Managing the timing and speed of vehicles reduces wildlife-transport collision risk

Casey Visintin¹, Nick Golding², Rodney van der Ree³, Michael A. McCarthy⁴

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

²Quantitative and Applied Ecology Group, School of BioSciences, University of Melbourne, Parkville, VIC 3010, Australia - Email: nick.golding.research@gmail.com

³Ecology and Infrastructure International and School of BioSciences, University of Melbourne, Parkville, VIC 3010, Australia - Email: rvdr@unimelb.edu.au

⁴Quantitative 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 114, University of Melbourne, Parkville, VIC 3010, Australia - Email: cvisintin@student.unimelb.edu.au, Phone: +61 4 34424084

Word Count: 6951

Abstract: 354

Main Text: 4291

Acknowledgements: 53

Literature Cited: 1651

Tables: 238

Figure Captions: 253

Number of Tables: 2

Number of Figures: 7 (3 supplemental, 1 colour)

Number of References: 64

Abstract

Understanding wildlife-vehicle collision risk is critical to mitigating its negative impacts on wildlife conservation, human health and economy. Research often focuses on collisions between wildlife and road vehicles, but collision risk factors for other types of vehicles, less examined in the literature, may be informative.

We studied spatial and temporal variation in wildlife-train collision risk in the State of Victoria, Australia. We quantified train movements in space and time, and mapped species occurrence likelihood, across the railway network. Using spatially- and temporally-resolved collision data, we fitted a model to analyse collisions between trains and kangaroos; accounting for time of day, train frequency and speed, and kangaroo occurrence. We then predicted collision rates on the passenger railway network under three management scenarios relating to train speed and occurrence of kangaroos near the railway lines.

Temporal variation in animal activity was the strongest predictor of collision risk. Train speed was the second most influential variable, followed by spatial variation in likelihood of species occurrence. Reducing speeds in areas of high predicted species occurrence and during periods of peak animal activity (early morning and evening for kangaroos) was predicted to reduce collision risk the most.

Our results suggest mechanisms that might improve existing wildlife-transport collision analyses. The model can help managers decide where, when and how best to mitigate collisions between animals and transport. It can also be used to predict high-risk locations or times for (a) timetable/schedule changes (b) proposals for new routes or (c) disused routes considered for re-opening.

23 Keywords

24 crepuscular; railway; risk; species distribution model; temporal; WTC

1.1 Introduction

Wildlife collisions with transport vehicles pose a serious global problem (Litvaitis & Tash, 2008) and inspire research to investigate causes and propose solutions. For example, deer-vehicle collisions on roads are common and well-studied in North America (Huijser et al., 2007; Romin & Bissonette, 1996) and Europe (Sáenz-de-Santa-María & Tellería, 2015; Seiler, 2004). 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). Collision events disrupt transportation services, and are therefore also an economic burden. Management of wildlife-vehicle collisions will become increasingly important as new transportation networks are constructed and existing networks are expanded.

Information about the spatial and temporal distribution and magnitude of wildlife-vehicle collisions is useful to managers because it can more effectively help mitigate impacts (Mountrakis & Gunson, 2009). For example, identifying a collision hotspot along a transportation network for a particular species can inform the most appropriate form of mitigation (e.g. animal exclusion or change in vehicle speed – see Visintin et al., 2016). Collision data can also be used in statistical modelling which helps to predict the probability of wildlife-vehicle collisions (Gunson et al., 2011).

47

48 The majority of wildlife-vehicle collision modelling deals with roads and traffic (Popp & Boyle,
49 2017, van der Ree et al, 2015), yet, the problem extends to other forms of vehicular
50 transportation such as air (van Belle et al., 2007), railway (Belant, 1995; Onoyama et al., 1998;
51 Wells et al., 1999) and shipping (Laist et al., 2001) operations. Regardless of the mode of
52 transport, the modelling of collisions share some common attributes (Forman et al., 2003). The
53 movements or presence of animals are often considered in the models and may include species-
54 specific habitat variables (Roger & Ramp, 2009) or behavioural traits (Lee et al., 2010). Vehicle
55 presence or movements are also considered and may be grouped into a larger category of human
56 behaviour as humans ultimately control speeds and trajectories of vehicles (Ramp & Roger,
57 2008).

58

59 Some drawbacks of existing road ecology studies are that the time of collision is often not known
60 precisely, the volume of traffic varies temporally and is difficult to accommodate in analyses
61 (especially when time of collision is not known), and collisions might be under-reported or
62 temporally and spatially biased. These drawbacks might obscure drivers of collision risk or
63 provide inaccurate predictions of risk upon which mitigation might be implemented. These
64 considerations are especially important for studies that explore temporal variation in animal
65 activity (e.g. Dussault et al., 2006, Thurfjell et al., 2013) as reporting errors in collision times
66 may produce spurious correlations (or lack of).

67

68 Wildlife-train collisions provide a different dataset for considering factors that influence

collision risk that are often difficult to obtain. Firstly, the timetabling of trains means that their movements are much more precisely documented than those of cars. Secondly, collisions with large animals are more consistently reported for trains than cars – although under-reporting is still an issue (Dorsey et al., 2015). And thirdly, the time of collision is also reported more precisely. These benefits mean that examining factors that influence wildlife-train collision risk may help uncover factors that influence wildlife-vehicle collisions more generally.

Herein, we develop a modelling framework to predict the rate of wildlife-train collisions across a large railway network. We use kangaroos and a regional passenger railway network in south-east Australia as an example to demonstrate our methods. In addition to informing railway operators in Australia of potential kangaroo collision risks, our approach can be generalised to other species (e.g. deer) and transportation modes (e.g. road vehicles, watercraft) elsewhere in the world by incorporating parameters that accommodate patterns of species behaviour.

2.1 Materials and Methods

2.1.1 Study Area

Our study area encompassed a 1712-kilometre passenger railway network from regional Victoria, Australia (operated by V/line, a government-owned corporation) in south-east Australia (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.

2.1.2 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 kangaroo occurrence and quantified train movements and speeds to generate predictor variables (Table 1); each is described in more detail in subsequent sections.

Collisions with large mammals (e.g. domestic livestock or kangaroos) must be reported by train drivers to allow trains to be inspected and maintenance performed as required. Eastern grey kangaroos (*Macropus giganteus*, Shaw, 1790; "kangaroos" hereafter) are frequently struck in regional Victoria and large enough – up to 85kg - to cause noticeable damage to composite body panels or require the train to be removed from service for cleaning. In this study, we assumed that 'kangaroos' reported by train drivers could also include other large macropods (two other

kangaroo species), albeit these would be rare due to the limited overlap of the train network on these species' ranges. We also assumed that collisions were perfectly detected at all times of the day, however, visibility issues during non-daylight hours may have affected reporting rates.

V/line provided records of all driver-reported collisions with kangaroos spanning a seven-year period between 1 January 2009 and 31 December 2015, a total of 439 collisions. These were reduced to 404 when spatial and temporal duplicates were removed. Whilst 404 collisions over a seven-year period suggests rarity, issues with data availability suggested that this number under-represented actual collision numbers. For example, not all records were available in electronic format (V/line, personal communication), therefore we only received records that were entered into, and could be queried from, a train operator database. Nonetheless, we considered this dataset useful to test our model and its sensitivity to operational changes on the train network. Each record included incident date and time, name of service line (unique route between two towns), and estimated distance to the nearest regularly spaced 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 given the one kilometre spacing of the distance markers along the railway network.

2.1.3 Species Occurrence

Kangaroos are widespread (Dawson, 2012) and abundant in many parts of Victoria, and Eastern Grey Kangaroos, in particular, are known to occur throughout the entire extent of the regional train network, yet comprehensive distribution records are lacking in many areas. To represent risk of collision by exposure to threat, we required continuous distributional data across the entire study area and used species distribution modelling to predict relative likelihood of Eastern Grey Kangaroo occurrence. Habitat predictors - often explicitly included in wildlife-vehicle collision models - were alternatively used in a model to determine the relative occurrence of the species across the train network. 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 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. The distance-based variables accounted for geographic bias that often occurs in opportunistic or museum-specimen occurrence records that were collected in easily accessible locations. To correct for areas that demonstrated potentially high sampling effort, we included the spatial coordinates of the occurrence records.

2.1.4 Characteristics of Railway Network

To determine train movements across space and time, we accessed publicly available locations of stops and times along train routes from V/Line general transit feed specification (GTFS) data (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, for each month of the year, where trains occurred.

2.1.5 Statistical Modelling

All analyses were carried out using the software ‘R’, version 3.4.1 (R Core Team, 2017). We developed a quantitative risk model to evaluate the capacity for kangaroo activity and occurrence and railway network characteristics to predict collision likelihoods. To account for temporal variation in collision risk throughout the day, we considered periods of train movements and estimated animal activity in relation to time of day. By adding a function that allowed a flexible response of collision rate to hour of day across all seasons, the crepuscular lifestyle of kangaroos (most active at dawn and dusk; see Dawson, 2012) was incorporated. The likelihood that a collision occurred in a given grid cell i at hour j in month k ($p_{ijk} = \Pr(Y_{ijk}=1)$) depended on species occurrence O , average number of trains V , average train speed S , length of track L (offset term), and a term T_{ijk} that accounted for the activity pattern of kangaroos:

$$\text{cloglog}(p_{ijk}) = \beta_0 + \beta_1 \log(O_i) + \beta_2 \log(V_{ij}) + \beta_3 \log(S_{ij}) + T_{ijk} + \log(L_i) \quad (1)$$

The temporal kangaroo activity term T_{ijk} (Table 1) has three components:

$$T_{ijk} = \gamma_1 \sin(\pi(j - 6) / 12) + \gamma_2 \sin(\pi(j - 6) / 12)^2 + \gamma_3 A_{ijk} \quad (2)$$

171

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

$$\begin{aligned} 176 \quad A_{ijk} = & \quad (3) \\ 177 \quad & \exp(-(j - U_{ik}) / ((24 - D_{ik}) + U_{ik}) / 2)^2) - \exp(-(j - D_{ik} - 24) / ((24 - D_{ik}) + U_{ik}) / 2)^2) & \text{if } j < U_{ik}; \\ 178 \quad & \exp(-(j - U_{ik}) / ((D_{ik} - U_{ik}) / 2)^2) - \exp(-(j - D_{ik}) / ((D_{ik} - U_{ik}) / 2)^2) & \text{if } U_{ik} \leq j < D_{ik}; \\ 179 \quad & \exp(-(j - U_{ik} + 24) / ((24 - D_{ik}) + U_{ik}) / 2)^2) - \exp(-(j - D_{ik}) / ((24 - D_{ik}) + U_{ik}) / 2)^2) & \text{if } j \geq D_{ik}; \end{aligned}$$

180

181 Dawn U and dusk D times for civil twilight (six degrees below horizon) were determined for
182 each observation in the dataset using a National Oceanic & Atmospheric Administration
183 (NOAA) astronomical algorithm in the R package *maptools* (Bivand & Lewin-Koh, 2017). We
184 selected the 15th day of each month k at the centroid of each respective grid cell i to calculate
185 dawn and dusk times.

186

187 The coefficients $\gamma_1, \gamma_2, \gamma_3$ were estimated during model fitting (to collision data) and measured the
188 relative influence of the three components (Equation 2) on the shape of the curve. This functional
189 form enables the risk profile to match known activity patterns for different species (diurnal,
190 nocturnal, crepuscular, or random), account for differing activity near dawn and dusk, and is
191 cyclical over a twenty-four hour period (Figure S2) regardless of the values of $\gamma_1, \gamma_2, \gamma_3$.

192

193 Prior to modelling, we centred the species occurrence, train frequency, and train speed variables

by subtracting their means (after log-transforming the variables indicated in Equation 1). Using pairwise analysis, all predictors (O_i , V_{ij} , S_{ij} , & T_{ijk}) exhibited Pearson's product moment correlation coefficients of less than 0.4 indicating low potential effects of multi-collinearity. We also examined the correlation between the untransformed variables on the data subset to cells with collisions and subset to cells without collisions (Figure S1) - both demonstrated low values.

We fitted this generalised linear model (McCullagh & Nelder, 1989) to the data (n=291,036 one-kilometre, hourly grid cells along the train network with at least one train present) using maximum likelihood estimation with a Bernoulli 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). We also tested interactions between all variables, however, none were statistically significant and thus omitted from our study.

To assess the influence of predictors on model fit we constructed five versions of the model, each using different combinations of variables (Table 1) and performed likelihood ratio tests to assess the importance of each predictor. To assess performance of each model version, 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, applying the method ten times, each time holding out a different fold. For each assessment we obtained two performance

216 metrics; area under the receiver operator characteristic (ROC) curve (Metz, 1978) and regression
217 of observations on predictions (Cox, 1989; Miller, 1991). We repeated this procedure for 100
218 iterations producing a total of 1000 sets of performance metrics and compared them with those
219 from the model fitted to all data. We repeated this procedure for each of the five models. To
220 examine model fit using all variables, we generated randomised quantile residuals (Dunn &
221 Smyth, 1996) and plotted them against collision probability, kangaroo occurrence, number of
222 trains per hour, average train speed, and hour. Note, we intentionally refrained from using
223 reduction in deviance as a performance measure as it is not appropriate for rare event encounter
224 models.

225

226 Using 1000 simulations of the best model fitted on all data, we predicted the number of expected
227 kangaroo-train collisions in the study area under different management scenarios:

228 A) no change to operations,

229 B) reduced train speeds in moderate to high kangaroo occurrence areas, and

230 C) exclusions of kangaroos in areas with highest average speed of trains.

231 Scenario B involved capping train speeds at 60 km h⁻¹ in grid cells with kangaroo relative
232 occurrence likelihoods of 0.3 or above during the hours of 5 a.m. to 9 a.m. and 4 p.m. to 8 p.m.

233 This change affected 548 out of 3,024 (~18%) unique weekly routes across a total of 171
234 kilometres of railway. In scenario C, relative kangaroo occurrence was reduced to nearly zero
235 (0.01) in all grid cells with average train speeds of more than 120 km h⁻¹ (total track length = 125
236 km). This simulates reducing access of kangaroos to the railway network (e.g. by fencing). The
237 values for reducing kangaroo abundance were used for all hours of the day as this management

238 strategy would most likely involve exclusion or reduction in animal populations which operate
239 irrespective of temporal variation.

3.1 Results

The signs of model coefficients (positive or negative) indicated plausible relationships between collision likelihood and each predictor variable. All variables except train frequency demonstrated strong influence on collision risk (Table 2). Dropping the temporal animal activity term had the largest negative effect on model fit, changing the Akaike information criteria (AIC) score from 5528 to 5740. The model residuals (using all variables) demonstrated no strong patterns with respect to predicted collision likelihoods and model predictors (Figure S3).

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. A strong predictor was train speed; collision risk increased exponentially with considerable increases at speeds above 85 km hr⁻¹. Increasing train speed from 105 to 125 km hr⁻¹ doubled collision risk (Figure 3a). Kangaroo occurrence was also an influential predictor of risk of collision. Collision risk increased rapidly at low values of occurrence and more slowly at higher values. Collision risk doubled between values of 0.25 and 0.75 relative likelihood of kangaroo occurrence (Figure 3b). Train frequency had very little influence on collision likelihood or model fit, compared to the other predictors (Table 2; Figure 3c) and was excluded from all models except one (full).

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 5:30 a.m. and 6:30 p.m. with a higher risk occurring in the morning. The response of collision risk over time was most uncertain in the morning peak. Both peaks showed similar shape - the density and spread of

262 collision risk around the maxima - and the lowest collision risk occurred around noon.
263
264 The models fitted on all data and fitted on the subsets of the data during cross-validation
265 performed differently (Figure 4). The best ROC values were for both the full model and model
266 excluding train frequency (0.82). However, the calibration statistics (intercept and slope of
267 regression line between observations and predictions) between all model specifications were
268 similar – slopes varied between 0.990 and 0.996 and intercepts varied between -0.080 and -
269 0.046. Dropping the temporal animal activity term from the best model did not affect the
270 calibration statistics, however it reduced the ROC value from 0.82 to 0.76.
271
272 Both of the simulated management scenarios reduced the predicted number of collisions from the
273 baseline (Scenario A) estimated with no management (409 collisions +/-39). Scenario C reduced
274 expected collisions by approximately 8.3% (375 collisions +/-34) whilst scenario B reduced
275 collisions by 10.0% (368 collisions +/-35).

4.1 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), road (Lao et al., 2011; Roger et al., 2012, Visintin et al., 2016) and marine collisions (Conn & Silber, 2013; van der Hoop et al., 2012) with other species. Our results also suggest that changing the timing and speed of vehicles, regardless of the mode of transportation, reduces collisions between wildlife and moving vehicles.

Train speed is 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). 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, although more accurate than road vehicle movements, there is still some uncertainty in both the location and trajectories of actual trains. Further study using automated global positioning system (GPS) waypoints of train (or other vehicular) movements would further reduce this uncertainty.

Species occurrence is also a useful predictor for collision risk, however, as previously

mentioned, it is difficult to accurately predict. One feature of the model framework is the flexibility of choice in how to represent species occurrence. We employed published methods to determine species 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 wildlife collision on railways (Seiler, 2005) and roads (Schwabe et al., 2002; Seiler, 2004). However, population control of kangaroos (via government-permitted culling) occurs periodically in certain areas and our model framework can accommodate changes in occurrence driven by these factors by including relevant predictor variables such as density indexes. Further, we assumed that the frequency of trains did not affect kangaroo occurrence near the railway network although this has been shown in other work (Dorsey et al., 2015; Seiler et al., 2011). A final note is that the predicted spatial distribution of species did not account for temporal variation as done for the other predictor and dependent variables. This is a common limitation in species occurrence models, however, it can be partially addressed by using temporally and spatially complete survey data (Erwig et al., 1999) and spatio-temporal models (Hooten, 2011).

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 (Huijser et al., 2007), and law-enforcement agencies (e.g. accident databases), to collect and archive carcass data, however, the coverage is often sparse due to the large spatial extent of road networks and insignificance of, or temporal uncertainty of collision events. Technology has been shown to assist with data collection (Olson et al., 2014; Shilling, 2015) but these methods do not address temporal lags (i.e. reporting the carcass discovery and not the collision time). One potential solution is to employ automatic detection systems in vehicles to record both presence of and collisions with species.

We extend our analysis by presenting a conceptual modelling tool to assist managers in creating safer (for people and wildlife) and more ecologically-sustainable transportation networks. Our framework supports management decisions in two distinct areas: reduction of animal presence (e.g. deterrents or exclusions) or reduction of vehicle 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. Several extremely 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. However, most other passenger (and freight) railway networks in the world, including those in Victoria, are not fenced. 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 nearly zero.

Each of our management scenarios reduced collisions, albeit with potentially different cost-benefit ratios. Collisions with kangaroos in Australia result in costs to the railway operator through removal of trains from service for cleaning and repair (V/line, personal communication). We were not able to obtain these costs for this study, however, this information would be useful to extend the application of our work. The benefit from reductions in annual collisions will have varying impacts 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 enormous costs. In these situations, avoiding a single collision event has remarkable benefit. When the costs of different management strategies (i.e. mitigation) are also considered, a cost-benefit analysis may be performed. This has been applied to collisions with road vehicles (Huijser et al., 2009) and can also be applied to railway, air and marine transport (see costs in Davenport & Davenport, 2006).

364

365 Herein, we have demonstrated a model framework that functions as an effective management
366 support tool. It utilises existing sources of data, is logically organised, and is
367 transferable/scalable to other transportation networks and species. Other potential uses of the
368 framework may include an ongoing implementation where the model is updated based on new
369 information and reports risk to operators in real-time. Further, our framework suggests important
370 considerations for managing wildlife-vehicle collisions on any transportation network
371 irrespective of scale or location. It is our hope that these insights will contribute to improving
372 transportation and reducing its negative global effects on humans and wildlife populations.

373 Acknowledgements

374 We thank V/Line for their generous collaboration; in particular, Sam Parsons, who queried and
375 provided the collision statistics used for the study. Michaela Plein provided valuable comments
376 on the manuscript. This project was supported by a University of Melbourne International
377 Research Scholarship and the Australian Research Council Centre of Excellence for
378 Environmental Decisions.

379 Data Accessibility

380 Model dataset - Archived on GitHub; original collision data is V/Line's Intellectual Property (IP)

381 and authorisation for additional use must be granted by V/Line

382 R code – archived and available on GitHub (https://github.com/cvisintin/coll_model_trains)

383 Literature Cited

- 384 Andreassen, H. P., Gundersen, H. & Storaas, T. (2005) The effect of scent-marking, forest
385 clearing, and supplemental feeding on moose-train collisions. *The Journal of Wildlife*
386 *Management*, 69, 1125-1132
- 387 Belant, J. L. (1995) Moose collisions with vehicles and trains in northeastern Minnesota. *Alces*,
388 31, 45-52
- 389 Bivand, R. & Lewin-Koh, N. (2017) maptools: Tools for Reading and Handling Spatial Objects.
390 R package version 0.9-2. <https://CRAN.R-project.org/package=maptools>
- 391 Campos, J. & de Rus, G. (2009) Some stylized facts about high-speed rail: A review of HSR
392 experiences around the world *Transport Policy*, 16, 19-28
- 393 Chivers, C., Leung, B. & Yan, N. D. (2014) Validation and calibration of probabilistic
394 predictions in ecology. *Methods in Ecology and Evolution*, 5, 1023-1032
- 395 Conn, P. B. & Silber, G. K. (2013) Vessel speed restrictions reduce risk of collision-related
396 mortality for North Atlantic right whales *Ecosphere*, 4, 1-16
- 397 Cox, D. R. & Oakes, D. (1984) *Analysis of survival data*. CRC Press
- 398 Cox, D. R. & Snell, E. J. (1989) *Analysis of binary data*. CRC Press
- 399 Davenport, J. & Davenport, J. L. (2006) *The Ecology of Transportation: Managing Mobility for*
400 *the Environment*, Springer, Netherlands
- 401 Dawson, T. (2012) *Kangaroos*. CSIRO Publishing
- 402 Dorsey, B., Olsson, M. & Rew, L. (2015) Ecological effects of railways on wildlife. In: van der
403 Ree R., Smith D., Grilo C. (eds) *Handbook of Road Ecology*
- 404 Dunn, P. K. & Smyth, G. K. (1996) Randomized Quantile Residuals. *Journal of Computational*

405 *and Graphical Statistics*, 5, 236-244

406 Dussault, C., Poulin, M., Courtois, R. & Ouellet, J. (2006) Temporal and spatial distribution of
 407 moose-vehicle accidents in the Laurentides Wildlife Reserve, Quebec, Canada. *Wildlife Biology*,
 408 12, 415-425

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

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

414 Erwig, M.; Gu, R. H., Schneider, M., Vazirgiannis, M. & others (1999) Spatio-temporal data
 415 types: An approach to modeling and querying moving objects in databases *GeoInformatica*, 3,
 416 269-296

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

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

422 Guillera-Aroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E.,
 423 McCarthy, M. A., Tingley, R. & Wintle, B. A. (2015) Is my species distribution model fit for
 424 purpose? Matching data and models to applications. *Global Ecology and Biogeography*, 24, 276-
 425 292

426 Gundersen, H. & Andreassen, H. P. (1998) The risk of moose *Alces alces* collision: A predictive

427 logistic model for moose-train accidents. *Wildlife Biology*, 4, 103-110
 428 Gunson, K. E., Mountrakis, G. & Quackenbush, L. J. (2011) Spatial wildlife-vehicle collision
 429 models: A review of current work and its application to transportation mitigation projects.
 430 *Journal of Environmental Management*, 92, 1074-1082
 431 Hillsman, E. L. & Barbeau, S. J. (2011) Enabling cost-effective multimodal trip planners through
 432 open transit data. Technical Report No. USF 21177926
 433 Hooten, M. B. (2011) The State of Spatial and Spatio-Temporal Statistical Modeling. In: Drew,
 434 C. A.; Wiersma, Y. F. & Huettmann, F. (eds.) *Predictive Species and Habitat Modeling in*
 435 *Landscape Ecology: Concepts and Applications*, Springer, New York, 29-41
 436 Huijser, M. P., Wagner, M. E., Hardy, A., Clevenger, A. P. & Fuller, J. A. (2007) Animal-
 437 vehicle collision data collection throughout the United States and Canada. *ICOET 2007*
 438 *Proceedings: Wildlife and Terrestrial Ecosystems*
 439 Huijser, M. P., Duffield, J. W., Clevenger, A. P., Ament, R. J. & McGowen, P. T. (2009) Cost-
 440 benefit analyses of mitigation measures aimed at reducing collisions with large ungulates in the
 441 United States and Canada: a decision support tool. *Ecology and Society*, 14, 15
 442 Hurley, M. V., Rapaport, E. K. & Johnson, C. J. (2009) Utility of Expert-Based Knowledge for
 443 Predicting Wildlife-Vehicle Collisions. *The Journal of Wildlife Management*, 73, 278-286
 444 Jaarsma, C. F., van Langevelde, F. & Botma, H. (2006) Flattened fauna and mitigation: Traffic
 445 victims related to road, traffic, vehicle, and species characteristics. *Transportation Research Part*
 446 *D: Transport and Environment*, 11, 264 - 276
 447 Jones, M. E. (2000) Road upgrade, road mortality and remedial measures: impacts on a
 448 population of eastern quolls and Tasmanian devils. *Wildlife research*, 27, 289-296

449 Laist, D. W., Knowlton, A. R., Mead, J. G., Collet, A. S. & Podesta, M. (2001) Collisions
 450 between ships and whales. *Marine Mammal Science*, 17, 35-75
 451 Langley, R. L., Higgins, S. A. & Herrin, K. B. (2006) Risk factors associated with fatal animal-
 452 vehicle collisions in the United States, 1995-2004. *Wilderness & Environmental Medicine*, 17,
 453 229-239
 454 Lao, Y., Zhang, G., Wu, Y. & Wang, Y. (2011) Modeling animal-vehicle collisions considering
 455 animal-vehicle interactions. *Accident Analysis & Prevention*, 43(6), 1991-1998
 456 Lee, E., Croft, D. B. & Ramp, D. (2010) Flight response as a causative factor in kangaroo-
 457 vehicle collisions. In: Coulson, G. & Eldridge, M. (eds.) *Macropods: The Biology Of Kangaroos,*
 458 *Wallabies And Rat-kangaroos, CSIRO Publishing, Clayton, VIC, Australia*, 301-311
 459 Litvaitis, J. A. & Tash, J. P. (2008) An approach toward understanding wildlife-vehicle
 460 collisions. *Environmental Management*, 42, 688-697
 461 McCullagh, P. & Nelder, J. A. (1989) *Generalized linear models*. CRC press
 462 Metz C.E. (1978) Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8(4), 283-
 463 298
 464 Miller, M. E., Hui, S. L. & Tierney, W. M. (1991) Validation techniques for logistic regression
 465 models. *Statistics in Medicine*, 10, 1213-1226
 466 Mountrakis, G. & Gunson, K. (2009) Multi-scale spatiotemporal analyses of moose-vehicle
 467 collisions: a case study in northern Vermont International. *Journal of Geographical Information*
 468 *Science*, 23, 1389-1412
 469 Olson, D. D., Bissonette, J. A., Cramer, P. C., Green, A. D., Davis, S. T., Jackson, P. J. & Coster,
 470 D. C. (2014) Monitoring Wildlife-Vehicle Collisions in the Information Age: How Smartphones
 471 Can Improve Data Collection. *PLoS One*, 9, e98613

472 Onoyama, K., Ohsumi, N., Mitsumochi, N. & Kishihara, T. (1998) Data analysis of deer-train
 473 collisions in eastern Hokkaido, Japan. In: Hayashi, C., Yajima, K., Bock, H., Ohsumi, N.,
 474 Tanaka, Y. & Baba, Y. (eds.) *Data Science, Classification, and Related Methods*
 475 Phillips, S. J. & Elith, J. (2010) POC plots: calibrating species distribution models with presence-
 476 only data. *Ecology*, 91, 2476-2484
 477 R Core Team (2017). R: A language and environment for statistical computing. R Foundation for
 478 Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
 479 Ramp, D. & Roger, E. (2008) Frequency of animal-vehicle collisions in NSW. In: Lunney, D.,
 480 Munn, A. & Will, M. (eds) *Too close for comfort: contentious issues in human-wildlife*
 481 *encounters*
 482 Roger, E. & Ramp, D. (2009) Incorporating habitat use in models of fauna fatalities on roads.
 483 *Diversity and Distributions*, 15, 222-231
 484 Roger, E., Bino, G. & Ramp, D. (2012) Linking habitat suitability and road mortalities across
 485 geographic ranges. *Landscape Ecology*, 27, 1167-1181
 486 Romin, L. A. & Bissonette, J. A. (1996) Deer: vehicle collisions: status of state monitoring
 487 activities and mitigation efforts. *Wildlife Society Bulletin*, 24, 276-283
 488 Rowden, P., Steinhardt, D., & Sheehan, M. (2008) Road crashes involving animals in Australia.
 489 *Accident Analysis & Prevention* 40(6), 1865-1871
 490 Sáenz-de-Santa-María, A. & Tellería, J. L. (2015) Wildlife-vehicle collisions in Spain European.
 491 *Journal of Wildlife Research*, 61, 399-406
 492 Sainsbury, A., Bennett, P. & Kirkwood, J. (1995) The welfare of free-living wild animals in
 493 Europe: harm caused by human activities. *Animal Welfare*, 4, 183-206

494 Schwabe, K. A.; Schuhmann, P. W. & Tonkovich, M. (2002) A dynamic exercise in reducing
 495 deer-vehicle collisions: Management through vehicle mitigation techniques and hunting. *Journal*
 496 *of Agricultural and Resource Economics*, 27, 261-280
 497 Seiler, A. (2004) Trends and spatial patterns in ungulate-vehicle collisions in Sweden. *Wildlife*
 498 *Biology*, 10, 301-313
 499 Seiler, A. (2005) Predicting locations of moose-vehicle collisions in Sweden. *Journal of Applied*
 500 *Ecology*, 42, 371-382
 501 Seiler, A., Helldin, J. & Olsson, M. (2011) Klövviltolyckor på järnväg: kunskapsläge,
 502 problemanalys och åtgärdsförslag (in Swedish). Trafikverket rapport 2011:048. In English –
 503 Ungulate-train collisions in Sweden – literature review, GIS-analyses of mortality data and
 504 train drivers' experiences.
 505 Shilling, F., Perkins, P. & Collinson, W. (2015) Wildlife/roadkill observation and reporting
 506 systems. In: van der Ree R., Smith D., Grilo C. (eds) *Handbook of Road Ecology*
 507 Skorka, P., Lenda, M., Moron, D., Kalarus, K. & Tryjanowski, P. (2013) Factors affecting road
 508 mortality and the suitability of road verges for butterflies. *Biological Conservation*, 159, 148-157
 509 Snow, N. P., Porter, W. F. & Williams, D. M. (2015) Underreporting of wildlife-vehicle
 510 collisions does not hinder predictive models for large ungulates. *Biological Conservation*, 181,
 511 44 - 53
 512 Thurfjell, H., Spong, G., Olsson, M. & Ericsson, G. (2015) Avoidance of high traffic levels
 513 results in lower risk of wild boar-vehicle accidents. *Landscape and Urban Planning*, 133, 98-104
 514 van Belle, J., Shamoun-Baranes, J., Van Loon, E. & Bouten, W. (2007) An operational model
 515 predicting autumn bird migration intensities for flight safety. *Journal of Applied Ecology*, 44,

516 864-874

517 van der Hoop, J. M., Vanderlaan, A. S. M. & Taggart, C. T. (2012) Absolute probability
518 estimates of lethal vessel strikes to North Atlantic right whales in Roseway Basin, Scotian Shelf
519 *Ecological Applications*, 22, 2021-2033

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

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

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

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

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

Figure Captions

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 and major towns are starred. Wildlife-train collisions (reported between 2009-2014) are shown as crosses.

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.

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. Note, the effect of hour is represented across all seasons (12 month period) as predicted collision risk from the model fit is determined using dawn and dusk times that vary based on month - represented by twelve replicates of weekly train movements. Shading indicates 95% confidence intervals around coefficient estimates.

Figure 4: Comparison between all five versions of collision model (the “wo” prefix indicates the variable dropped from the model) fitted to full data and to cross-validated subsets. ROC (receiver operating characteristic) measures the models’ ability to discriminate between collisions and no-collisions. Open circles represent ROC for the full data model, solid dots represent mean values

554 and solid lines the ranges for the 1000 cross-validated subsets.

555 **Table 1:** Predictor variables used in collision model; means and ranges are expressed on
556 untransformed scale. T1, T2 and T3 determine a curve shape representing relative species
557 activity (see Equation 2 and Figure S2). Note, the range of EGK occurrence values exclude zero
558 due to the link function used to express the likelihood as a probability.

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	3; 1:18
SPEED	Mean train speed in grid cell	km h ⁻¹	86.7; 16.5:147.7
T1	Temporal function linear component	-	0.25; -1.00:1.00
T2	Temporal function quadratic component	-	0.47; 0.00:1.00
T3	Temporal function astronomical component	-	-0.05; -0.98:0.98

559

560 **Table 2:** Summary of likelihood ratio tests (relative to the best model) and Akaike information
561 criterion (AIC) for five model specifications using all data (n=291,036 grid cells). “wo” prefix
562 indicates variables were omitted from the model specification. Note, “temp” refers to the
563 temporal component that is made up of three variables that model animal activity.

Model	egk (O _i)	trains (V _{ij})	speed (S _{ij})	temp (T _{ijk})	Difference in Log Likelihood	Degrees of Freedom (Difference)	X ²	p-value	AIC
wo_trains (best)	x		x	x	—	6	—	—	5528
wo_trains_egk			x	x	58.9	5 (-1)	117.8	<.00001	5644
wo_trains_speed	x			x	79.0	5 (-1)	157.9	<.00001	5684
wo_trains_temp	x		x		108.7	3 (-3)	217.5	<.00001	5740
full	x	x	x	x	0	7 (+1)	0.043	0.84	5530

564

Figure 1

Figure 2

Figure 3

Figure 4

Figure S1: Correlation between untransformed predictor variables in all cells with **a)** reported collisions (n=404) and **b)** no reported collisions (n=290632). ‘egk’ is relative likelihood of kangaroo occurrence in grid cell; ‘trains’ is train frequency in grid cell; ‘speed’ is mean train speed in grid cell; ‘hour’ is the time of day in grid cell. Values of one or negative one (black) indicate high correlation whilst values of zero (light gray) indicate low correlation. Visual representations of correlation are shown above the diagonal line and correlation values are reported below the diagonal line.

Figure S2: **a)** Components used to build temporal 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 (g_1 , g_2 , g_3) used to produce curves are shown above each

591 respective graph.

592 a)

593 b)

594 d)

e)

595 **Figure S3:** Binned randomised quantile (RQ) residuals plots - average residual versus the
 596 average predictor value for each bin. For all continuous predictors, bins are equally spaced and
 597 determined by taking the square root of the total number of unique values for each predictor; for
 598 the discrete predictor of hour, there are 23 total bins. The gray lines indicate ± 2 standard-error
 599 bounds which should contain 95% of the residuals if the model were actually true. Note some
 600 variables are log-transformed to assist with visual inspection of the residuals.