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

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: 6476  
 Abstract: 263  
 Main Text: 3820  
 Acknowledgements: 65  
 References: 1572  
 Tables: 239  
 Figure Legends: 517

Number of Tables: 3

Number of Figures: 6 (2 supplemental)

Number of References: 62

* 1. Abstract

Despite considerable research on collisions between and wildlife and vehicles, animal mortality from strikes by trains is 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.

* 1. Keywords

framework, risk, species distribution model, 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). 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 (Huijser et al., 2007; Romin & Bissonette, 1996) and Europe (Sáenz-de-Santa-María & 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 et al., 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 et al., 2006; Rowden et al., 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., 2007; Hurley et al., 2009). Vehicle collisions with deer, although smaller than moose, frequently kill humans in North America (Williams & Wells, 2005).

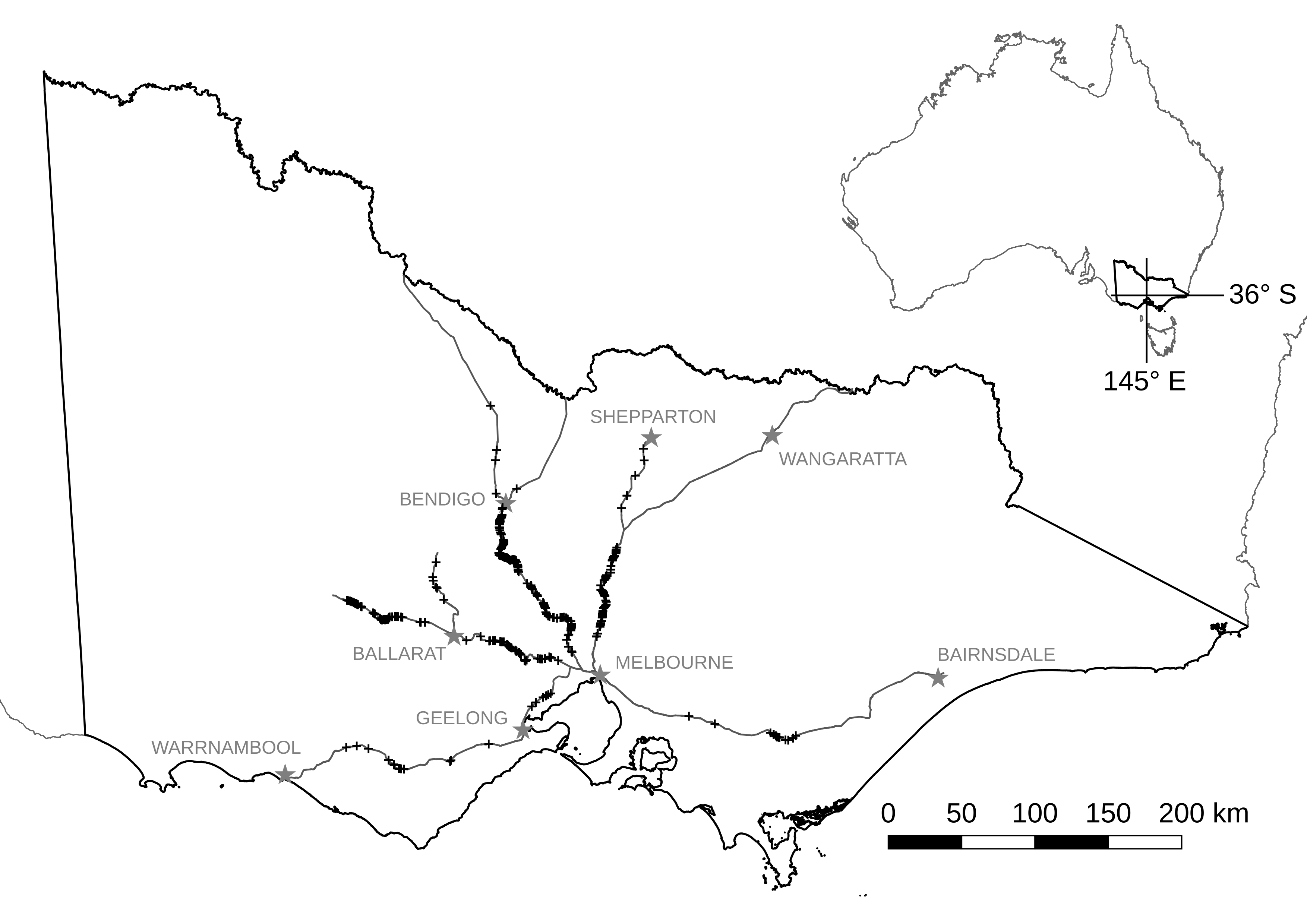
Information about the spatial and temporal distribution and magnitude of wildlife-vehicle collisions is useful to managers because it may help more effectively mitigate impacts (Mountrakis & Gunson, 2009). For example, knowing a collision hotspot along a transportation network for a particular species will inform the most appropriate form 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 et al., 2011, Visintin et al., 2016).

The majority of wildlife-vehicle collision modelling deals with roads and traffic (van der Ree et al, 2015), yet, the problem extends to other forms of vehicular transportation such as air (van Belle et al., 2007), railway (Wells et al., 1999) and shipping (Laist et al., 2001) operations. Regardless of the mode of transport, the modelling of collisions share some common attributes (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).

Extensive railway networks with considerable train movements exist on every inhabited 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 (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 et al., 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 southeast 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.

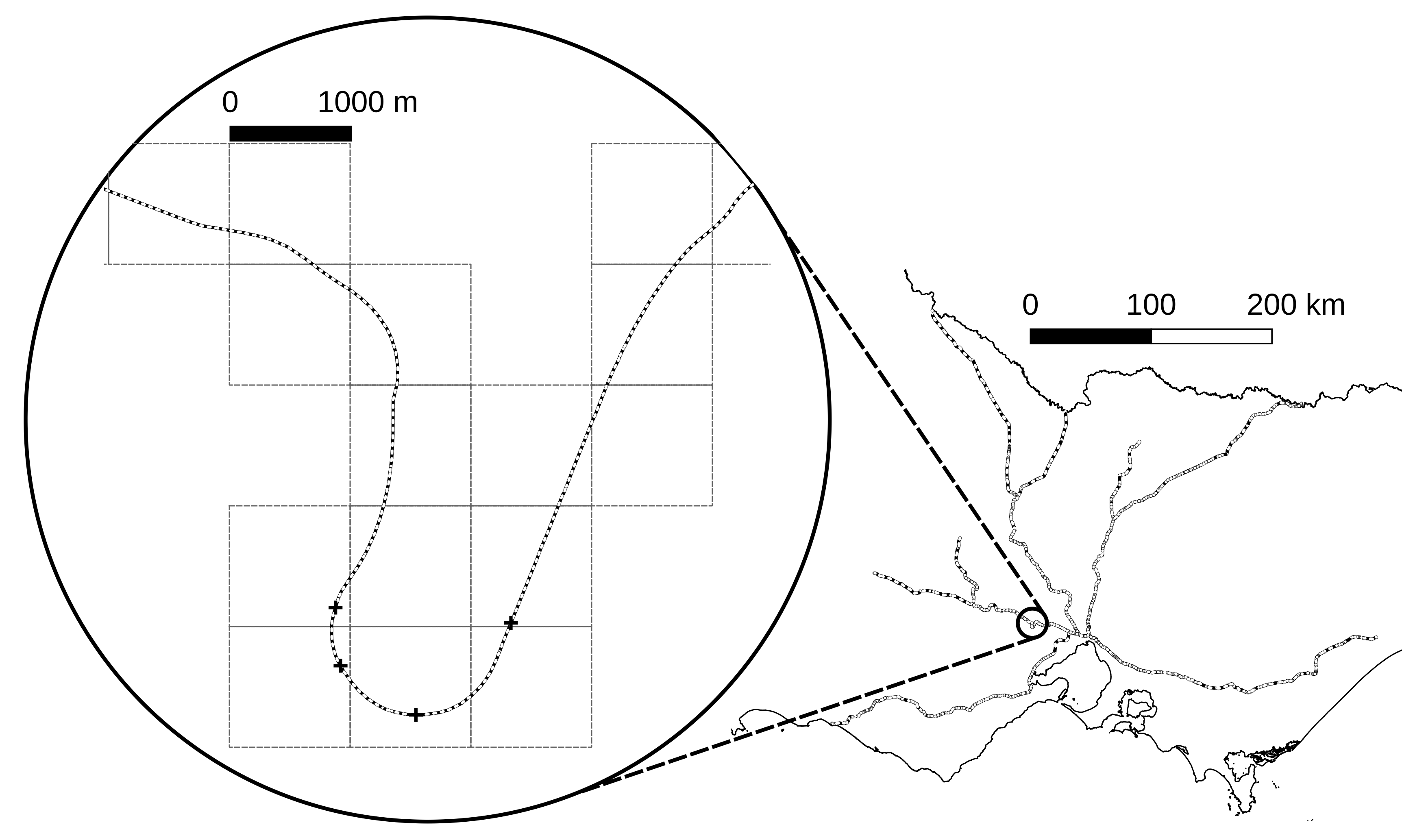
* 1. Materials and Methods
     1. *Study Area*

We used a 1712-kilometre passenger railway 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 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.



1. *Figure 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.*
   * 1. *Data Preparation*

To organise our data and modelling, we overlaid a 1-km2 cell grid on the railway network (Figure 2). In each grid cell we modelled species occurrences and quantified train movements and speeds.



***Figure 2:***  *One km2 grid framework used to organise modelling data: 2,015 total cells; extent coordinates 104000E,5741000N x 556000E,6084000N; GDA94 MGA zone 55 projection. The railway network is shown as a heavy dashed line and wildlife-train collisions are shown as crosses.*

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.

*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 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 (after Visintin et al., 2016). 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.

*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 (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 (see Hillsman, 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.

* + 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 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) is incorporated. 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*, length of track *T* (offset term), and a term *Cijk* that accounts for the crepuscular behaviour of kangaroos:

cloglog(*pijk*) = *β0* + *β1* log(*Oi*) + *β2*log(*Vij*) + *β3*log(*Sij*) + *Cijk* + log(*Ti*) (1)

The crepuscular term *Cijk* has three components:

*Cijk* = *γ1* sin(π(*j −* 6) / 12) + *γ2* sin(π(*j −* 6) / 12)2 + *γ3**Aijk* (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 *Aijk* 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:

*Aijk* = (3)

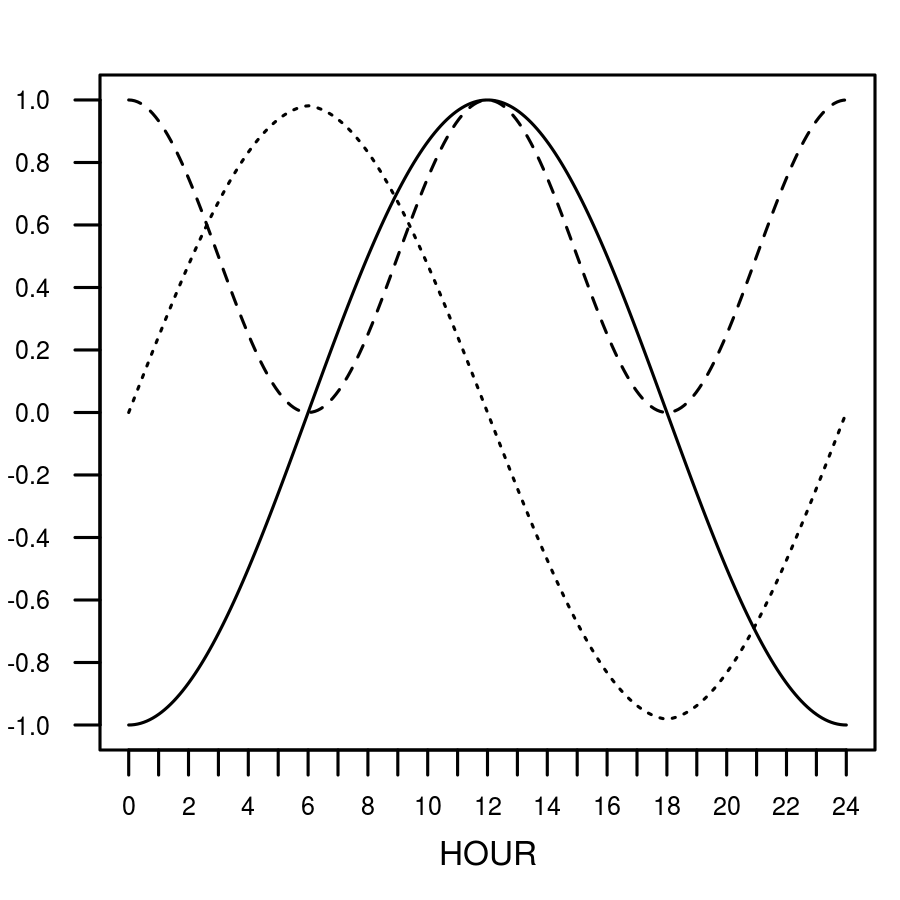
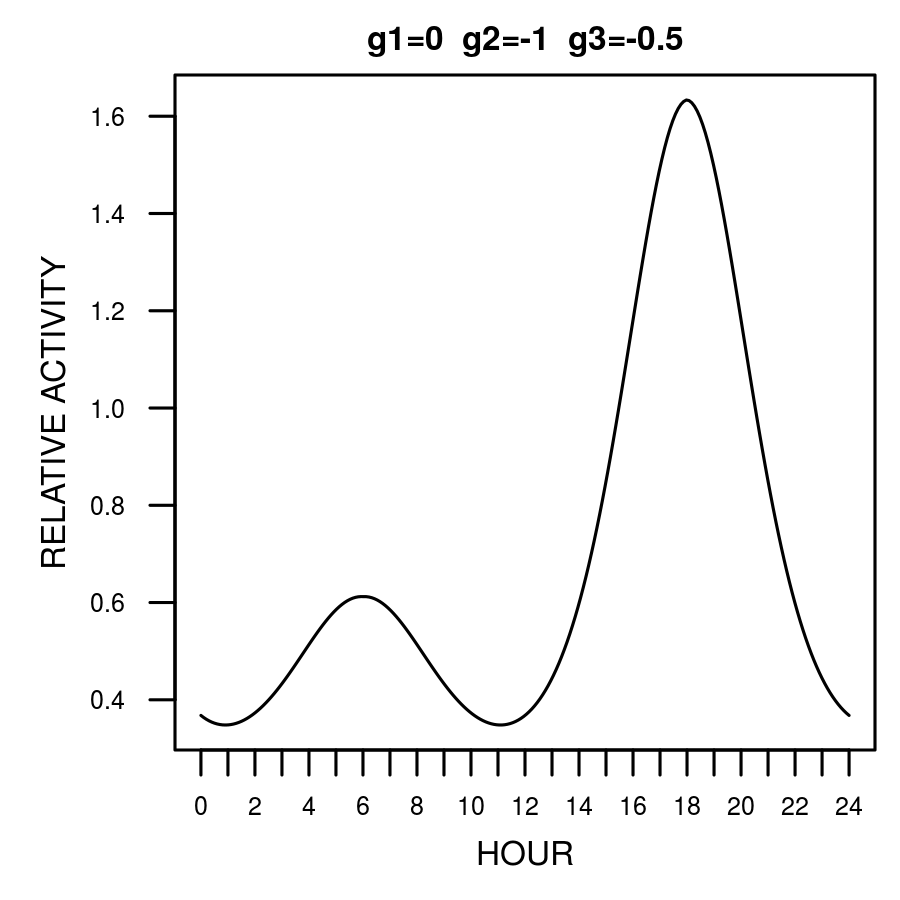
exp(−((*j* − *Uik*) / ((24 − *Dik*) + *Uik*) / 2)2) − exp(−((*j* − *Dik* − 24) / ((24 − *Dik*) + *Uik*) / 2)2) if *j* < *Uik* *;*

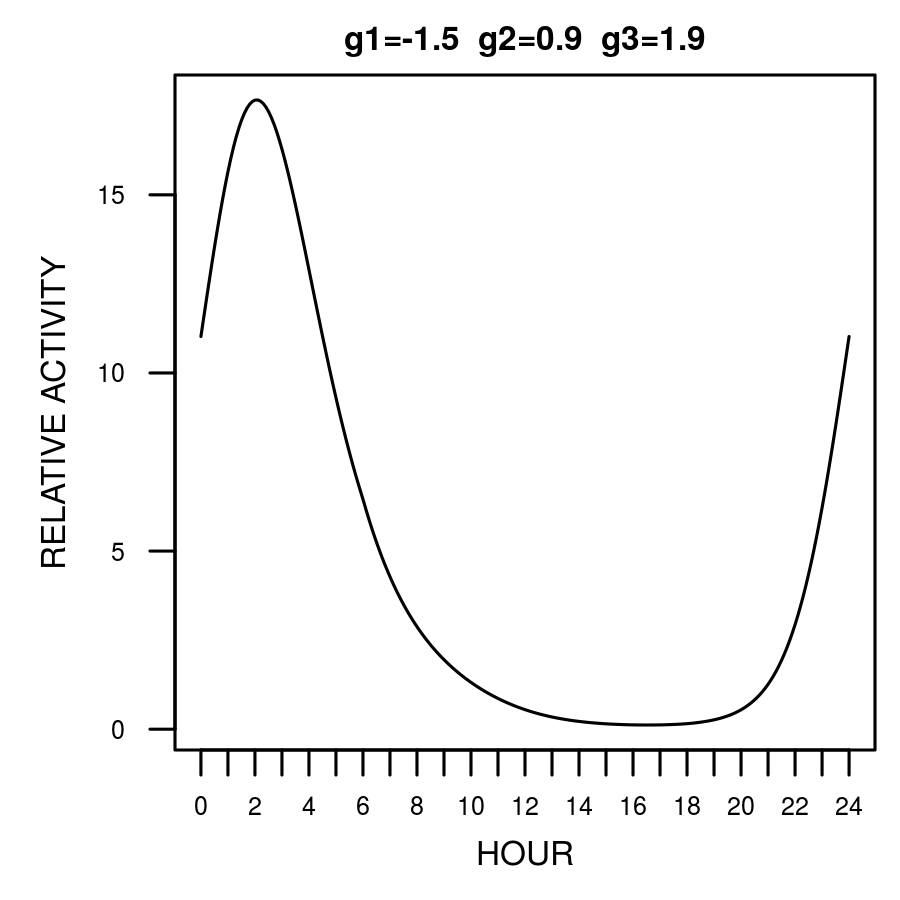
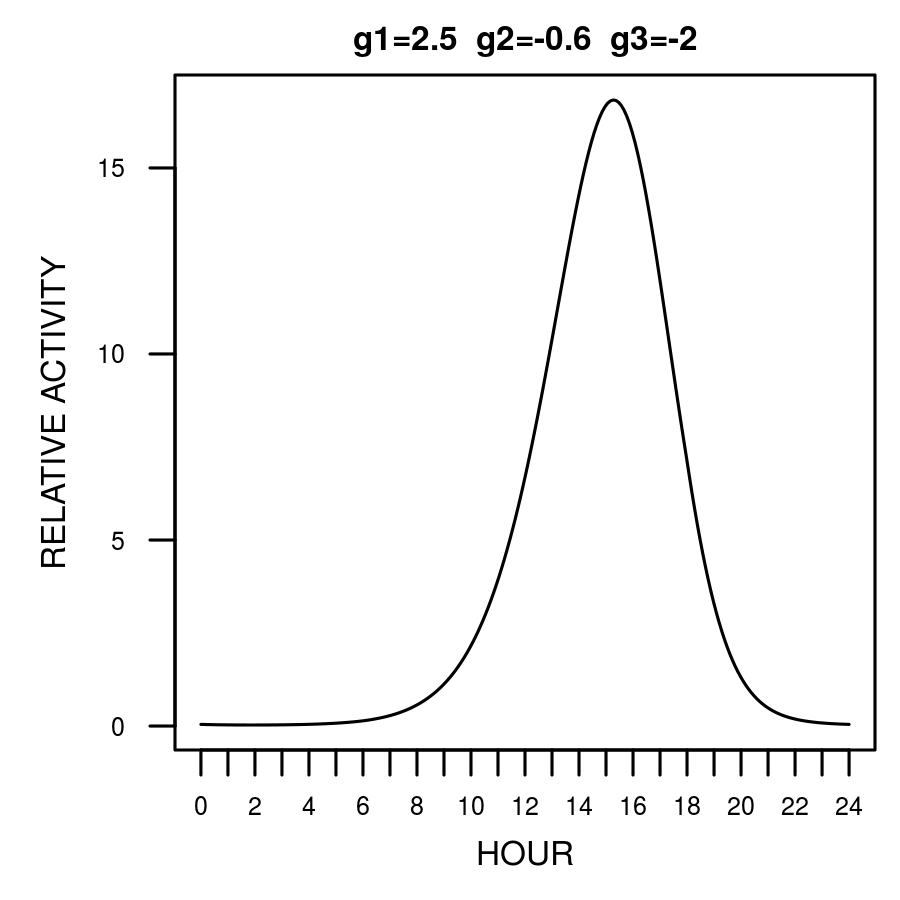
exp(−((*j* − *Uik*) / ((*Dik −* *Uik*) / 2)2) − exp(−((*j* − *Dik*) / ((*Dik −* *Uik*) / 2)2) if *Uik*  <= *j* < *Dik ;*

exp(−((*j* − *Uik + 24*) / ((24 − *Dik*) + *Uik*) / 2)2) − exp(−((*j* − *Dik*) / ((24 − *Dik*) + *Uik*) / 2)2) if *j* >= *Dik ;*

Dawn *U* and dusk *D* times for civil twilight (six degrees below horizon) were determined for each observation in the dataset using a National Oceanic & 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 *γ1*, *γ2*, *γ3*are estimated from the data and measure the relative influence of the three components (Equation 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 S1) regardless of the parameter estimates of *γ1*, *γ2*, *γ3*.

a)b)

c)d)

***Figure S1:*** ***a)*** *Components used to build crepuscular term in Equation 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 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 coefficient values (g1, g2, g3) 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 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 1 further describes the variables used in the model.

*Table 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 2 and Figure S1).*

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Description | Units | Mean; Range |
| EGK | Relative likelihood of kangaroo occurrence in grid cell | - | 0.13; 0.01:0.87 |
| TRAINS | Train frequency in grid cell | trains h-1 | 6; 1:80 |
| SPEED | Mean train speed in grid cell | km h-1 | 86.75; 3.92:147.73 |
| C1 | Crepuscular function linear component | - | 0.25; -1.00:1.00 |
| C2 | Crepuscular function quadratic component | - | 0.47; 0.00:1.00 |
| C3 | Crepuscular function astronomical component | - | -0.05; -0.98:0.98 |

We fit the data (n=291,120 one-kilometre grid cells along the train network) 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 examine model fit, we generated randomised quantile residuals (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 (Metz, 1978) and regression of observations on predictions (Cox, 1989; Miller, 1991). We repeated this procedure for 100 iterations producing a total of 1000 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.

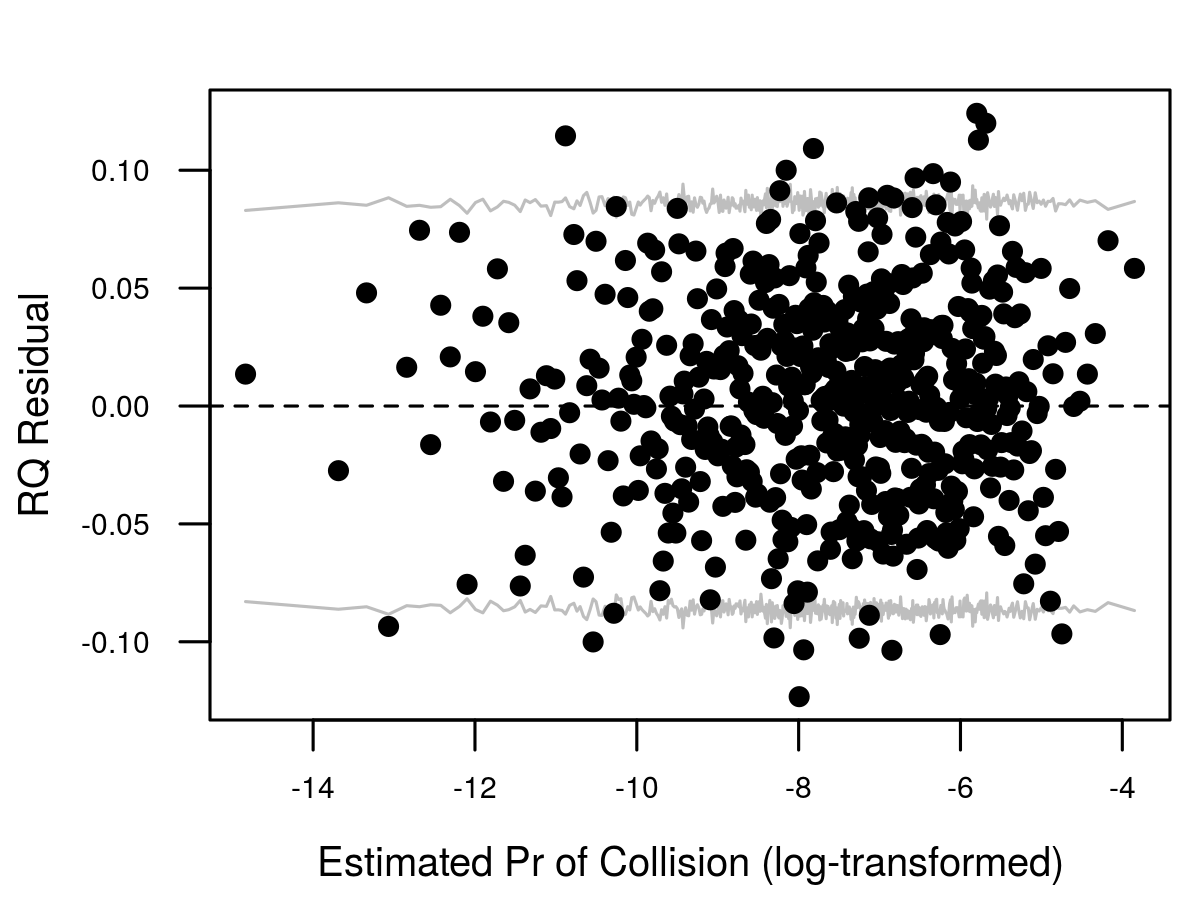
Scenario B involved capping train speeds at 80 km h-1 in grid cells with kangaroo relative occurrence likelihoods of 0.5 or above during the hours of 5 a.m. to 9 a.m. and 4 p.m. to 8 p.m. This change affected 42 cells and 275 out of 3,024 unique weekly routes. In scenario C, relative kangaroo occurrence 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.

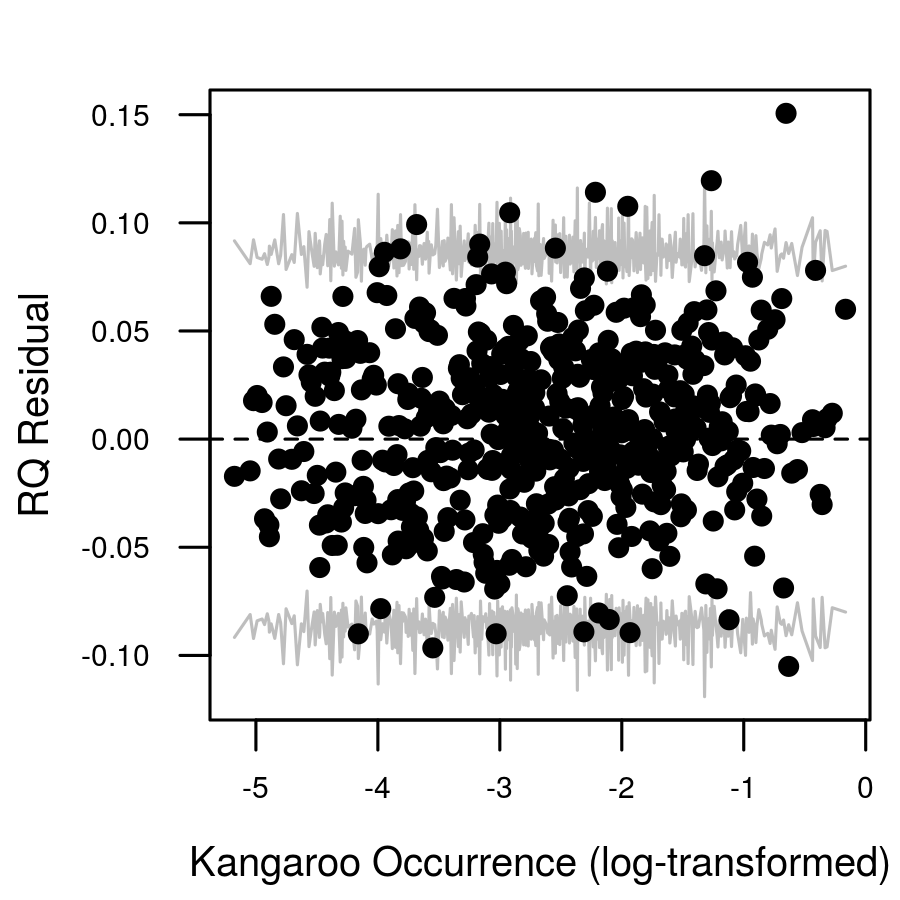
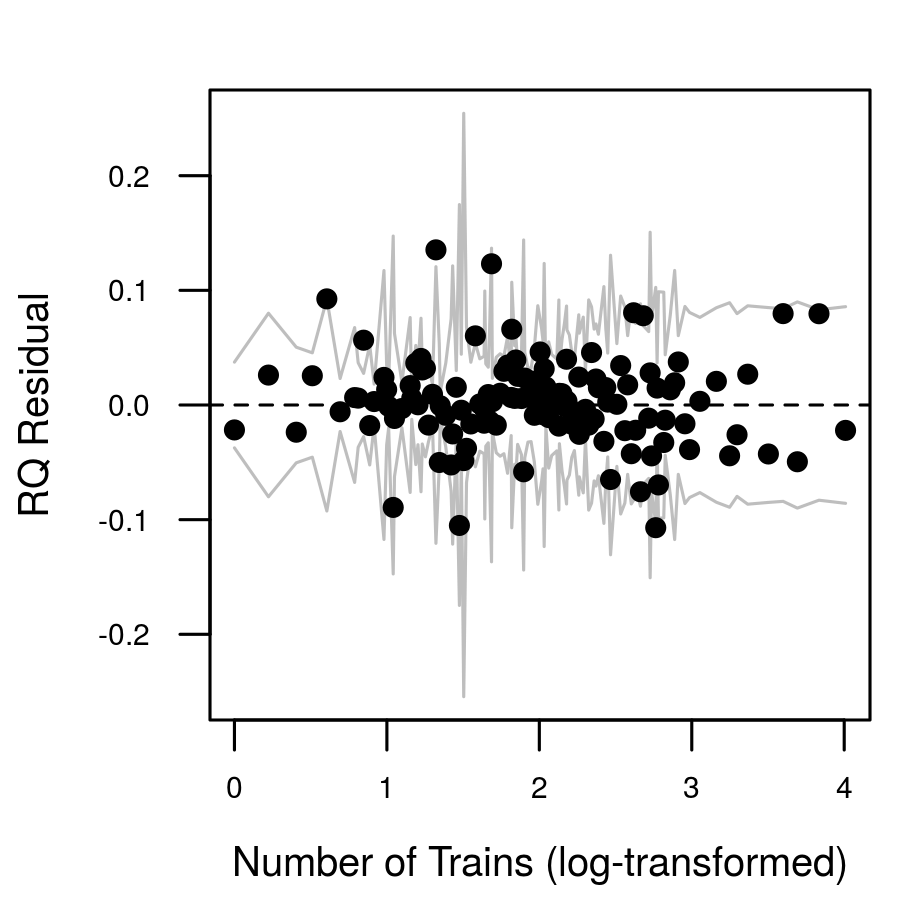
* 1. Results

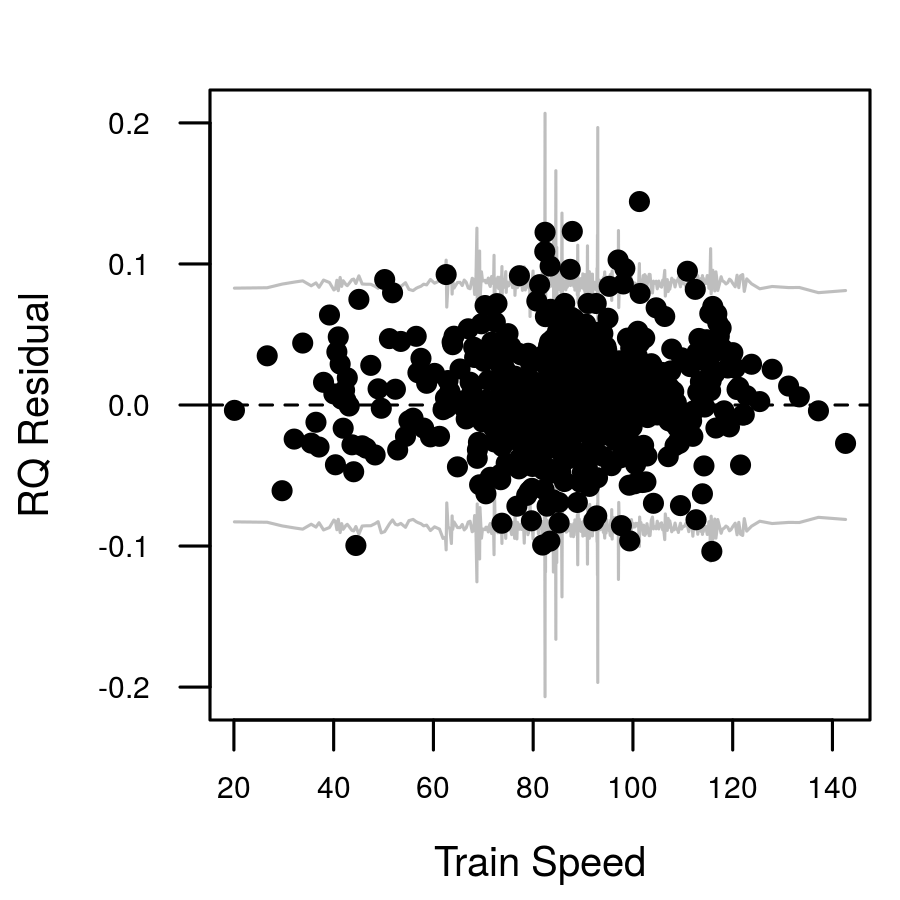
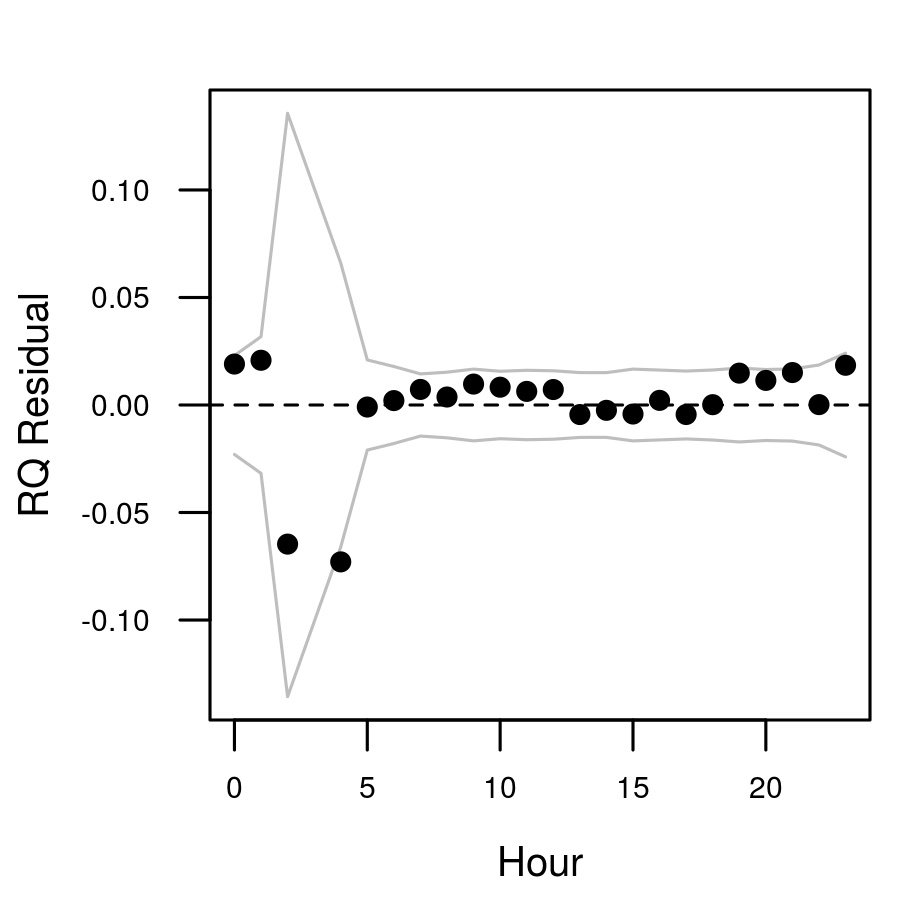
Our model fitted the data and the signs of model coefficients (positive or negative) indicated plausible relationships between collision likelihood and each predictor variable. All variables except train frequency demonstrated high statistical significance in the model fit (Table 2). The model residuals demonstrated no strong patterns with respect to predicted collision probabilities and model predictors (Figure S2).

*Table 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 | -0.66 | 0.11 | -5.83 | 5.7x10-7\* |
| C2 | -1.87 | 0.17 | -10.84 | <1.0x10-14\* |
| C3 | 0.25 | 0.07 | 3.77 | 1.7x10-5\* |

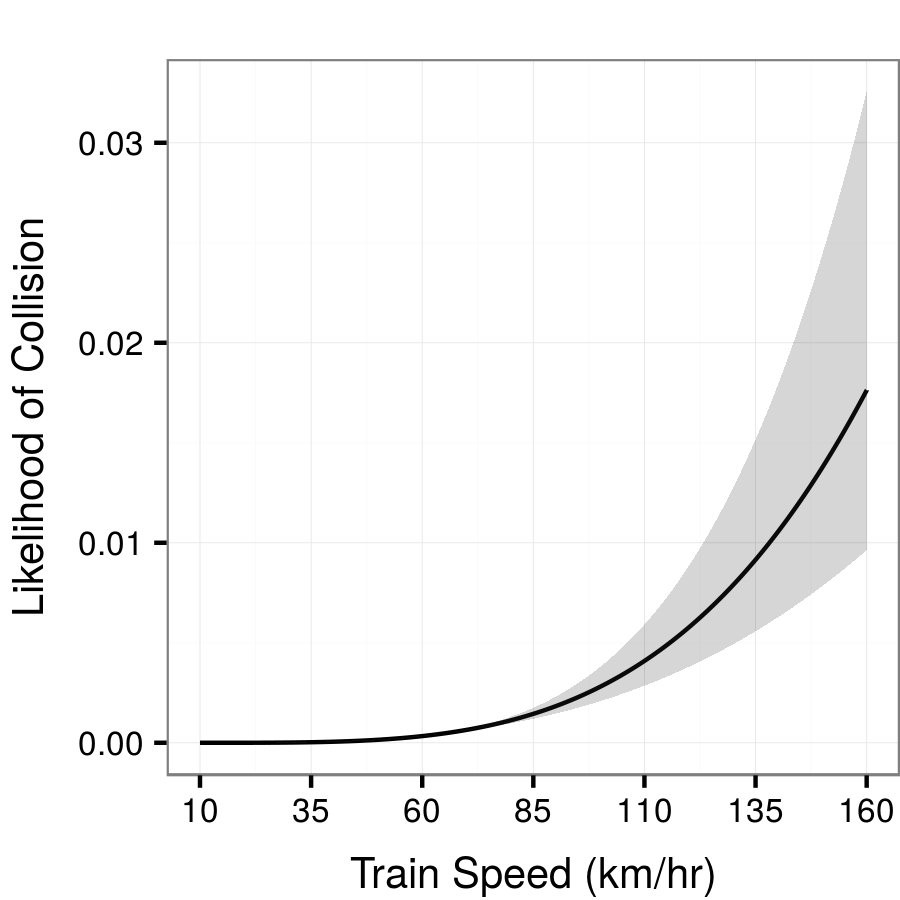
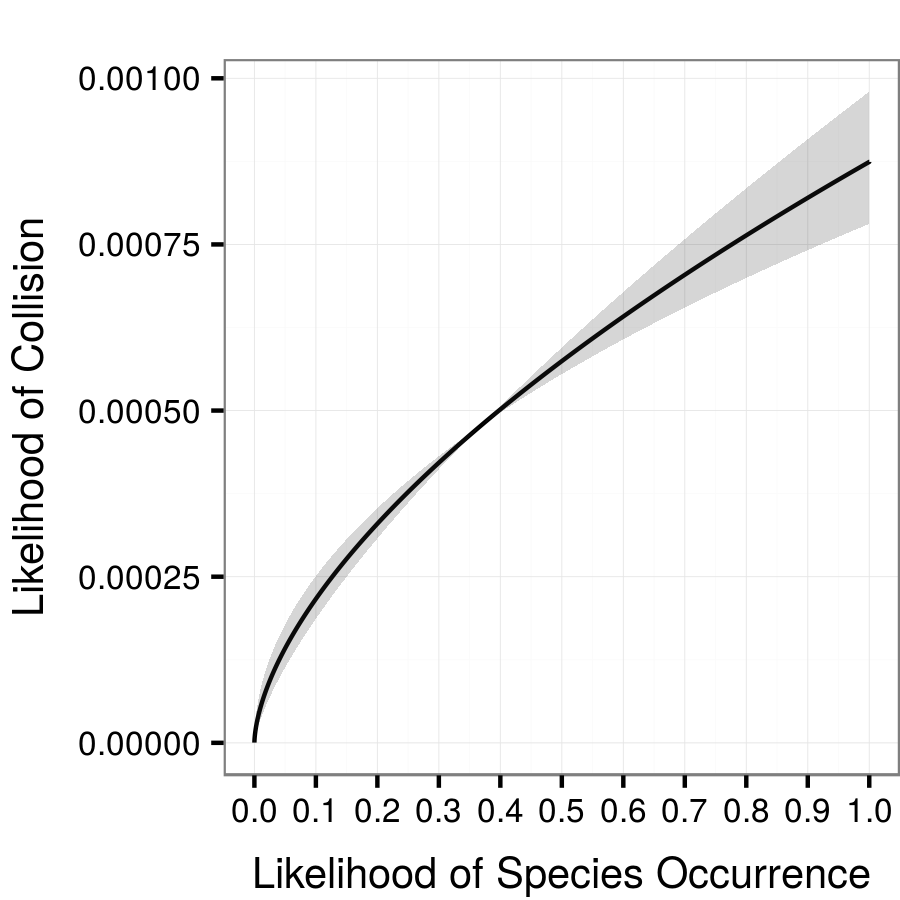
a)

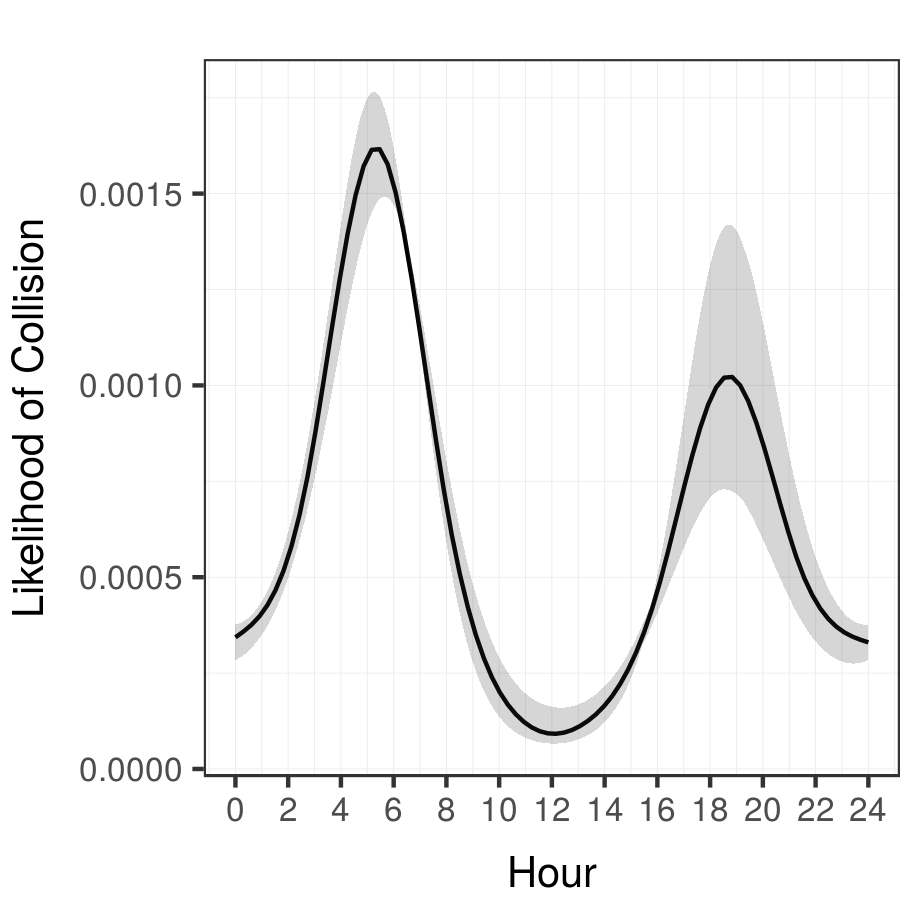
b) c)

d) e)

***Figure S2:*** *Binned randomised quantile (RQ) 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.*

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 3a). Kangaroo occurrence was the second most influential predictor of risk of collision. Collision risk increased rapidly at low values of occurrence and more slowly at higher values. Collision risk increased approximately 10-fold across the range of values for kangaroo occurrence (Figure 3b). Train frequency had very little influence on collisions, compared to the other predictors (Table 2; Figure 3c).

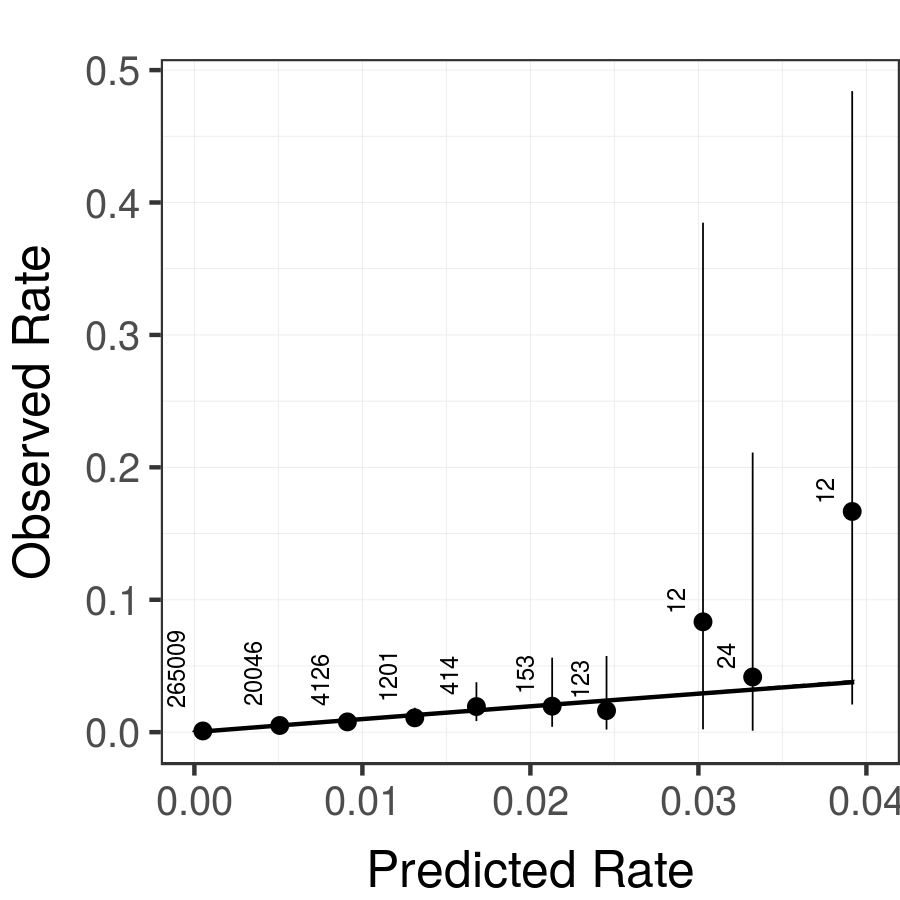
a)b)

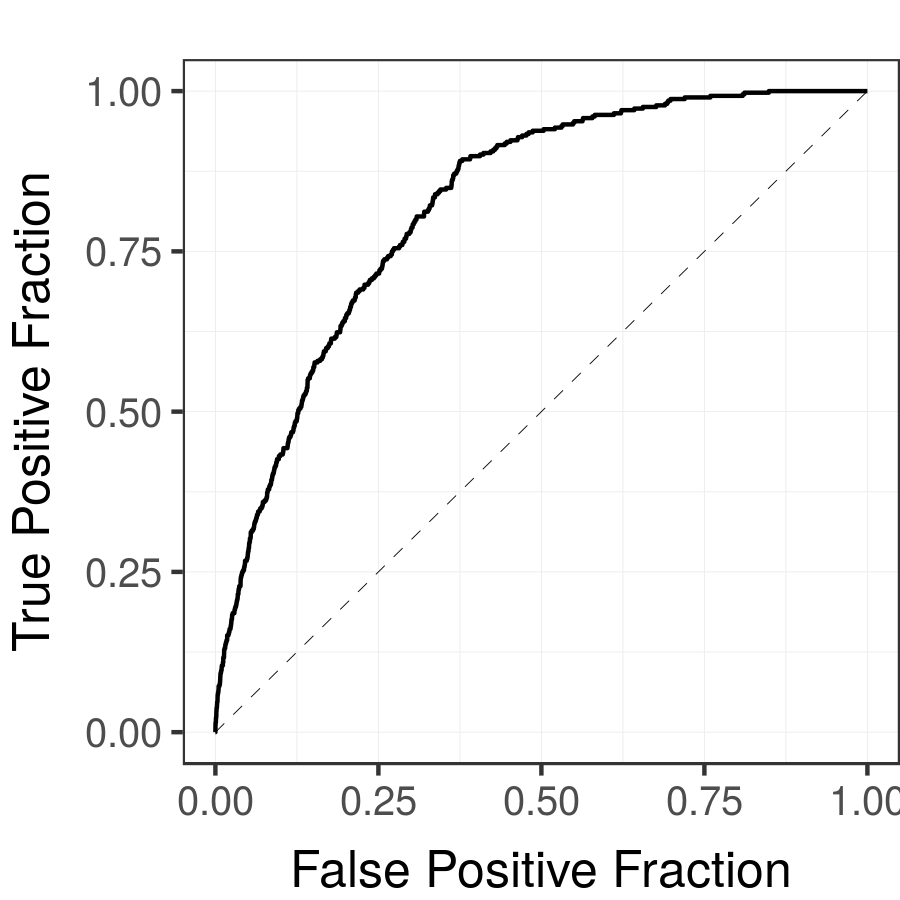
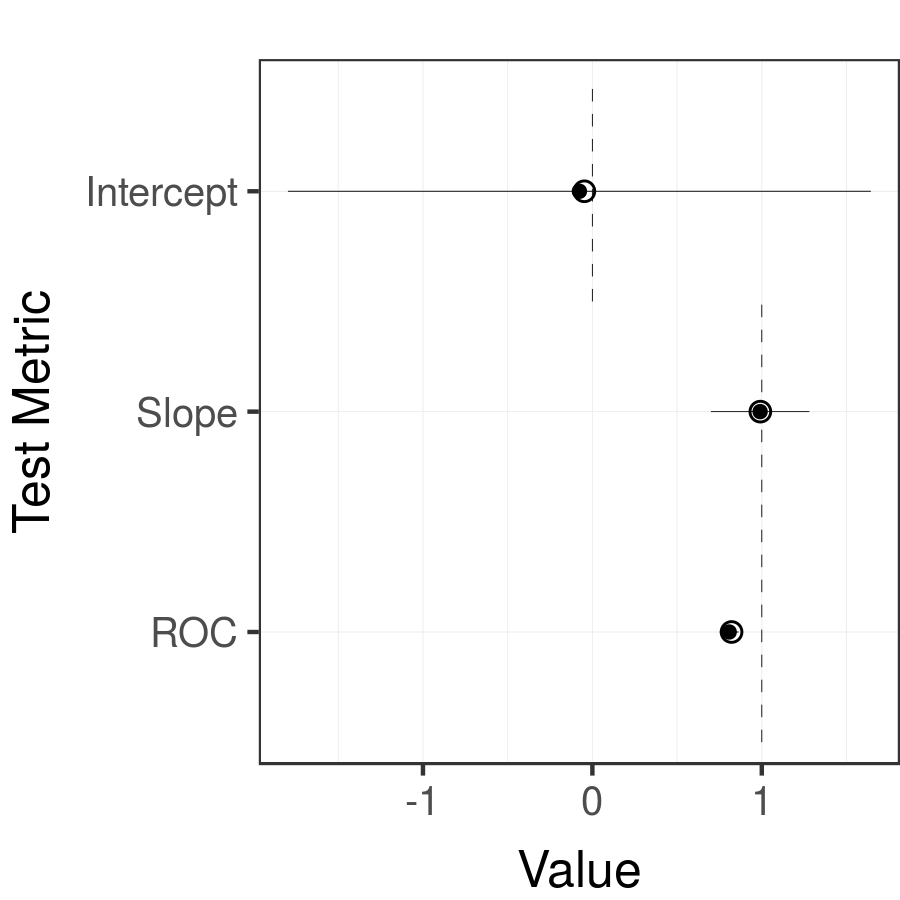
c)d)

***Figure 3:***  *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.*

The bi-modal functional form for the effect of kangaroo activity on collision risk demonstrated a plausible shape (Figure 3d). 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 4b) and mean of the cross-validated models (Figure 4c). Likewise, the calibration statistics (intercept and slope of regression line between observations and predictions) were similar for both the full data model (Figure 4a) and mean of the cross-validated models (Figure 4c). 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 4a).

a)

b) c)

***Figure 4:*** ***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 1000 cross-validated subsets. Dashed lines indicate the expected values for a perfectly calibrated and discriminatory model.*

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 3).

*Table 3: 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 | 400 |
| C | controlled kangaroo occurrence in areas with highest average speed of trains | 392 |

* 1. 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 (Gundersen & Andreassen, 1998) and road (Lao et al., 2011; Roger et al., 2012, Visintin et al., 2016) 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 our 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 et al., 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 (Lao et al., 2011; 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 et al., 2013), and validation (Chivers et al., 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 (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., 2013). 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 S1. 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 reporting bias and errors. Train drivers are obligated 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 collision studies (but see Snow et al., 2015). Moreover, spatial and temporal errors in reported collisions are assumed to be less as standardised mechanisms such as collision report forms, GPS devices, and distance signage are implemented in railway operations. Similar practices are used by some road authorities to collect and archive carcass data (Huijser et al., 2007), 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, 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 et al., 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 kangaroos). Mitigation choice may also be limited by operational objectives. Fencing may be chosen to exclude animals on railways when changes to train speed and frequencies 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 fencing 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 reduces collisions which is a positive outcome. Although many negative outcomes - not easily quantified - 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, we have 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 (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 et al., 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.

* 1. Data Accessibility

Model Dataset - Archived on GitHub; original collision data is V/Line’s Intellectual Property (IP) and authorisation for additional use must be granted by V/Line

R Code - Archived on GitHub

* 1. Acknowledgements

We thank V/Line for their generous collaboration; in particular, Sam Parsons, who queried and provided the collision statistics used for the study. Nick Golding advised 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*, 69, 1125-1132

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

Campos, J. & de Rus, G. (2009) Some stylized facts about high-speed rail: A review of HSR experiences around the world *Transport Policy,* 16, 19-28

Chivers, C., Leung, B. & Yan, N. D. (2014) Validation and calibration of probabilistic predictions in ecology. *Methods in Ecology and Evolution*, 5, 1023-1032

Clevenger, A. P., Wierzchowski, J., Chruszcz, B. & Gunson, K. (2002) GIS-generated, expert-based models for identifying wildlife habitat linkages and planning mitigation passages. *Conservation Biology*, 16, 503-514

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

Dawson, T. (2012) *Kangaroos.* CSIRO Publishing

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

Dorsey, B., Olsson, M. & Rew, L. (2015) Ecological effects of railways on wildlife. In: van der Ree R., Smith D., Grilo C. (eds) *Handbook of Road Ecology*

Dunn, P. K. & Smyth, G. K. (1996) Randomized Quantile Residuals. *Journal of Computational and Graphical Statistics*, 5, 236-244

Dussault, C., Poulin, M., Courtois, R. & Ouellet, J. (2006) Temporal and spatial distribution of moose-vehicle accidents in the Laurentides Wildlife Reserve, Quebec, Canada. *Wildlife Biology*, 12, 415-425

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

Fahrig, L. & Rytwinski, T. (2009) Effects of Roads on Animal Abundance: an Empirical Review and Synthesis. Ecology and Society, 14(1), 21

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*, 26, 593-611

Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., McCarthy, M. A., Tingley, R. & Wintle, B. A. (2015) Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, 24, 276-292

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

Hillsman, E. L. & Barbeau, S. J. (2011) Enabling cost-effective multimodal trip planners through open transit data. Technical Report No. USF 21177926

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

Huijser, M. P., Duffield, J. W., Clevenger, A. P., Ament, R. J. & McGowen, P. T. (2009) Cost-benefit analyses of mitigation measures aimed at reducing collisions with large ungulates in the United States and Canada: a decision support tool. Ecology and Society, 14, 15

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

Jaarsma, C. F., van Langevelde, F. & Botma, H. (2006) Flattened fauna and mitigation: Traffic victims related to road, traffic, vehicle, and species characteristics. Transportation Research Part D: Transport and Environment, 11, 264 - 276

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

Laist, D. W., Knowlton, A. R., Mead, J. G., Collet, A. S. & Podesta, M. (2001) Collisions between ships and whales. Marine Mammal Science, 17, 35-75

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*, 17, 229-239

Lao, Y., Zhang, G., Wu, Y. & Wang, Y. (2011) Modeling animal-vehicle collisions considering animal-vehicle interactions. *Accident Analysis & Prevention*, 43(6), 1991-1998

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

Neumann, W., Ericsson, G., Dettki, H., Bunnefeld, N., Keuler, N. S., Helmers, D. P. & Radeloff, V. C. (2012) Difference in spatiotemporal patterns of wildlife road-crossings and wildlife-vehicle collisions. *Biological Conservation*, 145, 70-78

Olson, D. D., Bissonette, J. A., Cramer, P. C., Green, A. D., Davis, S. T., Jackson, P. J. & Coster, D. C. (2014) Monitoring Wildlife-Vehicle Collisions in the Information Age: How Smartphones Can Improve Data Collection. *PLoS One*, 9, e98613

Onoyama, K., Ohsumi, N., Mitsumochi, N. & Kishihara, T. (1998) Data analysis of deer-train collisions in eastern Hokkaido, Japan. In: Hayashi, C., Yajima, K., Bock, H., Ohsumi, N., Tanaka, Y. & Baba, Y. (eds.) *Data Science, Classification, and Related Methods*

Phillips, S. J. & Elith, J. (2010) POC plots: calibrating species distribution models with presence-only data. *Ecology*, 91, 2476-2484

Ramp, D. & Roger, E. (2008) Frequency of animal-vehicle collisions in NSW. In: Lunney, D., Munn, A. & Will, M. (eds) *Too close for comfort**: contentious issues in human-wildlife encounters*

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

Roger, E., Bino, G. & Ramp, D. (2012) Linking habitat suitability and road mortalities across geographic ranges. *Landscape Ecology*, 27, 1167-1181

Romin, L. A. & Bissonette, J. A. (1996) Deer: vehicle collisions: status of state monitoring activities and mitigation efforts. *Wildlife Society Bulletin*, 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. (2005) Predicting locations of moose–vehicle collisions in Sweden. *Journal of Applied Ecology*, 42, 371-382

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*

Shilling, F., Perkins, P. & Collinson, W. (2015) Wildlife/roadkill observation and reporting systems. In: van der Ree R., Smith D., Grilo C. (eds) *Handbook of Road Ecology*

Skorka, P., Lenda, M., Moron, D., Kalarus, K. & Tryjanowski, P. (2013) Factors affecting road mortality and the suitability of road verges for butterflies. *Biological Conservation*, 159, 148-157

Snow, N. P., Porter, W. F. & Williams, D. M. (2015) Underreporting of wildlife-vehicle collisions does not hinder predictive models for large ungulates. *Biological Conservation*, 181, 44 - 53

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

Thurfjell, H., Spong, G., Olsson, M. & Ericsson, G. (2015) Avoidance of high traffic levels results in lower risk of wild boar-vehicle accidents. *Landscape and Urban Planning*, 133, 98-104

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

van Strien, A. J., van Swaay, C. A. & Termaat, T. (2013) Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology*, 50, 1450-1458

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*

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

Wintle, B. A., Elith, J. & Potts, J. M. (2005) Fauna habitat modelling and mapping: a review and case study in the Lower Hunter Central Coast region of NSW. *Austral Ecology*, 30, 719-738