Developing Proactive Solutions to Animal Strikes on Regional Roads

Suyash Joshi

Gopalan Nair
Department of Mathematics and Statistics
University of Western Australia

Inge Koch
Department of Mathematics and Statistics
University of Western Australia

Maria Drysdale
Central and Northern Regions
Main Roads WA

Abstract

Animal strikes by vehicles on regional roads have led to damaged property, loss of animals and pose a serious risk to drivers, particularly cattle strikes (CS), which are the primary focus of this project. This project provides Main Roads WA (MRWA) the ability to identify and investigate a variety of potential pre-emptive factors that may contribute to CS. This is achieved by applying data science techniques to transform existing data within MRWA, along with supporting data into a format that allows for a comparison to be made between regions with and without CS. Enabling MRWA to better target existing resources to prevent future CS through measures such as new animal warning signs and pastoral animal fencing. Furthermore, this transformed data is applied to create the foundations for a predictive model that MRWA can use to find high-risk areas for CS. The factors of vehicle speed, traffic volume, and whether a road is permitted to be used for concessional loading by heavy vehicles, were all found to be strongly correlated with determining whether an area is high-risk for CS.

1. Introduction

Animal vehicle collisions (AVCs) have gradually increased as new infrastructure develops into existing natural habitats and pastoral lands. AVCs hinder MRWA in providing road users with a safe, reliable, and consistent road experience which is the primary organisational aspiration (Main Roads Western Australia, 2020). However, many initiatives to curb AVCs cannot be implemented without the consent of the pastoralists (Department of Planning, Lands and Heritage, 1997) as most regional roads run through pre-existing pastoral lands created before any roads passed through them. Thus, any proposed solution must be evidence-based and demonstrate its necessity, to gain the cooperation of these stakeholders. As the pastoralists are mainly concerned with the impact their animals can have on AVCs, this project initially concentrated on pastoral animal strikes. However, it was quickly found that CS made up the vast majority of pastoral animal strikes in regional areas. Therefore, CS became the specific focus of this project. Regional areas in the context of this project are Kimberley, Pilbara, Midwest-Gascoyne, and Goldfields-Esperance. This does not include the Metropolitan, South West, Great Southern, or Wheatbelt regions.

1.1 Current method used by MRWA

In 2013 the Animal Hazard Mitigation Program (AHMP) was commenced by MRWA to address the issue of AVCs (Main Roads Western Australia, 2020). As part of this program, MRWA quantitatively defined what an AVC-hotspot was to help target existing resources.
It was defined as any stretch of road where the 10-kilometer rolling average of AVCs over a 3-year period exceeded 0.25, as seen in figure 1.

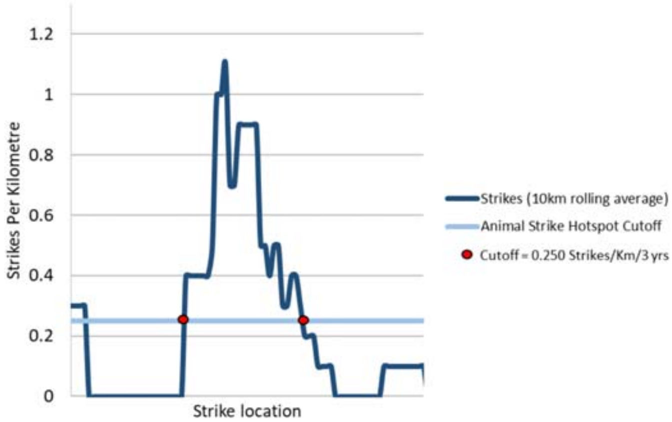

Figure 1 Sample Hotspot Identification Figures, Sample Station – Great Northern Hwy – Kimberley which uses a 10km rolling average to find the strike rate (Main Roads Western Australia, 2020).

The aforementioned method will be referred to as in the MRWA AVC identifier. This method may have been relatively effective as the years following its implementation, AVCs in pastoral regions have noticeably reduced. However, it cannot be ruled out that these trends are due to unrelated factors that have changed between 2013 and today.

1.2 Project aims

The primary aim of this project is to build upon the currently used MRWA AVC identifier, by integrating existing data to find future AVC risk areas, specifically CS, through the use of pre-emptive factors. This should allow MRWA to test theories about what may contribute to CS in a systematic and quantitative way. By applying a data-driven approach, the findings can be effectively conveyed to stakeholders as to why certain areas are being targeted in the following year which may not have been prone to CS in the previous years. This overcomes the shortfall of the MRWA AVC identifier which cannot identify new risk areas as a result of changing conditions.

2. Process

2.1 Data collection process

At the start of the project, meetings with MRWA staff were conducted to discuss how AVC data was recorded and stored at MRWA. From this, two main sources of data were identified, the AHMP dataset, and Detailed Crash History (DCH) dataset, both had observations from

2016 to 2020 and only included main roads and highways, which fall under the authority of MRWA. The AHMP dataset was then used to map all the CS from 2016 to 2020 as seen in figure 2. This was done using Microsoft PowerBI, which is a software used for data visualisation.

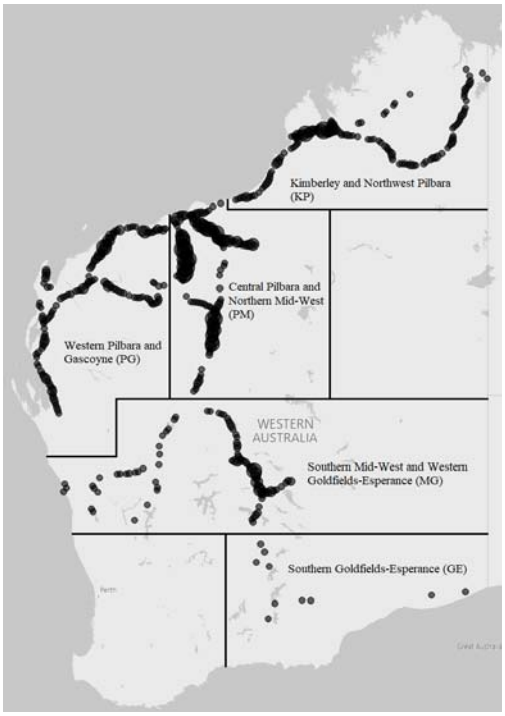

Figure 2 Mapped cattle strikes using AHMP data, each bubble identifies a 10km segment of the road and its size is proportionate to the number of CS in that segment. Labeled are the 5 sub-regions being analysed.

Five sub-regions were defined with boundaries as seen in figure 2 to conduct separate analyses for each sub-region. This was done because WA has a variety of geographic conditions, which could result in certain factors only being important in some sub-regions. Note, these sub-regions are built around road networks that have clusters of CS, and do not follow the exact boundaries of the regional areas mentioned earlier. These sub-regions are referred to as Kimberley and Northwest Pilbara (KP), Central Pilbara and Northern Mid-West (PM), Western Pilbara and Gascoyne (PG), Southern Mid-West and Western Goldfields Esperance (MG), and Southern Goldfields-Esperance (GE).

Further meetings were later conducted to learn about what factors MRWA believes contribute to CS and the existing data on these factors that can be implemented as part of the project. These factors alongside factors used by similar studies in Edmonton Canada (Found et al., 2011), Texas USA (Burton et al., 2014), and Sweden (Seiler, 2005) were used to gather 8 broad factor categories which are Concessional Loading, Pastoral Fencing, Pastoral Station, Number of Culverts, Number of Floodways, Vehicle Speed, Vehicle Traffic and Adjacent Material. These factor categories were then used to gather specific datasets from MRWA of interest.

2.2 Processing the data

Using R, a programming language for data science and statistical computing, the road network was segmented into 10km sections using MRWAs road network dataset. Next, using the roads code and Straight Line Kilometer (SLK) value the number of CS for each year in that 10km segment were recorded. For example, if a CS occurred 500km along Great Northern Highway, with a road code of H006, the road code and SLK would be matched in that section of the road in the road network dataset. This was done to create a response variable which was defined as the number of CS that occurred in that 10km segment for each year from 2016 to 2020 for the linear regression. As well as a binary variable that tracked whether at least 1 CS had occurred in that 10km section for the logistic regression, which was defined as a high-risk area for CS. A similar approach was taken to integrate each of the potential pre-emptive factors that were collected.

2.3 Statistical analysis and modeling

A statistical analysis using the methods of linear and logistic regression was then conducted on the processed data. Regression analysis is used to find whether there exists a relationship between a response variable and predictor variables. These two regression methods were chosen due to their ease of implementation and interpretability for finding statistically significant variables impacting CS. A statistically significant variable that impacts CS is selected at a 5% rejection level for the linear and logistic regressions. This is measured by the p-value, which if it is smaller than 5%, the null hypothesis is rejected. For this project, the specific null hypothesis would be that the predictor variables do not impact the variation in the number of CS for the linear regression. For the logistic regression, the null hypothesis would be that the predictor variables do not impact the variation in whether a 10km segment is classified as a high-risk area for CS. For the linear regression, there is also the Adjusted R-Squared, which measures the proportion of the variation in the response variable that is explainable by the predictor variable, adjusting for the number of predictors.

Finally, a predictive model was constructed using the two regression methods to make predictions on whether a road segment was a high-risk area for CS. This predictive model was then compared to a baseline model that was created using the average number of CS from 2016 to 2019 to predict the number of CS for 2020. Although not identical, the baseline models approach is similar to that used in the MRWA AVC identifier. If the predicted CS for this baseline model were greater than 1, it was classified as a high-risk area. The comparison between the baseline model and the predictive model was made using the measure of balanced accuracy for both models. Balanced accuracy takes into account the relative imbalance in the number of segments with and without CS and is a better measure for uneven data than just accuracy.

3. Results and Discussion

The adjusted R-Squared values found were ~11% for the entire region of the study. With the highest being ~19% in the PM sub-region and the lowest being ~1% in the GE sub-region. Table 1 shows the 8 broad factor categories and the regions in which these were statistically significant.

Vehicle speed and traffic volume are both consistently important across most of the sub-regions. Within these factors, the sub-factors of the 85th percentile speed of vehicles, and the percentage of total vehicles that travel during the night are positively correlated with CS. Where concessional loading permits are allowed for heavy vehicles, there is also a strong positive correlation with CS, especially in the PM sub-region. Together, these two observations suggest that areas of the road which have a high level of heavy vehicle traffic, particularly during the night, could be important to consider when deciding where to implement targeted preventative measures. For example, if a stretch of road is expected to be used for concessional loading for heavy vehicles in the following year, it may be important to consider whether fencing exists and is in good repair in that stretch of road.

Factor	Sig. in Linear	Sig. in Logistic
Concessional Loading	Com, MG	Com
Pastoral Fencing		
Pastoral Station	Com, PG, MG	Com, KP, PG
Number of Culverts	MG	KP, MG, GE
Number of Floodways	PG, MG	MG
Vehicle Speed	Com, KP, PM, PG, MG	Com, KP, PM, MG, GE
Vehicle Traffic	Com, KP, PM, PG	Com, KP, PM, PG, MG, GE
Adjacent Material	PM, PG	Com, GE

Table 1 The statistically significant (Sig.) factors in the Logistic and Linear regressions and the sub-regions in which they are significant. With "Com" referencing all the 5 sub-regions combined which looks at the entire region of study

Note, pastoral fencing did not have consistent useable data for the year of fence completion or data on the condition of existing fencing. This means that it is cannot be used to determine whether new fencing is positively, negatively, or not at all correlated with CS. The variables of Culverts and Floodways were moderately statistically significant and positively correlated with CS in some sub-regions. This observation could support the theory that areas where water is likely to pool attracts animals looking for water and vegetation, which was a specific theory MRWA staff were interested in.

The predictive model that was developed using the logistic regression model, had a balanced accuracy between 60% to 70% in its predictive accuracy across the entire region of study after model tuning. When the baseline model was tested for its predictive accuracy, it consistently outperformed the predictive model that was developed using pre-emptive factors by ~10% across the entire region of study. This large difference suggests that given the variables we have available for this project, it is unlikely that we can outperform this baseline model. As this baseline model is similar to that used by MRWA, it cannot be justified to use this projects developed predictive model over the existing MRWA AVC identifier.

4. Conclusion and Future Work

This project has demonstrated that using a host of existing datasets within MRWA, we can quantitatively investigate theories proposed by domain experts regarding factors may contribute to CS. This can assist MRWA in providing an evidence-based justification to influence the decisions of pastoralists and the PLB in approving MRWAs initiatives, such as pastoral fencing which is built on pastoral lands. However, predictions on the number of CS and whether a particular segment is a high-risk area for CS cannot yet be achieved as the predictive model is still in its early stages. The next steps would be to fine-tune it and also test it using linear regression and other statistical approaches. In addition, the DCH dataset includes a variety of variables giving context to how severe CS were. Investigating this can allow for a better understanding of what conditions contribute to CS being classified as severe.

Beyond this project, future work should integrate more of the vast amounts of potential variables to test more theories to gain a better insight into the issue. These additional variables should work towards developing a predictive model that is reliable enough to consistently outperform the existing approach. This existing approach is the MRWA AVC identifier that is similar to the baseline model. This projects method for segmenting the road network to investigate CS can also be extended to other issues faced by MRWA and allow them to implement a data-driven approach to understanding them.

5. Acknowledgements

Thanks to Maria Drysdale, Gopalan Nair and Inge Koch for their guidance as my mentors and supervisors for this project. As well as all the assistance provided by Jeremy Leggoe and Amanda Bolt at the CEED Office. Also, thanks to the MRWA staff involved in this project. Particular mentions include Anthony Maroni, Simon Beard, Omar Sadi, the Business Intelligence team, and all staff involved in the meetings conducted to understand the problem of AVCs on regional roads. As well David Colvin and Teong Chuah from the Department of Planning Lands and Heritage.

6. References

Burton, Prozzi, Buddhavarapu (2014). Predicting animal-vehicle collisions for mitigation in texas. *Proceedings of the International Safer Roads Conference, Cheltenham, UK*. https://saferroadsconference.com/wp-content/uploads/2016/05/Monday-pm-SandC-7-Burton_Maria_127_V1_201427-Predicting-Animal-Vehicle-Collisions-for-Mitigation-in-Texas.pdf

Department of Planning, Lands and Heritage (1997, October). *Land Administration Act*. Minister for Lands. https://www.legislation.wa.gov.au/legislation/statutes.nsf/main_mrtitle_509_homepage.html

Found, Boyce. (2011). Predicting deer–vehicle collisions in an urban area. *Journal of environmental management*, 92(10):2486–2493. https://www.sciencedirect.com/science/article/pii/S0301479711001654

Main Roads Western Australia (2020, November). *Annual Report Pastoral Animal Hazard Advisory Group*. Government of Western Australia. https://www.mainroads.wa.gov.au/globalassets/travel-information/safety/pastoral-animal-hazard-advisory-group-report-2020.pdf

Main Roads Western Australia. (2021). *Our organisation*. Government of Western Australia. https://www.mainroads.wa.gov.au/about-main-roads/our-organisation/#:~:text=Our%20Aspiration%3A%20To%20provide%20world,sustainable%20road%2Dbased%20transport%20system

Seiler. (2005). Predicting locations of moose-vehicle collisions in Sweden. *Journal of Applied Ecology*, 42(2):371–382. https://doi.org/10.1111/j.1365-2664.2005.01013.x