Wildlife-train collisions: modelling and analysis in space and time

Casey Visintin1, Rodney van der Ree2, Michael A. McCarthy3

1Quantitative and Applied Ecology Group, School of BioSciences, University of Melbourne, Parkville, VIC 3010, Australia - Email: cvisintin@student.unimelb.edu.au

2Australian Research Centre for Urban Ecology, Royal Botanic Gardens Victoria and School of BioSciences, University of Melbourne, Parkville, VIC 3010, Australia - Email: rvdr@unimelb.edu.au

3Quantitative and Applied Ecology Group, School of BioSciences, University of Melbourne, Parkville, VIC 3010, Australia - Email: mamcca@unimelb.edu.au

Corresponding Author: Casey Visintin, School of BioSciences, Bldg 122 - Rm 106A, University of Melbourne, Parkville, VIC 3010, Australia - Email: cvisintin@student.unimelb.edu.au, Phone: +61 4 34424084

Running Title: WTC model with kangaroos

Word Count: XXXX  
 Summary: XXXX  
 Main Text: XXXX  
 Acknowledgements: XXXX  
 References: XXXX  
 Tables: XXXX  
 Figure Legends: XXXX

Number of Tables: 5

Number of Figures: 6

Number of References: XXXX

* 1. Summary

Collisions between and wildlife and vehicles has been widely studied, however, animal mortality from strikes by rail-based networks remains under-represented in the literature.

To assess the risk of collisions, we developed methods to quantify regional train movements in space and time, determine likelihoods of species occurrence, and fit a model to reported collision data.

Model results...

Predictions from the model can help managers decide where, when and how best to mitigate strikes.

* 1. Keywords

animal, framework, train, risk, species distribution model, speed, track, crepuscular, WTC

Introduction

Linear infrastructure that carries moving vehicles such as road and rail developments cause both direct and indirect disruptions to ecological systems (Seiler & Helldin, 2006; van der Ree et al, 2015). Although transportation networks support human civilisations by moving goods, expanding services, and enabling recreational activities, they also introduce environmental impacts that must be managed (Spellerberg, 1998). One of the most visible impacts is the direct mortality of wildlife resulting from strikes that occur on transportation networks (Forman et al., 2003).

Collisions between animals and moving vehicles is a common problem throughout the world (Litvaitis & Tash, 2008). Larger wildlife are more problematic as they often pose safety concerns in addition to animal welfare and conservation issues (Langley et al., 2006; Rowden et al., 2008). For example, deer-vehicle collisions on roads are well-studied in North America (Huijser et al., 2007; Romin & Bissonette, 1996) and Europe (Sáenz-de-Santa-María & Tellería, 2015; Seiler, 2004). Moreover, as under-developed countries create infrastructure to match the intensity of developed nations in North America and Europe, management of wildlife-vehicle collisions will become increasingly important.

Information about the spatial and temporal distribution and magnitude of wildlife-vehicle collisions is useful to managers (Mountrakis & Gunson, 2009). Information derived from empirical, simulated, or modelled data may help to plan more effective mitigation. For example, knowing a hotspot location along a transportation network for a particular species will assist managers to select and implement the most appropriate form of mitigation (e.g. animal exclusion or change in network activity). Statistical modelling has become a common method of predicting wildlife-vehicle collisions to inform management (Gunson et al., 2011).

The majority of wildlife-vehicle collision modelling deals with road networks (van der Ree et al, 2015), however, the problem extends to other forms of vehicular networks such as air (van Belle et al., 2007) and rail (Wells et al., 1999) operation. Regardless of the mode of transport, the modelling of collisions share some common attributes. First, the movements or presence of animals are often considered in the models which may include behaviour traits (Roger & Ramp, 2009). Second, the presence or movements of vehicles are also considered and may be grouped into a larger category of human behaviour as humans ultimately control speeds and trajectories of vehicles (Ramp & Roger, 2008). These distinctions have been established previously in the road ecology literature (see Forman et al., 2003).

Extensive rail networks with considerable activity exist on every continent in the world, and although broader ecological effects have been discussed (De Santo, 1993; Givoni, 2006) and analysed (Waller & Servheen, 2005), very few studies model wildlife-train collisions. Perhaps due to their size, moose have gained attention in both North America (Belant, 1995) and Europe (Gundersen & Andreassen, 1998) and have also been the focus of management (Andreassen et al., 2005). Deer have also been studied in Japan (Onoyama et al., 1998). In Oceania, train collisions with kangaroos (analogues to deer in road collision modelling) and other Australian species have yet to be modelled. We aim to develop a modelling framework that predicts the rate of collisions across the regional passenger train network in Victoria. We envisage these methods to inform rail operators of potential dangers as well as generalise to other species (e.g. wombats) and rail operations (e.g. freight transport).

* 1. Materials and Methods
     1. *Study Area*

We used a 1712-kilometre Victorian regional passenger rail network (operated by V/line, a government-owned corporation) in south-east Australia to conduct our study (Figure 1). Trains operate on all sections of the network between the hours of 4am and 2am with the largest volume occurring Monday through Friday between the hours of 7am and 9am and 4pm and 6pm. Regional train activity has been steadily increasing due to population growth in outer suburbs and small towns and more residents opting to commute into the Melbourne metropolitan area.

* + 1. *Data Preparation*

To organise our data and modelling, we overlaid a spatial grid of one square kilometre resolution on the rail network (Figure 2). Each grid cell was a modelling unit for species occurrence and a micro-site for quantifying the regional train movements and speeds.

Eastern grey kangaroos (Macropus giganteus, Shaw; "kangaroos" hereafter) are the most frequently struck wildlife and large enough to cause significant damage to trains. We obtained 439 kangaroo collision records spanning a six year period between 1 January, 2009 and 31 December, 2015 (V/line, unpublished data). Each record had a corresponding incident date, incident time, name of service line, and nearest fraction of a kilometre post. Using geographic information system (GIS) data on the regional rail network, we determined spatial coordinates (GDA94 MGA zone 55 projection) from the reported kilometre post and service line and uploaded them into the spatial database (Postgres version 9.6; PostGIS version 2.3.0).

*Species Occurrence*

We required distributional data for kangaroos and used species distribution modelling to predict relative likelihood of kangaroo occurrence. We emulated methods by Elith et al. (2008) to model and predict occurrence to each cell in the spatial grid for the State of Victoria. The model was trained on Victorian Biodiversity Atlas data sourced online (VBA, 2014) and used several environmental variables relating to the biology and behaviour of kangaroos (details in Visintin et al, 2016). To reduce the effects of sampling bias, two anthropogenic variables and spatial coordinates were also included in the models as predictors.

*Characteristics of Rail Network*

To determine train movements across space and time, we used a spatial database to interpolate locations and times of unique train routes from V/Line general transit feed specification (GTFS) data (Public Transport Victoria, accessed online 3 March, 2016). GTFS is a standard publishing format for public transport agencies for scheduling and spatial data. This data allows software developers to write applications for mobile devices that track and report the locations of public transportation. Our query returned the average 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.

*Temporal Variation*

We also considered the coincidence of peak periods of train movements with daylight by adding variables corresponding to a crepuscular (most active at dawn and dusk) functional form. Relative daylight intensity and duration of sunrise to sunset data were included to test a bimodal response of collision rate to hour of day across all seasons.

* + 1. *Statistical Modelling*

We adapted a single-species quantitative risk model (see Visintin et al, 2016) to fit and compare the relationship of species presence, characteristics of the rail network, and temporal phenomena to collision likelihood expressed as:

*cloglog(pijk) = β0 + β1 ln(****O****ijk) + β2 ln(****V****ijk) + β3 ln(****S****ijk) + β4 ln(****L****ijk) + β5 ln(****L****ijk)2 + β6 ln(****D****ijk) + ln(****K****ijk)* (1)

where *pijk = Pr(Yijk=1)* is the relative likelihood of a collision equal to one, **O** is species occurrence, **V** is average number of trains, **S** is average train speed, **L** is relative daylight intensity, **D** is duration of daylight, **K** is length of track, in a given grid cell **i** at hour **j** in month **k**. Prior to modelling, all explanatory variables were centred by subtracting their means. All predictors exhibited Pearson's product moment correlation coefficients of less than 0.4 using pairwise analysis.

We fit the data (n=291120) using maximum likelihood estimation. We specified a generalised linear model (McCullagh & Nelder, 1989) using a binomial distribution with a complementary log-log link on the linear predictor.

To assess model performance, we used 10-fold internal cross-validation. The data was randomly split into K=10 partitions (n~29112) with nine subsets used for model fitting and the remainder used for assessing accuracy. For each assessment, we obtained several performance metrics (Cox, 1989; Harrell, 1996; Metz, 1978; Miller, 1991; Somers, 1962) using the function *val.prob* in the R (version 3.3.1) package *rms*. We repeated this procedure for 100 iterations producing a total of 1000 sets of metrics and compared them with those from the model fit to all data (Figure 3).

Using the model fit on all data, we generated predictions under different management scenarios. Whilst holding all other predictors constant, we calculated the aggregated number of expected collisions in the study area for one-year for A) no change to operations, B) moderated train speeds in high kangaroo occurrence areas, and C) controlled kangaroo occurrence in areas with highest average speed of trains. Scenario B involved reducing the speeds of trains in grid cells with kangaroo relative occurrence likelihoods of 0.5 or above during the hours of 5am to 9am and 4pm to 8pm. We capped train speeds at 80 km h-1 for these areas. Relative kangaroo occurrence was reduced by approximately half in all grid cells with average train speeds of more than 120 km h-1 in scenario C. The values were modified 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.

* 1. Results
  2. Discussion

Model fit...

Train speed...

EGK presence...

Crepuscular activity and temporal patterns...

Detection issues...

Implied management...

Use of existing data...

Reduce animal presence (e.g. deterrents or exclusions) or reduce train threat (e.g. adjust schedules or speeds)...

Costs of management in sensitivity analysis...

* 1. Data Accessibility

Model Dataset - Archived on GitHub

R Code - Archived on GitHub

* 1. Acknowledgements

Sam Parsons, on behalf of V\Line, queried and provided the collision statistics used for the study. Nick Golding consulted on the mathematics and statistics used in the analysis. This project was supported by a University of Melbourne International Research Scholarship and the Australian Research Council Centre of Excellence for Environmental Decisions.

References

Andreassen, H. P.; Gundersen, H. & Storaas, T. (2005) The effect of scent-marking, forest clearing, and supplemental feeding on moose-train collisions The Journal of Wildlife Management, BioOne, , 69, 1125-1132

Belant, J.L. (1995) Moose collisions with vehicles and trains in northeastern Minnesota. Alces, 31, 45-52

Cox, D. R. & Snell, E. J. (1989) *Analysis of binary data* CRC Press

De Santo, R. S. & Smith, D. G. (1993) An introduction to issues of habitat fragmentation relative to transportation corridors with special reference to high-speed rail (HSR) Environmental Management, 17, 111-114

Elith, J.; Leathwick, J. R. & Hastie, T. (2008) A working guide to boosted regression trees Journal of Animal Ecology, 77, 802-81

Forman, R. T.T., Sperling, D., Bissonette, J. A., Clevenger, A. P., Cutshall, C. D., Dale, V. H., Fahrig, L., France, R., Goldman, C. R., Heanue, K., Jones, J. A., Swanson, F., Turrentine, T., & Winter, T. C. (2003) Road ecology: Science and solutions. Island Press, Washington, D.C.

Givoni, M. (2006) Development and impact of the modern High-speed train: A review Transport reviews, Taylor & Francis, , 26, 593-611

Gundersen, H. & Andreassen, H. P. (1998) The risk of moose Alces alces collision: A predictive logistic model for moose-train accidents Wildlife Biology, 4, 103-110

Gunson, K. E., Mountrakis, G. & Quackenbush, L. J. (2011) Spatial wildlife-vehicle collision models: A review of current work and its application to transportation mitigation projects.

Journal of Environmental Management 92: 1074-1082

Harrell, F. E.; Lee, K. L. & Mark, D. B. (1996) Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors  
Statistics in Medicine, 15, 361-387

Huijser, M. P.; Wagner, M. E.; Hardy, A.; Clevenger, A. P. & Fuller, J. A.  
(2007) Animal-vehicle collision data collection throughout the United States and Canada  
ICOET 2007 Proceedings: Wildlife and Terrestrial Ecosystems

Langley, R. L.; Higgins, S. A. & Herrin, K. B. (2006) Risk factors associated with fatal animal-vehicle collisions in the United States, 1995--2004 Wilderness & Environmental Medicine, Elsevier, , 17, 229-239

Litvaitis, J. A. & Tash, J. P. (2008) An approach toward understanding wildlife-vehicle collisions. Environmental Management 42: 688-697

McCullagh, P. & Nelder, J. A. (1989) *Generalized linear models* CRC press

Metz C.E. (1978) Basic principles of ROC analysis. Seminars in Nuclear Medicine 8(4): 283-298

Miller, M. E.; Hui, S. L. & Tierney, W. M. (1991) Validation techniques for logistic regression models Statistics in medicine, 10, 1213-1226

Mountrakis, G. & Gunson, K. (2009) Multi-scale spatiotemporal analyses of moose--vehicle collisions: a case study in northern Vermont International Journal of Geographical Information Science, 23, 1389-1412

Onoyama, K.; Ohsumi, N.; Mitsumochi, N. & Kishihara, T. (1998) Data analysis of deer-train collisions in eastern Hokkaido, Japan Data Science, Classification, and Related Methods, Springer, , 746-751

Ramp, D. & Roger, E. (2008) Frequency of animal-vehicle collisions in NSW in *Too close for comfort*, 118-126

Roger, E. & Ramp, D. (2009) Incorporating habitat use in models of fauna fatalities on roads Diversity and Distributions 15, 222-231

Romin, L. A. & Bissonette, J. A. (1996) Deer: vehicle collisions: status of state monitoring activities and mitigation efforts Wildlife Society Bulletin, JSTOR, 24, 276-283

Rowden, P., Steinhardt, D., & Sheehan, M. (2008) Road crashes involving animals in Australia. Accident Analysis & Prevention 40(6): 1865-1871

Sáenz-de-Santa-María, A. & Tellería, J. L. (2015) Wildlife-vehicle collisions in Spain European Journal of Wildlife Research, 61, 399-406

Seiler, A. (2004) Trends and spatial patterns in ungulate-vehicle collisions in Sweden Wildlife Biology, 10, 301-313

Seiler, A. & Helldin, J. O. (2006) Mortality in wildlife due to transportation. In: Davenport, J. & Davenport, J. L. (eds) The Ecology of Transportation: Managing Mobility for the Environment. Springer, Netherlands, pp. 165-189

Somers, R. H. (1962) A new asymmetric measure of association for ordinal variables American sociological review, 799-811

Spellerberg, I. (1998) Ecological effects of roads and traffic: a literature review. Global Ecology and Biogeography 7: 317-333

van Belle, J.; Shamoun-Baranes, J.; Van Loon, E. & Bouten, W. (2007) An operational model predicting autumn bird migration intensities for flight safety Journal of Applied Ecology, 44, 864-874

van der Ree, R., Smith, D. J. & Grilo, C. (2015) Handbook of Road Ecology. John Wiley & Sons Ltd, Chichester, UK

Visintin, C., van der Ree, R. & McCarthy, M. A. (2016) A simple framework for a complex problem? Predicting wildlife–vehicle collisions. Ecology and Evolution 6(17): 6409–6421

Waller, J. S. & Servheen, C. (2005) Effects of transportation infrastructure on grizzly bears in northwestern Montana Journal of Wildlife Management, 69, 985-1000

Wells, P.; Woods, J.; Bridgewater, G. & Morrison, H. (1999) Wildlife mortalities on railways: Monitoring methods and mitigation strategies Proceedings of the Third International Conference on Wildlife Ecology and Transportation , 85-88

* 1. Tables

Table 1: Predictor variables used in collision model. Note, prior to modelling all variables were centred by subtracting the means. The collision model includes the quadratic terms of LIGHT.

|  |  |  |
| --- | --- | --- |
| Variable | Description | Units |
| EGK | Relative likelihood of kangaroo occurrence in grid cell | – |
| TRAINS | Train frequency in grid cell | trains h-1 |
| SPEED | Mean train speed in grid cell | km h-1 |
| LIGHT | Relative intensity of ambient light in grid cell based on month | – |
| DAWNORDUSK | Hours between dawn and dusk in grid cell based on month | – |

Table 2: Summary of model fit using all data (n=291120). 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 | 0.00E+00\* |
| EGK | 0.61 | 0.06 | 10.46 | 0.00E+00\* |
| TRAINS | 0.01 | 0.09 | 0.16 | 8.71E-01 |
| SPEED | 3.62 | 0.31 | 11.53 | 9.29E-31\* |
| LIGHT | -0.65 | 0.11 | -5.82 | 6.02E-09\* |
| DAWNORDUSK | -1.87 | 0.17 | -10.85 | 2.06E-27\* |
| EGK | 0.25 | 0.07 | 3.76 | 1.73E-04\* |

Table 2: Summary of predicted collisions based on different management scenarios.

* 1. Figures

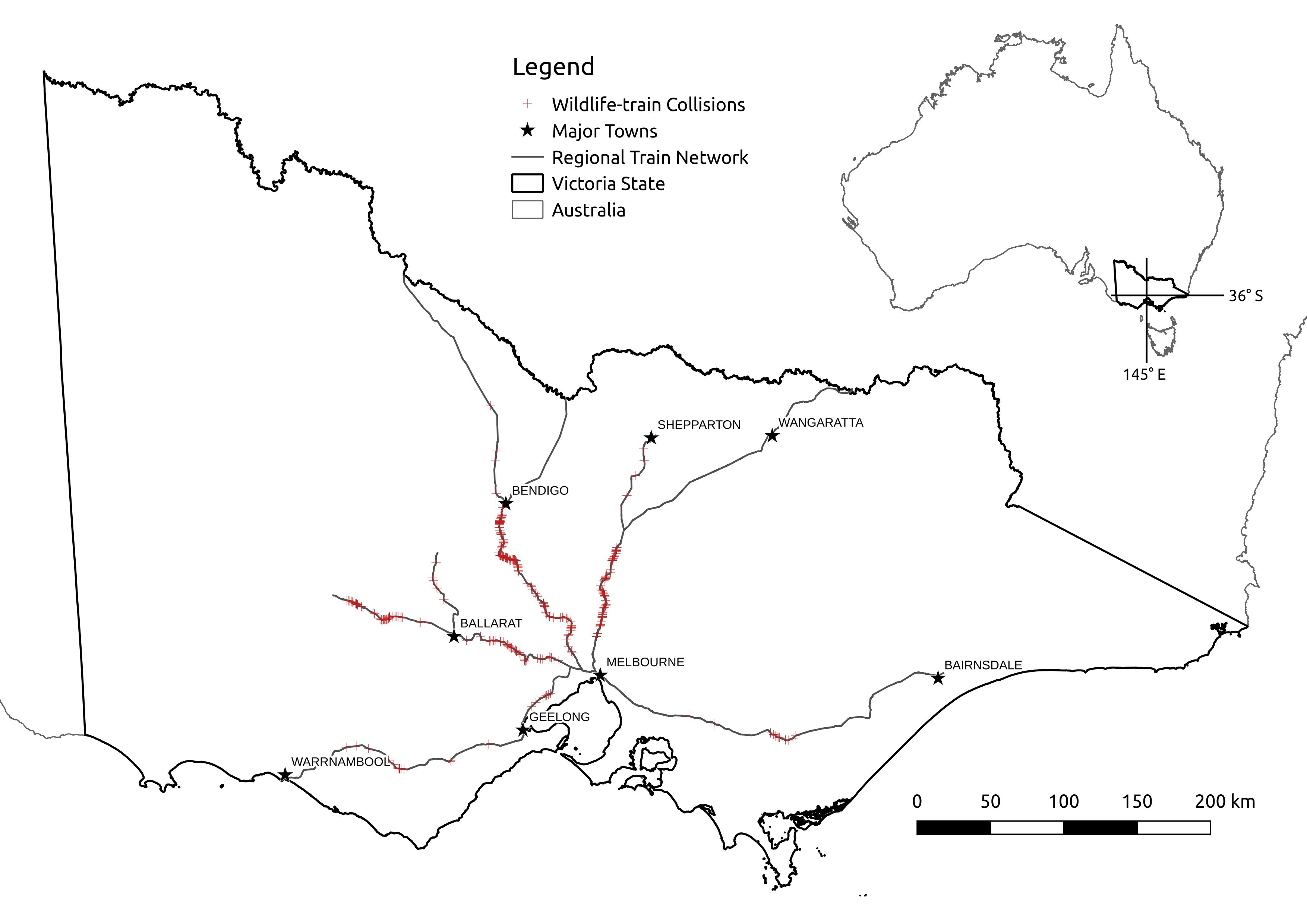
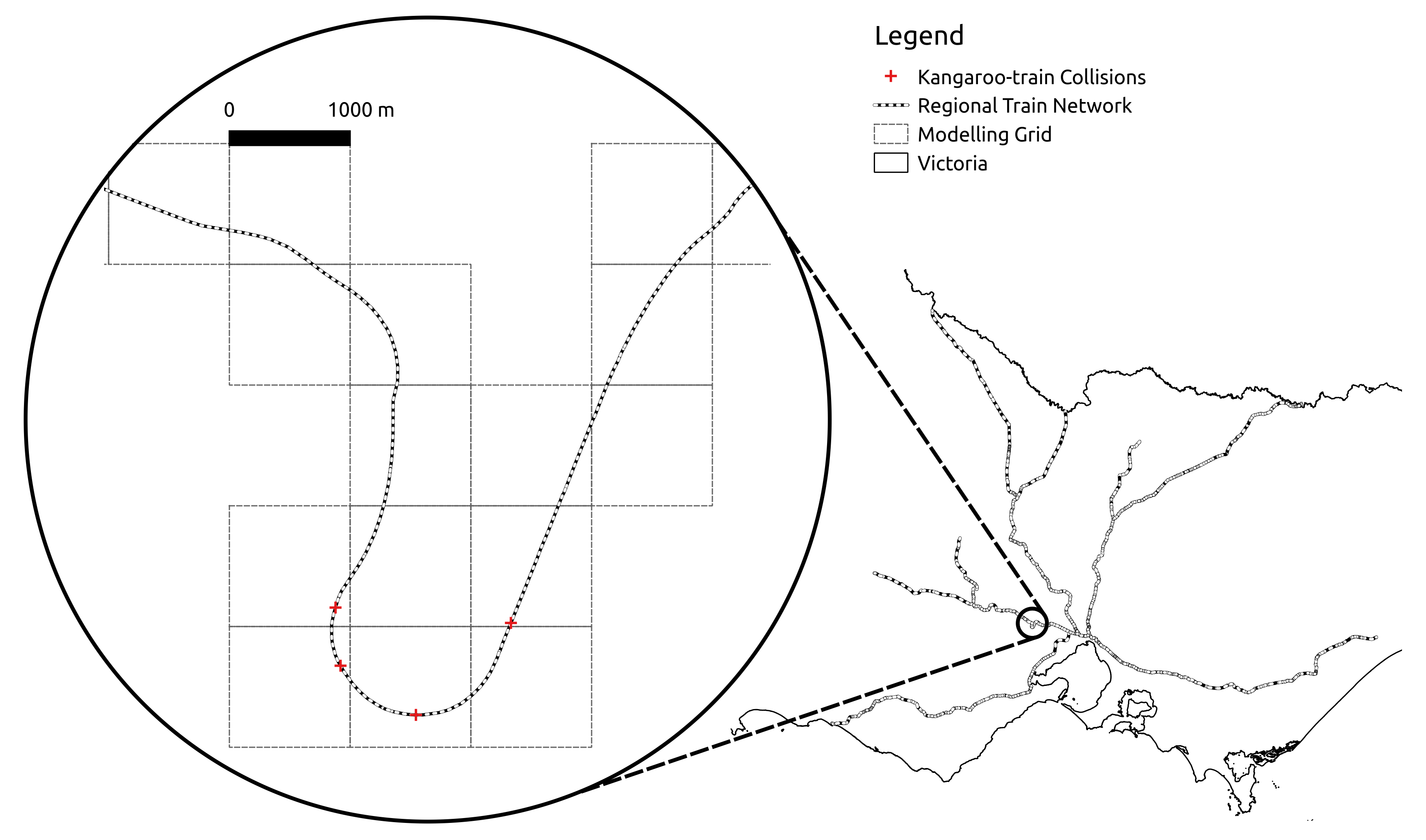
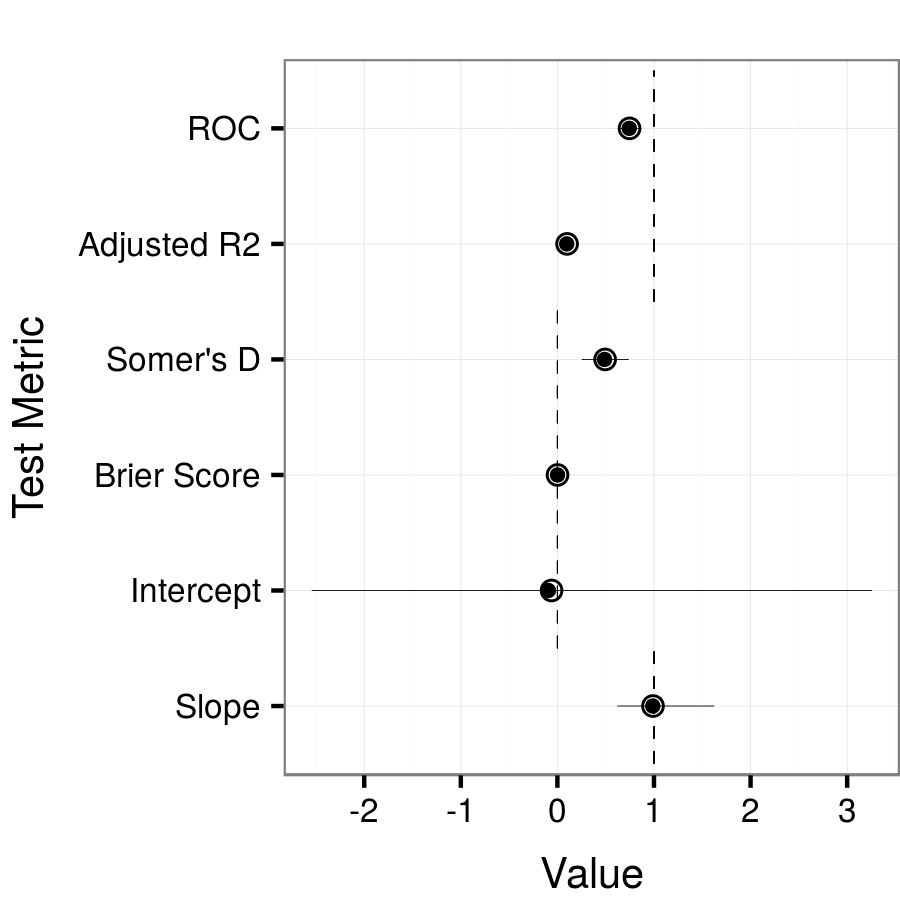


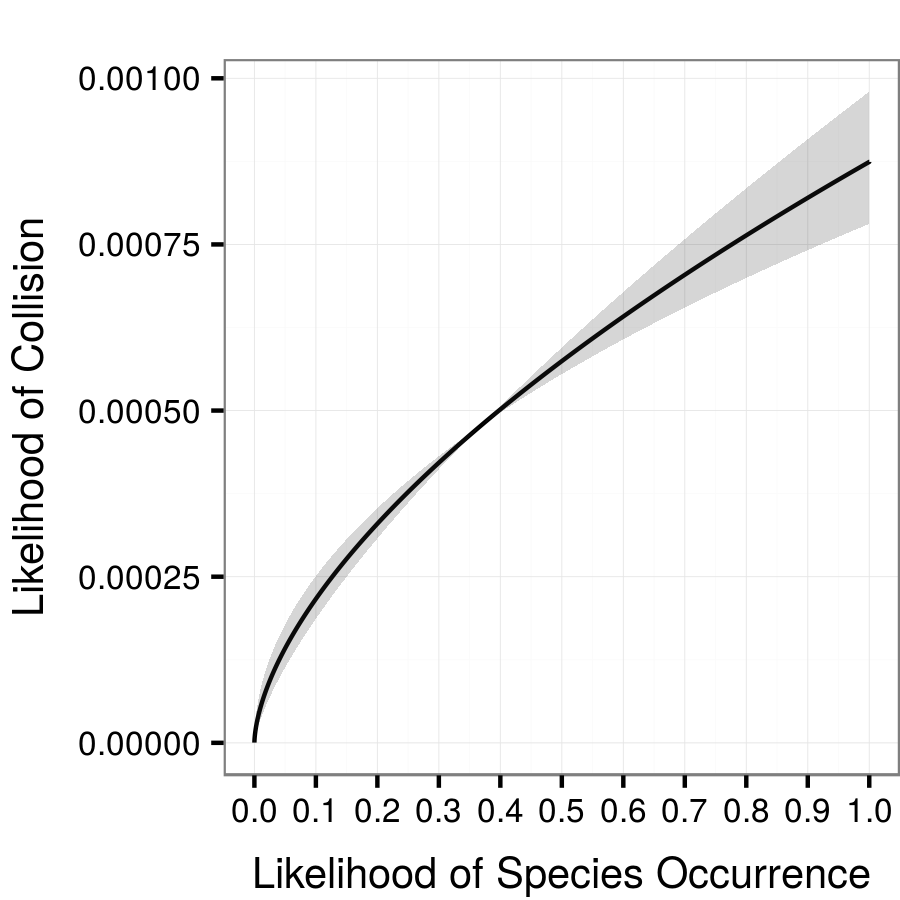
Figure 1: Wildlife-train collisions reported between 2009-2014 in Victoria.

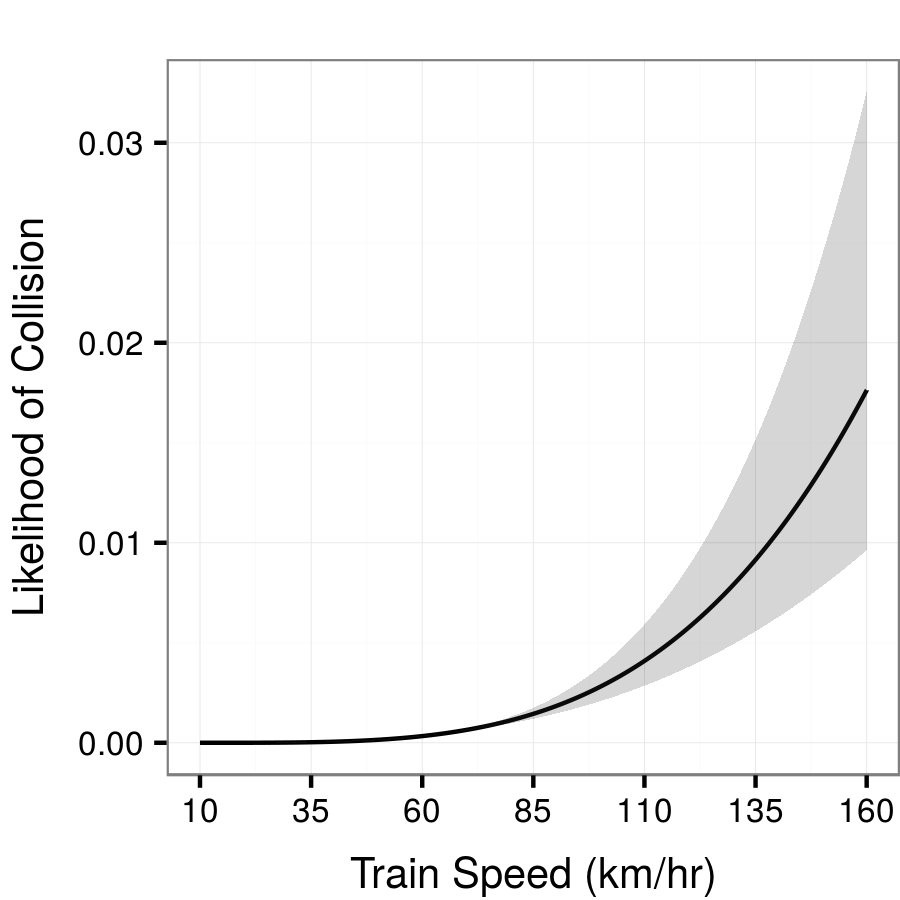
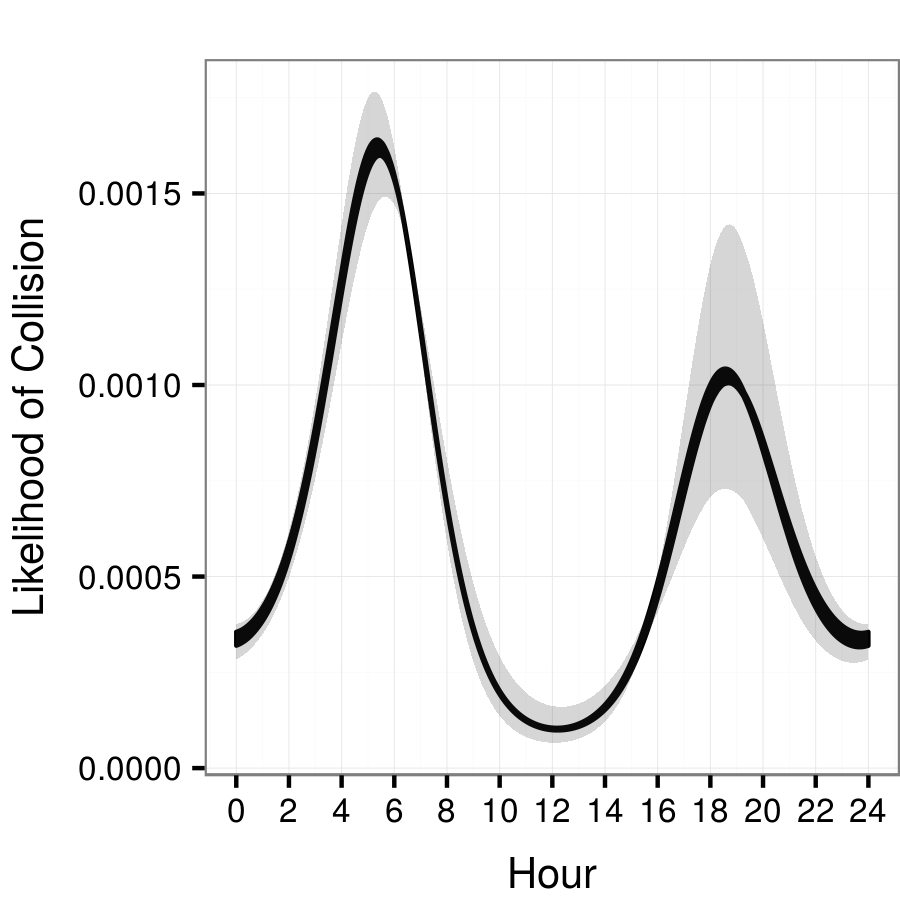


**Figure 2:**  Grid used to organise modelling data (number of cells: 2015; extents: 104000,5741000 x 556000,6084000; projection: GDA94 MGA zone 55)



**Figure 3:**  Comparison between the collision model fit on full data and on cross-validated subsets. "ROC" is area under the receiver operating characteristic curve, "Adjusted R2" is the Nagelkerke pseudo R-squared, and the intercept and slope are the results of regressing the dependent variable on the predicted values (link-scale). For each metric, open circles represent the full data model and solid dots represent mean values - ranges shown as bars - for the cross-validated subsets (n=1000). Dashed lines indicate the expected values for a perfect model.

**Figure 4:**  Marginal effects of model predictors on collision likelihood. For each plot, non-target variables are held constant at mean values. Shading indicates 95% confidence intervals.