
|                       | TVOL     | 12.46       | 2.013      | 6.193   | <.001*  | 6.234 |
|                       | TVOL\(^2\) | −0.7598     | 0.128      | −5.938  | <.001*  | 27.77 |
|                       | TSPD     | 1.877       | 0.5203     | 3.607   | <.001*  | 5.966 |
| Common Wombat         | Intercept| −64.38      | 7.182      | −8.965  | <.001*  | —     |
|                       | WOM      | 0.8487      | 0.06144    | 13.81   | <.001*  | 75.19 |
|                       | TVOL     | 12.01       | 1.763      | 6.815   | <.001*  | 5.406 |
|                       | TVOL\(^2\) | −0.7186     | 0.1107     | −6.491  | <.001*  | 13.50 |
|                       | TSPD     | 2.302       | 0.4135     | 5.567   | <.001*  | 5.903 |
| Koala                 | Intercept| −60.98      | 9.01       | −6.768  | <.001*  | —     |
|                       | KOA      | 0.5093      | 0.07744    | 6.576   | <.001*  | 53.63 |
|                       | TVOL     | 11.75       | 2.311      | 5.085   | <.001*  | 10.28 |
|                       | TVOL\(^2\) | −0.7203     | 0.149      | −4.834  | <.001*  | 30.73 |
|                       | TSPD     | 1.663       | 0.6261     | 2.656   | 0.008   | 5.366 |

the least amount of variation in shape (Figure 3.6) compared to traffic volume (Figure 3.7) and traffic speed (Figure 3.8). Koalas had the highest increases in collision risk at lower values of occurrence likelihood and gradually reduced risk at higher
Consistent patterns of vehicle collision risk for six terrestrial mammal species

Figure 3.4: Effects of three most significant predictor variables on relative likelihood of occurrence per species.

values. This response was consistent for all species except ringtail possums which had an exponential relationship, albeit minor (coefficient of 1.08), between risk and occurrence.

All of the models had highly significant coefficients for both linear and quadratic terms of traffic volume. All of the model fits demonstrated curvilinear responses to AADT with consistent peaks around 3,000–5,000 vehicles day$^{-1}$ for all species except the two possums which peaked between 9,000–12,000 vehicles day$^{-1}$. Between the two possums, collision risk peaked at higher values of AADT for the brushtail possum and did not increase or decrease as rapidly across all values of AADT. Koalas and swamp wallabies had very similar response shapes and also showed the quickest rise and fall of collision risk between 0–10,000 vehicles day$^{-1}$ compared to all other species.
Collision risk had an exponential relationship to traffic speed for all species except the two possums. For brushtail possums, collision risk increased rapidly for low values of traffic speed and increased slowly for values above 30 km hr$^{-1}$. It should be noted, however, that the predictor of traffic speed was not significant for brushtail possums and had less than 0.1% relative influence on the model fit. As with species occurrence, ringtail possums demonstrated a nearly linear relationship between collision risk and traffic speed. Wombats demonstrated the highest rate of increase in collision risk for increasing collision speed compared to all other species.

The aggregated predictions of collision likelihood were highest on several sections of the Hume Highway and portions of the Calder Freeway just north of the urban centre of Melbourne (Figure 3.3). There were 60 road segments that were identified with the highest 0.1% of collision risk for all six species and 3,678 road segments for at least one species.
Consistent patterns of vehicle collision risk for six terrestrial mammal species

(a) Eastern Grey Kangaroo
(b) Common Brushtail Possum
(c) Common Ringtail Possum
(d) Swamp Wallaby
(e) Common Wombat
(f) Koala

Figure 3.5: Spatial autocorrelation in occupancy models residuals for each species grouped by distance between observations. Moran’s I measures the amount of similarity between values at each spatial distance grouping - zero indicating no correlation and one indicating perfect correlation.
Figure 3.6: Marginal effect of species occurrence on relative likelihood of collision per species. An overall plot (below) shows relative collision rates for all species and six sub-panels (above) depict response shapes with rescaled collision risk (relative minimum to maximum) for each species. Species symbols are: 🦘—‘Eastern Grey Kangaroo’; 🐨—‘Common Brushtail Possum’; 🦥—‘Common Ringtail Possum’; 🦏—‘Swamp Wallaby’; 🦨—‘Wombat’; 🍀—‘Koala’.
Figure 3.7: Marginal effect of traffic volume on relative likelihood of collision per species. An overall plot (below) shows relative collision rates for all species and six sub-panels (above) depict response shapes with rescaled collision risk (relative minimum to maximum) for each species. Species symbols are: –‘Eastern Grey Kangaroo’; –‘Common Brushtail Possum’; –‘Common Ringtail Possum’; –‘Swamp Wallaby’; –‘Wombat’; –‘Koala’.
3.3 Results

Figure 3.8: Marginal effect of traffic speed on relative likelihood of collision per species. An overall plot (below) shows relative collision rates for all species and six sub-panels (above) depict response shapes with rescaled collision risk (relative minimum to maximum) for each species. Species symbols are: 

- ‘Eastern Grey Kangaroo’;
- ‘Common Brushtail Possum’;
- ‘Common Ringtail Possum’;
- ‘Swamp Wallaby’;
- ‘Wombat’;
- ‘Koala’.
Consistent patterns of vehicle collision risk for six terrestrial mammal species

Figure 3.9: Marginal effect of species occurrence on relative likelihood of collision for six mammal species by season.
3.3 Results

(a) Eastern Grey Kangaroo
(b) Common Brushtail Possum
(c) Common Ringtail Possum
(d) Swamp Wallaby
(e) Common Wombat
(f) Koala

Figure 3.10: Marginal effect of traffic volume on relative likelihood of collision for six mammal species by season.
Consistent patterns of vehicle collision risk for six terrestrial mammal species

Figure 3.11: Marginal effect of traffic speed on relative likelihood of collision for six mammal species by season.
3.4 Discussion

Our study successfully applied a collision risk model framework to six different mammal species in south-east Australia and identified individual species’ and cumulative collision risk for all road segments. We acknowledge that our study is not the first to model collision risk for multiple taxa; others have studied mammals (e.g. Clevenger et al., 2002; Cserkesz et al., 2013; Jaarsma, van Langevelde & Botma, 2006), reptiles (e.g. Gunson, Ireland & Schueler, 2012; Langen et al., 2012) and mixed taxonomic groups (e.g. Clevenger et al., 2002; Garrah et al., 2015; Langen, Ogden & Schwarting, 2009; Litvaitis & Tash, 2008). And like ours, all these studies used mixtures of specific environmental and anthropogenic variables in multivariate models. What our research provides is a useful analytical framework for managers that generalises well and allows ongoing scrutiny and tuning of the working parts. For example, if species occurrence is a strong predictor of collisions but the performance of the SDMs are not satisfactory, managers may choose to further improve the sub-models or re-consider how much emphasis is placed on installing mitigation measures that influence animal occurrence. Similarly, spatial autocorrelation, detectability, bias, and other factors affecting model calibration and robustness may be assessed accordingly. Moreover, our methodology can be applied to single road segments or across an entire linear network, and on a single species, taxonomic class or mixed group.

The conceptual framework utilised predictors that connect to clear management objectives – species occurrence on road, traffic volume, and driver behaviour (i.e. traffic speed) – and are also consistent with theories of road ecology (see Forman et al., 2003). The models identified common trends in the relationships between the predictors and collision risk. Collision risk increased consistently with higher likelihoods of species occurrence suggesting the usefulness of separating out the exposure parameter via SDMs. Roger & Ramp (2009) demonstrated the importance of habitat variables (a surrogate for species occurrence) for modelling collisions. The work was extended to use an SDM for wombats (Roger, Bino & Ramp, 2012), indicating higher predicted likelihoods of species occurrence resulted in higher likelihoods of collisions.

Collision risks for ringtail possums increased more slowly with occurrence likelihood than for the other mammal species. This makes ecological sense considering the species spends most of the time elevated (mostly in highly connected urban development via power lines and street trees) and would come into contact with moving vehicles less frequently than ground-dwelling species. However, it is difficult to make direct comparisons between species due to uncertainties in the data. Potential reporting biases (e.g. not reporting smaller species) and errors (e.g. possum was assumed hit by car due to discovery on or near road verge) may result in artificially increased or decreased sample sizes. Of the 6,546 collision records from the Wildlife Victoria database used in this study, ringtail possums were the third most reported species (Table 3.1). Thus, possums could be much more abundant than wombats or wallabies. Grilo et al. (2014) explore collision risk among species based on differences in abundance and other traits that might pre-dispose species to collisions. Studies on reporting bias (e.g. Snow, Porter & Williams, 2015) and spatial error (e.g. Gunson
Consistent patterns of vehicle collision risk for six terrestrial mammal species

et al., 2009) in WVC also examine some of these issues in more detail.

Collision likelihoods also increased with higher traffic speeds for all species. With the exception of the brushtail and ringtail possums, the magnitudes of increase were large, suggesting high vehicle speed is a hazard for wildlife. One plausible explanation is that higher traffic speeds reduce reaction distances for both wildlife and drivers. For example, the distance covered is increased whilst processing information and responding accordingly during animal or vehicle movements. Although, particular combinations of traversing speeds may actually avoid collisions, stationary species on roads may be more likely to be affected by reduced driver reaction distances. Simulation analysis has been used to model the relationship of speed and collisions by using species traits to determine crossing success (Jaarsma, van Langevelde & Botma, 2006). Several other studies have shown similar results on the effects of traffic speed on road mortality of wildlife (e.g. Farmer & Brooks, 2012; Gkritza et al., 2013; Lao et al., 2011b; Ramp, Wilson & Croft, 2006; Seiler, 2005; Seiler & Heldin, 2006; Sudharsan, Riley & Campa, 2009; van Langevelde, van Dooremalen & Jaarsma, 2009).

The consistent quadratic shape observed in the responses to traffic volume for all species (Figure 3.7) indicates a possible threshold effect. Varying levels of traffic have been shown to affect animal presence on or near roads (Jaeger et al., 2005; Rhodes et al., 2014). Aside from species behavioural responses, traffic volumes above a certain level may affect species abundance due to mortality caused by wildlife-vehicle collisions. Lower abundances result in less animals on the road and, therefore, less collisions. Under this hypothesis, traffic volume is especially problematic for species that are attracted to roads (Forman et al., 2003). Nonetheless, identifying traffic volumes that have high relative collision risks for multiple species may help reduce roadkill. For example, high likelihood of collisions at particular traffic speed and volume levels may warrant further examination to determine potential mitigation such as redistribution or re-routing of vehicles and traffic-calming infrastructure.

As mentioned previously, our model framework is correlative and disregards temporal patterns of collisions. A qualitative comparison of the effects plots revealed similar trends between seasons for all species except ringtail and brushtail possums, however, the small sample sizes used in modelling warrant a cautionary interpretation. None of our species demonstrate migratory patterns and have relatively stable patterns of activity throughout the year. For species that exhibit seasonal patterns of movement, it would be useful to incorporate this information into the collision model (see Neumann et al., 2012). Partitioning the temporal component at the hourly scale may be more useful as systematic patterns of collisions exist throughout the day (e.g. Litvaitis & Tash, 2008; Rhodes et al., 2014). This was beyond the scope of this work as a robust temporal analysis would require the expansion of all explanatory variables to a complementary temporal scale, significantly increasing the computational requirements. Temporal differences in species movements and lags in collision reporting must also be considered.

The spatial autocorrelation in the occurrence models’ residuals was high at small distances and similar for all species. Any spatial correlation or patterning in the residuals violates statistical independence assumptions and several methods exist in
the SDM literature to deal with this issue (e.g. Augustin, Mugglestone & Buckland, 1996; Dormann et al., 2007, 2013). Patterning and higher values of spatial autocorrelation are plausible in certain ecological data and for some species. For example, correlations in model residuals based on the locations of koalas may be expected as they are selective in their choice of tree species, which are often found in homogeneous patches. However, high values of spatial autocorrelation or patterning should be carefully reviewed and appropriately addressed when used in statistical models (Wintle & Bardos, 2006). We included covariates to represent both sampling bias and the spatial autocorrelation arising from such in our models. ‘Distance to road’ was selected by all species models and ‘distance to town’ was selected by all models except brushtail possum and koala. Moreover, boosted regression trees are equipped to partially address spatial autocorrelation by reducing bias and variance (Elith, Leathwick & Hastie, 2008).

We did not find significant spatial autocorrelation or patterning in the collision model residuals and, thus, we did not choose to address it as part of the scope of this project. As mentioned previously, our model framework easily allows for this correction by the inclusion of auto-covariates in the model specification (e.g. Dwyer et al., 2016; Farmer & Brooks, 2012; Gomes et al., 2008), or by selecting alternative statistical modelling methods (e.g. Zhang, Gove & Heath, 2005).

3.4.1 Conclusions & Implications for Management

Our study analysed reported vehicle collisions for six terrestrial mammal species in Australia to test the generality of a model framework and to help infer cause and effect. Species occurrence modelling is useful to determine wildlife exposure to risk (on a relative scale). Species occurrence, traffic volume and traffic speed, as single modelled predictors, allow managers to improve models as new data becomes available. Managers may also predict effects by varying predictor values that reside in their authoritative control such as species occurrence or speed limit; here, the model framework functions as a decision planning tool. For example, the relationship between species occurrence and collision risk was similar for the six species in our study. Thus, managers may simulate changes in expected collisions resulting from fencing parts of the road network by reducing values of occurrence on road segments and predicting with the model. Further, the model predictions may also be spatially aggregated to suggest areas on the road network that require general mitigation or further study. In the case of Victoria, high-speed freeways and highways north of Melbourne had the highest overall collision risk. Our framework may also be extended to explore additional threats and predict risk to other taxonomic groups and vulnerable species (i.e., focus of conservation efforts) arising from the use of linear infrastructure.
A generalised predictive model for wildlife-vehicle collision risk with large mammals
Abstract

Existing wildlife-vehicle collision risk models do not generalise well to other geographic areas or species. Our study is the first to apply a generalised model framework to analyse and predict wildlife-vehicle collision risk in two continents. Despite being different species occurring in different geographic areas, we examined whether collision risk correlated similarly with animal behaviour (occurrence of the species) and human behaviour (traffic volume and speed).

Our case study species were eastern grey kangaroos in Victoria, south-east Australia and mule deer in California, USA. We trained our collision models with citizen-science collected, collision/carcass records for each species and predicted collision risk for all road segments in each study area. We used a very different set of data, police records of road accidents, to evaluate the predictions.

All predictor variables were highly significant in the collision models. Relative collision risk increased with increased species occurrence, traffic volume, and traffic speed. Speed had the highest relative importance followed by species occurrence. The shape of the responses to species occurrence and traffic characteristics were similar for both species, suggesting that responses to management will be similar across species despite having different biology and geography. For both species, collision model predictions using the validation datasets were highly correlated with independent observations, indicating well calibrated models.

Our model generalised well to species in different geographical areas. The modelling framework enables managers to conduct sensitivity analyses and calculate overall reductions in expected collisions based on applying mitigation strategies on different road segments. Our analysis also suggests appropriate mitigation for both species may include reduced speeds or fencing on road segments where species occurrence is predicted to be high.
4.1 Introduction

Wildlife-vehicle collisions (WVC) result from human development and activity. Approximately one-million vertebrates are killed per day in the United States (Forman & Alexander, 1998) and up to 27-millions birds are killed annually in select European countries (Erritzoe, Mazgajski & Rejt, 2003). Costs of WVC, including vehicle repair, human medical costs and the value of a human life, are estimated to exceed eight billion dollars annually in the United States (Huijser et al., 2007a) with deer collisions comprising the largest share (see Bissonette, Kassar & Cook, 2008). These costs and impacts will continue to worsen globally as new roads are constructed and existing roads are upgraded in developing countries; with worldwide traffic volume doubling by 2050 and increasing fivefold in developing countries (van der Ree, Smith & Grilo, 2015).

Mitigation strategies aim to reduce the rate and severity of WVC by influencing either human or animal behaviour (see Huijser & McGowen, 2010). Traffic planning (e.g. the magnitude and routing of vehicles), speed control (e.g. signage, traffic calming mechanisms and penalty enforcement), and education/training can influence the operation of vehicles on roads. Exclusion (e.g. fencing) and re-routing, either over or under the road, can influence the presence and movement of wildlife on the roads. Other strategies involve visual or auditory discouragement such as flashing lights and ultra-sonic whistles, however, these are largely ineffective (Reeve & Anderson, 1993; Bender, 2003; Scheifele, Browning & Collins-Scheifele, 2003; Ramp & Croft, 2006). Choices to implement mitigation strategies are influenced by an understanding of the major factors contributing to WVC in a particular area.

Decisions about where and when to mitigate are important because applying ineffective strategies incurs costs. To address this issue, many studies examine patterns of collisions and related variables to inform management, often involving spatial modelling and prediction (see Gunson, Mountrakis & Quackenbush, 2011). Most research is specific to an area (i.e. section of road) or problem and is not easily transferable due to variation in scale, geographic location and/or species traits. Clevenger et al. (2015) found that factors influencing deer-vehicle collisions were largely context dependent and highlighted a critical need for a conceptual framework in road ecology. Although many variables are difficult to generalise – partly due to the uncertainty present in ecological systems – predictors may be grouped into larger functional categories in analytical frameworks to assist managers to deal with uncertainty and make inferences regarding wildlife-vehicle collision risk (see Chapter 2).

We tested the general applicability of the conceptual model framework in Chapter 2 for predicting collision risk in California and Victoria. We had two aims; first we were interested in the applicability of our methods to planning for WVC mitigation irrespective of geographical location or scale (see van der Ree et al., 2011). We propose that our analytical methods will supplement management efforts by allowing clearer inferences about the factors related to collision risk, identifying potential biases and uncertainties in the analysis (thereby suggesting further research or data collection), and suggesting appropriate mitigation for WVC. Second, collisions of vehicles occur
at a similar frequency for both kangaroos and deer, have similar impacts (Langley & Mathison, 2008), and are subject to similar management practices (McShea & Underwood, 1997; Croft, 2004). So while models of collision risk should differ between species because of the different contexts, we expect that our framework should reveal similar patterns of collision risk for the two species as a function of traffic attributes and species occurrence. To our knowledge, this work represents the first study to test a broadly applicable model for two analogous species on different continents for the purpose of satisfying contrasting needs of management.

4.2 Materials & Methods

4.2.1 Overview of Workflow

Our analysis is comprised of the following overall workflow for each of the two study areas:

1. Define the study area and divide it into a regular spatial grid.
2. Split all roads in the study area into segments using the spatial grid.
3. Develop traffic models and predict relative traffic volume and speed for all road segments in the study area.
4. Develop a species distribution model and predict relative occurrence for each road segment based on the grid cell in which it occurs.
5. Classify road segments based on occurrence of collision (‘1’ for one or more collisions; ‘0’ otherwise) for the training data.
6. Develop a regression model of the collisions (training data) with species occurrence, traffic volume and traffic speed as explanatory variables.
7. Use the regression model to predict relative rate of collisions for each road segment.
8. Compare the predicted rate of collision to a new set of collision data (validation data) to evaluate the regression model.

Each of these steps is explained in more detail in the following sections.

4.2.2 Study Areas

We used two unique geographic locations to test our framework; the 227,819 km² State of Victoria in south-east Australia and twenty-nine counties comprising a 146,478 km² section of central California in North America (Figure 4.1). The case study regions were chosen because of availability of data on animal occurrence, collision records, and the road networks. Both California and Victoria had single agencies responsible for
4.2 Materials & Methods

Figure 4.1: Locations of wildlife-vehicle collisions for Eastern Grey kangaroos in Victoria (above) and Mule deer in central California (below). Insets (not to scale) show the state of Victoria in Australia and portion of the state of California; both are referenced with latitude and longitude coordinates. Sealed roads are shown as light gray lines and locations of reported collisions/carcasses are shown as black crosses.

the major roads (e.g. high-speed, high-capacity, or both) whilst all other roads were managed by municipal districts. We used all sealed roads within the study areas.
4.2.3 Data

To organise our spatial data and modelling, we overlaid spatial grids of 1-km² resolution on the study area. Each grid cell was used as the modelling unit for species occurrence. All roads in the study area were bisected by the grid resulting in road segments that were approximately one kilometre or less in length. We used 655,348 road segments for Victoria and 646,705 road segments for California as our modelling units for each respective collision model.

We selected one species from each study area that are frequently involved in WVC. Eastern grey kangaroos (*Macropus giganteus*, Shaw) are the second largest mammal in Australia – up to 85 kilograms for males (Van Dyck & Strahan, 2008) – and have similar management issues as ungulates found in North America and Europe (Croft, 2004; Coulson & Eldridge, 2010). Mule deer (*Odocoileus hemionus*, Rafinesque) are common across western North America and adults range in size up to 120 kilograms (Kays & Wilson, 2009). See Appendix B for more information.

To train our collision models, we obtained spatially unique collision/carcass records for each species from two citizen-science databases: the Wildlife Victoria database for kangaroos, and the California Roadkill Observation System (CROS, see Shilling & Waetjen, 2015) for deer. More information on these data sources are in Appendix C. The collision/carcass records spanned a six year period between 1 January, 2010 and 1 January, 2016 for kangaroos and a ten year period between 1 June, 2006 and 1 June, 2016 for deer. We limited our datasets to records with associated global positioning system (GPS) coordinates. Kangaroo records were initially inspected for spatial accuracy and we only retained observations with less than three-hundred metre error. Deer records from CROS have greater than ninety-nine percent species identification accuracy and less than one-hundred metre spatial error (Shilling and Waetjen, unpublished observations).

For each target species, we selected road segments that intersected with reported species’ collision records and coded them with ones. We coded all other road segments in each study area with zeros, to represent background data, and combined them with the collision record segments. After removing spatial duplicates, there were 4,245 presence and 640,470 background points in the kangaroo collision modelling dataset and 933 presence and 644,296 background points in the deer collision modelling dataset.

Traffic volume and speed values for all road segments in both Victoria and California were predicted following the methods of Chapter 2. We regressed annual average daily traffic (AADT) and speed on explanatory variables of distance to anthropogenic development (derived from remotely-sensed land use), distance to highway/freeway, road class, road density within a one kilometre radius of each road segment mid-point, and population density (from the Australian Bureau of Statistics and the United States Census Bureau, respectively) in random forest models (see Breiman, 2001) (see Table 4.1). We developed species distribution models to predict occurrence for each species across their respective geographic areas (Section 4.5) and used the mid-points of the road segments to sample each species occurrence prediction grids. This resulted in two modelling datasets, each with a binary dependent variable of collision (1) or
4.2 Materials & Methods

background (0) and three continuous predictors of species occurrence, traffic volume, and traffic speed.

Table 4.1: Performance of traffic models. We used Random Forest regression with 500 trees and two variables tried at each split. The traffic volume models used the variables Distance to Development, Distance to Highway, Population Density, Road Class and Road Density; the traffic speed models used Road Class and Road Density. $R^2$ reports the variation in the training data explained by the model. Training observations are the total data points used to train each model.

<table>
<thead>
<tr>
<th>Study Area</th>
<th>Predicted Variable</th>
<th>$R^2$</th>
<th>Training observations</th>
<th>Most influential variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Victoria</td>
<td>Traffic Volume</td>
<td>0.54</td>
<td>3,174</td>
<td>Road Class</td>
</tr>
<tr>
<td></td>
<td>Traffic Speed</td>
<td>0.59</td>
<td>42,438</td>
<td>Road Class</td>
</tr>
<tr>
<td>Central California</td>
<td>Traffic Volume</td>
<td>0.43</td>
<td>68,474</td>
<td>Distance to Highway</td>
</tr>
<tr>
<td></td>
<td>Traffic Speed</td>
<td>0.60</td>
<td>7,292</td>
<td>Road Class</td>
</tr>
</tbody>
</table>

We used police records of road accidents to evaluate our model fits. For each species, we extracted spatially-unique records where the incidents were collisions with our target species. For kangaroos, we used data reported between 1 January, 2010 and 1 October, 2016. For deer, the reporting period was 1 February 2015 to 1 December 2016. Again, we used all road segments as background and coded the segments with collisions as ones and the remainder with zeros. The final datasets had 481 presence and 644,234 background points for kangaroo collision model validation and 1,795 presence and 643,434 background points for deer. As with the training data, each road segment had values for species occurrence, traffic volume and traffic speed.

4.2.4 Collision Modelling & Validation

We used a quantitative risk model introduced in Chapter 2 to fit and compare the relationship of species presence and road threat to collision likelihood, and the open-source software package R version 3.3.0 (R Core Team, 2016) to perform all statistical analyses. We regressed collisions on the predictor variables with an added quadratic term for traffic volume and an offset term to scale risk. Our model is expressed as:

\[
cloglog(p_i) = \beta_0 + \beta_1 \log(O_i) + \beta_2 \log(V_i) + \beta_3 \log(S_i) + \log(L_i \times T) \tag{4.1}
\]

where $p_i = \Pr(Y_i = 1)$ is the relative likelihood of a collision occurring on a road segment $i$, $O_i$ is species occurrence, $V_i$ is traffic volume, $S_i$ is traffic speed, $L_i$ is an offset term for road segment length and $T$ is the time span of collision training data in years. We chose the complementary log-log link on the linear predictor due to the mathematical theory underpinning our model – risk being measured by the rate of collisions (see Chapter 2).

To test for spatial autocorrelation, we calculated normally-distributed, randomised quantile residuals (see Dunn & Smyth, 1996) using 5,000 simulations from each model fit. We then calculated Moran’s I on the residuals at 20 spatial lags at intervals of both one kilometre and 250 metres. For each spatial resolution, we repeated the
Moran’s I calculation twenty times for 5,000 randomly sampled road segments (due to computational limitations) and plotted the twenty trend lines for visual inspection at both spatial scales (Figure 4.2).

![Figure 4.2](image)

**Figure 4.2:** Spatial autocorrelation in randomised quantile model residuals for each species at two spatial lags (1-km and 250-m). In each plot, trend lines (20 total) are for randomly selected subsets of the data (5,000 observations per subset).

The two models were used to predict relative collision risk on all road segments in each respective study area (Figure 4.3). The collision risks were expressed as expected annual collisions per kilometre on each road segment. An analyst or manager may
use this information to determine: (a) relative risk on roads by summing the segment values and comparing to other problematic roads; (b) locations to prioritise mitigation; or (c) locations for additional investigation such as monitoring.

We tested the calibration strength and discrimination ability of our models with their respective independent validation datasets. To assess calibration strength, we regressed the collision observations in the validation sets on predictions made using the validation data in the trained models. An intercept coefficient of ‘zero’ and slope coefficient of ‘one’ indicates a perfectly calibrated model (see Miller, Hui & Tierney, 1991). To measure ability to discriminate between true positive and false positives we used a receiver operating characteristic (ROC) score – a score of one indicating perfect discrimination ability whilst 0.5 suggesting a performance no better than random (see Metz, 1978).
A generalised predictive model for wildlife-vehicle collision risk with large mammals

Figure 4.3: Predicted relative collision risk expressed as expected annual collisions per kilometre for all sealed roads in Victoria (above) and central California (below). Enlarged sections from each map show the predicted risk on road segments in the selected areas. Darker, heavier lines indicate higher predicted values (up to 0.14 annual collisions per kilometre in Victoria and 0.05 in California) whilst light gray lines indicate low values (less than .02 annual collisions per kilometre in Victoria and .01 in California).
4.3 Results

All of the predictor variables demonstrated plausible relationships to collision likelihood in the partial dependency plots. The probability of collision for both species changed similarly with occurrence of the species (Figures 4.4a and 4.4b), traffic volume (Figures 4.4c and 4.4d), and traffic speed (Figures 4.4e and 4.4f).

Traffic speed and species occurrence were highly significant variables for kangaroos and deer (Table 4.2). Speed had the highest relative importance for both kangaroos and deer based on an analysis of variance for each predictor. Increasing speed from 80 to 100 km hr$^{-1}$ approximately doubled the relative collision risk for both species. Collision risk with deer was less than kangaroos at lower speeds and increased at a faster rate at higher speeds. Species occurrence influenced collision risk the second most for both species. The response shape of collision risk against likelihood of occurrence was similar for both species, however, deer had larger confidence intervals around the marginal response (Figures 4.4a and 4.4b).

**Table 4.2:** Summary of collision models for kangaroos and deer. Highly significant variables ($p < .001$) are marked with asterisks.

<table>
<thead>
<tr>
<th>Species</th>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Z-value</th>
<th>Pr(Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern Grey Kangaroo</td>
<td>Intercept</td>
<td>$-42.30$</td>
<td>$1.07$</td>
<td>$-39.66$</td>
<td>$&lt;1.0x10^{-14}$*</td>
</tr>
<tr>
<td></td>
<td>EGK</td>
<td>$0.70$</td>
<td>$0.02$</td>
<td>$47.77$</td>
<td>$&lt;1.0x10^{-14}$*</td>
</tr>
<tr>
<td></td>
<td>TVOL</td>
<td>$5.55$</td>
<td>$0.25$</td>
<td>$22.68$</td>
<td>$&lt;1.0x10^{-14}$*</td>
</tr>
<tr>
<td></td>
<td>TVOL$^2$</td>
<td>$-0.33$</td>
<td>$0.02$</td>
<td>$-21.76$</td>
<td>$&lt;1.0x10^{-14}$*</td>
</tr>
<tr>
<td></td>
<td>TSPD</td>
<td>$3.93$</td>
<td>$0.07$</td>
<td>$54.49$</td>
<td>$&lt;1.0x10^{-14}$*</td>
</tr>
<tr>
<td>Mule Deer</td>
<td>Intercept</td>
<td>$-39.61$</td>
<td>$3.05$</td>
<td>$-12.98$</td>
<td>$&lt;1.0x10^{-14}$*</td>
</tr>
<tr>
<td></td>
<td>DEER</td>
<td>$0.61$</td>
<td>$0.02$</td>
<td>$29.12$</td>
<td>$&lt;1.0x10^{-14}$*</td>
</tr>
<tr>
<td></td>
<td>TVOL</td>
<td>$2.49$</td>
<td>$0.66$</td>
<td>$3.80$</td>
<td>$.0001*$</td>
</tr>
<tr>
<td></td>
<td>TVOL$^2$</td>
<td>$-0.14$</td>
<td>$0.04$</td>
<td>$-3.44$</td>
<td>$.0001*$</td>
</tr>
<tr>
<td></td>
<td>TSPD</td>
<td>$6.29$</td>
<td>$0.21$</td>
<td>$30.67$</td>
<td>$&lt;1.0x10^{-14}$*</td>
</tr>
</tbody>
</table>

Relative collision risk for kangaroos increased with traffic volume until approximately 5,000 vehicles day$^{-1}$ and then declined (Figure 4.4c); the relationship was similar for deer although the peak was closer to 10,000 vehicles day$^{-1}$ (Figure 4.4d). Collision likelihood of kangaroos decreased more quickly than deer at higher traffic volumes, but deer had larger confidence intervals around the marginal response curve indicating the possibility of a shape more consistent with kangaroos (Figure 4.4d). The actual magnitudes of relative risk are not directly comparable between species because sizes of the collision and background datasets were not standardised across the two species.

Spatial correlation in the randomised quantile residuals was low and showed no patterns across distances at the two spatial scales analysed (Figure 4.2), suggesting assumptions of statistical independence were reasonable. All of the predictor variables were highly significant and the signs of the coefficients indicated predictor-response relationships that were plausible (Table 4.2). The deer model reduced deviance compared to the null model by nearly 150% that of the kangaroo model.

For each species, the models had similarly high calibration statistics when independent data (police reports) were used for validation. The kangaroo model had a
A generalised predictive model for wildlife-vehicle collision risk with large mammals

A generalised predictive model for wildlife-vehicle collision risk with large mammals

Each model had a different intercept coefficient due to varying numbers of collisions used for model fitting. Both models discriminated well between true positives and false positives; the kangaroo model had an ROC value of 0.84, and the deer model had a value of 0.91.

**Table 4.3:** Statistical model performance metrics shown as percent reduction of deviance on the null model using training data and receiver operating characteristic (ROC) scores using independent validation data. Regression coefficients indicate model calibration when using the independent validation data; a coefficient of ‘one’ suggests a perfectly calibrated model. All regression coefficients were highly significant (p<.001).

<table>
<thead>
<tr>
<th></th>
<th>Eastern Grey Kangaroo</th>
<th>Mule Deer</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Reduction in Deviance</td>
<td>11.8</td>
<td>18.6</td>
</tr>
<tr>
<td>ROC Score</td>
<td>0.79</td>
<td>0.88</td>
</tr>
<tr>
<td>Regression Intercept</td>
<td>−2.22 (0.17 standard error)</td>
<td>0.35 (0.08 standard error)</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.99 (0.04 standard error)</td>
<td>0.94 (0.01 standard error)</td>
</tr>
</tbody>
</table>

The predictions for relative collision risk across all road segments were plausible in most regions of the two study areas (Figure 4.3). In Victoria, predicted collision risk was highest on segments along the Hume Highway to the north of Melbourne – where collisions frequently occur – and on segments along the Princes Highway to the east of Melbourne. In contrast, predictions in California were highest on road segments along Interstate 280 in the San Francisco Bay Area, where deer collisions are common, and along Highway 50 in the foothills of the Sierra Nevada mountain range. Table 4.4 lists the road segments with the highest predicted risk in both study areas.
### 4.3 Results

#### Table 4.4: Ten highest ranked road segments for predicted collision risk for kangaroos and deer. Road segments are listed by unique identification (UID), name, and northing and easting coordinates of the centroid (projected to GDA94 MGA zone 55 for Victoria and NAD83 UTM zone 10N for California). Collision risk is relative to all of the road segments in the respective study area and expressed as the expected annual number of collisions per kilometre on the segment.

<table>
<thead>
<tr>
<th>UID</th>
<th>Name</th>
<th>Collision Risk</th>
<th>X and Y Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>407754</td>
<td>Hume Highway</td>
<td>0.1444</td>
<td>403963.159042543, 5951629.83565613</td>
</tr>
<tr>
<td>407760</td>
<td>Hume Highway</td>
<td>0.1401</td>
<td>403071.710826330, 5951397.77864573</td>
</tr>
<tr>
<td>407762</td>
<td>Hume Highway</td>
<td>0.14</td>
<td>403925.196291583, 5951593.74234382</td>
</tr>
<tr>
<td>407755</td>
<td>Hume Highway</td>
<td>0.1343</td>
<td>404600.149299800, 5951610.35738630</td>
</tr>
<tr>
<td>621992</td>
<td>Hume Highway</td>
<td>0.1339</td>
<td>322625.570078198, 5856999.38576165</td>
</tr>
<tr>
<td>621994</td>
<td>Hume Highway</td>
<td>0.13</td>
<td>322662.033521619, 5856987.78800383</td>
</tr>
<tr>
<td>407747</td>
<td>Hume Highway</td>
<td>0.1299</td>
<td>406048.292349148, 5951020.16537145</td>
</tr>
<tr>
<td>404199</td>
<td>Hume Highway</td>
<td>0.1299</td>
<td>405074.058515895, 5951472.28462078</td>
</tr>
<tr>
<td>407749</td>
<td>Hume Highway</td>
<td>0.1299</td>
<td>404847.642617257, 5951565.85124194</td>
</tr>
<tr>
<td>407185</td>
<td>Hume Highway</td>
<td>0.128</td>
<td>407014.420745575, 5950554.12249437</td>
</tr>
<tr>
<td>46648</td>
<td>Highway 101</td>
<td>0.0512</td>
<td>545171.623088090, 4188210.43687450</td>
</tr>
<tr>
<td>46646</td>
<td>Redwood Highway</td>
<td>0.0505</td>
<td>545162.417260965, 4188160.2128528</td>
</tr>
<tr>
<td>46647</td>
<td>Highway 101</td>
<td>0.0485</td>
<td>545142.154540168, 4188117.62343993</td>
</tr>
<tr>
<td>108268</td>
<td>Interstate 280</td>
<td>0.0484</td>
<td>574988.367401353, 4137976.64649412</td>
</tr>
<tr>
<td>105697</td>
<td>Interstate 280</td>
<td>0.0466</td>
<td>573127.177281627, 4138775.43354061</td>
</tr>
<tr>
<td>46643</td>
<td>Highway 101</td>
<td>0.0465</td>
<td>545174.043806978, 4188058.63361548</td>
</tr>
<tr>
<td>100045</td>
<td>Frontal Rd</td>
<td>0.044</td>
<td>570614.122713676, 4191363.28459612</td>
</tr>
<tr>
<td>73842</td>
<td>Interstate 280</td>
<td>0.0435</td>
<td>562033.376408464, 4148111.35353173</td>
</tr>
<tr>
<td>100046</td>
<td>Frontal Rd</td>
<td>0.0435</td>
<td>570628.521461339, 4191337.49247383</td>
</tr>
<tr>
<td>101059</td>
<td>Interstate 280</td>
<td>0.0434</td>
<td>571986.800791011, 4139982.42749593</td>
</tr>
</tbody>
</table>
A generalised predictive model for wildlife-vehicle collision risk with large mammals

Figure 4.4: Effects of predictor variables on relative likelihood of collision per species. Each variable is expressed with all other variables set at mean values. Likelihood of collision is expressed as a rate across all road segments for the total period of the observation data. To convert relative collision rate to expected annual number of collisions, multiply rate by total road segments divided by years of data.
4.4 Discussion

Our results suggest that the proposed conceptual framework may have utility irrespective of locality, spatial scale, or species. Species occurrence, traffic volume and speed, and environmental data are publicly available for many jurisdictions, therefore, an analyst only requires data on vehicle collisions with wildlife species to train the models in the conceptual framework, make inferences and predict risk. Moreover, the collection of data on collisions with vehicles by road authorities (via carcass collection operations) or citizen scientists such as wildlife groups is increasing due to human safety or animal welfare concerns and technological innovations (see Olson et al., 2014; Shilling, Perkins & Collinson, 2015). Analysis from our framework may guide mitigation actions in several problematic areas to reduce collisions.

Forman et al. (2003) identifies both animal and human behaviour as major drivers for WVC. Our framework operates on this hypothesised relationship and indicates maximum relative risk where animal presence and vehicles moving at speed co-occur in space. Road authorities mitigate collision risk by managing human activity (traffic volume and speed) or animal behaviour (occurrence on or movement across the road) and these have different costs. Our results demonstrated that speed is an important predictor of collision risk for both kangaroos and deer and therefore mitigation that either addresses animal presence near high speed road segments (e.g. fencing and crossing structures) or traffic speeds (e.g. controls and enforcement) in collision hotspot areas should be considered. This pattern is consistent with other studies on deer collisions (e.g. Gkritza et al., 2013; Meisingset et al., 2014; Sudharsan, Riley & Campa, 2009), kangaroo collisions (e.g. Rowden, Steinhardt & Sheehan, 2008), and other taxa (e.g. Gunson, Mountrakis & Quackenbush, 2011). Moreover, the traffic models demonstrated plausible fits to the posted speed training data (Table 4.1) therefore suggesting greater reliability in their predictions. However, when assessing where to place mitigation for speed, analysts or managers should also consider the uncertainty around predictions at each road segment for both speed and animal presence. If there is higher confidence around speed predictions, for example, more emphasis (and funds) may be placed on controlling vehicle speeds, rather than excluding animals.

The conceptual framework with sub-models allows a more clear identification of bias and uncertainty in the analysis. For example, the two occurrence models are produced using presence-only data that explicitly assumes perfect detection and includes potential sampling bias. If these assumptions are deemed unsatisfactory, an analyst may choose to use alternative statistical methods that allow incorporation of uncertainty into the model (see Dorazio, 2014) or improve the training data (e.g. increase accuracy standards or collect additional records that include true absences). The deer occurrence predictions may also be less reliable than the kangaroo predictions due to a smaller number of training observations. Likewise, the lower statistical significance and wide confidence intervals in the marginal effects of traffic volume on collision likelihood for deer may warrant further analysis. The wide confidence intervals suggest that reducing vehicles on problematic road segments may not have the desired effect. Moreover, reported or observed collision data are subject to the same
limitations as species occurrence data (e.g. reporting biases or data deficiencies such as under-reporting of actual presences) and should be taken into account.

As expected, collision risk for kangaroos and deer were similar in response to species occurrence, traffic volume and traffic speed. Collision risk increased monotonically with increasing species occurrence for both kangaroos and deer, however, there was an order of magnitude greater change for kangaroos due to differences in ratios of collision to background observations. Assuming species occurrence is of equal importance as the other collision model predictors, the smaller increase in collision risk for deer may be attributed to lower predictions of species occurrence arising from a mis-specified species distribution model (see Guillera-Arroita et al., 2015). Collision risk response to traffic volume was uni-modal for both species, however, the change in collision risk was an order of magnitude greater per unit of traffic volume for kangaroos. This response shape can arise from different mechanisms. First, there may be a potential observer bias meaning people are less likely to stop or report collisions on high-volume roads out of concern for their own safety or simply not seeing the dead animal. Second, there may be road avoidance effects as some species are repelled by roads when traffic volume increases beyond a threshold value due to high noise loads or small crossing intervals (see van Langevelde & Jaarsma, 2004; Seiler, 2005; Gagnon et al., 2007). Our results are consistent with other studies on behavioural responses to roads that include kangaroos and deer (see Jacobson et al., 2016). At higher values of traffic speed, collision risk increased monotonically with increasing speed with a similar magnitude for both species (Figures 4.4e and 4.4f). Kangaroos were exposed to higher collision risk at lower traffic speeds than deer. This may be due to Melbourne’s suburbs expanding into natural kangaroo habitat, thereby increasing wildlife-vehicle collisions.

We did not explore temporal effects in this study, however, this may be incorporated into our framework. One such method is to add a function-based term to the model that allows collision risk to vary as a function of time and season (see Chapter 5). The functional form may be expressed in relation to the known activity periods of the target species. Kangaroos are crepuscular, with peaks of activity at dawn and dusk, and deer are either crepuscular or nocturnal depending on the season. Hour of day and time of year are useful predictors of collisions with ungulates (Meisingset et al., 2014; Mountrakis & Gunson, 2009), however, not yet fully tested for kangaroos.

The inferences made from the conceptual framework will be more robust as more observations are added to train the models. We currently use a mixture of professional and citizen-science collected data – each with unique advantages, disadvantages, and implications for analysis. Citizen-science data can supplement data-deficient ecological studies, and are effective in road ecology (Dwyer et al., 2016; Paul et al., 2014). Moreover, citizen-science data is relatively low-cost and can cover large spatial scales. For example, the collision records in both the Wildlife Victoria database and the California Roadkill Observation System are systematically collected and stored in databases and employ mechanisms to elicit data from unrestrained spatial distances. As innovative and standardised data collection techniques are implemented (see Aanensen et al., 2009; Donaldson & Lafon, 2010; Shilling, Perkins & Collinson, 2015), inferences and predictions from the model framework will improve.
4.4.1 Management Implications

Our analysis suggests appropriate mitigation to reduce wildlife-vehicle collisions for both species may include reducing speeds or fencing road segments where species occurrence is predicted to be high. Although omitted for this work, our methods enable analysts or managers to conduct sensitivity analyses and calculate overall reductions in expected collisions based on mixed applications of mitigation on different road segments (assuming the mitigation is 100% effective, but see Huijser et al., 2009). As costs are often known for specific mitigation measures, simulations can determine optimal uses of resources to maximise reductions of collisions and minimise costs. We are excited to apply these methods to other species and geographical areas.

4.5 Supplemental Information

4.5.1 Species Distribution Modelling

To establish occurrence probabilities for Eastern Grey kangaroos and Mule deer in their respective native geographic areas, we sourced presence records from two online, publicly-available databases; the Victorian Biodiversity Atlas and the Global Biodiversity Information Facility (see Appendix C for more information). We obtained presence records that satisfied the following criteria: survey date between 1 January 2000 and 1 December 2015 and spatial coordinate certainty of 500 metres or less. As our occurrence models are correlative, we grouped the records based on identical spatial coordinates regardless of observation dates; thus, multiple observations were aggregated to single presence observations in space. For each of the two study species, we selected all spatially unique records of occurrence. To reduce potential effects of spatial dependency, we thinned each species occurrence points to maintain a minimum separation distance of 1,000 metres between observations. As we did not have access to recorded absence data, we generated 10,000 randomly sampled points across each study area to represent background data. After omitting null values from sampling predictor variable grids, there were 700 presence and 9957 background points in the kangaroo occurrence modelling dataset and 366 presence and 9,986 background points in the deer occurrence modelling dataset.

We developed species distribution models using tools and methods by Elith, Leathwick & Hastie (2008). We chose Boosted Regression Trees (BRT, see Friedman, 2002) as the statistical method. BRT fit complex non-linear relationships and automatically incorporate interaction effects between predictors (Elith & Leathwick, 2009). We selected a tree complexity of five (limit on number of terminal nodes per tree used to regulate interactions), a learning rate of 0.005 (contribution of each tree to the model), and bag fraction of 0.5 (decimal percent of data values used to cross validate the model predictions). Using the model fits, we predicted relative likelihood of occurrence to all grid cells for each species.

We selected predictor variables that influenced the biology, behaviour, and characteristics of each species based on existing literature and ecological principles (e.g.
Coulson & Eldridge, 2010; Ferguson, 2005). Although each species is unique in its respective biology, we considered similar behaviours (sheltering and movements) and traits (foraging requirements) as the basis for choosing broad environmental variables that covered both species (Table 4.5).

Table 4.5: Predictor variables used in species occurrence models for kangaroos in Victoria and deer in California.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELEV</td>
<td>Elevation of terrain above sea level</td>
<td>m</td>
</tr>
<tr>
<td>GREEN</td>
<td>Mean seasonal change in vegetation greenness (relative)</td>
<td>—</td>
</tr>
<tr>
<td>LIGHT</td>
<td>Remote-sensed artificial light intensity (relative)</td>
<td>—</td>
</tr>
<tr>
<td>MNTMPWQ</td>
<td>Mean temperature of wettest quarter</td>
<td>°C + 10</td>
</tr>
<tr>
<td>PRECDM</td>
<td>Precipitation of driest month</td>
<td>mm</td>
</tr>
<tr>
<td>SLOPE</td>
<td>Slope of terrain</td>
<td>decimal %</td>
</tr>
<tr>
<td>TREEDENS</td>
<td>Tree cover</td>
<td>decimal %</td>
</tr>
<tr>
<td>X</td>
<td>X spatial coordinate of intersecting 1-km² grid centroid</td>
<td>m</td>
</tr>
<tr>
<td>Y</td>
<td>Y spatial coordinate of intersecting 1-km² grid centroid</td>
<td>m</td>
</tr>
</tbody>
</table>

As bioclimatic variables exhibit spatial gradients, we also included the spatial coordinates of grid cell centroids – X (Easting) and Y (Northing) – as predictor variables in the models. This reduces biases in the influence of variables with high spatial correlation. It should be noted that our framework allows any combinations of data deemed appropriate for species modelling, however, using the same environmental variables for both kangaroos and deer demonstrates the generality of our framework and allows a more direct comparison in the analysis.

The models explained 26.7% of the uncertainty in the recorded occurrences of kangaroos and 30.8% of the uncertainty in the recorded occurrences of deer (deviance). Both models had good discriminative ability; the internal cross-validated ROC scores were 0.88 for kangaroos and 0.91 for deer.

Although the mechanism of gradient boosting (BRT in this case) does not assume independence in the dependent variable, we reviewed spatial trends in the model fit for each species by calculating Moran’s I from model residuals and spatial coordinates and plotting against distance in one kilometre bins. Spatial patterning was evident in the species occurrence model residuals between one and nine kilometres for both deer and kangaroos (Figure 4.5) and both species showed similar trends. Kangaroos demonstrated higher values of spatial autocorrelation than deer at closer spatial distances. The spatial covariates consistently ranked high among the most influential variables in both occurrence models (ranked 2nd and 3rd for kangaroos and 1st and 4th for deer). The X and Y variables accounted for 27.8% of the influence in the kangaroo model and 33.3% in the deer model.

All of the environmental variables were influential for both species. The three most influential non-spatial predictors were slope (14.9% relative influence), artificial light (13.5%) and vegetation greenness (9.5%) for deer and artificial light (16.5%), elevation (12.5%) and vegetation greenness (10.5%) for kangaroos (Figure 4.6). Kangaroo occurrence showed a potential bi-modal response to artificial light with large peaks at both 25 and 55 units of relative light intensity – the scale is unitless and relative to a range between 0 (no light) and 63 (brightest spot detected across all artificial light...
Figure 4.5: Spatial autocorrelation in occupancy model residuals by distance grouping (spatial lag) for Eastern Grey kangaroos in Victoria (triangle) and Mule deer in central California (dot).

visible on earth). Deer occurrence did not have the same response shape, rather, there was a general increasing trend in occurrence probability with increasing light intensity and a sharp peak between 45 and 50 units. Kangaroo occurrence was higher at low levels of elevations (<500m) with a peak at approximately 200m. Deer occurrence increased with higher slope aspects.

The species occurrence models produced ecologically-reasonable predictions about the relative likelihood of species presence across each study area (Figure 4.7). Eastern Grey kangaroo occurrence was predicted to be lower in north-eastern Victoria which is consistent with historical knowledge. Predictions of deer occurrence were higher in areas of topographically-varied natural space (parks and undeveloped) around the San Francisco Bay Area which is also consistent with recorded observations of deer. However, lower relative occurrence was predicted in the eastern portion of California (Sierra Nevada foothills) where deer are also known to be present (Kucera, 1988).
Figure 4.6: Marginal effects of the three most significant predictor variables on the relative likelihood of occurrence for kangaroos and deer.
Figure 4.7: Predicted relative likelihood of occurrence for kangaroos in Victoria (above) and deer in central California (below). Darker shading indicates higher relative likelihood of occurrence. Study boundaries are shown as dashed lines. Victoria: extent (−58000E, 5661000N) x (764000E, 6224000N); projected to GDA94 MGA zone 55; 462,786 cells. California: extent (445000E, 3962000N) x (1165000E, 4329000N); projected to NAD83 UTM zone 10N; 264,240 cells.
Predicting wildlife-train collisions across space and time to inform railway operations
Abstract

Despite considerable research on collisions between wildlife and road vehicles, animal mortality resulting from train strikes has been rarely studied. In addition to animal welfare and conservation concerns, costs from train strikes may be considerable, and railway authorities have a vested interest to manage the problem. As railways expand spatially and operationally throughout the world, predicting and mitigating wildlife-train collisions will become increasingly important.

To assess the risk of wildlife-train collisions, we quantified regional train movements in space and time, and determined the likelihood of occurrence of Eastern Grey kangaroos in the State of Victoria, Australia. We then fitted a model to data on collisions between trains and kangaroos, accounting for time of day, train occurrence and speed, and kangaroo occurrence. We predicted collision rates on a passenger railway network based on three management scenarios that influenced train speed and occurrence of kangaroos near the railway lines.

The model fit and predictions were plausible. Train speed was the most influential variable followed by presence of kangaroos. Reducing speeds in areas of high predicted kangaroo occurrence, during periods of peak animal activity within a twenty-four hour cycle, reduced collision rate the most.

Predictions from the model can help managers decide where, when and how best to mitigate collisions between animals and trains. It can also be used to predict high-risk locations or times for (a) timetable/schedule changes (b) proposals for new railway lines or (c) disused lines considered for re-opening. The model framework is easily adaptable to other species and railway operations, and it allows managers to assess model performance and calibrate/update analyses accordingly.
5.1 Introduction

Roads and railways support human civilisations by facilitating economic and recreational activities. However, transportation networks may directly or indirectly disrupt ecological systems (Seiler & Helldin, 2006; van der Ree, Smith & Grilo, 2015) and their environmental impacts must be managed (Spellerberg, 1998). Animal mortality due to collision with vehicles is one of the most visible impacts (Forman et al., 2003) and a primary cause of reductions in population size (Fahrig & Rytwinski, 2009). Wildlife-vehicle collisions pose a serious global problem (Litvaitis & Tash, 2008), spawning a new discipline (road ecology) and inspiring research to develop solutions. For example, deer-vehicle collisions on roads are well-studied in North America (e.g. Huijser et al., 2007b; Romin & Bissonette, 1996) and Europe (e.g. Sáenz-de-Santamaría & Tellería, 2015; Seiler, 2004). Moreover, management of wildlife-vehicle collisions will become increasingly important as new transportation networks are constructed and existing networks are expanded, perhaps most significantly in developing countries.

In addition to concerns about animal welfare (Sainsbury, Bennett & Kirkwood, 1995) and conservation status of threatened species (Dwyer et al., 2016; Jones, 2000), collisions with large animals pose direct risks to the life of humans (Langley, Higgins & Herrin, 2006; Rowden, Steinhardt & Sheehan, 2008). For example, moose are one of the largest animals struck by vehicles in North America and Europe, causing significant damage and injuries (Huijser et al., 2007b; Hurley, Rapaport & Johnson, 2009). Vehicle collisions with deer, although smaller than moose, frequently kill humans in North America (Williams & Wells, 2005).

Information about the spatial and temporal distribution and magnitude of wildlife-vehicle collisions is useful to managers because it may help more effectively mitigate impacts (Mountrakis & Gunson, 2009). For example, knowing a collision hotspot along a transportation network for a particular species will inform the most appropriate choice of mitigation (e.g. animal exclusion or change in vehicle speed). Data can also inform statistical modelling which helps to predict the probability of wildlife-vehicle collisions (Gunson, Mountrakis & Quackenbush, 2011).

The majority of wildlife-vehicle collision modelling deals with roads and traffic (van der Ree, Smith & Grilo, 2015), yet, the problem extends to other forms of vehicular transportation such as air (e.g. Van Belle et al., 2007), railway (e.g. Wells et al., 1999) and shipping (e.g. Laist et al., 2001) operations. Regardless of the mode of transport, the modelling of collisions share some common attributes (see Forman et al., 2003). The movements or presence of animals are often considered in the models and may include behavioural traits (Roger & Ramp, 2009), and vehicle presence or movements can also be considered and may be grouped into a larger category of human behaviour as humans ultimately control speeds and trajectories of vehicles (Ramp & Roger, 2008).

Whilst the formulation of collision models are similar for both trains and road vehicles, wildlife-train collisions provide a different dataset for considering factors that influence collision risk that are often difficult to obtain. Firstly, the timetabling of
trains means that their movements are much more precisely documented than those of cars. Secondly, collisions with large animals are more consistently reported for trains than cars – although under-reporting is still an issue (Dorsey & Rew, 2015). And thirdly, the time of collision is also reported more precisely. These benefits mean that examining factors that influence wildlife-train collision risk may help uncover factors that influence wildlife-vehicle collisions more generally.

Extensive railway networks with considerable train movements exist on every inhabited continent in the world, and although broader ecological effects have been discussed (e.g. De Santo & Smith, 1993; Givoni, 2006) and analysed (e.g. Waller & Servheen, 2005), very few studies analyse wildlife-train collisions (but see Belant, 1995; Onoyama et al., 1998). Many anecdotal records/observations illustrate the severity of wildlife-train collisions such as group fatalities (e.g. herds of ungulates or flocks of birds) due to browsing on railways or train derailments from very large species (Dorsey & Rew, 2015). These serious events suggest a growing need for more research into predicting wildlife-train collisions; however, we are only aware of one published study (see Gundersen & Andreassen, 1998). Here, we develop a modelling framework to predict the rate of kangaroo collisions on the regional passenger railway network in south-east Australia. In addition to informing railway operators in south-east Australia of potential kangaroo collision risks, our approach can be generalised to other species (e.g. deer) and railway operations (e.g. freight transport) elsewhere in the world.

5.2 Materials & Methods

5.2.1 Study Area

We used a 1,712-kilometre passenger railway network from regional Victoria, Australia (operated by V/line, a government-owned corporation, see Appendix C) in south-east Australia to conduct our study (Figure 5.1). Trains operate on all sections of the network between 4 a.m. and 2 a.m. (the following day), with the largest volume occurring Monday through Friday between 7 a.m. and 9 a.m., and 4 p.m. and 6 p.m. Most trains operate at speeds of 100 km h\(^{-1}\) or less, however, on some sections of track commuter trains operate at maximum speeds of 160 km h\(^{-1}\). Due to limited data available, we did not include freight operations in our study.
5.2 Materials & Methods

Figure 5.1: Regional passenger train network in the state of Victoria. Inset shows location of Victoria in Australia. The railway network is shown as thin lines through major towns (stars). Wildlife-train collisions (reported between 2009–2014) are shown as crosses.

5.2.2 Data Preparation

To organise our data and modelling, we overlaid a 1-km$^2$ cell grid on the railway network (Figure 5.2). In each grid cell we modelled species occurrences and quantified train movements and speeds.

Eastern Grey kangaroos (*Macropus giganteus*, Shaw, 1790; “kangaroos” hereafter) are frequently struck in regional Victoria and large enough to cause damage and subsequent maintenance (e.g. cleaning) of trains. V/line provided records of all driver-reported collisions with kangaroos spanning a six-year period between 1 January 2009 and 31 December 2015, a total of 439 collisions. Collisions with large mammals (e.g. domestic livestock or kangaroos) must be reported to allow trains to be inspected and maintenance performed as required. Each record included incident date and time, name of service line (unique route between two towns), and nearest fraction of a kilometre post (physical sign markers indicating distance along a railway line). Using geographic information system (GIS) data on the regional railway network, we determined spatial coordinates (GDA94 MGA zone 55 projection) for all collisions from the reported kilometre post and service line. Although uncertainty in estimates of location was not explicitly reported, we assumed a maximum possible error of ±500 metres.
Predicting wildlife-train collisions across space and time to inform railway operations

5.2.3 Species Occurrence

Kangaroos are widespread (Dawson, 2012) and abundant in many parts of Victoria, yet comprehensive distribution records are lacking in many areas. To represent risk of collision by exposure to threat, we required distributional data across the entire study area and used species distribution modelling to predict relative likelihood of kangaroo occurrence. We emulated methods by Elith, Leathwick & Hastie (2008) to model and predict occurrence in each grid cell for the whole State of Victoria. The model was trained on data from the online Victorian Biodiversity Atlas (see Appendix C for more information) and included several environmental variables relating to the biology and behaviour of kangaroos (see Chapter 2). To reduce the effects of sampling bias, we also included four additional predictors: 1) distance to urban areas, 2) distance to roads, and the 3) easting and 4) northing spatial coordinate of the grid cell centroid.

5.2.4 Characteristics of Railway Network

To determine train movements across space and time, we accessed publicly available locations of stops and times along train routes from V/Line general transit feed specification (GTFS) data (www.ptv.vic.gov.au/about-ptv/ptv-data-and-reports/digital-products, accessed online 3 March, 2016). GTFS is a standard publishing format developed and maintained by a community of public transport agencies for scheduling and spatial data. Since it is publicly available, it also allows software developers
to write applications for mobile devices that track and report the locations of public transportation (see Hillsman & Barbeau, 2011). We used a spatial database (Postgres version 9.6; PostGIS version 2.3.0) to process this information and report the average scheduled number of trains, the total length of track, and average train speed in each grid cell for each hour of the day where trains occurred.

### 5.2.5 Statistical Modelling

We adapted the single-species quantitative risk model in Chapter 2 to fit and compare the relationship of kangaroo presence and characteristics of the railway network to collision likelihood. To account for temporal variation in collision risk throughout the day, we considered periods of train movements and animal activity in relation to time of day. By adding a function that allowed a bimodal response of collision rate to hour of day across all seasons, the crepuscular lifestyle of kangaroos (most active at dawn and dusk; see Dawson, 2012) was incorporated. The likelihood that a collision occurs in a given grid cell \( i \) at hour \( j \) in month \( k \) (\( p_{ijk} = \Pr(Y_{ijk} = 1) \)) depends on species occurrence \( O_i \), average number of trains \( V_{ij} \), average train speed \( S_{ij} \), length of track \( T_i \) (offset term), and a term \( C_{ijk} \) that accounts for the crepuscular behaviour of kangaroos:

\[
c\log(p_{ijk}) = \beta_0 + \beta_1 \log(O_i) + \beta_2 \log(V_{ij}) + \beta_3 \log(S_{ij}) + C_{ijk} + \log(T_i) \tag{5.1}
\]

The crepuscular term \( C_{ijk} \) has three components:

\[
C_{ijk} = \gamma_1 \sin\left(\frac{\pi(j-6)}{12}\right) + \gamma_2 \sin\left(\frac{\pi(j-6)}{12}\right)^2 + \gamma_3 A_{ijk} \tag{5.2}
\]

The first two terms are the linear and quadratic terms of a sine function that varies relative to 6 a.m. The third term \( A_{ijk} \) models the influence of time since dawn and dusk. This function takes one of three forms depending on whether the time is before dawn, after dusk, or between dawn and dusk:

\[
A_{ijk} = \begin{cases} 
  e^{-\left(\frac{2(j-U_{ik})}{(24-D_{ik})+U_{ik}}\right)^2} - e^{-\left(\frac{2(j-D_{ik}-24)}{(24-D_{ik})+U_{ik}}\right)^2}, & \text{if } j < U_{ik} \\
  e^{-\left(\frac{2(j-U_{ik})}{(24-D_{ik})+U_{ik}}\right)^2} - e^{-\left(\frac{2(j-D_{ik})}{V_{ik}}\right)^2}, & \text{if } U_{ik} \leq j < D_{ik} \\
  e^{-\left(\frac{2(j-D_{ik})}{(24-D_{ik})+U_{ik}}\right)^2} - e^{-\left(\frac{2(j-D_{ik})}{V_{ik}}\right)^2}, & \text{if } j \geq D_{ik}
\end{cases} \tag{5.3}
\]

Dawn \( U_{ik} \) and dusk \( D_{ik} \) times for civil twilight (six degrees below horizon) were determined for each observation in the dataset using a National Oceanic and Atmospheric Administration (NOAA) astronomical algorithm in the ‘R’ package maptools. We selected the 15th day of each month \( k \) at the centroid of each respective grid cell \( i \) to calculate dawn and dusk times.

The coefficients \( \gamma_1, \gamma_2, \gamma_3 \) are estimated from the data and measure the relative influence of the three components (Equation (5.2)) on the shape of the curve. One advantage of our parameterisation of the crepuscular term is that it is cyclical over a
twenty-four hour period (Figure 5.3) regardless of the parameter estimates of $\gamma_1$, $\gamma_2$, $\gamma_3$.

**Figure 5.3:** a) Components used to build crepuscular term in Equation (5.2); the solid curve controls activity relative to day or night, the dashed line controls uni-modal (e.g. diurnal) or bi-modal (e.g. crepuscular) activity, the dotted line controls activity closer to dawn or dusk. The model estimates a coefficient $\gamma$ in Equation (5.2) – for each component that controls the final curve shape. The resulting curves may reflect b) crepuscular, c) nocturnal, or d) diurnal activity patterns. Example *gamma* coefficient values ($g_1$, $g_2$, $g_3$) used to produce curves are shown above each respective graph.

Prior to modelling, we centred all explanatory variables by subtracting their means (after log-transforming the variables indicated in Equation (5.1)). Using pairwise analysis, all predictors exhibited Pearson’s product moment correlation coefficients of
less than 0.4 indicating low potential effects of multi-collinearity. Table 5.1 further describes the variables used in the model.

Table 5.1: Predictor variables used in collision model. Means and ranges are expressed on untransformed scale. C1, C2 and C3 determine a curve shape representing relative species activity – see Equation (5.2) and Figure 5.3.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Units</th>
<th>Mean; Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGK</td>
<td>Relative likelihood of kangaroo occurrence in grid cell</td>
<td>—</td>
<td>0.13; 0.01:0.87</td>
</tr>
<tr>
<td>TRAINS</td>
<td>Train frequency in grid cell</td>
<td>trains h(^{-1})</td>
<td>6; 1.80</td>
</tr>
<tr>
<td>SPEED</td>
<td>Mean train speed in grid cell</td>
<td>km h(^{-1})</td>
<td>86.75; 3.92:147.73</td>
</tr>
<tr>
<td>C1</td>
<td>Crepuscular function linear component</td>
<td>—</td>
<td>0.25; −1.00:1.00</td>
</tr>
<tr>
<td>C2</td>
<td>Crepuscular function quadratic component</td>
<td>—</td>
<td>0.47; 0.00:1.00</td>
</tr>
<tr>
<td>C3</td>
<td>Crepuscular function astronomical component</td>
<td>—</td>
<td>−0.05; −0.98:0.98</td>
</tr>
</tbody>
</table>

We fit the data (n=291,120 1-km\(^2\) grid cells along the train network) to a generalised linear model (see McCullagh & Nelder, 1989) using maximum likelihood estimation with a binomial distribution and a complementary log-log link on the linear predictor. The complementary log-log link was selected over the more common logit link due to the mathematical theory underpinning our model – risk being measured by the rate of collisions (see Chapter 2). The model is similar to a proportional hazards model (discrete censored time) often used in survival analysis and epidemiological studies (see Cox & Oakes, 1984).

To examine model fit, we generated randomised quantile residuals (see Dunn & Smyth, 1996) and plotted them against collision probability, kangaroo occurrence, number of trains per hour, average train speed, and hour. To assess performance, we cross-validated the model by randomly splitting the data into \( K = 10 \) partitions. We used nine of these subsets for model fitting and one for assessing model accuracy. For each assessment we obtained two performance metrics; area under the receiver operator characteristic (ROC) curve (see Metz, 1978) and coefficients of observations regressed on predictions (see Cox & Snell, 1989; Miller, Hui & Tierney, 1991). We repeated this procedure for 100 iterations producing a total of 1,000 sets of performance metrics and compared them with those from the model fitted to all data.

Using the model fitted on all data, we predicted the number of expected kangaroo-train collisions in the study area under different management scenarios: A) no change to operations; B) moderated train speeds in high kangaroo occurrence areas; and C) reduced kangaroo occurrence in areas with highest average speed of trains.

Scenario B involved capping train speeds at 80 km h\(^{-1}\) in grid cells with kangaroo relative occurrence likelihoods of 0.5 or above during the hours of 5 a.m. to 9 a.m. and 4 p.m. to 8 p.m. This change affected 42 cells and 275 out of 3,024 unique weekly routes. In scenario C, relative kangaroo occurrence was reduced by approximately half in all grid cells with average train speeds of more than 120 km h\(^{-1}\) (n = 154 cells, total track length = 121 km). This simulates reducing access of kangaroos to the railway network (e.g. by fencing). The values for reducing kangaroo abundance were used for all hours of the day as this management strategy would most likely involve exclusion or reduction in animal populations which operate irrespective of temporal variation.
5.3 Results

Our model fitted the data and the signs of model coefficients (positive or negative) indicated plausible relationships between collision likelihood and each predictor variable. All variables except train frequency demonstrated high statistical significance in the model fit (Table 5.2). The model residuals demonstrated no strong patterns with respect to predicted collision probabilities and model predictors (Figure 5.4).

Table 5.2: Summary of model fit using all data (n=291,120 grid cells). Highly significant variables are marked with an asterisk.

| Variable | Beta Coefficient Estimate | Standard Error of Coefficient Estimate | Z-value | Pr>|Z| |
|----------|---------------------------|----------------------------------------|---------|---------|
| Intercept | −7.2                      | 0.09                                   | −79.03  | <1.0x10^{-14} |
| EGK (β₁) | 0.61                      | 0.06                                   | 10.46   | <1.0x10^{-14} |
| TRAINS (β₂) | 0.02                    | 0.09                                   | 0.17    | 8.7x10^{-2}  |
| SPEED (β₃) | 3.62                      | 0.31                                   | 11.53   | <1.0x10^{-14} |
| C₁ (γ₁)   | −0.66                     | 0.11                                   | −5.83   | 5.7x10^{-7}  |
| C₂ (γ₂)   | −1.87                     | 0.17                                   | −10.84  | <1.0x10^{-14} |
| C₃ (γ₃)   | 0.25                      | 0.07                                   | 3.77    | 1.7x10^{-5}  |

The relative risk of collisions increased with higher average train speeds, kangaroo occurrence, train frequency and during hours of high kangaroo activity in grid cells. The strongest and most significant predictor was train speed; collision risk increased exponentially with considerable increases at speeds above 85 km hr⁻¹. Increasing train speed from 110 to 130 km hr⁻¹ doubled collision risk, however, the effect of speed also demonstrated large uncertainty, with wide confidence intervals (Figure 5.5a). Kangaroo occurrence was the second most influential predictor; collision risk increased rapidly at low values of occurrence and more slowly at higher values. The increase was approximately 10-fold across the range of values for kangaroo occurrence (Figure 5.5b). Train frequency had very little influence on collisions, compared to the other predictors (Table 5.2; Figure 5.5c).

The bi-modal functional form for the effect of kangaroo activity on collision risk demonstrated a plausible shape (Figure 5.5d). Collision risk peaked at approximately 5:45 a.m. and 6:15 p.m. with a higher risk occurring in the morning. The response of collision risk over time was most uncertain in the evening peak. Both peaks showed similar shape – the density and spread of collision risk around the maxima – and the lowest collision risk occurred around noon.

The models fitted on all data and fitted on the subsets of the data during cross-validation performed similarly. The ROC value was 0.82 for both the full data model (Figure 5.6b) and mean of the cross-validated models (Figure 5.6c). Likewise, the calibration statistics (intercept and slope of regression line between observations and predictions) were similar for both the full data model (Figure 5.6a) and mean of the cross-validated models (Figure 5.6c). The uncertainty in the calibration metrics was higher than that of the ROC values as shown by the 95% confidence intervals. The overall calibration of the full data model was good for low collision rates where the uncertainty around the observed rates was also low. Observed rates with higher uncertainty were evident at higher predicted values (Figure 5.6a).
Both of the simulated management scenarios reduced the predicted number of collisions from the baseline estimated with no management (Scenario A). Scenario C reduced expected collisions by approximately 3.0% whilst scenario B reduced collisions by 1.0% (Table 5.3).

**Table 5.3:** Summary of predicted collisions based on different management scenarios. Expected collisions are expressed as a total across the entire regional train network for a period of one year.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Description</th>
<th>Expected Total Collisions</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>no change to current operations or infrastructure</td>
<td>404</td>
</tr>
<tr>
<td>B</td>
<td>moderated train speeds in high kangaroo occurrence areas during peak travel times</td>
<td>400</td>
</tr>
<tr>
<td>C</td>
<td>controlled kangaroo occurrence in areas with highest average speed of trains</td>
<td>392</td>
</tr>
</tbody>
</table>
Figure 5.4: Binned randomised quantile residuals plots – average residual versus the average predictor value for each bin. For all continuous predictors, bins are equally spaced and determined by taking the square root of the total number of unique values for each predictor; for the discrete predictor of hour, there are 23 total bins. The gray lines indicate ±2 standard-error bounds which should contain 95% of the residuals if the model were actually true. Note, some variables are log-transformed to assist with visual inspection of the residuals.
Figure 5.5: Marginal effects of model predictors on collision risk. For the effects of species occurrence, number of trains, and train speed on collision risk, non-target variables were held constant at mean values. For the effect of hour on collision risk, species occurrence, number of trains, and train speed were held at mean values. Shading indicates 95% confidence intervals around coefficient estimates.
Figure 5.6: a) Calibration plot showing rate of observed collisions against predicted rate of collisions. Dots represent the observed rate with 95% confidence intervals at the medians of each bin of predictions (10 total). Labels indicate the total observations in each bin. A regression line is shown between the dependent variable and the predicted values (response-scale) of the model; perfect calibration has an intercept of 0 and slope of 1; b) ROC (receiver operating characteristic) curve measuring discrimination ability of model at all threshold values. Values shown are averages of all cross-validations; c) Comparison between the collision model fit on full data and on cross-validated subsets. "Intercept" and "Slope" result from regressing the dependent variable on the predicted values and measure calibration (see plot a); ROC measures discrimination between collisions and no-collisions (see plot b). For each metric, open circles represent the full data model and solid dots represent mean values (95% confidence intervals shown as bars) for the 1,000 cross-validated subsets. Dashed lines indicate the expected values for a perfectly calibrated and discriminatory model.
5.4 Discussion

Our model demonstrates that kangaroo-train collisions are related to train speed, kangaroo exposure to moving trains, and the coincidence of periods of high train and kangaroo activity. All of these relationships are consistent with our initial expectations and also shown in related studies on railway (e.g. Gundersen & Andreassen, 1998) and road (e.g. Lao et al., 2011b; Roger, Bino & Ramp, 2012) collisions. Our work extends further by presenting a conceptual modelling tool to assist managers create safer and more cost effective railways.

Train speed was an important predictor for collision risk. As trains increased speed, the risk of collisions increased rapidly, up to the typical maximum speed of 160 km h$^{-1}$. Collision risk relating vehicle speed to the size and velocity of animals has been demonstrated through simulation (Jaarsma, van Langevelde & Botma, 2006) concluding that smaller and slower-moving species are more vulnerable (also see Fahrig & Rytwinski, 2009). Moreover, as these relationships are often exponential, high-speed vehicles may become significantly problematic regardless of species trait. It should be noted that our study utilised published schedule data to interpolate train movements in space and time. Therefore, there is uncertainty in both the location and trajectories of actual trains. Further study using global positioning system (GPS) waypoints of train movements would reduce some of this uncertainty.

Kangaroo occurrence is also a useful predictor for collision risk. Collision risk consistently increased with predicted relative occurrence, as found for roads (e.g. Lao et al., 2011b; Roger & Ramp, 2009). One feature of the model framework is the flexibility of choice in how to represent species occurrence. We employed published methods to determine kangaroo occurrence, however, our framework is not limited to data derived using only this type of model. The species distribution modelling literature is vast and covers topics relating to model choice (Guillera-Arroita et al., 2015), calibration and bias (Phillips & Elith, 2010), sources of data (van Strien, van Swaay & Termaat, 2013), and validation (Chivers, Leung & Yan, 2014). Our framework can also incorporate data from population viability analyses to test the effects of population dynamics on collision risk. For example, collisions can correlate with expected counts of single species in a given area (see Skorka et al., 2013). Kangaroos in Victoria are not subject to hunting pressure to the same extent as ungulates in parts of North America and Europe, which has been shown to affect the rate of train-wildlife collision (Seiler, 2005). However, population control of kangaroos occurs periodically in certain areas and our model framework can accommodate changes in occurrence driven by these factors by including relevant predictor variables.

Temporal patterns, such as the crepuscular activity of wildlife, have implications for collision risk. Hourly and seasonal patterns have been applied differently in wildlife collision research. Some studies treat temporal predictors as a categorical variable (Dussault et al., 2006), whilst others have explicitly defined cyclic functions (Thurfjell et al., 2015). Our model uses a term with three components to define a functional form that is flexible to the crepuscular (bi-modal) nature of kangaroo activity; which also happens to coincide with peak train activity in particular seasons. The use of
control parameters estimated from the data allows this temporal function to suit the
behaviour of any target species (e.g. nocturnal, diurnal, crepuscular and random) –
see Figure 5.3. Moreover, seasonal movements such as migration (see Neumann et al.,
2012) may also be included in the model specification, however, kangaroos do not
migrate (Dawson, 2012).

The collision data used for this study has unique properties with respect to re-
porting bias and errors. Train drivers are obliged to report the time and location of
large animal strikes as they usually result in damage to or required cleaning of trains.
Therefore, these data are less subject to reporting bias compared to many road colli-
sion studies (but see Snow, Porter & Williams, 2015). Moreover, spatial and temporal
errors in reported collisions are assumed to be less as standardising mechanisms, such
as collision report forms, GPS devices, and distance signage, are implemented in rail-
way operations. Similar practices are used by some road authorities to collect and
archive carcass data (see Huijser et al., 2007b), however, the coverage is often sparse
due to the spatial extent of road networks and temporal uncertainty of collision events.
Technology has been shown to assist with data collection (Olson et al., 2014; Shilling,
Perkins & Collinson, 2015) and similar approaches may also be applied to railway
networks.

We used existing data to create predictors for the model framework. The data
are publicly-available online and, in some cases, maintained and updated regularly.
This reduces potential costs involved in the collection of data. Moreover, the model
framework may be easily updated as new information becomes available. As some of
the data in the framework result from modelling (species occurrence) or interpolation
(train movements), potential inherent uncertainty should be considered when drawing
inferences. For example, the relative effect of each predictor may be weighted according
to associated uncertainty or experts may be used to assess sub-model predictions (e.g.
species occurrence – see Clevenger et al., 2002; Wintle, Elith & Potts, 2005).

Our framework supports management decisions in two distinct areas: reduction
of animal presence (e.g. deterrents or exclusions) or reduction of train threat (e.g.
adjusted schedules or speeds). The choice of mitigation may be influenced by the
effects of each predictor on collision likelihood (e.g. if speed is more correlated or has
a stronger influence). This is determined by examining the model fit or predicting
responses based on changes in parameter values (e.g. increasing likelihood of kan-
garoos). Mitigation choice may also be limited by operational objectives. Fencing
may be chosen to exclude animals on railways when changes to train speed and fre-
cuencies are not possible or desirable, regardless of the effect in the model. Most
high-speed railways are fenced (Campos & de Rus, 2009) as the speed of trains is the
dominant technological characteristic and reducing it may be contrary to its public
service objective. The scenario we did not model – installing 100% effective fenc-
ing along the entire network – would have the effect of reducing train collisions with
ground-traversing species to zero, but is not realistic and thus omitted in this study.

Each of our management scenarios reduced collisions which is a positive outcome.
Although many negative outcomes – not easily quantifiable – are related to collisions
(e.g. trauma, loss of life), monetary costs are a useful metric for comparing management options. From a transportation authority perspective, collisions with kangaroos in Australia incur costs through removal of trains from service for cleaning and repair. These costs will vary by railway operator and are useful to compare with costs of mitigation. For example, let us assume that a Victorian regional passenger train must be taken out of service following a collision with a kangaroo and the cost of this activity is $20,000. By reducing the annual number of collisions by 10, or approximately 2.5% of reported collisions in this study, there is an estimated savings of $200,000 per year. If the costs of different management strategies are also calculated, a cost-benefit analysis may be performed. This has been applied to collisions with road vehicles (see Huijser et al., 2009) and the concepts are similar for railway transport. Further, the benefit from reductions in annual collisions will have varying magnitudes depending on the costs incurred from wildlife-train collisions. In India, for example, collisions with elephants may de-rail trains (Dorsey & Rew, 2015) resulting in significant costs. In these situations, avoiding a single collision event has remarkable benefit.

Herein, we have demonstrated a model framework that functions as an effective management support tool. It utilises existing sources of data, is logically organised, and is transferable/scalable to other networks and species. Other potential uses of the framework may include an ongoing implementation where the model is updated based on new information and reports risk to operators in real-time.
Validating collision model predictions with disparate datasets: a case for a unified system
Abstract

Wildlife-vehicle collisions are often costly due to property damage and injury and have been estimated to kill billions of vertebrate fauna each year. Many studies use predictive modelling to determine where and when WVC will occur. These models are often trained on collision data that is sourced from non-government organisations, private-sector entities, academic institutions, government agencies, community groups, and individual citizen scientists. These data are not easily attainable and often have implicit biases and errors.

We analysed change in model performance by combining four disparate datasets of collisions in the State of Victoria in south-east Australia. For each combination of training datasets, we fitted a model, made predictions, and validated the predictions with an independent dataset not used to train the model. We also used a collision dataset of total insurance claims per town to validate the predictions of models using different combinations of data.

Among the four datasets, the discrimination ability of the model did not increase as data were added to models. Using the town-level independent data to validate models resulted in an increase in model calibration as more combinations of training data were used. Likewise, the variation in the data explained by the model also increased as more combinations of data were used.

We demonstrate that increasing the amount of collision data used for modelling, by combining disparate datasets, has the potential to improve model performance and predictive strength. These results highlight the importance of uniting the currently disparate sources of collision data with a comprehensive and systematic archive. We also suggest qualities that are integral to the success of such a system.
6.1 Introduction

Roads are critical infrastructure, but their development and use has ecological impacts (Forman et al., 2003). Perhaps one of the most directly visible and confronting issues is vehicle strikes to animals, especially large terrestrial vertebrates. These collision events are often costly due to property damage, injury, and death. Globally, wildlife-vehicle collisions (WVC) on transportation networks have been estimated to kill billions of vertebrate fauna each year (Seiler & Helldin, 2006). In North America, millions of animals are estimated to be killed by vehicle strikes each day (Forman & Alexander, 1998) and annual economic costs due to collisions exceed eight billion dollars (Huijser et al., 2007a). However, these adverse effects are not unique to North America and occur at all locations where roads dissect habitat for wildlife. Due to such large impacts, governments, scientists, private-sector companies and citizens throughout many nations have undertaken data collection and analysis of WVC (see van der Ree, Smith & Grilo, 2015).

A large amount of research seeks to determine the frequencies and magnitudes of wildlife-vehicle collisions. This often involves examining several environmental (e.g. landscape, climate, road characteristics), anthropogenic (e.g. human behaviour, vehicle movements) and biotic (e.g. animal traits) factors at different spatial and temporal scales (Litvaitis & Tash, 2008). More recently, computer modelling has been employed to predict where and when collisions are most likely to occur based on a large range of predictor variables (see review in Gunson, Mountrakis & Quackenbush, 2011). These predictions aim to help environmental managers and road authorities to identify and mitigate problematic areas and better construct new roads. Effective decisions made from these predictions rely on robust, calibrated models as incorrect inferences may result in costly outcomes such as ineffective mitigation (see Huijser et al., 2009) or continuing effects of unmitigated collisions (see Bissonette, Kassar & Cook, 2008). Data deficiencies generate uncertainty in models and result from data collected with specific procedures or spatial/temporal trends (e.g. bias), measurement error or inappropriate sample size (Bean, Stafford & Brashares, 2012). Although other factors, such as model misspecification, may also result in biased and uncertain predictions, we argue that the quality and quantity of data used to train WVC predictive models are of paramount importance. It is therefore recommended that data collection includes information about the extents and intensity of surveyed areas. For example, the total length of roads and the temporal span of collection included in a survey area are both useful to determine any sampling biases in the data. A centralised system can employ forms and guides to prompt for this information.

Many studies that use predictive modelling to determine where and when WVC will occur are limited to single sources of data – either field collected based on predetermined project requirements (e.g. Langen, Ogden & Schwarting, 2009; Roger & Ramp, 2009) or pre-existing from independent surveys or routine collections (e.g. Hothorn, Brandl & Müller, 2012; Malo, Suárez & Díez, 2004). Bias and error may arise from the data sources and collection methods, however, these are not often explicitly reported or accounted for in the modelling methods. Further, few studies
utilise independently-obtained data to validate model performance; often this is due to limitations in availability. Incorporating detailed descriptions of survey methods (e.g., observer travel speed, record inclusion criterion such as ‘on paved section’ or ‘on shoulder’) into metadata enables analysts to develop comparative measures to improve modelling. As previously mentioned, a centralised system with standardised forms can assist with this task. Weighted regression or Bayesian analysis that accounts for missing observations or underreporting can also help to accommodate combinations of data with different survey effort – in cases where existing data are not based on consistent survey and reporting effort. Modelling that accounts for covariance between disparate data sources can determine relative added value (Pacifici et al., 2017).

In response to the importance of consistent data collection, technology has been proposed to augment field activities. For example, mobile phones have been proposed to assist with data collection of wildlife-vehicle collisions (see Aanensen et al., 2009; Olson et al., 2014). Electronic tablets and other small devices have also been proposed to assist road authorities to record collisions (Ament et al., 2007). Cataloguing and reporting systems are used to store data on WVC, however, few are open-access or incorporate data from other institutions or previous surveying events. Perhaps one of the largest systems in current use is the California Roadkill Observation System (CROS), maintained by the Road Ecology Center at the University of California, Davis (see Shilling, Perkins & Collinson, 2015). This system offers many features such as public-accessibility, an intuitive user interface, and collection of important metadata on reliability and precision of reporting, but has yet to achieve full uptake by agencies outside the academic and citizen-science networks it supports.

Worldwide, collision data is collected by non-government organisations, private-sector entities (e.g., insurance companies), academic institutions, government agencies (e.g., road authorities and environmental managers), community groups, and individual citizen scientists but these data are not always easily attainable due to privacy restrictions or data formats. We assert that modelling and prediction of WVC will benefit from extensive, open-access data repositories that catalogue collision records from these entities. These mechanisms exist in other areas of research. For example, contemporary species distribution modelling draws from worldwide atlases of species occurrence; the Global Biodiversity Information Facility (www.gbif.org) contains more than 700 million records from over 800 data publishers and the Atlas of Living Australia (www.ala.org.au) provides over 50 million records from a wide range of data providers.

In our study, we analyse the predictions of a wildlife-vehicle collision risk model fitted to a citizen-science dataset joined with four external independent datasets in different combinations. As the sample size and variation in collection techniques (e.g., spatial scale, field expertise) of our training data increases, we analyse the performance of our models. Our overall aim is to demonstrate the value of a unified, publicly-accessible data storage system to support analysis and prediction of WVC and determine characteristics of data that are important to ensure robust inference.
6.2 Materials & Methods

Our study analyses risk of WVC over approximately 150,000 kilometres of sealed roadway (i.e. only bitumen surfaces) in the State of Victoria in south-east Australia (Figure 6.1). The Victorian road authority (VicRoads) manages 25,256 kilometres of major roadway and seventy-nine municipal districts manage the remaining roads. We split the sealed road network into 644,715 road segments by intersecting the roads with a 1-km$^2$ spatial grid. We predicted traffic volume and speed to each road segment using linear regression and predicted species occurrence to the spatial grid using species distribution models. Species occurrence was sampled using the midpoints of the road segments resulting in each segment having predicted values of kangaroo occurrence, traffic speed and traffic volume in the final dataset – more details are provided in Chapter 2.

Eastern Grey kangaroos (*Macropus giganteus*, Shaw) are the second largest native mammal (up to 85 kilograms for males) in Australia (Coulson & Eldridge, 2010), and the most commonly reported species struck by motor vehicles in Victoria (Rowden, Steinhardt & Sheehan, 2008), prompting several organisations to record carcass collection and collision locations. We obtained four datasets of spatially-explicit collisions with Eastern Grey kangaroos in the state of Victoria in south-east Australia for our study (Table 6.1). The data were sourced from: the City of Bendigo, a large regional center located in north-east Victoria; the Western District, a large VicRoads administrative boundary located in the center of the western portion of Victoria; Crashstats, an online repository of police-reported WVC; and Insurance Australia Group, an insurance company that records WVC reported by their members. See Appendix C for more information on data sources.

Data from the City of Bendigo were based on carcasses reported by in the 3,001-km$^2$ municipality by members of the public. The Western District data were comprised of reported carcasses, along 1,417 kilometres of major roads within an area of 24,373 km$^2$, made by a VicRoads-authorised, litter-abatement contractor. Crashstats data covered the entire state and were based on reports made by the state police regarding injuries from WVC. Similarly, the Insurance Australia Group data were also based on statewide reports of WVC, however, originating from insurance claims made by motorists. Each of the road segment-based datasets (City of Bendigo, Western District, and Crashstats) were used to augment a dataset obtained from Wildlife Victoria for modelling and to validate predictions resulting from these models.

To develop each modelling dataset for our study, we selected road segments that intersected with reported kangaroo collision records to represent presences. All other road segments without collisions represented background data. When combining datasets, multiple collisions on the same road segments were reduced to single presences. Thus, each modelling dataset was comprised of 644,715 road segments having either a collision or no collision. Eight permutations were developed by combining independent data with Wildlife Victoria (baseline) data (codes shown in brackets):

1. Wildlife Victoria (o)
Validating collision model predictions with disparate datasets: a case for a unified system

Figure 6.1: Study area (state of Victoria in south-east Australia) showing all sealed road segments as light gray lines. The inset shows the geographic location of Victoria in Australia. The darker shaded region is the City of Bendigo and the lighter shaded region is the VicRoads Western District. Major towns (>100,000 residents) are starred and labelled accordingly.

2. Wildlife Victoria + City of Bendigo (ob)
3. Wildlife Victoria + Western District (ow)
4. Wildlife Victoria + Crashstats (oc)
5. Wildlife Victoria + City of Bendigo + Western District (obw)
6. Wildlife Victoria + Western District + Crashstats (owc)
7. Wildlife Victoria + Crashstats + City of Bendigo (ocb)
8. Wildlife Victoria + City of Bendigo + Western District + Crashstats (obwc)

To test the effects of combining independent datasets, we fitted a collision model based on Equation (2.8) in Chapter 2 to all permutations of data. We predicted collision risk from a model that regressed reported collisions on species occurrence, traffic volume (linear and quadratic terms) and traffic speed:

\[
c log  \log(p_i) = \beta_0 + \beta_1 \log(O_i) + \beta_2 \log(V_i) + \beta_3 (\log(V_i))^2 + \beta_4 \log(S_i) \quad (6.1)
\]

where \( p_i = \Pr(Y_i = 1) \) is the relative likelihood of a collision occurring on a road segment \( i \), \( O_i \) is species occurrence, \( V_i \) is traffic volume, \( S_i \) is traffic speed.
To test model sensitivity to different combinations of data, we validated the predictions from our models using held-out datasets. For each combination of training datasets, we fitted a model, made predictions, and validated the predictions with an independent dataset not used to train the model. To assess model calibration, we regressed the observed collisions in each independent dataset on values predicted from models using the independent data (see Miller, Hui & Tierney, 1991). A perfectly calibrated model would report an intercept coefficient of zero and a slope coefficient of one. To measure the ability of each model to discriminate between false positive and false negative values, we also calculated receiver operator characteristic (ROC) scores (see Metz, 1978). A score of one indicates perfect discrimination while 0.5 indicates no better than random selection; 0.75 or above is often considered acceptable performance.

In addition to the four datasets with data on individual collisions, we also had access to a dataset in which the number of collisions that involved insurance claims was recorded for towns. The Insurance Australia Group ("iag" hereafter) collision data detailed counts of collisions within each town boundary in Victoria. From the predictions made to all road segments by the models using different combinations of data, we calculated the mean collision risk within each town boundary. Using each town as a datapoint, we regressed the mean collision risk prediction (log-transformed) on the total collisions reported. Due to the response data being positive integers, we used the Poisson link:

$$\log(C_j) = \beta_0 + \beta_1 \log(P_j)$$

(6.2)

where $C_j$ is the count of reported collisions and $P_j$ is the mean predicted collision risk, in a town $j$. As with the road segment-based data analysis, calibration was assessed using the intercept and slope of the model fits. We also calculated the reduction in deviance (unexplained variation in the data) for each model.

To further investigate how each independent dataset contributed to the models utilising combined data, we fitted the model described in Equation (6.1) to each road segment-based independent dataset (City of Bendigo, Western District, and Crashstats). We added an offset term of road length multiplied by the data period in years to standardise the model outputs for comparison. For each dataset, we restricted road segments to a domain that corresponded to the respective survey or reporting area. The City of Bendigo data were restricted to road segments that fell within the city boundary as carcass reports outside of the jurisdiction were not catalogued. Similarly, only road segments that were surveyed in the Western District were included for analysis. Crashstats covers the entire state and so all 644,715 road segments were used.
Table 6.1: Datasets used to fit models and validate predictions.

<table>
<thead>
<tr>
<th>Source</th>
<th>Data Code</th>
<th>Type</th>
<th>Records</th>
<th>Reporting Method</th>
<th>Spatial Unit</th>
<th>Spatial Coverage</th>
<th>Temporal Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>City of Bendigo</td>
<td>b</td>
<td>Public municipality</td>
<td>395</td>
<td>Public reported</td>
<td>Road segment</td>
<td>City (5,001 km²)</td>
<td>8 years: 2007–2014</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>carcasses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VicRoads Western</td>
<td>w</td>
<td>State government organisation</td>
<td>815</td>
<td>Independent contractor recorded carcasses</td>
<td>Road segment</td>
<td>District (24,373 km²)</td>
<td>2 years: 2014–2015</td>
</tr>
<tr>
<td>District</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crashstats</td>
<td>c</td>
<td>State government organisation</td>
<td>487</td>
<td>Police reported</td>
<td>Road segment</td>
<td>State of Victoria (227,819 km²)</td>
<td>7 years: 2010–2016</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>wildlife strikes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wildlife Victoria</td>
<td>o</td>
<td>Non-government organisation</td>
<td>4245</td>
<td>Public reported</td>
<td>Road segment</td>
<td>State of Victoria (227,819 km²)</td>
<td>6 years: 2010–2015</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>observations of</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>wildlife strikes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insurance Australia</td>
<td>iag</td>
<td>Private corporation</td>
<td>1344</td>
<td>Public reported</td>
<td>Town</td>
<td>State of Victoria (227,819 km²)</td>
<td>3 years: 2011–2013</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
<td>claims of wildlife strikes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6.3 Results

All three models fitted to independent road segment-based datasets (City of Bendigo, Western District, and Crashstats) had mixed results when compared to the model fitted to the original dataset (Table 6.2). Both of the models using statewide data (Wildlife Victoria and Crashstats) performed better than the two using localised data (City of Bendigo and Western District). The worst performing model was fitted to data from the Western District of Victoria – only traffic speed was a significant predictor. In contrast, the only non-significant predictor was traffic speed in the model fit to data from the City of Bendigo. All other predictors in every model were significant. The reduction of unexplained variation in the data were remarkably similar for the two statewide data models (≈11.8%) and the localised data models (≈5.2%).

Most of the marginal effects of the predictors on collision risk had similar shapes for all of the model fits. Collision risk increased with increasing species occurrence for all models, however, the Western District data model reached nearly 75% of the maximum collision likelihood at low values of species occurrence (≈0.2). Collision risk had a quadratic relationship to traffic volume, and peaked between 2,000 and 7,000 vehicles day⁻¹ for all models except the Western District data model. The Western District data model showed a monotonically decreasing collision risk with increasing traffic volume beginning at a value of ≈0.95 at zero vehicles day⁻¹. Collision risk increased with increasing traffic speed for all models. Collision risk increased nearly linearly between 40 and 110 km hr⁻¹ in the model fit to the City of Bendigo data whilst collision risk increased slowly – the maximum increase of risk per unit speed
6.3 Results

Table 6.2: Summary of model fits using original and validation data in solidarity (i.e. no combinations of data are used). Deviance explained is the percentage of unexplained variation in the data reduced by the model. Highly significant variables (p < .0001) are marked with asterisks.

| Dataset          | Variable | Coefficient | Standard Error | Z-value | Pr(>|Z|) | Deviance Explained |
|------------------|----------|-------------|----------------|---------|----------|-------------------|
| Wildlife Victoria | Intercept | −42.30      | 1.07           | −39.66  | <.0001*  | 11.83             |
|                  | EGK      | 0.70        | 0.01           | 47.77   | <.0001*  |                   |
|                  | TVOL     | 5.55        | 0.24           | 22.68   | <.0001*  |                   |
|                  | TVOL2    | −0.33       | 0.02           | −21.76  | <.0001*  |                   |
|                  | TSPD     | 3.93        | 0.07           | 54.49   | <.0001*  |                   |
| City of Bendigo  | Intercept | −53.81      | 4.87           | −11.05  | <.0001*  | 5.16              |
|                  | EGK      | 0.49        | 0.09           | 5.28    | <.0001*  |                   |
|                  | TVOL     | 11.71       | 1.20           | 9.74    | <.0001*  |                   |
|                  | TVOL2    | −0.74       | 0.08           | −9.60   | <.0001*  |                   |
|                  | TSPD     | 1.19        | 0.39           | 3.08    | 0.0021   |                   |
| Western District | Intercept | −10.19      | 3.47           | −2.93   | 0.0034   | 5.15              |
|                  | EGK      | 0.02        | 0.05           | 0.36    | 0.7207   |                   |
|                  | TVOL     | 0.37        | 0.80           | 0.47    | 0.6408   |                   |
|                  | TVOL2    | −0.07       | 0.05           | −1.32   | 0.1885   |                   |
|                  | TSPD     | 2.03        | 0.45           | 4.48    | <.0001*  |                   |
| Crashstats       | Intercept | −51.74      | 2.89           | −17.88  | <.0001*  | 11.75             |
|                  | EGK      | 0.55        | 0.04           | 12.89   | <.0001*  |                   |
|                  | TVOL     | 4.89        | 0.69           | 7.13    | <.0001*  |                   |
|                  | TVOL2    | −0.32       | 0.04           | −7.19   | <.0001*  |                   |
|                  | TSPD     | 6.52        | 0.27           | 24.43   | <.0001*  |                   |

Figure 6.2: Marginal effects of each predictor on relative likelihood of collisions. Codes for data combinations are: ‘o’–Original (Wildlife Victoria); ‘b’–City of Bendigo; ‘w’–Western District; ‘c’–Crashstats. Note, collision likelihoods have been rescaled for comparison as all datasets have different numbers of data points and, thus, vary in their respective ranges of predicted values.

occurring above 95 km hr\(^{-1}\) – in the model fit to the Crashstats data.

The ability of models to discriminate between false-positive and true-positive rates of collisions were stable using different training data and validation data (Table 6.3); discrimination ability did not increase as data were added to models. Among the
Validating collision model predictions with disparate datasets: a case for a unified system

Table 6.3: Discrimination ability of models expressed as receiver operator characteristic scores. Data combinations used to model and make predictions are shown as column headings. Data used to validate model predictions are shown as row headings.

<table>
<thead>
<tr>
<th></th>
<th>Wildlife Victoria (original)</th>
<th>original + City of Bendigo</th>
<th>original + Western District</th>
<th>original + Crash-stats</th>
<th>original + City of Bendigo + Western District</th>
<th>original + Western District + Crash-stats</th>
<th>original + Crash-stats + City of Bendigo</th>
</tr>
</thead>
<tbody>
<tr>
<td>City of Bendigo</td>
<td>0.87</td>
<td>0.85</td>
<td>0.87</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western District</td>
<td>0.76</td>
<td>0.75</td>
<td>0.78</td>
<td>0.77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crashstats</td>
<td>0.8</td>
<td>0.8</td>
<td>0.83</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

different validation datasets, variation in the ROC scores were less than 4% between different combinations of data used to fit models. All of the models discriminated best when validated with the City of Bendigo dataset (ROC scores of 0.85–0.87) and worst with the Western District dataset (ROC scores of 0.75–0.78).

There was no observable trend in calibration when validating model predictions made with increasing amounts of data (Figure 6.3). The lowest calibration scores were for models validated with the Western District (w) collision data. The model fit on the Wildlife Victoria data (o) and validated with data from Crashstats (c) had the highest calibration score which was close to one. Moreover, there was no apparent improvement in the relationship between observed and predicted values when increasing the amounts of data used for modelling. All of the models demonstrated similar trajectories for observed-versus-predicted rates at low values and then diverged at high values (Figure 6.4).

Using the town-level independent data (iag) to validate models resulted in a slight increase in model calibration as more combinations of training data were used (Figure 6.5). The lowest calibration coefficient (0.943) occurred using only the original (o) data to fit the model and the highest (0.945) occurred using a combination of the original and all independent (obwc) data. Higher reductions of unexplained variation in the data also occurred with more combinations of training data (Figure 6.6). Deviance explained by the model fit on the original (o) data was 33.2% whilst the model fit on a combination of the original and all independent (obwc) data explained 33.3% of the deviance.
6.3 Results

Figure 6.3: Model calibration for all combinations of data. Codes for data combinations are: ‘o’–Original (Wildlife Victoria); ‘b’–City of Bendigo; ‘w’–Western District; ‘c’–Crashstats. Characters before the hyphen represent the datasets used for training the model and making predictions; characters after the hyphen indicate the data used for validation. Estimated calibration coefficients are shown as dots with bars representing standard errors.

Figure 6.4: Comparisons of observations versus model predictions for all combinations of data. Codes for data combinations are: ‘o’–Original (Wildlife Victoria); ‘b’–City of Bendigo; ‘w’–Western District; ‘c’–Crashstats. Characters before the hyphen represent the datasets used for training the model and making predictions; characters after the hyphen indicate the data used for validation.
Validating collision model predictions with disparate datasets: a case for a unified system

Figure 6.5: Model calibration for all combinations of data using the town-level independent data (iag) for validation. Codes for data combinations are: 'o'—Original (Wildlife Victoria); 'b'—City of Bendigo; 'w'—Western District; 'c'—Crashstats. Characters before the hyphen represent the datasets used for training the model and making predictions; the same town-level data (iag) were used for all validation. Estimated calibration coefficients are shown as dots with bars representing standard errors.

Figure 6.6: Model performance for all combinations of data using the town-level independent data (iag) for validation. Codes for data combinations are: 'o'—Original (Wildlife Victoria); 'b'—City of Bendigo; 'w'—Western District; 'c'—Crashstats. The percent of variation in the training data explained by the model (deviance) are shown as dots.
6.4 Discussion

This study demonstrated that increasing the amount of collision data used for modelling, by combining disparate datasets, improved model performance and predictive strength. The improvements were marginal, most likely due to the limited amount and significant variation of the data used in this study. However, even small, incremental improvements can be quite valuable when comparing to the costs of WVC. For example, if the model trained on more data can predict and prevent a single additional WVC from occurring, it has tremendous value when considering the high costs of building roads, a human life, or other unquantifiable effects such as species extinctions. From this perspective, the results of this study suggest that the development of a unified data system has high utility given its relatively inexpensive costs to implement. Further, such a system would be able systematically catalogue disparate sources of data and enable more robust analyses of wildlife-vehicle collisions, however, there are several important factors to consider. Large quantities of data may help to overcome issues of under-reporting (see Snow, Porter & Williams, 2015), and limited temporal/spatial coverage that affect generality (see Clevenger et al., 2015), but equally important is the quality of the data.

It is important to identify and address spatial biases when developing predictive models using presence-only data (Kramer-Schadt et al., 2013). Models trained on observations that are spatially biased (e.g. collected at particular places/times or dependent on characteristics of reporting individuals) may cause model predictions to reflect patterns of collection rather than distributions of collision risk. Each of our datasets had varying levels of spatial bias but these were difficult to identify. For example, the original Wildlife Victoria dataset appears to have a reporting bias that is spatially correlated with proximity to the urban centre of Melbourne. Other datasets can be expected to exhibit less spatial bias due to the nature of reporting. Crashstats, for instance, is comprised of state-wide police records of kangaroo-vehicle collisions that is more likely to be free of spatial reporting biases. However, Crashstats data is subject to under-reporting because it only includes records involving human injuries or deaths resulting from WVC. Regardless of differences, implicit biases in data may be considered in modelling efforts if they are adequately described in metadata (see Warton, Renner & Ramp, 2013) but generally, collision data does not include this information.

Metadata should also include the survey effort involved in collecting collision data as this will have modelling implications and an effect on predictions of road risk. Despite its importance, this information is seldom reported and varies considerably amongst and within jurisdictions (Huijser et al., 2007b). For example, the Western District data was collected using a combination of systematic surveys and opportunistic methods. Road cleaning contractors regularly patrolled select sections of major roads (Freeways, Highways, Major Arterials) in the district and recorded carcasses collected but were also irregularly dispatched based on carcass observations reported by the public. It should be noted that contractors only collected carcasses that were
in close proximity to the road, and deemed a risk to motorists, which leads to under-reporting. It is reasonable to assume our four datasets had different surveying efforts, however, because these were not explicitly defined, we did not incorporate them into the modelling. Information on survey effort can improve models by generating parameters that account for detectability (see Dorazio, 2014).

Citizen-science collected data has the potential to contribute large amounts of observational data on wildlife-vehicle collisions (e.g. Cosentino et al., 2014; Dwyer et al., 2016; Paul et al., 2014), however, controlling the quality of the data is critical to ensure sound statistical analyses. To augment collision or carcass observation data, some systems store information on reliability and accuracy (e.g. Shilling & Waetjen, 2015). This is often reported by the user and requires additional review by a system administrator to determine its merit but is a good foundation for assessing and comparing the effects of data quality on model outputs. In the Victorian data, we did not have access to such metadata and assumed all datasets had the same reliability and accuracy. Yet, we suspect there is some variation between datasets as some data, such as insurance or police reports, may be more reliable than observations reported by citizens. Moreover, our data only included records on collisions with kangaroos, which are quite identifiable; smaller species may be harder to identify and introduce additional uncertainty. Nonetheless, reliability may be incorporated into modelling if properly reported (see Gunson et al., 2009).

Changes in relative risk over different temporal scales is important to reflect in collision model predictions (see Chapter 5). Many species exhibit different patterns of activity, as do humans, and thus collision risk will ultimately vary hourly (e.g Joyce & Mahoney, 2001), seasonally (e.g da Rosa & Bager, 2012; Beaudry, Demaynadier & Hunter, 2010; Grilo, Bissonette & Santos-Reis, 2009), or both (e.g Mizuta, 2014; Morelle, Lehaire & Lejeune, 2013). Most data collected on carcasses or injured animals incorporates this information but does not report the uncertainty around reported times. This is problematic due to potential temporal lags between actual collision events and subsequent observations. Similar to issues resulting from spatial bias, models trained on data with inaccurate times will result in predictions that correspond to collection times rather than collision times. Smaller species are subject to more reporting error, as compared to larger species, as they often become unrecognisable after a few hours or days. Although kangaroos are large species, we did not incorporate hourly effects due to a lack of information about the reported times and associated uncertainty. Seasonal effects are less prone to this shortcoming due to larger temporal scales, however, we excluded seasonal effects from this analysis as kangaroos do not exhibit seasonal patterns.

Our road segment-based datasets were represented by ‘zeros’ or truncated counts of collisions as ‘ones’ based on the mathematical framework used for modelling. This was appropriate considering that more than one collision on a road segment rarely occurred in the data. More importantly, we were concerned about potential duplicate records arising from different data sources (i.e. the same collision reported twice). But it is beneficial to maximise the use of available collision data and some studies take advantage of count data to model and predict wildlife-vehicle collision risk (e.g.
Cserkesz et al., 2013), and explicitly reconcile disparate data sources within the model (e.g. Lao et al., 2011a). However, duplication of recorded collision events or carcass observations may affect model predictions and should be scrutinised. If duplicates could be identified reliably, then those records might help to examine biases in the reported collisions in a way that is similar to using multiple observers to estimate detection probabilities (see Nichols et al., 2000).

At time of writing, we are not aware of a unified data storage system that is accessible to and used by governments, organisations, companies and individual citizens to catalogue wildlife-vehicle collisions anywhere in the world. Although this study is limited to the State of Victoria in south-east Australia, the results and general discussion suggest qualities that a centralised storage system should have and we recommend the development and implementation of a WVC database to support future modelling and prediction efforts. Although systems maintained at the global scale, such as the Global Biodiversity Information Facility, may reduce potential redundancies and ambiguities, national systems are also useful to achieve this objective. Of paramount importance is uniting the currently disparate sources of data with an archive that is both comprehensive and systematic. Such a repository will enable informed analyses of WVC, with improved model calibration, to ultimately make more reliable predictions of wildlife-vehicle collision risk.
7

General Discussion
7.1 Synthesis of work

In this thesis, I presented a modelling framework for determining wildlife-vehicle collision risk that can be used across taxa and locations. Drawing on well-established concepts of road ecology and risk theory, this conceptual framework is the first of its kind to address the disparate factors that contribute to wildlife mortality on roads. Clevenger et al. (2015) has recently identified this critical need in the discipline of road ecology. Perhaps the most useful extension to previous collision modelling work is the grouping of predictors that influence wildlife-vehicle collision risk into two categories that more directly relate to management activities. Most mitigation measures used by transport managers control either animal presence on a transportation network (e.g. fencing) or the operation of vehicles (e.g. speed) (Glista, DeVault & DeWoody, 2009). Using species occurrence to represent exposure and traffic volume/speed to represent hazard creates variables that can be more readily interpreted (see Figure 2.1). Further, all other predictors used in past wildlife-vehicle collision modelling may be incorporated into the sub-models in each respective group, thereby reducing uncertainty about potentially confounding effects. From this simplified conceptual framework, managers can compare the relative influences of exposure and hazard, as well as the uncertainties of each one. If costs of mitigation are known for each category, the conceptual framework is useful for performing cost-benefit analyses – this was explored in Chapter 5.

Another useful development, applicable to previous work that uses generalised linear regression to model wildlife-vehicle collisions, is the choice of link function. Previous studies have used a logit link (i.e. logistic regression, see Seo et al., 2015; Gundersen & Andreassen, 1998), meaning their parameters and predictions cannot be applied to different spatial scales. This thesis is the first study to employ the use of a complementary log-log (cloglog) link function for modelling wildlife-vehicle collision events. This is important because the choice of a link function has theoretical underpinnings which are normally based on assumptions about how events arise in a system. I used the cloglog link as it assumes that collisions arise from a Poisson process which is reasonable for collision events occurring over a domain bounded by time. Another attractive feature of the cloglog is that it is not influenced by the spatial scale of the observations (e.g. road or track length) as it considers the rate of collision events per unit area or time. Model predictions can therefore be re-scaled to compare models fitted to different datasets. In this thesis, it was possible to compare parameter estimates for different models as they were invariant to the spatial scale of sampling (i.e., the slope parameters were unaffected by the sampling area, and the intercept parameters differed by the log of the ratio of the two areas). This demonstrates a robust and flexible modelling method that will enhance new and existing road ecology work. Moreover, these modelling methods are also applicable to other research involving ecological risk as discussed further in Section 7.5.

Beyond demonstrating the broad utility of the conceptual framework, my thesis also highlighted three key results. First, higher speeds of moving vehicles was a considerable hazard for wildlife having access to roads or railways. Chapters 3 and 4
demonstrated similar patterns of collision risk with respect to traffic speed for seven mammals. Nearly all species had rising marginal increases in risk with increasing speed. Collision risk also increased exponentially with increasing train speed (Chapter 5), suggesting this phenomenon is not isolated to road networks. These results were consistent with the broader literature on wildlife-vehicle collision analysis, and also other collision analyses involving pedestrians and vehicles (Rosén & Sander, 2009) and only vehicles (Liu & Popoff, 1997). It should be noted that although mitigation to reduce animal proximity to, or exclude entirely from, transportation networks is costly (Huijser et al., 2009) – and modelling results suggest that speed reductions may contribute to reduced collision risk – reducing posted speeds is often counter to the main directive of transportation agencies. In particular, major highways are designed to be safe and efficient transportation routes and to minimise dangerous activities (e.g. mixes of slow and fast moving vehicles). The conceptual framework I have developed in this thesis is able to assist managers to make more informed decisions about road design and upgrades by simulating the effects on wildlife-vehicle collision risk using different levels of traffic volume and speed. Further, the predictions of risk from the conceptual framework may be used by technology to influence motorist behaviour (e.g. voluntary speed reductions) in areas, or at times, of high risk.

Second, temporal variation was important to incorporate into models that predict collision risk. Both animal and human activities have distinct patterns that correspond to time. In Chapter 5, I introduced a function into the model framework that accounted for cyclical patterns of animal activity at the hourly scale. This was a statistically significant variable and useful for extending predictions of risk from the spatial domain to the spatio-temporal domain. Temporal variation is rarely accounted for in predictive modelling of wildlife-vehicle collisions (but see Beaudry, Demaynadier & Hunter, 2010; Gundersen & Andreassen, 1998; Neumann et al., 2012). The function used in Chapter 5 generalises to any species, contrasting with existing work that only characterises expected activity patterns for single species. Further, the function also used sunrise and sunset times (relative to geographic location) to consider time of day with respect to ambient light – daylight being a factor known to influence animal activity.

Uncertainty is one reason that modellers who analyse collisions may be reluctant to utilise fine resolution temporal data (e.g. hourly scale). Indeed, all of the collision data used in Chapters 2 to 4 and 6 had temporal information at an hourly scale, however, differences between actual collision times and reported collision (or carcass collection) times were difficult to determine. Due to this uncertainty, I chose to exclude this data for modelling and prediction on the road networks. Patterns were evident from the histograms of reported collisions for all six species (Figure 7.1) and suggested temporal components may be useful for modelling and predicting collisions if the uncertainty can be reduced or accounted for. In contrast, the data used in Chapter 5 had a high degree of certainty around the collision times reported by the train drivers. This was likely due to the systematic (i.e. standardised forms) and procedural (i.e. obliged to report collisions) nature of train operations. Reporting of road collision events are not subject to the same stringencies thereby making it more necessary to document
uncertainty in the data and represent it in the modelling. One strategy would be to model both the collision events and the reporting. To do this would require some collateral data to help calibrate possible temporal variation in reporting rate so that the pattern of temporal variation in collisions could be elucidated. Precedents exist in ecology, such as state-space models that consider both biological processes and observation processes and account for biological stochasticity and measurement error. These include occupancy detection models (e.g. MacKenzie et al., 2002) and models of population dynamics that include imperfect observation (e.g. MacKenzie et al., 2009).

Lastly, collision models and predictions are likely to benefit from increases in data quantity and quality. Chapter 6 introduced the first known study in the discipline of road ecology to compare the effects of combining data from different data sources on collision risk modelling and predictions. This complements work that evaluates the use of citizen science to collect collision data (Paul et al., 2014; Dwyer et al., 2016), and proposes technological advances for collecting data (Olson et al., 2014). Moreover, these studies call for the development of a centralised data storage system to improve road ecologists’ understanding of where and when collisions occur. As demonstrated in Chapter 6, as more disparate collision datasets were combined together, model performance and calibration improved. Prospects for centralising collision data are discussed in a subsequent section.
7.1 Synthesis of work

(a) Eastern Grey Kangaroo
(b) Common Brushtail Possum

(c) Common Ringtail Possum
(d) Swamp Wallaby

(e) Wombat
(f) Koala

Figure 7.1: Histograms showing total collisions (reported between the years 2010 and 2015) by hour for six Australian mammal species. Note, records indicate the time that wildlife-vehicle collision events were reported and may not accurately reflect actual times due to reporting lags.
7.2 Benefits of spatial modelling

Most, if not all, environmental problems occur in a geographic context. Therefore, research involving ecological phenomena and issues often involves spatial analyses and outputs; such as species distribution modelling (e.g. Elith & Leathwick, 2009). The use of spatial modelling in this thesis was fundamental to its intended purpose – to represent collision risk across large transportation networks. Further, the use of spatial modelling in this work also supports its future application as a decision support tool for management or public outreach.

Currently, there is a remarkable amount of publicly-accessible spatial data available on the Internet (Ma et al., 2015) and it is reasonable to expect increases in future years. Data derived by scientific or government organisations often include additional metadata that detail reliability and data-generating procedures. In this thesis, nearly all of the predictor data used in modelling were derived from the public domain. This allows methods to be reproducible and transferable; for example, I developed several of the predictors in Chapter 4 using the same source of satellite-derived data (see Appendix C) but for different geographic locations.

The use of geographic information systems (GIS) to catalogue and view spatial features is widespread in management. Contemporary versions of these software applications incorporate functionality to perform spatial analyses. As most managers are familiar with spatial data, I employed spatial methods in this thesis that would easily integrate into GIS software. For example, all computer code was written in R (R Core Team, 2016), statistical software that can be called to perform operations on spatial data from within programs such QGIS (www.qgis.org). Further, I developed computer code to convey collision risk as outputs that managers would be able to most readily utilise (e.g. maps). Generalised methods for analysing and predicting wildlife-vehicle collisions have not been available in road ecology and, therefore, systems to support management have not been developed. Moreover, some tools that have been proposed for analysing collisions utilise proprietary software and therefore limit potential uptake and use. Open-source applications (that interface well) are essential to maximise the availability of methods to management organisations. This broad reach will make the analysis of wildlife-vehicle collision more accessible, potentially creating a larger impact on the problem.

Conveying spatial information is an effective communication tool (Kingston et al., 2000); humans respond to visual stimuli. Most of the general public is familiar with geographic information through daily exposure to maps and other spatial data. Spatial data and analyses are natural candidates for data visualisation (O’Sullivan & Unwin, 2014), a technique used to communicate information through the use of visual objects, and may increase public engagement through interactive content. Although the spatial outputs included in this thesis are static, the methods used to create them are easily adaptable to interfaces that enable users to perform analysis on collision risk by varying model parameters.
7.3 Limitations of models and caveats

Models are simplified abstractions of reality and therefore subject to error (Burgman, 2005). Models are also useful as conceptual tools for organising complex problems and supplementing decision-making. However, care must be taken when using models; although they may be internally logical, models are only as good as the assumptions, data, and specifications they are based on. Management decisions based on inferences from a mis-specified model or a model trained on poor data can have negative consequences (e.g. ineffective outcomes or costs). Levins (1966) asserts that models can only be built to maximise two of three characteristics, namely generality, realism, and/or precision. The primary intent of the conceptual model framework introduced in this thesis was to maximise generality, or applicability to any ecological issue involving threats to wildlife (from humans and related hazards, such as use of vehicles). The model also exhibits realism in its theoretical treatment of collisions resulting from the coincidence of animals and moving vehicles in space and time. Much of the literature in road ecology that focuses on collision modelling operates on this assumption (Forman et al., 2003; Gunson, Mountrakis & Quackenbush, 2011), but does not make this explicit in the modelling methods. Therefore, it is difficult to draw inferences from the models to directly support management. For example, high collision risk may be predicted with variables that are not logically related to collisions, such as tree cover.

One strategy to explore model uncertainty is to perform collision modelling analyses using Bayesian inference, which allows prior information to be incorporated into a model (McCarthy, 2007). In Bayesian analyses, posterior probability distributions represent the uncertainty that remains after combining prior information with new data. With respect to the work presented in the thesis, this may involve using expert-derived species distribution data as priors in the species occurrence models (see Murray et al., 2009) or expert judgements on the reliability of the predictions from the occurrence models as priors in the collision model. For example, I had experts in macropod ecology examine the distribution of kangaroos predicted by the species occurrence model. Although there were some minor discrepancies, the overall patterns were realistic; but this information was not included in the final modelling. Since managers may base decisions about mitigation on the relative strength of exposure (animal presence) or hazard (vehicle movement), it may be useful to incorporate relative confidence, based on both prior information and the data, into model predictions and decisions that can reflect different risk tolerances.

The data used to train the collision model in this thesis are referred to as ‘presence-only’ data which require special assumptions and treatments (Warton, Renner & Ramp, 2013). True prevalence is not able to be inferred from presence-only data without a systematic and exhaustive search of the entire spatial domain with 100% detectability – if absences were recorded, this would become ‘presence-absence’ data (see Hastie & Fithian, 2013). Thus, the conceptual model in this thesis makes assumptions about prevalence based on the selection of background data (used to represent road segments where no collisions occurred). In this respect, predictions from the
models are relative to reporting rates, as opposed to absolute. In Chapter 2, I randomly selected background data that was equivalent to twice that of reported collisions resulting in predictions that represented a higher relative rate of collisions on road segments. Although this is still informative for ranking road segments based on relative risk, the risk is not directly comparable to total reported collisions across the entire network. In contrast, Chapters 3 to 6 all use the entire set of linear segments without reported collisions as background sets. In these cases, the models will predict reported collision rates, but actual collision rates will be higher, depending on the degree to which collisions remain unreported. A particular special case may be in Chapter 5 where the reporting rates were assumed to be quite good and, therefore, the modelling dataset may better reflect of true collision rates.

Another potential issue of presence-only data is sample selection bias which refers to the manner in which data are collected and reported (Phillips et al., 2009). In a spatial context, this may bias model predictions by over-representing and under-representing collision risk due to patterns of reporting, such as proximity to urban areas. Recent work in species distribution modelling has suggested the utility of spatial and temporal cross-validation for improving model performance (Wenger & Olden, 2012; Roberts et al., 2017). These methods have important implications for reducing bias and ensuring the generality of model predictions. The work in this thesis does not incorporate these methods, however, doing so would improve the analysis and resulting outputs.

7.4 Novel future research

In this thesis I presented several methodological advances not employed in previous wildlife-vehicle collision research, but also identified several important areas for future developments. Species traits, such as body size, are becoming increasingly employed in mechanistic modelling to predict species niches (Kearney & Porter, 2009), and also inform potential prediction errors (e.g. Seoane et al., 2005). Although risk rates did not perfectly correspond to mammal sizes in Chapter 3, body size may influence reporting rates as smaller species are less likely to be observed, and potentially reported, due to reduced detectability and faster rates of decay or predation. Body size can also play a role in susceptibility to collisions based on movement speed (Jaarsma, van Langevelde & Botma, 2006) or visibility to motorists. Body size could be included in the collision model as an additional exposure term, however, care must be taken when interpreting results due to potential confounding effects. For example, small body size may cause the animal to be more difficult to see by motorists and/or reduce road crossing time. Species behaviour may also influence exposure to wildlife vehicle collisions, such as flight response to vehicles (DeVault et al., 2014; Lee, Croft & Ramp, 2010) or feeding patterns (e.g. near roads). As with species traits, this information can be used as an additional exposure term. In road ecology, there are very few studies that compare predicted risk between species based on traits or behaviour (but see Litvaitis & Tash, 2008). As such, this could be a potentially powerful extension of the work presented in this thesis.
In a similar manner, incorporating additional characteristics of perceived threats (hazards) into the risk model is also useful. In this thesis, I represented hazard as the frequency and speed of vehicles across transportation networks. Only passenger vehicles were considered in the analysis, however, freight transport comprises approximately 20% of road usage (Australian Bureau of Statistics, 2016). Large trucks have considerable momentum and reduced stopping distances and are, therefore, potentially more threatening. One novel extension of the model would be to represent hazard with kinetic energy calculated by combining all masses and velocities of objects on each road segment. General collision risk literature has made use of this concept (Aarts & Van Schagen, 2006), but road ecology has yet to apply it to wildlife-vehicle collision modelling.

The work in this thesis models risk exclusively for common species, which raises the question of whether the methods are also applicable to rare species. Indeed, this would be an important feature for managers, and also more directly support conservation efforts. The theory underpinning the modelling framework enables analysis irrespective of species commonality, however, a primary difficulty is obtaining adequate amounts of data with which to train the models. Additional research that uses the framework to compare model performance between rare species and more common surrogate species, collected using a small targeted study, would be useful. If the results are deemed comparable, extrapolating to other areas where data on the surrogate species are available but data on rare species are more deficient would improve wildlife-vehicle collision modelling, prediction, and management. This may involve the use of population viability analysis which helps to understand changes in species persistence due to threatening processes (Rhodes et al., 2014). Although a few wildlife-vehicle collision modelling studies have addressed conservation of rare or endangered species (Dwyer et al., 2016), none have explored the use of surrogate species to suggest risk to rare species. This strategy has established utility in conservation biology (Caro & O’Doherty, 1999) and ecological modelling (Wenger, 2008).

Several concepts in wildlife-vehicle collision modelling parallel those used in pedestrian-vehicle collisions, a vastly broader literature. One common feature is that activity patterns of humans and animals are quite important in determining when and where collision risk may be high (Miranda-Moreno, Morency & El-Geneidy, 2011; Lao et al., 2011b). For the reasons discussed previously, the analysis in Chapters 2 to 4 and 6 only included spatial patterns and disregarded temporal variation altogether. However, Chapter 5 demonstrated the utility of accounting for both spatial and temporal patterns of animal activity. Future collision modelling work would benefit from additional development of sub-models that simulate the motion of animals (see Jaarsma et al., 2007) – borrowing concepts from pedestrian-vehicle collision literature (see Löhner, 2010).

### 7.5 Generalising to other threats to wildlife

The conceptual framework developed in this thesis may be used to analyse risk for other human-wildlife interactions, thereby increasing its potential contribution to
other environmental research questions. Its modular nature allows analysis of other adverse risks to wildlife (Figure 7.2) by replacing the hazard component with an alternative threatening process of interest, such as attacks from introduced predators or human poaching. Further, the exposure of species to hazards may be represented by data generated by alternative models; for example, movement patterns from connectivity analyses (e.g. McRae et al., 2008) or local abundances from population viability analyses (e.g. Beissinger, 2002). Bauduin et al. (2013) used a similar approach to determine collision risk between seafaring vessels and manatees, although did not emphasise the generality in their approach.

![Figure 7.2: Conceptual risk model framework with other sub-models or data sources. The modular framework allows analyses of other risks to wildlife by replacing the hazard or exposure components with alternative processes (or data) of interest.](image)

The modularity of the conceptual framework in this thesis allows possibilities for future research projects. Any human-wildlife conflict can be conceptually framed to analyse risks to wildlife from humans, risks to humans from wildlife, or both. Examples from Figure 1.1 in Chapter 1, and others, are listed in Table 7.1.
Table 7.1: Examples of human-wildlife conflicts and conceptualised parameters. Each situation has a description of risk, exposure, and hazard (corresponding to data) for use in the conceptual model introduced in Chapter 2.

<table>
<thead>
<tr>
<th>Human-wildlife Conflict</th>
<th>Risk</th>
<th>Exposure</th>
<th>Hazard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elephants raiding agricultural crops</td>
<td>Damaged crops</td>
<td>Spatial extent of crops that elephants may target</td>
<td>Spatial distribution and movements of elephants</td>
</tr>
<tr>
<td>Raccoons scavenging rubbish bins</td>
<td>Nuisance and potential economic loss</td>
<td>Spatial distribution of unsecured rubbish bins with high scrap food loads</td>
<td>Spatial distribution and activity patterns of raccoons</td>
</tr>
<tr>
<td>Wildlife-vehicle collisions</td>
<td>Collision between wildlife and moving vehicles</td>
<td>Distribution and movements of wildlife</td>
<td>Movements of vehicles</td>
</tr>
<tr>
<td>Illegal poaching of rhinos</td>
<td>Rhino mortality</td>
<td>Spatial distribution of rhinos</td>
<td>Distribution and movements of human poachers</td>
</tr>
<tr>
<td>Mountain lion attacks on humans</td>
<td>Human injury or death</td>
<td>Spatial distribution of humans</td>
<td>Spatial distribution of mountain lions</td>
</tr>
<tr>
<td>Boat strikes to marine mammals</td>
<td>Marine mammal mortality</td>
<td>Spatial movements of marine mammals</td>
<td>Spatial movements of marine vessels</td>
</tr>
<tr>
<td>Domesticated dog attacks on possums</td>
<td>Possum mortality</td>
<td>Spatial distribution of possums</td>
<td>Spatial distribution and activity patterns of domesticated dogs</td>
</tr>
<tr>
<td>Electrocutions of flying foxes</td>
<td>Flying fox mortality</td>
<td>Spatial movements of flying foxes (based on fly-out trajectories and feeding trees)</td>
<td>Distribution of active high-voltage power lines</td>
</tr>
<tr>
<td>Bear incursions into homes</td>
<td>Property damage</td>
<td>Unsecured homes</td>
<td>Spatial distribution of bears</td>
</tr>
</tbody>
</table>

7.6 Potential applications of research

In an effort to maximise potential applications of this work to reducing wildlife-vehicle collisions, I have utilised publicly-accessible predictor data and open-source software tools, and developed reproducible computer code (Appendix A). This allows the methods to be used for developing additional resources to supplement management and engage the public. Given these qualities, there are several foreseeable future extensions of this work.

7.6.1 Centralised data collection and reporting system

Beyond offering incredible global interconnectedness, the information age has been marked by a massive amount of data generation that is publicly-accessible via Internet portals. One example is atlases of species occurrence data, such as the global biodiversity information facility (www.gbif.org), which holds nearly three-quarter of a billion records worldwide. Risk analysis to reduce human-wildlife conflicts would
benefit tremendously from the creation of an open-access, single-source repository of wildlife-vehicle collision records. Worldwide, wildlife-vehicle collision data are collected by governments, private-sector corporations, non-profit organisations, community groups, and individual citizen scientists. This information is stored in several different formats (e.g. electronic or hard copy) and rarely made publicly-accessible. Collision data that are made available through world-wide web portals are often read-only and do not provide access to detailed spatial information; often this is due to propriety or concerns about public use of unwarranted data.

I propose a centralised data system that not only catalogues collision records and metadata, but also provides other important functions (Figure 7.3). The system communicates with other data systems and external applications (shown in green and red, respectively, in Figure 7.3). This includes interfacing with existing data repositories to obtain information for analyses, and executing programs to model and predict wildlife-vehicle collision risk. The system also facilitates data collection via remote technology as shown in blue in Figure 7.3. An online portal or mobile device ‘app’ can enable users to provide information originating from any source (e.g. citizens, private-sector companies, governments). Most importantly, the system can report real-time wildlife-vehicle collision risk, based on model predictions, across geographic space to motorists (shown in gray in Figure 7.3).

**Figure 7.3:** Schematic diagram of centralised data collection and reporting system for wildlife-vehicle collisions. Arrows indicate directions of information flow. Additional collection of collisions data (in blue) is by both citizen scientists (top) and professionals (bottom).
7.6 Potential applications of research

7.6.2 Integration of thesis research with existing and new technology

In developed nations, technology pervades nearly every aspect of human lives. Every office desk is adorned with a computer and nearly all persons carry a mobile device (mostly smartphones). Mobile device use is also increasing rapidly in emerging economies. We use these tools to access information and to assist us with decision-making. Unbeknownst to a significant portion of the public, models are continuously running and recalibrating on servers throughout the world to deliver information to these devices. Some help to optimise our search results on a web page, whilst others determine efficient transportation routes. This approach to disseminating real-time, model-generated information can be applied to road safety, specifically wildlife-vehicle collisions, to provide decision support for managers or warnings to motorists. I envision my research to be easily extended to either application.

Decision support tools are used by managers in a wide range of fields (Shim et al., 2002). Figure 7.4 shows an unpublished web-based interface to the model framework presented in this thesis. Developed as a proof of concept, I wrapped the modelling methods with a graphical user interface (GUI) that allows managers to change values of species occurrence, traffic volume, and traffic speed on road segments and simulate expected collision risk across a network. Although this proof of concept is quite rudimentary, it illustrates the potential features that a decision support tool may provide to direct efforts on reducing wildlife-vehicle collisions.

Changing human behaviour is a difficult task and a large amount of sociological and psychological literature is devoted to it. Studies suggest that one important factor to precipitate human behaviour change is the convenience of attaining outcomes (i.e. least effort hypothesis proposed in human ecology, see Zipf, 1949). Further, technology has also been shown to be persuasive (Fogg, 2003; Lathia et al., 2013), and therefore, mobile devices may be an effective media for influencing human behaviour. As such, mobile phones and global-positioning units are ideal candidates for receiving relative collision risk (based on geographic location and time of day) from model predictions and disseminating that information to motorists. The dynamic nature of a warning application may exceed the utility of traditional warning devices, such as signage, which are considered largely ineffective (Bond & Jones, 2013). Smartphone applications that report locations of collisions (and also accept observations) exist and are in use today (Aanensen et al., 2009), however, do not utilise modelling mechanisms. In this respect, these applications are simply reporting mechanisms and, thus, may only be slightly more effective than road signs. Modelling risk and reporting it in real-time provides dynamic information to motorists that can be conveyed in understandable terms such as ‘relative likelihood of collision’ rather than dots (representing collisions) aggregated along a road segment.
7.7 Concluding remarks

Human-wildlife interactions, and thus wildlife-vehicle collisions, are a global phenomenon of immense proportion contributing to millions of wildlife deaths annually. Issues such as animal mortality on roads will be exacerbated by future human development as we increase our proximity to wildlife. To address these issues, we need to be strategic about where and when we mitigate. This thesis provided analytical tools for making better decisions about environmental management. It is my hope that the contents of this work will assist others in solving contemporary ecological issues.
References


References


R Core Team (2016) R: A language and environment for statistical computing.


Appendix A: Computer code
All *R* code used to perform the analyses in this thesis are archived and publicly available at *GitHub*. Note, most of the analyses utilised a spatially-enabled database (*PostGIS*, version 2.3) that is behind a university firewall and therefore cannot be accessed without appropriate credentials. For access to this data, I encourage the reader to contact me. The thesis work is organised by chapter and includes links to the respective online code.

### A.1 Links to *R* code used in Chapter 2

Prepare species occurrence model data:

Perform species occurrence modelling:

Create species occurrence plots:

Prepare traffic model data:

Perform traffic modelling:

Create traffic plots:

Perform collision modelling:

Perform alternative collision modelling:

Create collision plots:

### A.2 Links to *R* code used in Chapter 3

Prepare species occurrence model data:

Perform species occurrence modelling:

Create species occurrence plots:
A.3 Links to R code used in Chapter 4

Prepare collision modelling data:

Perform collision modelling:

Create collision plots:

---

A.4 Links to R code used in Chapter 5

Obtain train movement data:

Process train movement data:

Perform train collision modelling:

Create train collision plots:

Generate train model predictions:

---

A.5 Links to R code used in Chapter 6

Prepare validation model data:
Perform modelling and validation:

Create validation plots:
Appendix B: Species used in research
The following sections provide information on the species used in this research. Unless otherwise cited, all of the descriptions provided were synthesised from information found in Van Dyck & Strahan (2008). All of the images are used under the Creative Commons License, version 4.0.

B.1 Mule deer

Mule deer (*Odocoileus hemionus*) are placental mammals that belong to a family of hoofed species that annually shed their antlers. They are of considerable size, males averaging between seventy-five and one-hundred kilograms (females are approximately seventy-five percent the weight). The fur of mule deer changes colour seasonally, from blue-gray in winter to reddish-brown in summer. Mule deer travel in herds during migration, however, males are generally solitary. Their home ranges are influenced by habitat quality and season and may cover between 100–5,000 hectares. Due to limited ability to digest fibrous roughage, Mule deer require easily digestible forage that includes leaves, needles, stems, fruits/nuts, shrubs and grasses – both living and dead. Occurring across most of western North America, except in tundra, sub-tropic, and extreme desert regions, Mule deer are quite adaptable to most habitats, however, migrate seasonally. Mule deer are subject to population management through harvesting/hunting and culling, especially when high densities initiate human-wildlife conflicts. The proceeding information and more details about Mule deer can be found in Ferguson (2005).

B.2 Eastern Grey kangaroos

The Eastern Grey kangaroo (*Macropus giganteus*) is the second-largest native terrestrial mammal in Australia. Males can weigh up to eighty-five kilograms and have body lengths of over three metres (including tail). Despite the name, Eastern Greys may also have significant patches of light brown or cream colouring, especially on the underbelly. As with most other Australian mammals, they are marsupials and carry young in pouches. They travel in groups, called ‘mobs’, and have stable home ranges that vary between 30–1,600 hectares. Eastern Greys are grazing animals and consume primarily grasses. Their distribution is quite broad, extending across a significant portion of eastern Australia from Cape York to north-east Tasmania. This area covers a wide range of habitat types including woodlands, shrub and heathland, and also human-modified landscapes such as golf courses and urban parklands. If conditions are right, Eastern Grey kangaroos can increase to densities exceeding three individuals per hectare. Therefore, populations are often managed by government agencies through culling, sterilisation, and/or harvesting.
B.3 Swamp wallabies

The Swamp Wallaby (*Wallabia bicolor*) is the only member in genus *Wallabia* due to its unique physical and behavioural characteristics. It is a medium-sized marsupial, weighing between ten and twenty kilograms and well over a metre long (including tail). Swamp wallabies also go by the name 'black wallabies', possibly due to appearing much darker than other wallabies, but their fur is normally brownish-gray or reddish-brown. As they are strictly solitary animals, it is quite rare to observe Swamp wallabies in groups. Shrubs comprise the main source of food for Swamp wallabies, both native and exotic, although they also occasionally forage on underground fungi. They are widely distributed along the eastern coastline of Australia as four sub-species. Swamp wallabies are occasionally culled to control populations, but their course fur and small body size make them unattractive for commercial harvesting.

B.4 Wombats

Common wombats (*Vombatus ursinus*) are robust, medium-sized marsupial mammals weighing between twenty and forty kilograms and having a body length of around a metre. Unlike most other ground or tree-dwelling marsupials, wombats live in underground burrows (sometimes up to twenty metres long). Their fur color is grayish-brown and does not vary much between individuals. Common wombats are solitary and visit multiple burrows over several weeks. Fibrous grasses are the main food source of common wombats. Depending on the distribution of food resources, their home range will vary between two and eighty hectares. They occur predominantly in the south-east of Australia, beginning in the southern Queensland and continuing down through eastern Victoria. Although not officially managed wildlife, common wombats are often threatened by domesticated dogs and persecuted by humans.

B.5 Koalas

Koalas (*Phascolarctos cinereus*) are perhaps one of Australia's most iconic animal. Often mistakenly referred to a 'koala bears', koalas have no relationship to bears and are more closely related to wombats. They are the largest Australian arboreal marsupial, weighing between four and ten kilometres. Koalas in the north of the continent have short, light gray fur whilst their southern counterparts have longer grayish-brown fur. As they are solitary and territorial, it is uncommon to find several koalas in a single tree. The koala diet consists, almost exclusively, of eucalypt
foliage – nearly half of a kilogram per day. Home ranges of koalas, which vary based on food availability, are between one and two hectares. They have a broad distribution along the eastern and south-eastern Australian coastlines extending from the southern part of Cape York to just over the state border between Victoria and South Australia. For many reasons, including threats to persistence and charismatic status, koalas are a primary species involved in wildlife management operations in three states of Australia.

B.6 Brushtail possums

Common Brushtail possums (Trichosurus vulpecula) are small-medium marsupial mammals that are highly adapted to urban environments. Animals can weigh between one and five kilograms depending on geographic location, smaller sizes occurring in northern Australia and larger sizes in the south. Fur color also varies quite considerably by location but the most familiar is gray with a white underbelly and black brushy tail (hence the name). Possums are normally solitary and establish territories over relatively small home ranges. Their primary native diet is leaves, flowers, and fruits; although in highly modified landscapes, Brushtail possums have resorted to omnivorous and opportunistic additions such as bird eggs, human-discarded food scraps, and even invertebrates. Brushtail possums are widespread throughout Australia and occur in many habitats, albeit preferring dry eucalypt forests. In urban environments, their densities can exceed four individuals per hectare, thereby resulting in conflicts with human inhabitants.

B.7 Ringtail possums

Common Ringtail possums (Pseudocheirus peregrinus) are small marsupial mammals and often found in suburban gardens. They weigh less than one kilogram and are no more than seven-hundred centimetres long (half of which is tail). Ringtail possums use their distinctively curled tails as an effective ‘fifth limb’ similar to primates. They have relatively short fur that is brownish-grey in colour, with white on the underbelly and tip at the end of the tail. Ringtail possums are strictly vegetarian and typically forage on leaves, flowers, and fruit. Although spanning several different habitat types, the species only occurs along the eastern seaboard of Australia, with limited distribution inland of the Great Dividing Range, and within the state of Tasmania. Ringtail possums adapt well to urban environments and, whilst considered annoying to some residential gardeners, do not cause any significant issues.
Appendix C: Data sources used for research
The following sections provide information on the sources of electronic data used in this thesis. Whilst most of the data is publicly accessible, some require additional permission for acquisition and use; these are identified in the descriptions. Any data modifications required to conform with project requirements are also described.

C.1 Wildlife Victoria

Wildlife Victoria is a not-for-profit organisation that provides emergency response service for sick, injured, or orphaned native wildlife. It has operated under the vision of living in a community that cares about the welfare of Australian wildlife for nearly thirty years. The organisation uses a professional database system to record details of all incidents reported across the State of Victoria. This database has more than 100,000 records, most of which have spatial and temporal attributes. Data is obtainable from Wildlife Victoria but remains their sole intellectual property and all uses of the data are subject to approval (obtained for the work in this thesis). More information on Wildlife Victoria can be found at https://www.wildlifevictoria.org.au

C.2 California Roadkill Observation System

The California Roadkill Observation System (CROS) is a web-based system used to catalogue observations of wildlife killed on roads throughout the state. The system is maintained by the Road Ecology Center based at the University of California, Davis and allows members of the public to enter observations from the field. At this time of writing, the system has 52,925 observations of 421 unique species originating from 1,301 registered observers. More information on CROS can be found at http://www.wildlifecrossing.net/california

C.3 VicRoads

VicRoads is responsible for planning, developing, and maintaining more than 50,000 kilometres of major arterial roads in the State of Victoria. Further, the government organisation is also responsible for registering vehicles and licensing motorists to operate within the state. VicRoads hires independent cleaning contractors to remove debris from major sections of roads, including wildlife carcasses, and this information is recorded – albeit scarcely. More information on VicRoads can be found at https://www.vicroads.vic.gov.au

C.4 V/Line

V/Line is Australia’s largest regional public transport operator scheduling more than 1,700 train services between the urban centre of Melbourne and regional Victoria. V/line also maintains over 3,500 kilometres of railway used by passenger and freight
transport throughout the state. Collisions with large species are recorded by train drivers as they have implications for train maintenance and safety. Although this information is obtainable from V/Line, it remains their sole intellectual property and all uses of the data are subject to approval (obtained for this thesis). More information on V/Line can be found at https://www.vline.com.au

C.5 City of Bendigo

The City of Bendigo is a large town located in regional Victoria and the third most populated urban area in the State (>100,000 persons). The council is responsible for litter abatement in public spaces (e.g. nature strips) and responds to reports from citizens. Kangaroo roadkill are commonly reported and the council maintains records of locations and times of these incidents. This information may be requested from the public works department. More information on Bendigo can be found at https://www.bendigo.vic.gov.au

C.6 Crashstats

Crashstats is a web-accessible repository of statistics on road crashes maintained by VicRoads. These data are based on reports made to the Victorian Police and usually involve human injuries and/or deaths. More information on Crashstats, and access to the data, can be found at https://www.vicroads.vic.gov.au/safety-and-road-rules/safety-statistics/crash-statistics

C.7 Insurance Australia Group

Insurance Australia Group (IAG) is a large private-sector company that operates throughout Australia, New Zealand and Asia. They provide underwriting services for several motor-vehicle insurance agencies and therefore receive access to vast amounts of claim data resulting from collisions. One significant component of this data are records on collisions with wildlife, albeit spatial uncertainty is quite high and therefore records are aggregated to the town level. Although this information is obtainable from IAG, it remains their sole intellectual property and all uses of the data are subject to approval (obtained for this thesis). More information on IAG can be found at https://www.iag.com.au

C.8 Victorian Biodiversity Atlas

The Victorian Biodiversity Atlas (VBA) contains high-quality records of flora and fauna occurrences throughout the State and is maintained by the Arthur Rylah Institute, a scientific research branch of the Victorian State Government. Information is
C.9 Global Biodiversity Information Facility

The Global Biodiversity Information Facility (GBIF) is an international collaboration and government-funded open data infrastructure containing information about all types of life on Earth. It is the world’s largest online repository of information, detailing 1.6 million species collected over three centuries. The system provides spatial and temporal records of species occurrence based on observations from citizen scientists, researchers, automated monitoring programmes, and museum specimen archives. More information and data are available at http://www.gbif.org.

C.10 WorldClim

WorldClim is a collection of global climatic data. The data are represented as 1-km² raster grids and include information on bioclimatic conditions for the past, present, and future. Interpolated temperature and rainfall data are also included in the collection. Hijmans et al. (2005) details the methods used to construct the data. These data are made available at http://www.worldclim.org. This thesis uses WorldClim data on isothermality, mean temperature of the wettest quarter, precipitation of the driest month, seasonality of precipitation, and annual temperature range.

C.11 Moderate Resolution Imaging Spectroradiometer

The Moderate Resolution Imaging Spectroradiometer (MODIS) is a remote-sensing instrument carried on-board the two satellites, Terra and Aqua. These two satellites view the entire Earth’s surface every 1 to 2 days and improve our understanding of global dynamics and processes. This thesis uses one of the MODIS vegetation indices, the normalized difference vegetation index (NDVI), to represent relative quality of vegetation. More information and data are available at https://modis.gsfc.nasa.gov. These data are provided in 16-day intervals and at multiple spatial resolutions, however, 1-km² resolution (MOD13A2) is used in this thesis. To represent the seasonal change in quality of vegetation (greenness variable), I averaged the differences between January and July measures of NDVI over a ten-year period (2000–2010) that spanned drought and non-drought conditions in Australia.
C.12 Defense Meteorological Satellite Program

The Defense Meteorological Satellite Program (DMSP), operated by the National Aeronautics and Space Administration (NASA) and National Oceanic and Atmospheric Administration (NOAA) records global nighttime artificial lighting. The data for all available satellite-based, smooth resolution imagery are used to produce cloud-free composites. Sunlit and glare data are excluded based on solar elevation angles. More information and data are available at https://ngdc.noaa.gov/eog/dmsp.html. In this thesis, I used data for the year 2013.

C.13 Australian Grassland and Rangeland Assessment by Spatial Simulation

The Australian Grassland and Rangeland Assessment by Spatial Simulation (Aussie GRASS) is used to monitor, at regional scales, pasture degradation and recovery using modelling that accounts for daily rainfall, evaporation, temperature, vapour pressure and solar radiation. It also accounts for soil and pasture types, tree and shrub cover, and abundance of domestic livestock and other key herbivores. The methods are detailed in Carter et al. (2003). Although this information is obtainable from the Queensland Government, it remains their sole intellectual property and all uses of the data are subject to approval (obtained for this thesis). Information on Aussie GRASS can be found at https://www.longpaddock.qld.gov.au/about/researchprojects/aussiegrass

C.14 Geoscience Australia

Geoscience Australia provides information and scientific data on the geology and geography of Australia. Data can be downloaded from http://www.ga.gov.au/data-pubs. I used a 9-second digital elevation model (DEM) to represent elevation for Australia in this thesis; methods are detailed in Hutchinson & Dowling (1991) and Hutchinson et al. (2011). Slope aspect values were derived from the DEM by using a geographic information system (GIS).

C.15 United States Geologic Survey

The United States Geologic Survey (USGS) provides information and scientific data on the geology and geography of the United States with select extended datasets describing global features. More information and data are available at https://www.usgs.gov. I used data from the USGS to represent two variables in the thesis. Hydrological flow accumulation was obtained from HydroSHEDS (https://hydrosheds.cr.usgs.gov), a mapping product that provides hydrographic information at regional and global scales. This data was available at 15 arc-second resolution and was projected and re-scaled to a 1-km² resolution. I used Shuttle Radar Topography Mission (SRTM) 3
arc-second, void-filled elevation data (https://lta.cr.usgs.gov/SRTM1Arc) to generate elevation variables for both California and Australia in this thesis.

C.16 Global Land Cover Facility

The Global Land Cover Facility (GLCF) provides data and products to describe global environmental systems and is based on remotely-sensed satellite data with a focus on land cover at different scales. More information and data are available at http://glcf.umd.edu. I obtained a 30-metre resolution tree cover layer containing estimated percentage of horizontal ground in each pixel covered by woody vegetation greater than 5 meters in height; originally derived from Landsat Vegetation Continuous Fields (VCF) – details in Sexton et al. (2013). For the tree cover variable used in this thesis, I spatially aggregated the data to a 1-km\(^2\) resolution.

C.17 VicMap

The VicMap collection of electronic spatial information is maintained by the Victorian Government and contains GIS data on many geographic features used in urban and regional planning and research. Information on these offerings is provided at http://www.depi.vic.gov.au/forestry-and-land-use/spatial-data-and-resources/vicmap. For this thesis, I obtained VicMap spatial data for roads, water features, and administrative boundaries such as towns. Using these features, I calculated distances to towns, roads, and linear waterways with geoprocessing tools and represented the outputs as 1-km\(^2\) resolution spatial grids.

C.18 SoilGrids

SoilGrids is a system for global soil mapping based on predictions from machine learning and statistics, maintained by the International Soil Reference and Information Centre (ISRIC). The methods are described in Hengl et al. (2014) and data can be obtained at http://www.isric.org/content/soilgrids. Soil predictions are made at a 1-km\(^2\) resolution scale which I use to represent variables pertaining to soil density, conductivity, and content (e.g. sand or clay) in this thesis.

C.19 Australian Bureau of Statistics

The Australian Bureau of Statistics (ABS) is a government agency that provides data and reports on economic, social, population and environmental topics. More information and data are available at http://www.abs.gov.au. Data obtained from the ABS were joined to spatial data in this thesis to create a population density variable. I used pycnophylactic interpolation (see Tobler, 1979) to obtain pixel-based population estimates whilst accounting for irregular administrative boundaries (e.g. towns).
The United States Census Bureau is the leading source for demographic and economic data about the United States which guides a significant portion of government spending (≈400 billion USD). Information and data are available at https://www.census.gov. For the California-based work in this thesis, data obtained from the Census Bureau were joined to spatial data to create a population density variable. I used pycnophylactic interpolation to obtain pixel-based population estimates at 1-km$^2$ resolution (see Appendix C.19).
Appendix D: List of studies used for comparing modelling variables
## Table D.1: Frequency of variables used in wildlife-vehicle collision studies.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Total Studies Using Variable</th>
<th>Percent Representation (71 Studies)</th>
<th>Study Reference Number in Table D.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic Volume</td>
<td>39</td>
<td>54.93</td>
<td>4, 5, 6, 7, 9, 10, 14, 17, 18, 19, 21, 25, 27, 28, 29, 33, 34, 35, 36, 37, 39, 40, 43, 47, 49, 50, 54, 55, 59, 61, 62, 63, 64, 65, 66, 68, 69, 70, 71</td>
</tr>
<tr>
<td>Topography</td>
<td>26</td>
<td>36.62</td>
<td>2, 3, 8, 9, 10, 11, 13, 16, 20, 21, 23, 30, 32, 34, 36, 41, 44, 49, 51, 53, 54, 56, 57, 61, 62, 66, 69</td>
</tr>
<tr>
<td>Land Use</td>
<td>24</td>
<td>33.80</td>
<td>2, 3, 10, 11, 17, 19, 20, 21, 23, 24, 26, 32, 34, 35, 36, 44, 49, 61, 63, 66, 67, 68, 69, 70</td>
</tr>
<tr>
<td>Traffic Speed</td>
<td>23</td>
<td>32.39</td>
<td>3, 6, 10, 12, 15, 19, 22, 29, 31, 32, 36, 37, 38, 41, 44, 47, 54, 58, 61, 63, 66, 68, 69</td>
</tr>
<tr>
<td>Vegetation Cover</td>
<td>21</td>
<td>29.58</td>
<td>2, 8, 21, 23, 27, 30, 33, 34, 39, 41, 47, 48, 51, 54, 63, 65, 66, 67, 68, 70, 71</td>
</tr>
<tr>
<td>Road Structure</td>
<td>17</td>
<td>23.94</td>
<td>2, 8, 9, 10, 11, 18, 20, 21, 23, 33, 36, 41, 42, 51, 61, 62, 63</td>
</tr>
<tr>
<td>Habitat Type</td>
<td>16</td>
<td>22.54</td>
<td>3, 8, 9, 10, 15, 23, 24, 36, 37, 46, 48, 57, 64, 65, 66, 71</td>
</tr>
<tr>
<td>Road Sinuosity</td>
<td>15</td>
<td>21.13</td>
<td>3, 8, 10, 13, 15, 23, 31, 33, 34, 37, 39, 41, 53, 63, 69</td>
</tr>
<tr>
<td>Distance to Water</td>
<td>14</td>
<td>19.72</td>
<td>3, 8, 10, 15, 21, 23, 32, 41, 47, 53, 56, 57, 66, 71</td>
</tr>
<tr>
<td>Development Density</td>
<td>13</td>
<td>18.31</td>
<td>2, 3, 11, 15, 20, 21, 34, 37, 41, 61, 63, 65, 68</td>
</tr>
<tr>
<td>Hydrology</td>
<td>13</td>
<td>18.31</td>
<td>3, 11, 20, 23, 33, 34, 35, 39, 61, 62, 63, 68, 69</td>
</tr>
<tr>
<td>Tree Density</td>
<td>11</td>
<td>15.49</td>
<td>20, 23, 34, 41, 44, 48, 53, 54, 61, 62, 66</td>
</tr>
<tr>
<td>Road Verge</td>
<td>11</td>
<td>15.49</td>
<td>1, 8, 14, 19, 21, 36, 41, 54, 63, 65, 69</td>
</tr>
<tr>
<td>Road Width</td>
<td>10</td>
<td>14.08</td>
<td>3, 8, 19, 20, 23, 28, 36, 37, 65, 69</td>
</tr>
<tr>
<td>Road Topology</td>
<td>10</td>
<td>14.08</td>
<td>8, 9, 13, 20, 21, 23, 37, 41, 63, 71</td>
</tr>
<tr>
<td>Temperature</td>
<td>10</td>
<td>14.08</td>
<td>12, 13, 15, 16, 22, 27, 33, 39, 56, 64</td>
</tr>
<tr>
<td>Rainfall</td>
<td>9</td>
<td>12.68</td>
<td>13, 15, 27, 33, 39, 53, 54, 56, 64</td>
</tr>
<tr>
<td>Season</td>
<td>9</td>
<td>12.68</td>
<td>17, 18, 38, 43, 44, 49, 58, 64, 70</td>
</tr>
<tr>
<td>Road Class</td>
<td>8</td>
<td>11.27</td>
<td>21, 26, 38, 44, 46, 58, 63, 68</td>
</tr>
<tr>
<td>Species Presence</td>
<td>7</td>
<td>9.86</td>
<td>2, 4, 11, 26, 55, 56, 60</td>
</tr>
<tr>
<td>Species Traits</td>
<td>7</td>
<td>9.86</td>
<td>5, 25, 26, 28, 29, 38, 40</td>
</tr>
<tr>
<td>Distance to Vegetation</td>
<td>7</td>
<td>9.86</td>
<td>3, 9, 41, 47, 57, 61, 71</td>
</tr>
<tr>
<td>Time of Day</td>
<td>7</td>
<td>9.86</td>
<td>17, 22, 25, 38, 45, 46, 58</td>
</tr>
<tr>
<td>Species Density</td>
<td>6</td>
<td>8.45</td>
<td>12, 33, 42, 50, 52, 66</td>
</tr>
<tr>
<td>Taxon</td>
<td>6</td>
<td>8.45</td>
<td>18, 37, 38, 42, 54, 58</td>
</tr>
<tr>
<td>Distance to Town</td>
<td>5</td>
<td>7.04</td>
<td>9, 10, 20, 21, 53</td>
</tr>
<tr>
<td>Road Segment Length</td>
<td>5</td>
<td>7.04</td>
<td>19, 36, 49, 50, 59</td>
</tr>
<tr>
<td>Ecotone</td>
<td>4</td>
<td>5.63</td>
<td>3, 10, 47, 66</td>
</tr>
<tr>
<td>Distance to Development</td>
<td>4</td>
<td>5.63</td>
<td>8, 23, 66, 71</td>
</tr>
<tr>
<td>Year</td>
<td>4</td>
<td>5.63</td>
<td>16, 26, 42, 43</td>
</tr>
<tr>
<td>Vegetation Condition</td>
<td>3</td>
<td>4.23</td>
<td>47, 56, 57</td>
</tr>
<tr>
<td>Distance to Structure</td>
<td>3</td>
<td>4.23</td>
<td>3, 23, 41</td>
</tr>
</tbody>
</table>
Table D.1: Frequency of variables used in wildlife-vehicle collision studies.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Total Studies Using Variable</th>
<th>Percent Representation (71 Studies)</th>
<th>Study Reference Number in Table D.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance to Infrastructure</td>
<td>3</td>
<td>4.23</td>
<td>10, 21, 61</td>
</tr>
<tr>
<td>Road Density</td>
<td>3</td>
<td>4.23</td>
<td>47, 61, 62</td>
</tr>
<tr>
<td>Sex Ratio</td>
<td>2</td>
<td>2.82</td>
<td>37, 38</td>
</tr>
<tr>
<td>Harvesting</td>
<td>2</td>
<td>2.82</td>
<td>61, 62</td>
</tr>
<tr>
<td>Habitat Connectivity</td>
<td>2</td>
<td>2.82</td>
<td>8, 60</td>
</tr>
<tr>
<td>Wind</td>
<td>2</td>
<td>2.82</td>
<td>12, 27</td>
</tr>
<tr>
<td>Climate</td>
<td>2</td>
<td>2.82</td>
<td>18, 26</td>
</tr>
<tr>
<td>Air Pressure</td>
<td>2</td>
<td>2.82</td>
<td>13, 27</td>
</tr>
<tr>
<td>Soil</td>
<td>2</td>
<td>2.82</td>
<td>27, 56</td>
</tr>
<tr>
<td>Moon</td>
<td>2</td>
<td>2.82</td>
<td>22, 45</td>
</tr>
<tr>
<td>Light</td>
<td>2</td>
<td>2.82</td>
<td>12, 31</td>
</tr>
<tr>
<td>Road Risk</td>
<td>1</td>
<td>1.41</td>
<td>60</td>
</tr>
<tr>
<td>Registered Vehicles</td>
<td>1</td>
<td>1.41</td>
<td>66</td>
</tr>
<tr>
<td>Management</td>
<td>1</td>
<td>1.41</td>
<td>1</td>
</tr>
<tr>
<td>Distance to Shelter</td>
<td>1</td>
<td>1.41</td>
<td>57</td>
</tr>
<tr>
<td>Road Condition</td>
<td>1</td>
<td>1.41</td>
<td>31</td>
</tr>
<tr>
<td>Visibility</td>
<td>1</td>
<td>1.41</td>
<td>8</td>
</tr>
<tr>
<td>Humidity</td>
<td>1</td>
<td>1.41</td>
<td>27</td>
</tr>
<tr>
<td>Cloud Cover</td>
<td>1</td>
<td>1.41</td>
<td>45</td>
</tr>
<tr>
<td>Snow</td>
<td>1</td>
<td>1.41</td>
<td>22</td>
</tr>
<tr>
<td>Month</td>
<td>1</td>
<td>1.41</td>
<td>42</td>
</tr>
</tbody>
</table>

Table D.2: Published studies using wildlife-vehicle collision modelling.

<table>
<thead>
<tr>
<th>ID</th>
<th>Author(s)</th>
<th>Year</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Andreassen, H.P., Gundersen, H. et al.</td>
<td>2005</td>
<td>Journal of Wildlife Management</td>
</tr>
<tr>
<td>2</td>
<td>Barnum, S., Rinehart, K. et al.</td>
<td>2007</td>
<td>ICOET Proceedings</td>
</tr>
<tr>
<td>3</td>
<td>Barthelmes, E.L.</td>
<td>2014</td>
<td>Biodiversity and Conservation</td>
</tr>
<tr>
<td>4</td>
<td>Bauduin, S., Martin, J. et al.</td>
<td>2013</td>
<td>Biological Conservation</td>
</tr>
<tr>
<td>5</td>
<td>Beaudry, F., Demaynadier, P.G. et al.</td>
<td>2010</td>
<td>Journal of Wildlife Management</td>
</tr>
<tr>
<td>6</td>
<td>Bissonette, J.A. and Adair, W.</td>
<td>2008</td>
<td>Biological Conservation</td>
</tr>
<tr>
<td>8</td>
<td>Clevenger, A.P. et al.</td>
<td>2015</td>
<td>Ecosphere</td>
</tr>
<tr>
<td>9</td>
<td>Clevenger, A.P., Chruszcz, B. et al.</td>
<td>2003</td>
<td>Biological Conservation</td>
</tr>
<tr>
<td>11</td>
<td>Cserkesz, T., Ottlecz, B. et al.</td>
<td>2013</td>
<td>European Journal of Wildlife Research</td>
</tr>
<tr>
<td>13</td>
<td>Dussault, C., Foulin, M. et al.</td>
<td>2006</td>
<td>Wildlife Biology</td>
</tr>
<tr>
<td>14</td>
<td>Eigenbrod, F., Heenar, S.J. et al.</td>
<td>2008</td>
<td>Biological conservation</td>
</tr>
</tbody>
</table>
### Table D.2: Published studies using wildlife-vehicle collision modelling.

<table>
<thead>
<tr>
<th>ID</th>
<th>Author(s)</th>
<th>Year</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>Fonnesbeck, C. <em>et al.</em></td>
<td>2008</td>
<td>Endangered Species Research</td>
</tr>
<tr>
<td>18</td>
<td>Garriga, N., Santos, X. <em>et al.</em></td>
<td>2012</td>
<td>Biodiversity and Conservation</td>
</tr>
<tr>
<td>19</td>
<td>Gkritza, K. <em>et al.</em></td>
<td>2013</td>
<td>Journal of Transportation Engineering</td>
</tr>
<tr>
<td>20</td>
<td>Gomes, L., Grilo, C. <em>et al.</em></td>
<td>2008</td>
<td>Ecological Research</td>
</tr>
<tr>
<td>21</td>
<td>Grilo, C., Bissonette, J.A. <em>et al.</em></td>
<td>2009</td>
<td>Biological Conservation</td>
</tr>
<tr>
<td>22</td>
<td>Gundersen, H. and Andreassen, H.P.</td>
<td>1998</td>
<td>Wildlife Biology</td>
</tr>
<tr>
<td>23</td>
<td>Gunson, K.E., Clevenger, A.P. <em>et al.</em></td>
<td>2009</td>
<td>Environmental Management</td>
</tr>
<tr>
<td>24</td>
<td>Gunson, K.E., Ireland, D. <em>et al.</em></td>
<td>2012</td>
<td>North-Western Journal of Zoology</td>
</tr>
<tr>
<td>25</td>
<td>Hels, T. and Buchwald, E.</td>
<td>2001</td>
<td>Biological Conservation</td>
</tr>
<tr>
<td>26</td>
<td>Hothorn, T., Brandl, R. <em>et al.</em></td>
<td>2012</td>
<td>PloS one</td>
</tr>
<tr>
<td>28</td>
<td>Jaarsma, C., Van Langevelde, F. <em>et al.</em></td>
<td>2007</td>
<td>Ecological Informatics</td>
</tr>
<tr>
<td>29</td>
<td>Jaarsma, C.F., van Langevelde, F. <em>et al.</em></td>
<td>2006</td>
<td>Transportation Research Part D: Transport and Environment</td>
</tr>
<tr>
<td>31</td>
<td>Joyce, T.L. and Mahoney, S.P.</td>
<td>2001</td>
<td>Wildlife Society Bulletin</td>
</tr>
<tr>
<td>32</td>
<td>Kanda, L.L., Fuller, T.K. <em>et al.</em></td>
<td>2006</td>
<td>The American Midland Naturalist</td>
</tr>
<tr>
<td>33</td>
<td>Klicker, U., Croft, D.B. <em>et al.</em></td>
<td>2006</td>
<td>Wildlife Research</td>
</tr>
<tr>
<td>34</td>
<td>Kolowski, J.M. and Nielsen, C.K.</td>
<td>2008</td>
<td>Biological Conservation</td>
</tr>
<tr>
<td>36</td>
<td>Lao, Y. <em>et al.</em></td>
<td>2011</td>
<td>Accident Analysis and Prevention</td>
</tr>
<tr>
<td>37</td>
<td>Lao, Y., Zhang, G. <em>et al.</em></td>
<td>2011</td>
<td>Accident Analysis and Prevention</td>
</tr>
<tr>
<td>38</td>
<td>Lee, E., Croft, D.B. <em>et al.</em></td>
<td>2010</td>
<td>Book Chapter</td>
</tr>
<tr>
<td>40</td>
<td>Litvaitis, J.A. and Tash, J.P.</td>
<td>2008</td>
<td>Environmental Management</td>
</tr>
<tr>
<td>42</td>
<td>Markolt, F. and Szemethy, L. <em>et al.</em></td>
<td>2012</td>
<td>North-Western Journal of Zoology</td>
</tr>
<tr>
<td>43</td>
<td>Mazerolle, M.J.</td>
<td>2004</td>
<td>Herpetologia</td>
</tr>
<tr>
<td>45</td>
<td>Mizuta, T.</td>
<td>2014</td>
<td>The Wilson Journal of Ornithology</td>
</tr>
<tr>
<td>46</td>
<td>Neumann, W., Ericsson, G. <em>et al.</em></td>
<td>2012</td>
<td>Biological Conservation</td>
</tr>
<tr>
<td>47</td>
<td>Ng, J. <em>et al.</em></td>
<td>2008</td>
<td>Human-Wildlife Interactions</td>
</tr>
<tr>
<td>53</td>
<td>Ramp, D., Caldwell, J. <em>et al.</em></td>
<td>2005</td>
<td>Biological Conservation</td>
</tr>
<tr>
<td>54</td>
<td>Ramp, D., Wilson, V.K. <em>et al.</em></td>
<td>2006</td>
<td>Biological conservation</td>
</tr>
<tr>
<td>56</td>
<td>Roger, E., Bino, G. <em>et al.</em></td>
<td>2012</td>
<td>Landscape ecology</td>
</tr>
</tbody>
</table>
Table D.2: Published studies using wildlife-vehicle collision modelling.

<table>
<thead>
<tr>
<th>ID</th>
<th>Author(s)</th>
<th>Year</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>57</td>
<td>Roger, E. and Ramp, D.</td>
<td>2009</td>
<td>Diversity and Distributions</td>
</tr>
<tr>
<td>58</td>
<td>Rowden, P., Steinhardt, D. et al.</td>
<td>2008</td>
<td>Accident Analysis and Prevention</td>
</tr>
<tr>
<td>59</td>
<td>Saeki, M. and Macdonald, D.</td>
<td>2004</td>
<td>Biological Conservation</td>
</tr>
<tr>
<td>60</td>
<td>Santos, S.M. and Lourenço, R. et al.</td>
<td>2013</td>
<td>PloS One</td>
</tr>
<tr>
<td>62</td>
<td>Seiler, A.</td>
<td>2004</td>
<td>Wildlife Biology</td>
</tr>
<tr>
<td>63</td>
<td>Seo, C., Thorne, J. et al.</td>
<td>2015</td>
<td>Landscape and Ecological Engineering</td>
</tr>
<tr>
<td>64</td>
<td>Shepard, D.B., Dreslik, M.J. et al.</td>
<td>2008</td>
<td>Copeia</td>
</tr>
<tr>
<td>65</td>
<td>Skorka, P., Lenda, M. et al.</td>
<td>2013</td>
<td>Biological Conservation</td>
</tr>
<tr>
<td>66</td>
<td>Snow, N.P., Porter, W.F. et al.</td>
<td>2015</td>
<td>Biological Conservation</td>
</tr>
<tr>
<td>68</td>
<td>Sudharsan, K., Riley, S.J. et al.</td>
<td>2009</td>
<td>Human Dimensions of Wildlife</td>
</tr>
<tr>
<td>69</td>
<td>Taylor, B.D. and Goldingay, R.L.</td>
<td>2004</td>
<td>Wildlife Research</td>
</tr>
<tr>
<td>70</td>
<td>Thurfjell, H., Spong, G. et al.</td>
<td>2015</td>
<td>Landscape and Urban Planning</td>
</tr>
</tbody>
</table>
Author/s:
Visintin, Casey

Title:
Modelling and predicting collision risks between wildlife and moving vehicles across space and time

Date:
2017

Persistent Link:
http://hdl.handle.net/11343/207990

File Description:
Completed_PhD_Thesis_Visintin

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.