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

2Ecology and Infrastructure International 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. In addition to animal welfare and conservation concerns, costs from train strikes may be considerable and rail authorities have a vested interest to manage the problem.

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. We made predictions of collision rates on the total network based on three management scenarios.

The model fit and predictions were plausible. Speed was the most influential variable followed by presence of kangaroos. Reducing speeds in areas of high predicted kangaroo occurrence during peak animal activity resulted in the greatest reduction in collision rate.

Predictions from the model can help managers decide where, when and how best to mitigate strikes. The model framework is easily adaptable to other species and rail operations and allows managers to assess bias and uncertainty and calibrate/update accordingly.

* 1. Keywords

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

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 et al, 2015) and their environmental impacts must be managed (Spellerberg, 1998). One of the most visible impacts are animals struck by moving vehicles which directly influence species mortality rates (Forman et al., 2003).

Wildlife-vehicle collisions are a serious problem throughout the western world (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 (Huijser et al., 2007; Romin & Bissonette, 1996) and Europe (Sáenz-de-Santa-María & Tellería, 2015; Seiler, 2004). Moreover, management of wildlife-vehicle collisions in developing countries will become important as new transportation networks are constructed and existing networks are expanded.

In addition to concerns about animal welfare (Sainsbury et al., 1995) and conservation status of threatened species (Dwyer et al., 2016; Jones, 2000), larger animals can directly pose risks to the life of humans (Langley et al., 2006; Rowden et al., 2008). For example, moose are one of the largest animals struck by vehicles in North America and Europe and cause significant damage and injuries due to mass (Hurley et al., 2009). Deer, although, smaller than moose are frequently reported in wildlife-vehicle collisions resulting in human deaths 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 location along a transportation network for a particular species, such as kangaroos, will assist managers to select and implement the most appropriate form of mitigation (e.g. animal exclusion or change in network activity). Data can also inform statistical modelling which helps to predict the probability of wildlife-vehicle collisions (Gunson et al., 2011).

The majority of wildlife-vehicle collision modelling deals with road networks (van der Ree et al, 2015), yet, 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 (Forman et al., 2003). First, the movements or presence of animals are often considered in the models and may include behavioural traits (Roger & Ramp, 2009). Second, 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).

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 analyse wildlife-train collisions (see Belant, 1995; Onoyama et al., 1998). Moreover, we only found one published study predicting wildlife-train collisions (Gundersen & Andreassen, 1998). Here, we develop a modelling framework to predict the rate of kangaroo collisions on the regional passenger train network in Victoria. Our methods aim to inform rail operators of potential kangaroo collision risks and can be used to 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 passenger rail network from regional Victoria, Australia (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 (following day), with the largest volume occurring Monday through Friday between the hours of 7am and 9am and 4pm and 6pm.

* + 1. *Data Preparation*

To organise our data and modelling, we overlaid a grid of one square kilometre resolution on the rail network (Figure 2). In each grid cell we modelled species occurrences and quantified the train movements and speeds.

Eastern grey kangaroos (Macropus giganteus, Shaw, 1790; "kangaroos" hereafter) are frequently struck animals in regional Victoria and large enough to cause significant damage to trains. V/line provided records of 439 kangaroo collisions spanning a six-year period between 1 January, 2009 and 31 December, 2015. Each record included incident date and time, the name of service line (unique route between two towns), and nearest fraction of a kilometre post (physical sign markers indicating distance along train line). Using geographic information system (GIS) data on the regional rail network, we determined spatial coordinates (GDA94 MGA zone 55 projection) for all collisions from the reported kilometre post and service line.

*Species Occurrence*

Kangaroo occurrence data is sparsely recorded in regional Victoria. To represent exposure risk, 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 et al. (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 (VBA, 2014) and included several environmental variables relating to the biology and behaviour of kangaroos (see Visintin et al, 2016). To reduce the effects of sampling bias, we also included two anthropogenic variables (distance to urban areas and roads) and the spatial coordinates of grid cell centroids as predictors in the model.

*Characteristics of Rail Network*

To determine train movements across space and time, we accessed publicly available 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 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 (e.g., tramTRACKER). We used a spatial database (Postgres version 9.6; PostGIS version 2.3.0) to process this information and report 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*

To account for temporal variation in collision risk throughout the day, we considered peak periods of train movements in relation to daylight hours. By adding variables that allow a bimodal response of collision rate to hour of day across all seasons, the crepuscular lifestyle of kangaroos (most active at dawn and dusk) is tested. We included three additional variables to the model for this purpose; relative daylight intensity (both linear and quadratic terms) and time between sunrise and sunset.

* + 1. *Statistical Modelling*

We adapted a single-species quantitative risk model (see Visintin et al, 2016) to fit and compare the relationship of kangaroo presence, characteristics of the rail network, and temporal patterns (kangaroo movements during high activity of trains) to collision likelihood. The likelihood that a collision occurs in a given grid cell **i** at hour **j** in month **k** (*pijk = Pr(Yijk=1))*depends on species occurrence **O**, average number of trains **V**, average train speed **S**, relative daylight intensity **L**, duration of daylight **D**, and length of track **K**::

*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)

Prior to modelling, we centred all explanatory variables by subtracting their means. All predictors exhibited Pearson's product moment correlation coefficients of less than 0.4 using pairwise analysis - indicating low potential effects of multi-collinearity.

We fit the data (n=291120) to a generalised linear model (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 Visintin et al., 2016). The model is similar to a proportional hazards model (discrete censored time) often used in survival analysis and epidemiological studies (Cox, 1984).

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 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 performance metrics and compared them with those from the model fit to all data (Figure 3).

Using the model fit on all data, we predicted the number of expected train-kangaroo collisions in the study area for one year under different management scenarios:

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 (n=42 cells, total unique train trips=275). 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) 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

Our model fit the data with no unexpected estimates of coefficients (e.g. negative values where positive were expected, or vice-versa). The relative risk of collisions increased with higher average train speeds, predicted kangaroo occurrence, train frequency and during hours of high kangaroo activity in grid cells. Train counts, however, had very little influence on collisions - and significance - in comparison to the other predictors (Figure 4).

The strongest and most significant predictor was train speed; collision risk increased exponentially with considerable increases at speeds above 85 km hr-1. An increase of of train speed from 110 to 130 km hr-1 resulted in a doubling of collision risk, however, the effect of speed also demonstrated large uncertainty in the confidence intervals at high values (Figure 4).

Kangaroo occurrence was the second most influential predictor. Collision risk increased rapidly at low values of occurrence and more slowly at higher values. There was an approximate 10-fold increase in collision likelihood across the range of values for kangaroo occurrence and less uncertainty in the confidence intervals at lower values (Figure 4).

The coefficients estimated for the three variables describing a bi-modal functional form of kangaroo activity demonstrated a plausible shape when the marginal effect of hour was plotted against collision risk (Figure 4). Collision risk peaked at approximately 5:45am and 6:15 pm with a higher risk occurring in the morning period of the day. The highest amount of uncertainty around the response of collision risk to hour was in the evening peak. The lowest collision risk occurred at noon and both peaks showed similar spread and distribution.

The performances of the models fit on all data and fit on the 1000 subsets of the data during cross-validation were similar. Receiver operator characteristic (ROC) value was 0.82 for both the full data model (Figure 3b) and mean of the cross-validated models (Figure 3c). Likewise, the calibration statistics (intercept and slope of regression line between observations and predictions) were similar for both the full data model (Figure 3a) and mean of the cross-validated models (Figure 3c). The uncertainty in the calibration metrics were higher than 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, however, became less calibrated at higher rates (Figure 3a).

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 3.2% whilst scenario B only reduced collisions by 1.2% (Table 3).

* 1. Discussion

The plausible model fit suggests our framework may be useful for analysing wildlife collisions on operating train networks.

Train speed was an important predictor for collision risk.

Kangaroo occurrence is also a useful predictor for collision risk.

Temporal patterns, such as crepuscular activity of wildlife and train movements, have implications for collision risk.

The data used for this study has unique properties with respect to reporting bias and errors.

We used existing data to create predictors for the model framework.

Our framework allows management decisions to be made in two distinct areas: reduction of animal presence (e.g. deterrents or exclusions) or reduction of train threat (e.g. adjust schedules or speeds).

Each of our management scenarios reduces collisions which is a positive outcome. Although many costs are related to collisions (e.g. animal welfare, ecological), monetary costs are a useful metric for assessing management. From a transportation authority perspective, costs result from removal of trains from service for cleaning and repair.

* 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. Michaela Plein provided valuable comments on the manuscript. 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. & Oakes, D. (1984) *Analysis of survival data* CRC Press

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

Dwyer, R. G.; Carpenter-Bundhoo, L.; Franklin, C. E. & Campbell, H. A. (2016) Using citizen-collected wildlife sightings to predict traffic strike hot spots for threatened species: a case study on the southern cassowary *Journal of Applied Ecology*, 53, 973-982

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

Hurley, M. V.; Rapaport, E. K. & Johnson, C. J. (2009) Utility of Expert-Based Knowledge for Predicting Wildlife-Vehicle Collisions The *Journal of Wildlife Management*, 73, 278-286

Jones, M. E. (2000) Road upgrade, road mortality and remedial measures: impacts on a population of eastern quolls and Tasmanian devils *Wildlife research*, 27, 289-296

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

Sainsbury, A.; Bennett, P. & Kirkwood, J. (1995) The welfare of free-living wild animals in Europe: harm caused by human activities *Animal Welfare*, 4, 183-206

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

Williams, A. F. & Wells, J. K. (2005) Characteristics of vehicle-animal crashes in which vehicle occupants are killed. *Traffic Injury Prevention*, 6, 56-59

* 1. Tables

Table 1: Predictor variables used in collision model. Note, prior to modelling all variables were centred by subtracting their means. The collision model includes both the linear and quadratic term 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 | Time between dawn and dusk in grid cell based on month | hr |

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\* |
| LIGHT2 | -1.87 | 0.17 | -10.85 | 2.06E-27\* |
| DAWNORDUSK | 0.25 | 0.07 | 3.76 | 1.73E-04\* |

Table 2: Summary of predicted collisions based on different management scenarios. Expected collisions are a total across the entire regional network for a period of one year.

|  |  |  |
| --- | --- | --- |
| Scenario | Description | Expected Total Collisions |
| A | no change to current operations or infrastructure | 404 |
| B | moderated train speeds in high kangaroo occurrence areas during peak travel times | 399 |
| C | controlled kangaroo occurrence in areas with highest average speed of trains | 391 |

* 1. Figures

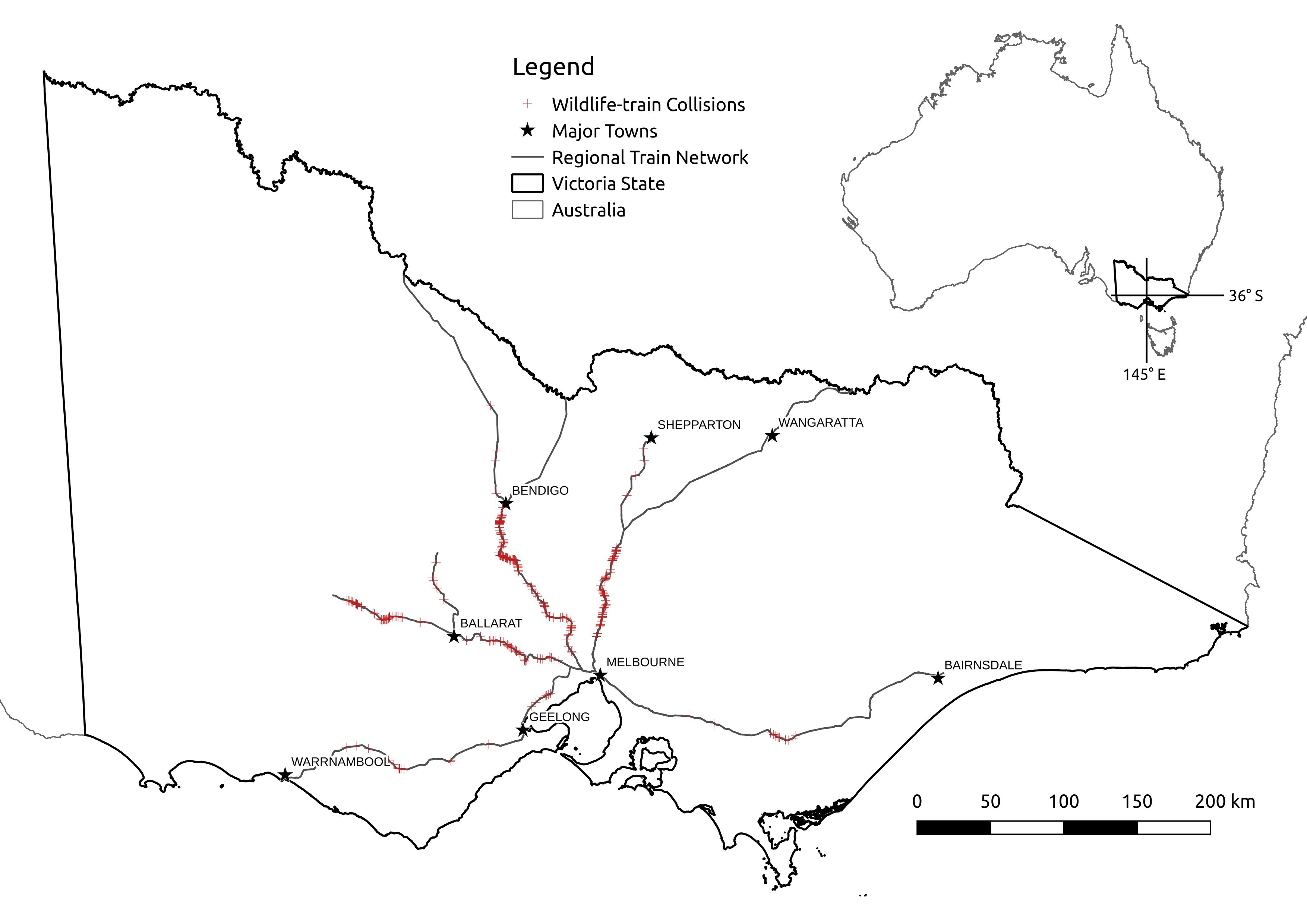
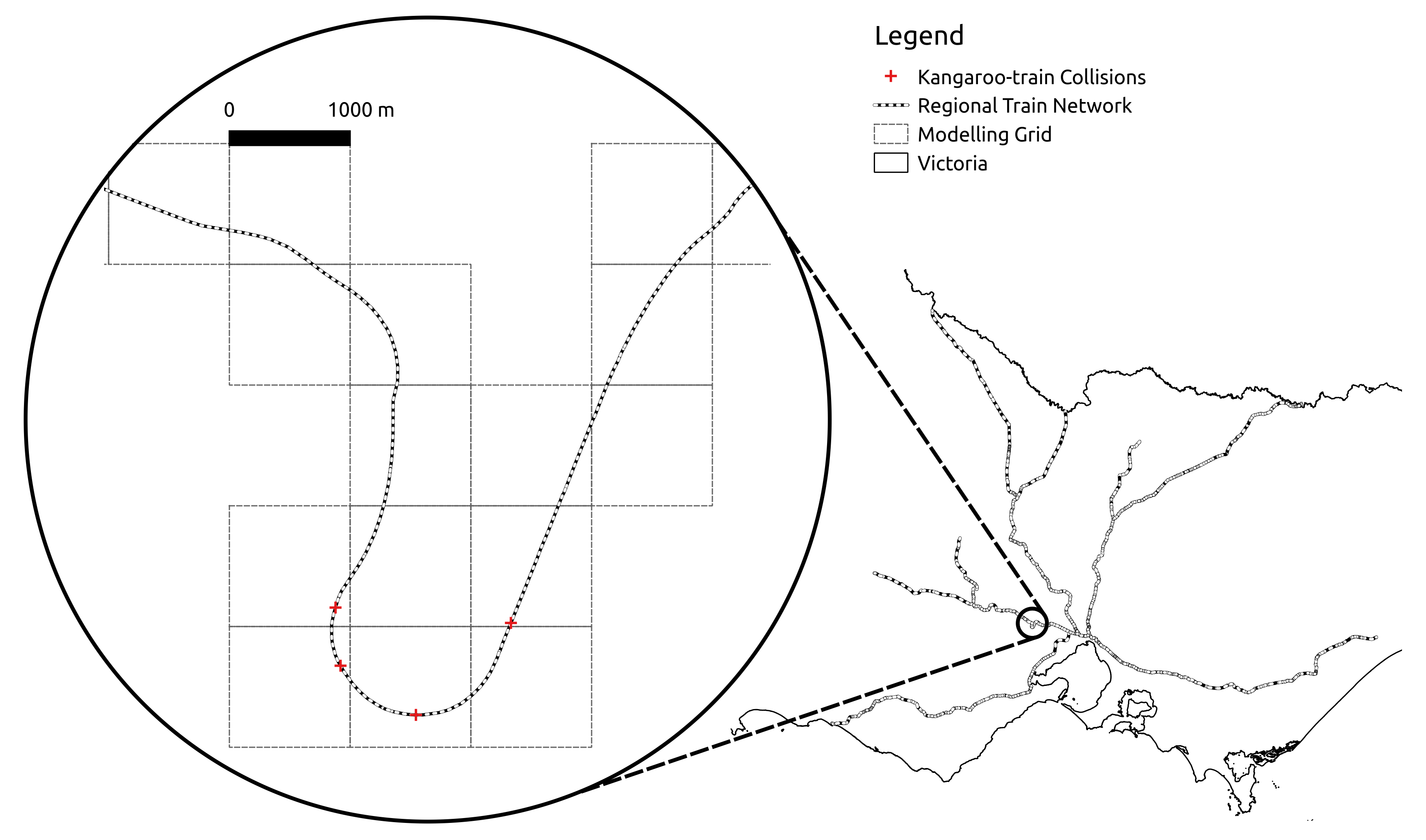
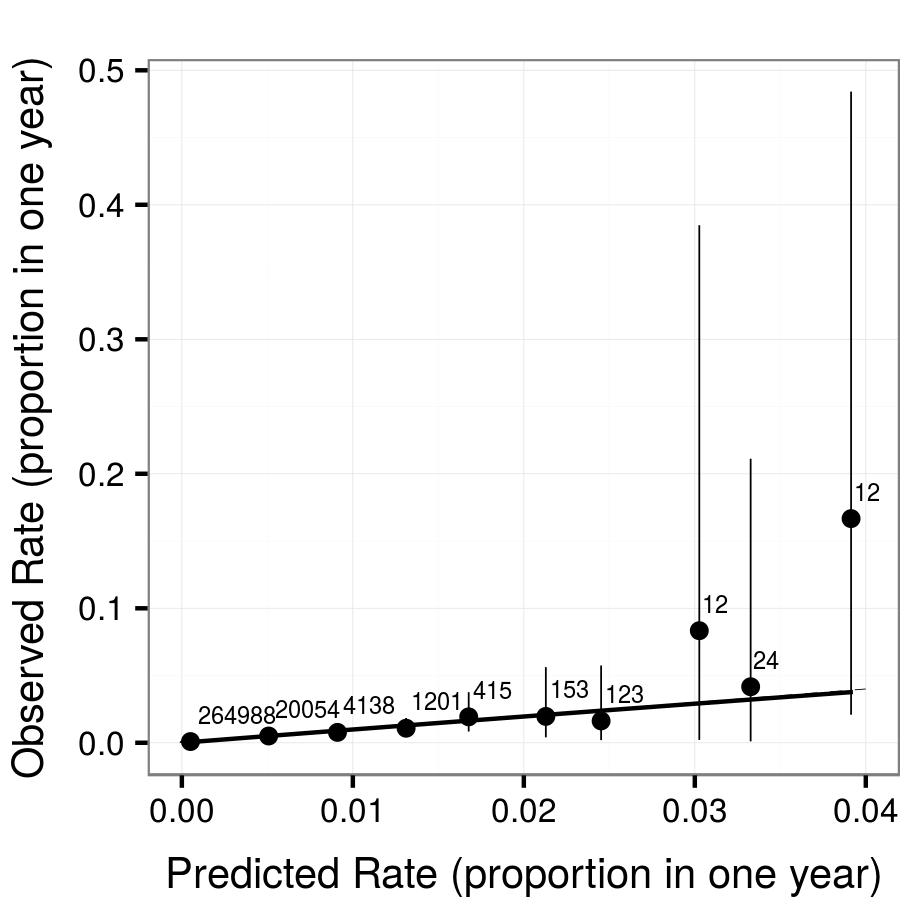
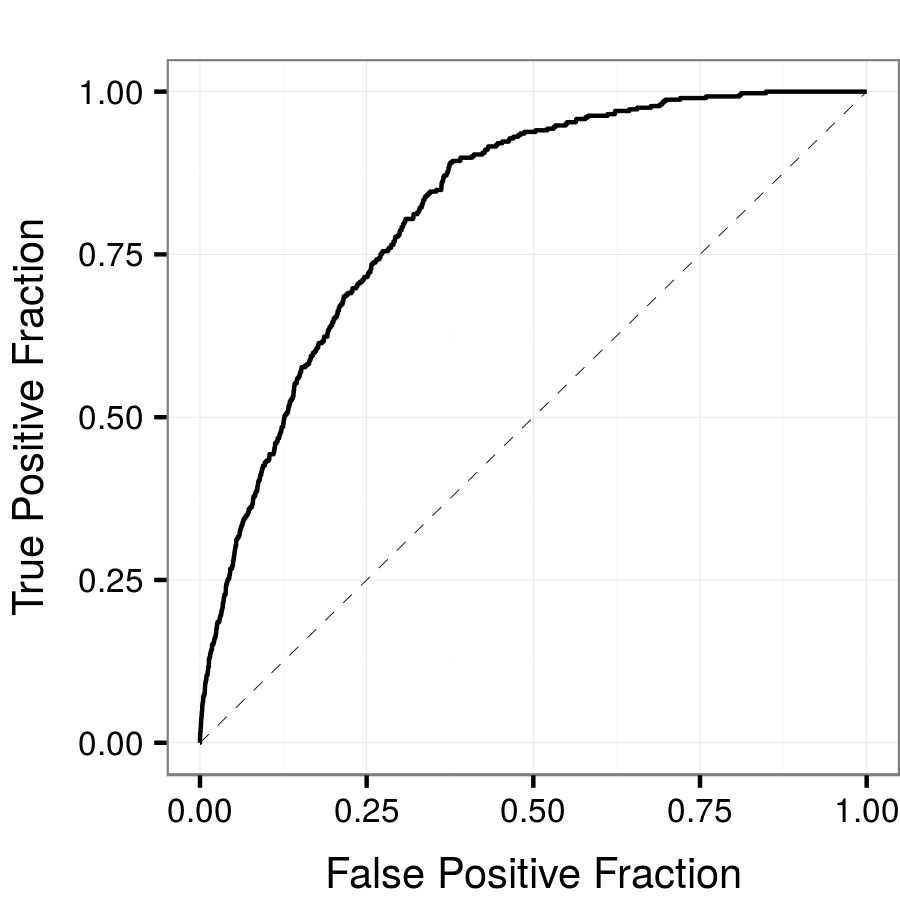
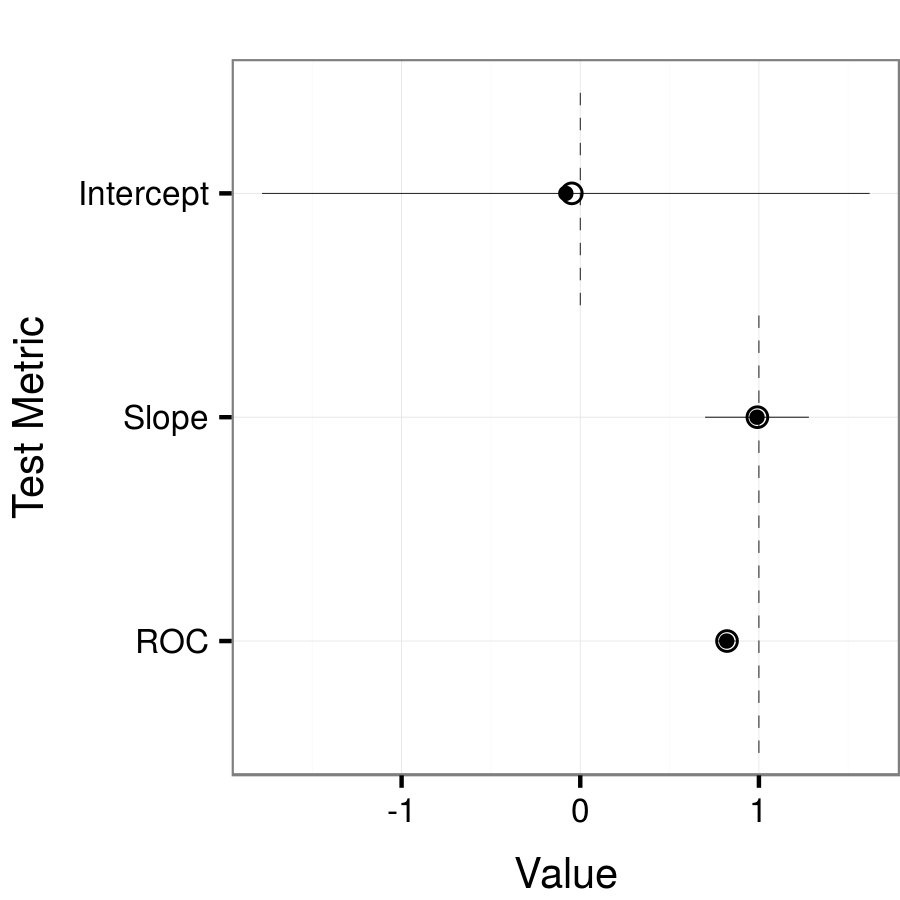
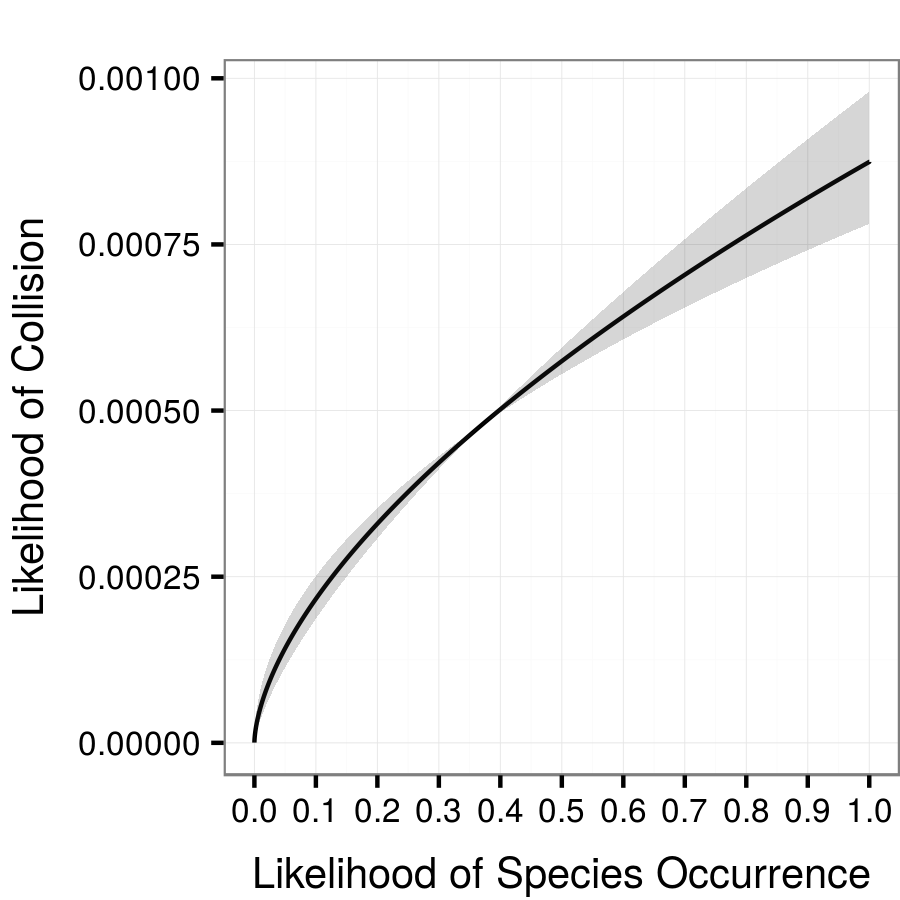


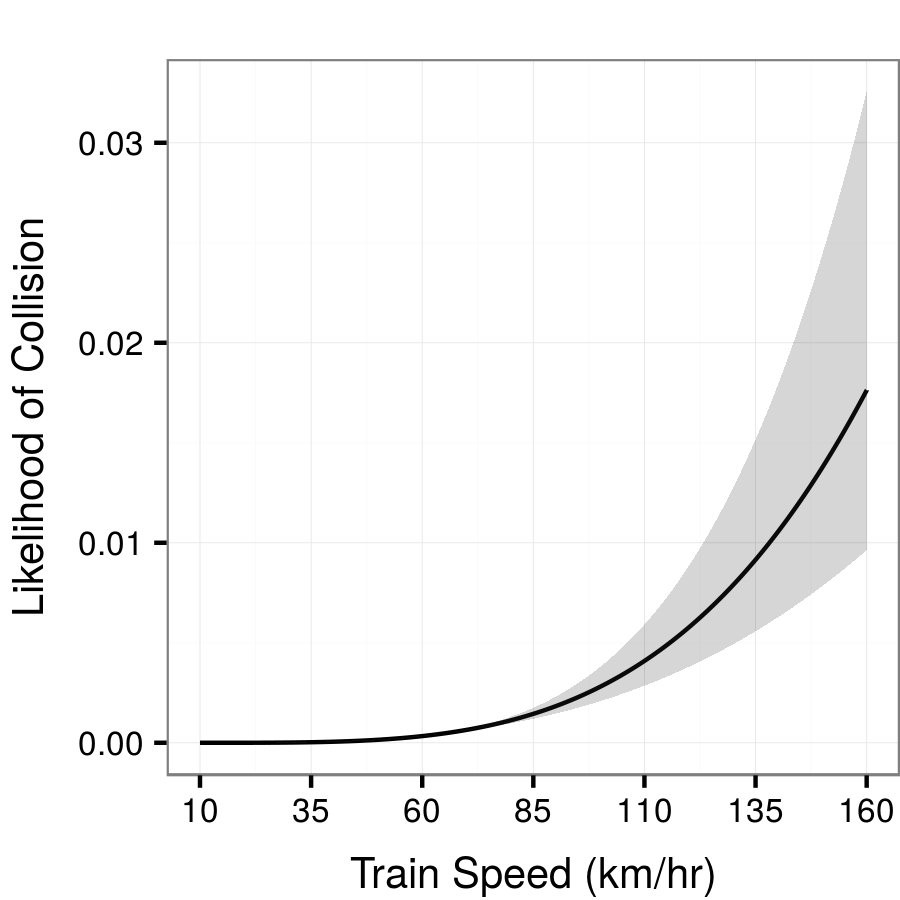
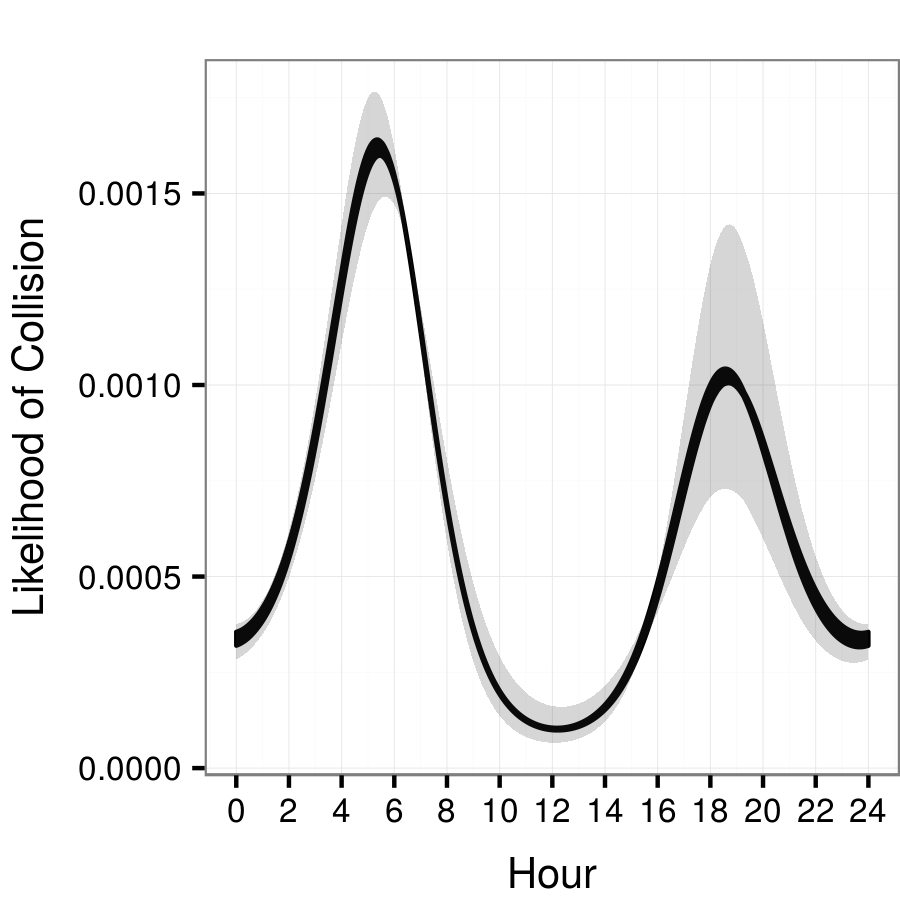
Figure 1: Wildlife-train collisions reported between 2009-2014 in Victoria.



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

a) b) c)**Figure 3:**  **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 is shown by the dashed line (intercept of 0 and slope of 1) **b)** ROC (receiver operating characteristic) curve measuring discrimination ability of model at all threshold values (see Metz, 1978). **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 1000 cross-validated subsets. Dashed lines indicate the expected values for a perfectly calibrated and discriminatory 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 around coefficient estimates.