WAGENINGEN
UNIVERSITY & RESEARCH

IDENTIFYING HIGH-RISK ZONES: PREDICTIVE FACTORS FOR WILDLIFE COLLISIONS IN THE NETHERLANDS. A GIS-BASED ANALYSIS

2025

Presented By
Ruud Roos
1033544

Supervised by Dr. Pim van Hooft
Wildlife Ecology and Conservation Group WEC80436
Wageningen University and Research

Inhoud

Abstract

This study examined the predictive factors of wildlife-vehicle collisions within the Netherlands, focusing on roe deer (*Capreolus capreolus*). Due to its high population density and extensive road network, the Netherlands has a large degree of habitat fragmentation. This road network leads to wildlife-vehicle collisions, posing a threat to both animal and human safety.

This study conducted three spatial analyses: one to identify roe deer occurrence in the Netherlands and two others to find potential vehicle-animal collision hotspots. The research utilized various data sources, including wildlife collision data, animal occurrence records, land-use information, road network characteristics, and mitigation works data, which were processed and analysed using R and MaxEnt software to create predictive models for both roe deer habitat suitability and collision risk areas.

The roe deer occurrence model showed bush cover, forest cover as well as the distance to the nearest forest, grass cover and dune cover to be of predicting value for habitat suitability. Established natural areas were shown to be highly suitable. The two wildlife-vehicle collision models showed: distance to the nearest forest, bush cover, roedeer occurrence, urban cover, road width, and maximum speed limit to be of predictive value. The east and south of the Netherlands showed the highest collision proneness and therefore the highest risk.

This research provides a better understanding of wildlife-vehicle collisions in the Netherlands and their predictive factors, identifying high-risk areas. The results can aid future mitigation strategies, potentially guiding more targeted and effective implementation of collision preventive measures within the Netherlands.

Introduction

Since the Industrial Revolution during the 18th century, the world has seen an increase in its human population (Milner & Boldsen, 2022; Zheng, 2020). The United Nations predicts that the world population will reach approximately 10.2 billion people at the end of the century (United Nations, 2024). The land demand of this growing population saw serious increases, seeing that all of these people need land to live on and land to cultivate food. This increase in human land use came at the cost of a decrease in nature, causing fragmentation in many natural habitats.

Fragmentation is defined as the process where continuous habitats are divided into smaller habitats or isolated patches, mainly due to human activities. Fragmentation is not only caused by the removal of designated natural habitats in favour of (mainly) agricultural land but also by the (newly) constructed road networks splitting the natural habitats into small fragments. This habitat fragmentation has huge consequences for these ecosystems, not only reducing their overall area but also leading to animals having a harder time moving between these fragmented habitats. Habitat fragmentation could potentially lead to serious effects, for example, the restriction in movement could lead to isolated populations which in turn could cause a decrease in genetic diversity, negatively affecting the survival of this population (Gibbs, 2001). Additionally, it could hinder animals more directly, finding enough food, water and other crucial resources becomes more difficult if there is a road that splits the habitat into multiple fragments (Haddad et al., 2015).

The restriction of movement by these road networks is caused by factors such as viewing the road as a foreign space which animals are less likely to cross (Coffin, 2007) lowering the risk of wildlife-vehicle collisions. However, when animals do decide to cross these roads, wildlife-vehicle collisions pose a real threat to these animals, leading to a significant mortality rate of the local wildlife (Rytwinski & Fahrig, 2015). Additionally, these accidents are not only a threat to the animals crossing the road but to the human motorists as well, resulting in at least substantial property damage but in more serious cases injuries and even fatalities (Huijser et al., 2009). To tackle this challenge of wildlife-vehicle collisions, policymakers and researchers have been implementing different strategies to reduce these wildlife-vehicle collisions. For instance, these strategies influence the behaviour of the wildlife, deterring the animals from crossing the road, or even redirecting them to a safer point to cross (Pagany, 2020). The latter can be done through fencing and under- or overpasses to make crossing the road safer for these animals (Huijser et al., 2009). Additionally, modifying driving behaviour by lowering their driving speed, by means of lowering the maximum speed limit, or installing signs reminding road users of the potential of crossing animals are examples of ways to influence the behaviour of motorists, leading to a reduction of wildlife-vehicle collisions (Huijser et al., 2009).

Within this context, the Netherlands forms an interesting challenge. The Netherlands is one of the countries within Europe with a high population density. In the Netherlands, approximately 18 million people inhabit only 41.543 square kilometres (WorldData.info, 2024). The country's high population density and limited land area make for a challenging task balancing their urban and agricultural expansion with preserving nature and its ecosystems. This expansion of the Dutch population and the populated area went hand in hand with an expansion of its road network (*Figure 1*) (De Vries, 2015).

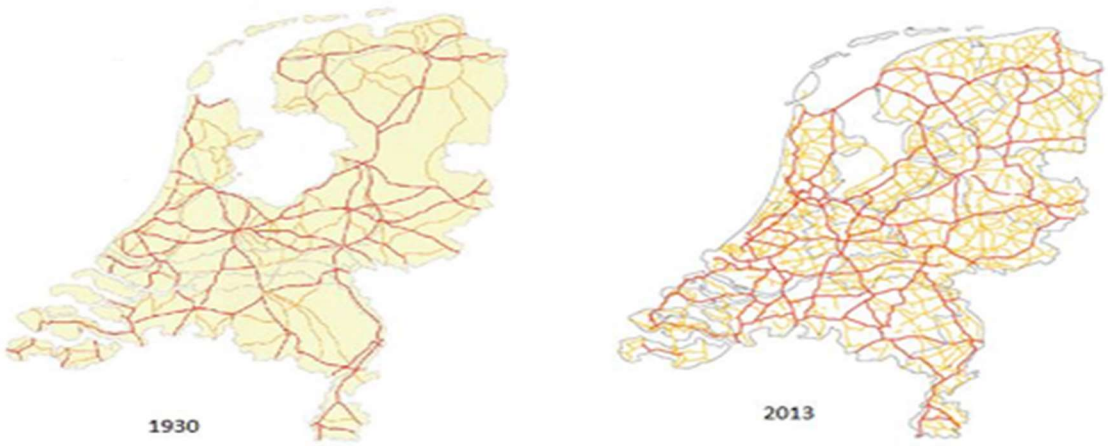

Figure 1. A depiction of the expansion of the Dutch road network during the last century. On the left is the situation in 1930 and on the right is the situation in 2013. (De Vries, 2015)

While aiming to prevent or at least limit wildlife-vehicle collisions as much as possible, the Netherlands has been and is constructing multiple measurements intending to make road crossings safer for both motorists as well as animals crossing the road. Nevertheless, a lot of factors are involved when it comes to determining which areas are most at risk of wildlife-vehicle collisions.

As this research will consist of a modelling study each of the different factors that could be involved in wildlife-vehicle collisions, which will be considered during this research, will be discussed below. **Land use** is the first of the potential predictive factors, where both anthropogenic and natural land use patterns play a role in predicting the occurrences of wildlife-vehicle collisions. The amount of urban landcover is shown to negatively affect the number of collisions (Madsen et al., 2001; Mayer et al., 2021). Regarding the natural lands that are expected to play a role in the number of collisions vegetation, forest bushes grasses, are shown to affect the number of collisions. The amount of forest cover was shown by Madsen to affect the number of collisions whereas Mayer suggests reducing the amount of vegetation near roads to reduce collision risk. Furthermore, it is known that, for example, roe deer (*Capreolus capreolus*) prefer the edge areas between agricultural land or grassland and forests, meaning that roads near or within these areas would likely see an increase in roe deer collisions (De Vries, 2015).

Linked to the land use mentioned above, but seen as a separate factor are **road type and traffic density**. The Netherlands has a diverse road network where highways are alternated with provincial roads. Where the national highways often are harder to access and cross due to the roads being raised or large noise barriers being placed along these roads. Provincial roads, however, lack these features and therefore are more likely to be crossed by animals. Additionally, there are links between the width of a road and its collision probability, in which roads from six meters and above see an increased chance of collision (De Molenaar & Henkens, 1998).

This increase is likely due to the fact that the animal has to spend a longer time on the road before it fully crosses it (De Molenaar & Henkens, 1998). In addition, traffic density seems to sometimes affect the number of wildlife-vehicle collisions. As the study of Tajchman showed, traffic density can lead to a decrease in the number of wildlife-vehicle collisions due to the high traffic density having an avoidance effect. However, traffic densities between 501-600 and 1101-1200 vehicles/day, saw an increase in collisions, something that could be linked to the higher driving speed used on these roads (Tajchman et al., 2020).

Animal population density will presumably also have an effect. Whether the density of animals will have a positive or negative effect will depend on a multitude of factors, for instance, the species involved or the type of road. Smaller animals may see a population increase due to the presence of a road, seeing that these smaller animals are prey for larger animals which are negatively affected by these same roads (Fahrig & Rytwinski, 2009). However, the overall trend shows a negative relationship between the number of roads and animal abundance. This means that higher population densities near roads are less common. Thus, to what extent wildlife-vehicle collisions increase in areas with higher animal density, is uncertain.

Lastly, the impact of **mitigation measures** will be considered. The Netherlands has already invested a lot in the mitigation of wildlife-vehicle collisions and thus the safekeeping of both animals as well as motorists. Around 70 million euros (10% of the Road Project Budget) was used to prevent wildlife-vehicle collisions (van der Grift et al., 2013). Properly implementing these mitigation measures could see a reduction in wildlife-vehicle collisions (Huijser et al., 2009). For example, when a properly constructed wildlife overpass is combined with fences it may reduce collisions by up to 87% for mammals (Huijser et al., 2009). However, the effectiveness of these mitigation measures varies depending on factors like landscape, intended species, and traffic patterns (Rytwinski et al., 2016).

Within this research, I aim to determine what the influence of these predictive factors are within the Netherlands. Furthermore, this research will aim to find where these predictive factors apply within the Netherlands. Therefore, the main research question of this study seeks to answer is:

What are the predicting factors for the number of wildlife-vehicle collisions, and where do these factors lead to a high-risk area, within the Netherlands?

Conceptual Model

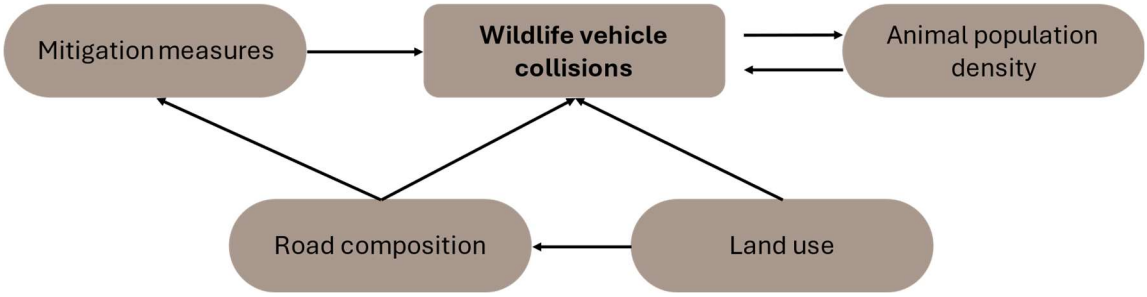

Figure 2. A conceptual model of the expected mechanisms contributing to wildlife-vehicle collisions. The model illustrates the interplay between animal factors, as the population density, mitigation measures as wildlife crossings, road composition as road width and maximum speed and landscape features as the different land-use cover classes. Each arrow indicates a relationship between the components of the full model.

To approach this, an analysis will be conducted, using the aforementioned factors in a geographic information system (GIS). Maps will be created that highlight areas where the risk of wildlife-vehicle collisions is highest based on the aforementioned predictive factors, allowing for a more effective allocation of future mitigation matters. This will be done by creating a model with the use of MaxEnt, modelling the effect of the expected predictive factors on wildlife collisions. Roe deer will be used as a model species during this research, as roe deer are the species involved in most (recorded) wildlife-vehicle collisions and are the only species that occur and have the ability to roam freely across the Netherlands.

Based on prior research I expect each of the aforementioned predictive factors will indeed affect the number of wildlife-vehicle collisions. I will go by each of the factors briefly explaining what I expect they will contribute to the number of wildlife-vehicle collisions.

Land use: as mentioned by earlier research, I expect the amount of collisions to be negatively associated with the amount of urban landcover (Grilo et al., 2009). I presume that this will be similar in the Netherlands. Furthermore, forest, bush and grass are expected to correlate with the number of collisions as suggested by earlier research (Madsen et al., 2001; Mayer et al., 2021).

Road composition: as suggested by the Conceptual model (*Figure 2*) the traffic density is likely to be affected by the road composition which itself is linked to land use. This means that it is expected that more and larger roads will be seen in more urbanized areas, meaning the traffic density will also increase. As Tajchman showed, traffic density could have both positive and negative correlations with the number of collisions (Tajchman et al., 2020). This means that predicting the effect of road type and traffic density will be difficult, however, I still expect to see the highest level of collisions on medium-sized roads. These roads will probably be too small to have a large deterring effect, but enough people use the road to make the chances of collisions a real threat. Likewise, **mitigation structures** most likely play a role in the number of wildlife-vehicle collisions. Structures such as fences along the roads which prevent animals from reaching the road would result in lower wildlife-vehicle collisions. Furthermore, I expect wildlife crossings to locally reduce the number of wildlife-vehicle collisions (Sugiarto, 2022).

Animal population density shows a mutual dependency, where the number of animals will influence the number of collisions. This means I expect to see more collisions in an area with higher population density. Furthermore, regarding the distribution of roe deer, I expect to see the highest population density to be around the forest edge, which is their natural habitat as proven by prior research (Zorzi et al., 2022).

Methods

Study Design

This study consisted of two large spatial analyses. The first analysis aimed to identify the occurrence of roe deer within the Netherlands, while the second analysis was designed to find potential vehicle-animal collision hotspots. The created roe deer occurrence map was used as one of the possible predictive factors within the collision model.

Data Collection and Processing

The data was imported into the newest version of R (R-4.4.1 for Windows) at the time of analysis and converted into spatial data using the Terra (version 1.7.78) and SF (version 1.0.18) packages within R itself. All the used scripts are provided in the appendix of this paper (*Appendix 1*).

Wildlife Collision Data

Faunabeheereenheid (FBE) (Fauna Control Unit in English), a Dutch institution responsible for the collection of wildlife-vehicle collisions throughout the Netherlands, provided all the data regarding wildlife-vehicle collisions. Each of the twelve provinces is responsible for the collection of data within their respective province, therefore FBE is split into 12 different organizations, one per province, which resulted in some small differences in the data format. However, all datasets used included at least the coordinates of the collision and the specific species that were involved in the collision. All received data was recorded between 2017 and 2024. However, the two provinces of Flevoland and Friesland never replied to email contact and therefore didn't provide any information regarding the collision points within their province. The location of these collision points was corrected to ensure that all collisions were on road cells only. This was done by calculating the distance of a collision point to the nearest road cell (based on the LGN, see section Land-use); each point with a distance of less than 25 meters was relocated to this nearest road cell. Collision points with a distance greater than 25 meters to the nearest road were excluded as data points, in order to prevent placing collisions on the wrong road or road section due to inaccurate or wrong coordinates.

Animal Occurrence

The data regarding animal occurrence was collected through Waarneeming.nl (*Welkom - Waarneming.Nl*, n.d.), a Dutch citizen initiative site, where citizens can register their local wildlife sightings, creating an estimated overview of the local wildlife populations. The data collected by Waarneemingen.nl was then downloaded via Gbif.org (*GBIF*, n.d.). Only the data of roe deer was used in the research. These occurrence datasets were downloaded via the website of Gbif.org and filtered to include the coordinates, which must be located in the Netherlands, and recorded between the research period of 2017

up to 2024. The resulting habitat suitability map that will be made via MaxEnt will be used as a proxy for roe deer density within the study area in the collision model.

Land-use Data

The land-use data was received from the Landelijk Grondgebruik Nederland (LGN)(Land-use the Netherlands) website (*LGN*, n.d.). For this research, the LGN2023 map was used. This data was loaded in R using R studio.

For each of the land-use categories (*Table 1*), a binomial map was created. This means a map with either a 0 or 1 was made, 1 if the cell was classified as the land-use in question and 0 if not. After this, an aggregation was performed to ensure that MaxEnt would be able to process all maps made. The maps were aggregated from a 5 by 5-meter grid structure to a 100 by 100-meter grid structure. In which 100 meters was chosen because of the inaccuracy of the collision location. The resulting maps now consist of a percentage of land-use class present in the cell instead of either 0 for absence or 1 for presence. These aggregated binomial maps were also used to calculate the distance, using the distance function within R, from the occurrence or collision point to the relevant land-use class (Coenen, 2022).

For the predictive factors regarding the collisions, a focal weight (see *Appendix 1*) was created to determine the percentage of occurrence of a land-use within the surrounding area (a circle of 300-meter) of each cell using the aggregated maps. This was done to ensure that the point of collision (a road cell) also had all the information about its surroundings. The distance maps earlier created were also used in the collision analysis.

Furthermore, information regarding the road network was included as a potential predictive factor for the risk of wildlife-vehicle collisions. These included two datasets which were also transformed into a raster using the same extent and coordinate reference system (CRS) as the other datasets i.e. 100x100 meter. The two datasets were the maximum speed limit of all roads within the Netherlands as well as the width of these roads. Both datasets were downloaded from Rijkswaterstaat-dataregister (*Dataregister Rijkswaterstaat*, n.d.).

Number in LGN database	Class	English translation of the class	combined class name in this analysis	Distance map created
1	Agrarisch grass	Pasture	Pasture	X
2	Maïs	Maize	Agroculture	
3	Aardappelen	Potato		
4	Bieten	Beetroot		
5	Granen	Grain		
6	Overige landbouw	Other agriculture		
8	Glastuinbouw	Greenhouse horticulture		
9	Boomgaarden	Orchards		
10	Bloembollen	Flower Bulbs		
11	Loofbos	Deciduous forest	Forest	X
12	Naaldbos	Coniferous forest		
16	Zoet water	Fresh water	Fresh water	X
17	Zout water	Salt water	Salt water	
18	Bebouwing in primair bebouwd gebied	Buildings in primary built-up area	Urban area	
19	Bebouwing in secondair bebouwd gebied	Buildings in secondary built-up area		
20	Bos in bebouwd gebied	Forest in built-up areas	Forest in urban area	
22	Bos in secundair bebouwd gebied	Forest in secondary built-up area		
23	Gras in primair bebouwd gebied	Grass in primary built-up area	Grass in urban area	
26	Bebouwing in buitengebied	Buildings in rural areas	Buildings in rural areas	
27	Overig grondgebruik in buitengebied	Other land use in rural areas	Other	
28	Gras in secundair bebouwd gebied	Grass in secondary built-up areas	Grass in urban area	
29	Zonneparken	Solar parks	Solarpark	
30	Kwelders	Salt marshes	Salt marshes	
31	Open zand in kust gebied	Open sand in coastal area	Open sand in coastal area	
32	Duienen met lage vegetatie	Dunes with low vegetation	Dunes	
33	Duinen met hoge vegetatie	Dunes with tall vegetation		
34	Duinheide	Dune heath		
35	Open stuifzand en/of rieverzand	Open drifting sand and/or river sand	Open drifting sand and/or river sand	
36	Heide	Heather	Heath	
37	Matig vergraste heide	Moderately grassed heath		
38	Sterk vergraste heide	Heavily grassed heath		
39	Hoogveen	Raised bog	Raised Bog	
40	Bos in Hoogveengebied	Forest in the raised moor area	Forest	X
41	Overige moeras vergetatie	Other swamp vegetation	Swamp	
42	Rietvegetatie	Reed vegetation	Reed	
43	Bos in moeras gebeid	Forest in swamp area	Forest	X
45	Natuurlijk beheerde agrarische grasslanden	Naturally managed agricultural grasslands	NMGA	
46	Gras in kustgebied	Grass in coastal area	Grass in Coastal area	
47	Overige gras	Other grass	Other Grass	
61	Boomkwekerijen	Tree nurseries	Nurseries	
62	Fruitkwekerijen	Fruit farms		
251	Hoofdinfrastructuur en spoorbaan lichaam	Main infrastructure and railway body	Roads	X
252	Halfverharde wegen, infrastructuur langzaam verkeer en overige infrastructuur	Semi-paved roads, slow traffic infrastructure and other infrastructure		
253	Smalle wegen	Narrow roads		
321	Struikvegetatie in hoogveengebeid laag	Shrub vegetation in raised bog areas low	Bushes	
322	Struikvegetatie in moerasgebied laag	Shrub vegetation in swamp area low		
323	Overige struikvegetatie laag	Other shrub vegetation low		
331	Struikvegetatie in hoogveengebeid hoog	Shrub vegetation in high moorland		
332	Struikvegetatie in moerasgebied hoog	Shrub vegetation in swamp area high		
333	Overige struikvegetatie hoog	Other shrub vegetation high		

Table 1. Land-use classes, their aggregated clusters, and associated distance map creation. Land-use classes are based on the LGN2023 map. The 'Combined class name in this analysis' column shows how individualclasses were combined into broader categories for analysis. The 'Distance Map Created' column indicates whether a distance raster was created for the respective land-use class (with X indicating a created distance map).

Mitigation Works Data

For the mitigation works, a list of the wildlife crossings within the Netherlands was compiled using all listed wildlife crossings on Wikipedia, each of which had its coordinates recorded through Google Earth. After translating the data to a raster, the distance to each wildlife crossing was calculated to be used as a potential predictive factor within the collision, to see whether these structures are predictive of the amount of wildlife-vehicle collisions that are recorded.

MaxEnt Analysis

All of the analyses were conducted with the software R and combined with the MaxEnt program (version 3.4.4). The spatial grid of the LGN (CRS = RDNew) was used to get the most accurate representation of the real-life situation. As mentioned, after all data was processed the resolution was reduced 400 times to reduce the size of all maps i.e. from 5x5 to 100x100 meters. All of the datasets were harmonised and aligned to this newly created spatial grid. Meaning that both the CRS, the extent and the resolution are identical. This was done to ensure a spatial analysis was possible using R to compare each grid cell for all its presumed predictive factors.

To improve the model, a mask was created within R, in which areas that could be considered unsuitable were removed from all different environmental layers. The mask included all areas outside of the borders of the Netherlands, seeing that all occurrences and collisions were recorded in the Netherlands. Furthermore, the city centres were removed since these areas can be presumed to be unsuitable for both roe deer occurrence and therefore also roe deer-vehicle collisions. This too goes for both freshwater as well as saltwater surfaces, hence these areas are also unsuitable for both occurrence and collisions (see *Appendix 1*).

All of the aforementioned different factors that were suspected to be predictive of either the roe deer occurrence or the number of wildlife-vehicle collisions of these roe deer were tested using a jackknife test. The jackknife test evaluated the importance of each of the predictive factors. These evaluations showed how much each variable contributed to the model's predictive power. After this, a collinearity test was performed to check for collinearity among the environmental variables. This is done in order to create the highest possible AUC (area under the curve (AUC) value. The higher this value is, the better the model predicts the real-life situation. Likewise, different models were created to find the most predictive one. The performance of these models was checked with both the AUC as well as a binomial test. Within this binomial test, firstly a binomial map was created, where values above the MaxEnt calculated threshold(a value that is different for each model), are given a value of 1 and those below the threshold are given a value of 0. After this, the original occurrences, (either roe deer sightings or collisions) 30% were used as testing data for the binomial test, these were compared to see whether they were within the areas above the threshold or in the areas below. Which resulted in a percentage of correctly predicted occurrences. This percentage makes it possible to check whether both false and correctly predicted occurrences are equally likely to occur. Whether or not a point is correctly predicted is based on the MaxEnt calculated “Maximum training sensitivity plus specificity”. The maximum training sensitivity plus specificity is based on the threshold that maximises both the sensitivity and the specificity of the model. Sensitivity is the proportion of correctly predicted

presences of all the presences, while specificity is the proportion of correctly predicted absences among all the absences.

Habitat Suitability Modelling

With all of the factors that were shown to be of predictive value, a habitat prediction model was created for the occurrence map of roe deer across the Netherlands as well as for the wildlife-vehicle collision model, in which the occurrence map was used as a predictive factor. The MaxEnt model produced a raster that indicates the suitability of each cell as a potential roe deer habitat. Since my focus is on the number of suitable areas near roads where collisions were recorded, a focal analysis was created, based on the MaxEnt occurrence output. This focal analysis was created in the same way as the focal analysis for the land use classes, only now using the Maxent occurrence output instead of the aggregated binomial land-use maps.

As mentioned, MaxEnt analysis is normally used to predict suitable habitats based on. However, in this specific case, instead of identifying a habitat suitable for roe deer, roads and their surrounding area that are prone to (suitable according to MaxEnt) for wildlife-vehicle collisions with roe deer were identified.

Results

The results will be described in two parts. First, the findings of the roe deer occurrence model will be examined. This will be followed by the results of the wildlife-vehicle collision model.

Roe Deer Occurrence

Final Model

The MaxEnt analysis of roe deer occurrence resulted in a distribution map, which is depicted in *Figure 3*. The analysis is based on 1,408 sighing points retrieved from Gbif.org (*GBIF*, n.d.). This map provides insights into the habitat suitability for roe deer within the Netherlands. As mentioned habitat suitability will be used as a proxy for roe deer density within the study area. It can be seen that the spots within the map that score the highest are areas that are established nature areas such as the Veluwe, De Utrechtse Heuvelrug and Oostvaardersplassen, also the dunes along the Dutch coast are shown as suitable. Furthermore, areas around the larger cities show a low suitability for roe deer occurrence.

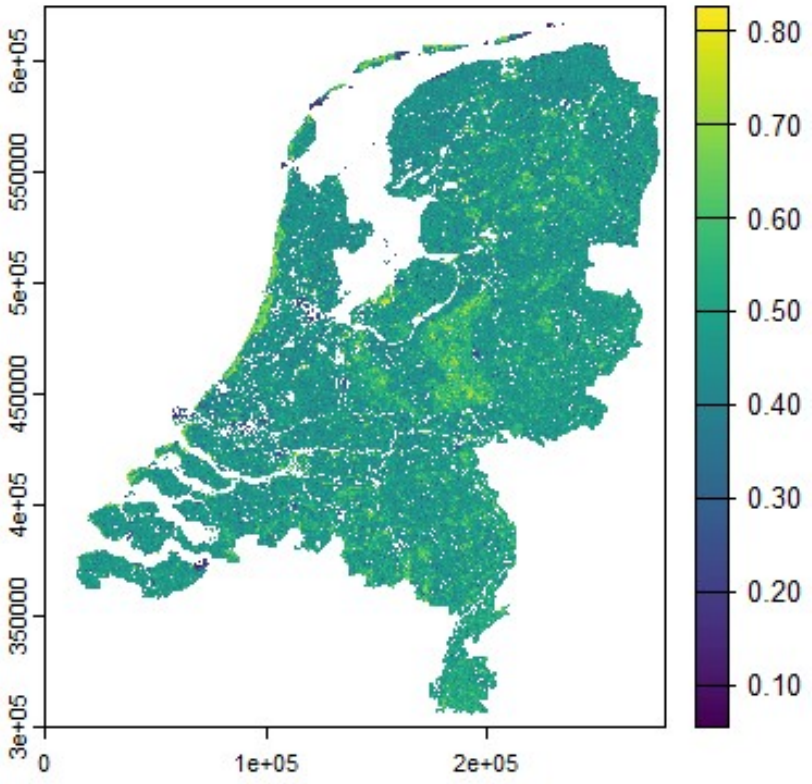

*Figure 3. A habitat suitability map for roe deer (*Capreolus capreolus*) generated using MaxEnt. In which 0 is not suitable and 1 is most suitable*, based on 1,408 sighting points. With the coordinates on the X and Y axis.

Variable	Percent contribution	Permutation importance
bushes	18.3	1.7
dunes	12.6	2.2
heath	8.9	2.3
forest	8.4	3.5
forest_distance_masked	8.3	5.2
other_grass	6.4	1.2
other	6.1	8.3
freshwater_distance_masked	4.4	9.2
swamp	3.8	2.3
fresh_water	3.1	5.5
pasture	3	7.7
roads_distance_masked	2.4	3.3
reed	2.4	1.5
pasture_distance_masked	2.3	10.2
agroculture	1.8	21.2
salt_water	1.6	1.2
solar	1.2	0.2
grass_in_urban	1.1	7.3
building_in_rural	0.8	1.2
open_sand_in_coast	0.8	0.5
nurseries	0.7	1.3
roads	0.7	0.2
raised_bog	0.4	0.5
NMAG	0.3	0.8
urban	0.2	0.2
salt_marsh	0.1	1.3
grass_in_coast	0.1	0.1
forest_in_urban	0	0.1
drifting_river_sand	0	0

Table 2. List of the environmental variables and their contribution to the model, used to create the roe deer occurrence model.

Predictive Factors

The following table shows the contribution of all the variables used to create the roe deer occurrence model and map (*Table 2*). The top predictive factors (>5%) were Bushes (18.3%), Dunes (12.6%), Forest (8.4%), Forest distance (8.3%), Other grass (6.4%), and Other (6.1%) (*Figure 4*).

Roe deer occurrence is positively associated with bushes(*Figure 4a*), dunes(*Figure 4b*) forest (*Figure 4d*) and other grass (*Figure 4e*). Forest distance(*Figure 4c*), shows that with an increasing distance to the nearest forest, the roe deer occurrence decreases. A negative correlation is also seen in the Other landuse class(*Figure 4f*), in which an increasing cover percentage decreases the occurrence suitability.

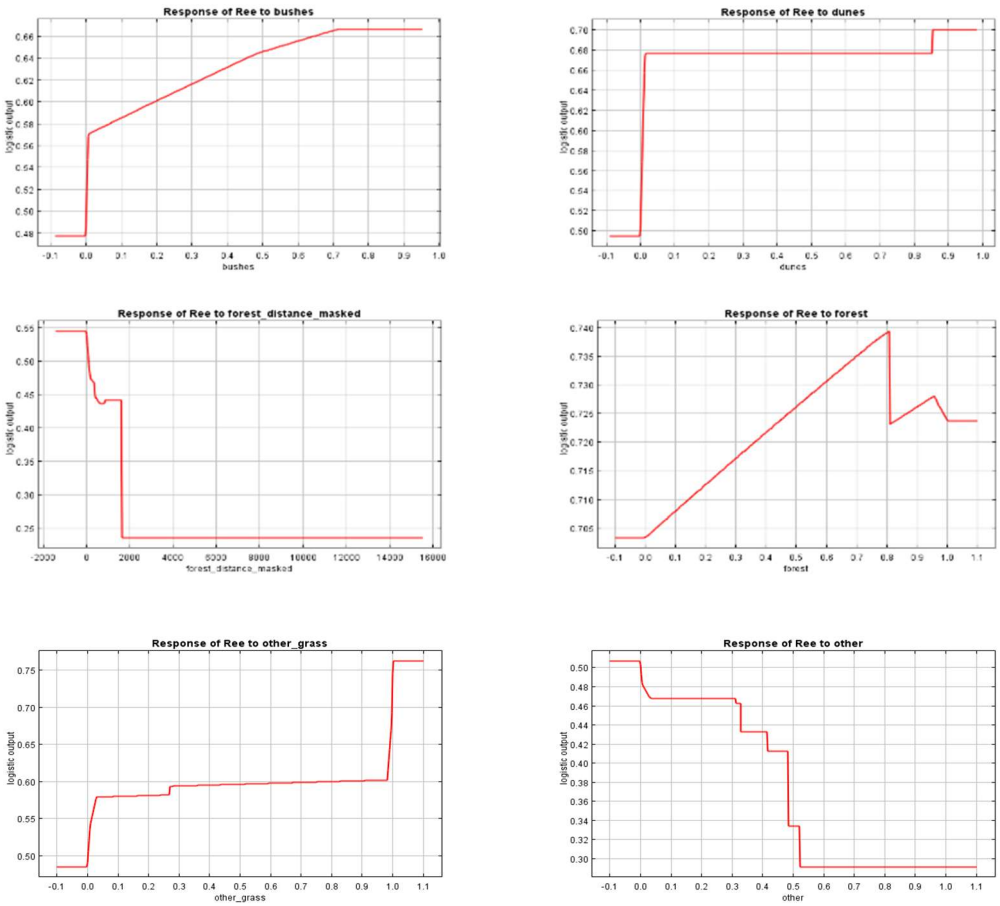

*Figure 4. The trajectory of each of the top predicters. On the Y axes the effect on roe deer (*Capreolus capreolus*) suitability and on the X axes the number of the environmental variable (percentage landcover for bushes, dunes, forest, other-grass and other) (distance in meters for distance to forest).*

A: Response of roe deer to Bushes. B: Response of roe deer to Dunes.
C: Response of roe deer to forest distance. D: Response of roe deer to forest.
E: Response of roe deer to other grass. F: Response of roe deer to other.

When looking at the jackknife results, there seems to be one more predictive factor that holds information: Pasture distance (*Figure 5*). Looking at the trajectory of the distance to pasture, it can be seen that at a little over 500 meters is the peak for roe deer occurrence (*Figure 6*).

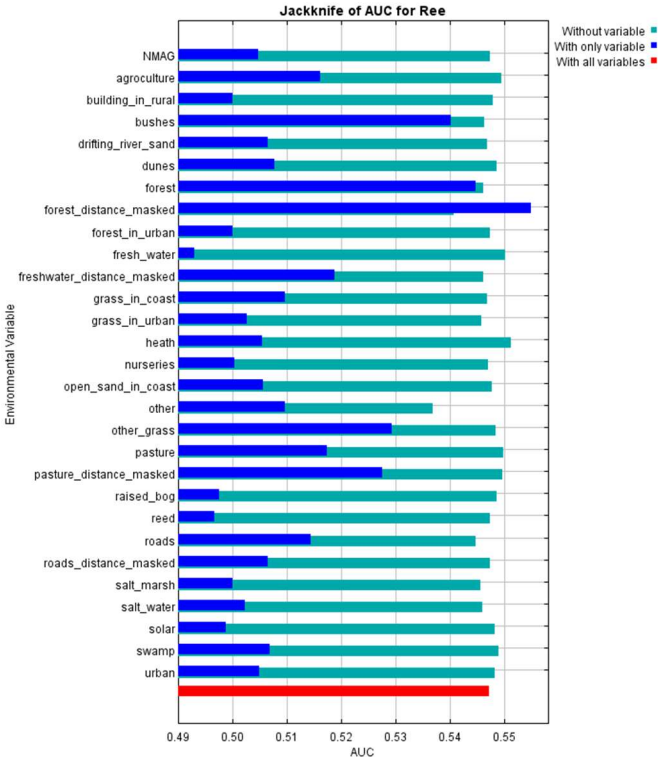

Figure 5. Jackknife test results for environmental variable contributions in the MaxEnt model. The bars represent model gain with each variable used in isolation (dark blue), and with that variable excluded (light blue). Longer dark blue bars indicate variables with high predictive power on their own. therefore a higher dark blue bar suggests that the variable contains information not present in other variables.

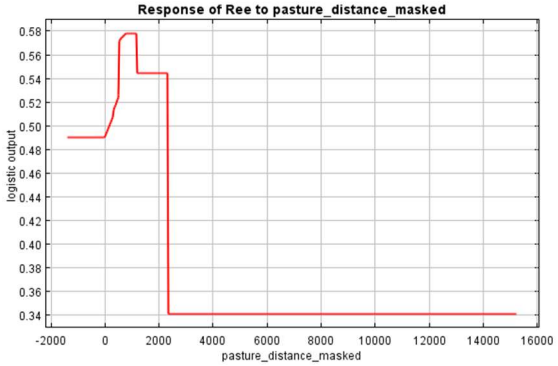

*Figure 6. Trajectory of the Pasture variable, with on the Y axis the roe deer (*Capreolus capreolus*) habitat suitability and the percentage of pasture cover on the X axis.*

Model Performance

The model's AUC value was 0.610 (*Figure 7*) indicating that the model is slightly better than a random prediction (AUC=0.5).

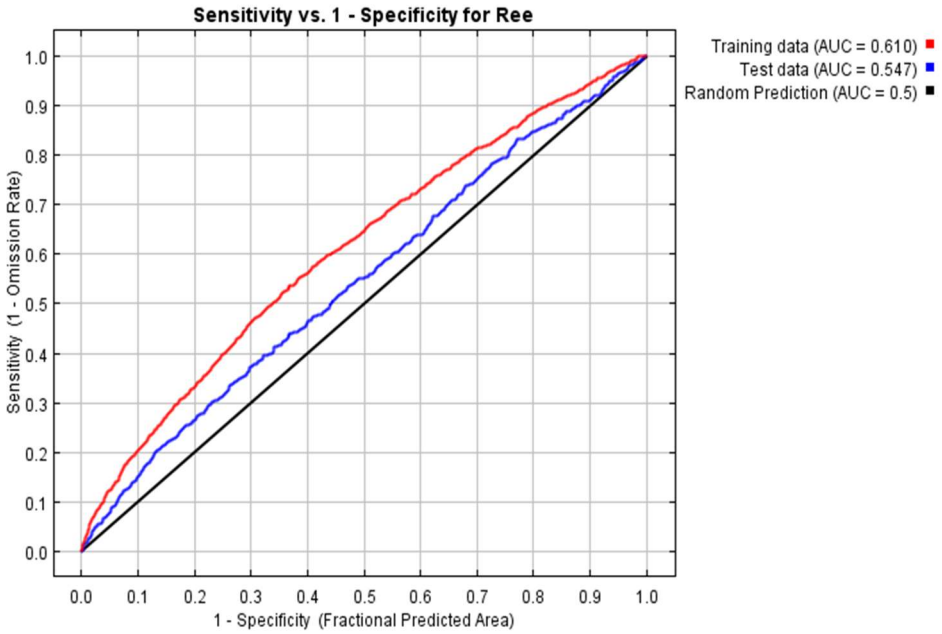

*Figure 7. The AUC of the roe deer (*Capreolus capreolus*) suitability model. The Training data AUC = 0.610.*

The occurrence model used a total of 1,408 points out of the total 2,375 occurrences gathered from the Gbif-database. The binomial test resulted in 43.9% correctly predicted occurrences (265 out of 603 trials), which is significantly more (p= 0.00023) than expected based on the average occurrence probability (36.6%). This is also proven by the 95% confidence interval [0.399, 0480] as the fraction of correctly predicted occurrences is set a 0.439. The Maxent calculated value at which both sensitivity and specificity are maximized is set at 0.485 for this model, meaning all cells with a value equal or higher were scored with a 1 and therefore suitable, and all cells with a lower value are scored unsuitable with a score of 0.

Wildlife-Vehicle collisions

The model depicting the occurrence of roe deer across the Netherlands can be used for the predictive factors for the wildlife-vehicle collision model. Using this occurrence map together with the focal maps of the land uses and road composition can be seen in *Figure 8*. This model used a total of 1,433 total collisions. This map shows how prone each road cell is to a collision, where a value of 0 means not at all prone and 1 means most prone. The map shows higher collision suitability in the east and south of the Netherlands, with areas further away from the larger cities with the highest suitability.

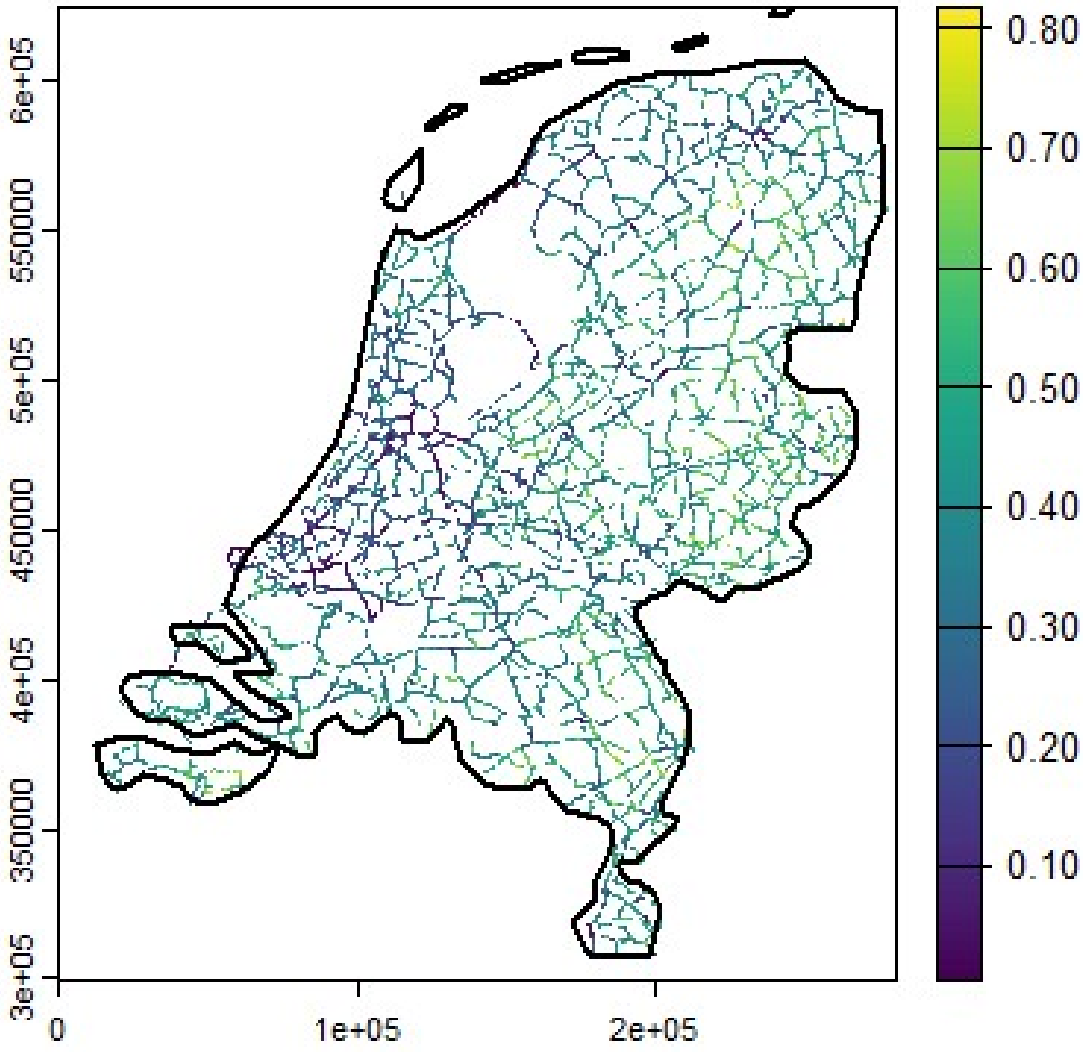

*Figure 8. A wildlife-vehicle collision suitability map for roe deer (*Capreolus capreolus*) generated using MaxEnt. In which 0 is not suitable and 1 is most suitable*, based on 1,433 collision points. With the coordinates on the X and Y axis.

Predictive Factors

When looking at *Table 3,* a table depicting the list of environmental variables and their contribution to the model, we see that: Forest (15.5%), Freshwater (14.8%), Forest distance (13.7%), Roe deer occurrence (9.4%), Grass in urban (9.2%), Bushes (8.8%), Road width (6.5%) and Other grass (6%) are the top predictors (>5%).

Forest (*Figure 9a.*), freshwater (*Figure 9b.*), forest distance (*Figure 9c.*) and urban grass (*Figure 9e.*) all showed a negative correlation with collision proneness. It needs to be noted that the freshwater trajectory has a large peak before going into a downward trajectory. The roe deer occurrence, bushes, road width, and other grass all show a hump-shaped trajectory (*Figure 9 d.f.g.h.*). Note that in contrast to the predictive factors in the occurrence model, the predictive factors are depicted as a focal, meaning it is the percentage of land-use cover within a 300 meter circle around each road point.

Variable	Percent contribution	Permutation importance
lgn_forest_focal	15.5	12.9
lgn_fresh_water_focal	14.8	16.6
lgn_forest_distance	13.7	8.5
Roedeer_occurrence_focal	9.4	2.1
lgn_grass_in_urban_focal	9.2	14.6
lgn_bushes_focal	8.8	1.5
dutch_road_width	6.5	6
lgn_other_grass_focal	6	4.5
lgn_agroculture_focal	2.3	4.5
dutch_road_speed_reclassified	2.2	1.3
lgn_other_focal	1.8	4.2
distance_to_ecoduct	1.7	3
lgn_pasture_focal	1.4	6.1
lgn_fresh_water_distance	1.2	0.5
lgn_raised_bog_focal	0.8	0.8
lgn_building_in_rural_focal	0.6	0.7
lgn_forest_in_urban_focal	0.5	2.8
lgn_heath_focal	0.5	0.6
lgn_pasture_distance	0.5	1.7
lgn_open_sand_in_coast_focal	0.5	1
lgn_reed_focal	0.5	1.3
lgn_nurseries_focal	0.4	0.9
lgn_urban_focal	0.4	0.1
lgn_swamp_focal	0.3	0.9
lgn_salt_marsh_focal	0.2	0.4
lgn_NMAG_focal	0.2	1
lgn_solar_focal	0.1	0.4
lgn_salt_water_focal	0.1	0
lgn_roads_focal	0.1	0.8
lgn_dunes_focal	0	0.2
lgn_drifting_river_sand_focal	0	0.1

Table 3. List of the environmental variables and their contribution to the model.

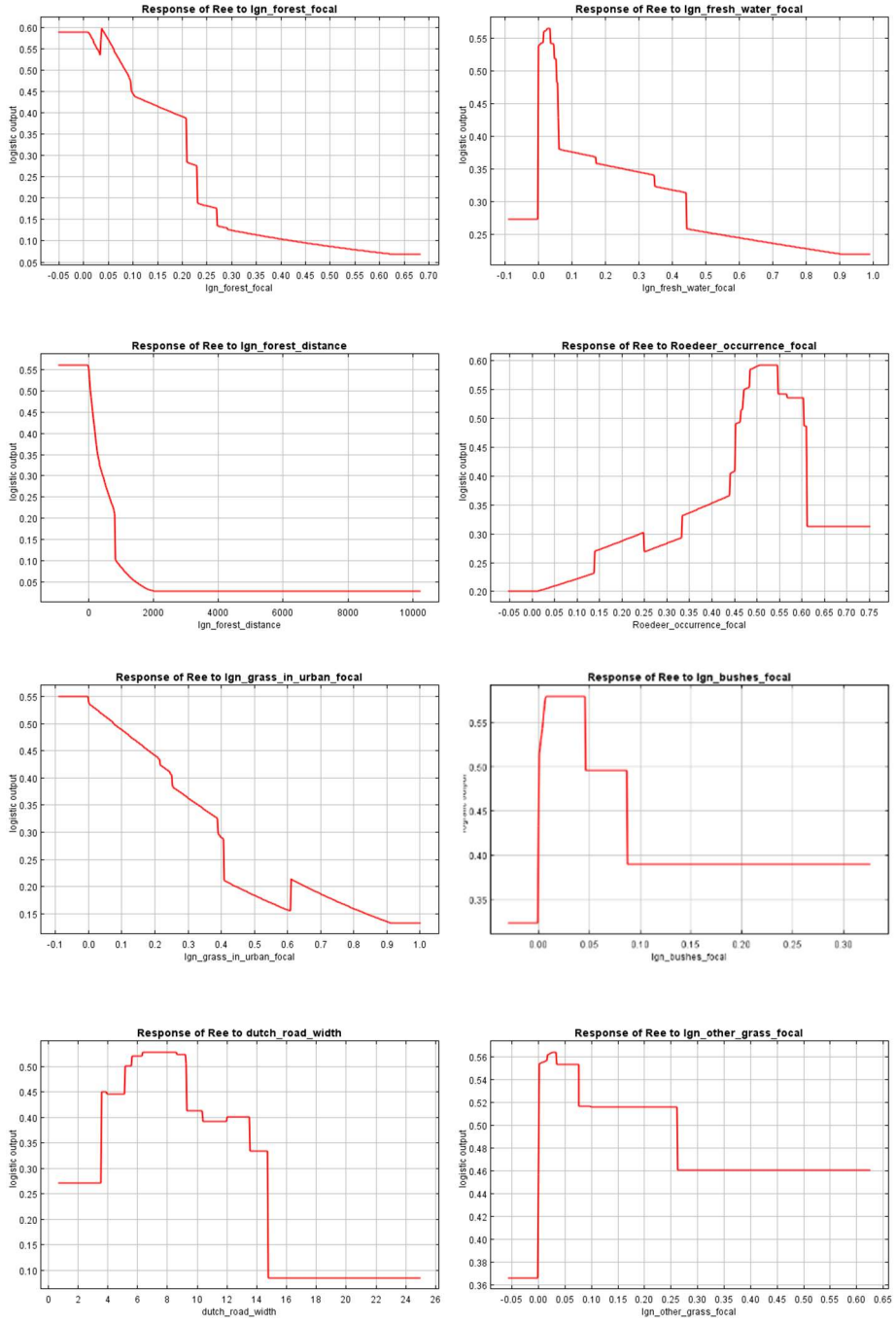

*Figure 9. The trajectory of each of the top predictors. On the Y axes the effect on roe deer (*Capreolus capreolus*) vehicle collision suitability and on the X axes the number of the environmental variable focal of landcover for forest freshwater roe deer occurrence, grass in urban, bushes and other grass. distance in meters for distance to forest, and meters for road width).*
A: Response of roe deer collision risk to forest. B: Response of roe deer collision risk to water. C: Response of roe deer collision risk to forest distance. D: Response of roe deer collision risk to occurrence.
E: Response of roe deer collision risk to grass in urban. F: Response of roe deer collision risk to bushes.
G: Response of roe deer collision risk to road width. H: Response of roe deer collision risk to other grass.

If we look at the jackknife of the model it can be seen that a few factors did not have a high contribution percentage but do seem to contribute something to the model (*Figure 10*). Both the urban as well as the roads focal seem to contribute something to the model. This could mean that although not important for the collisions overall, it has a predictive value in particular areas, likely in the area where the factor has a high percentage of cover.

Both of these predictive factors show a negative correlation, meaning higher percentages of either urban or road ground cover results in lower collision suitability (*Figure 11*). Something that is likewise shown by the low suitability values around the larger cities in the Netherlands.

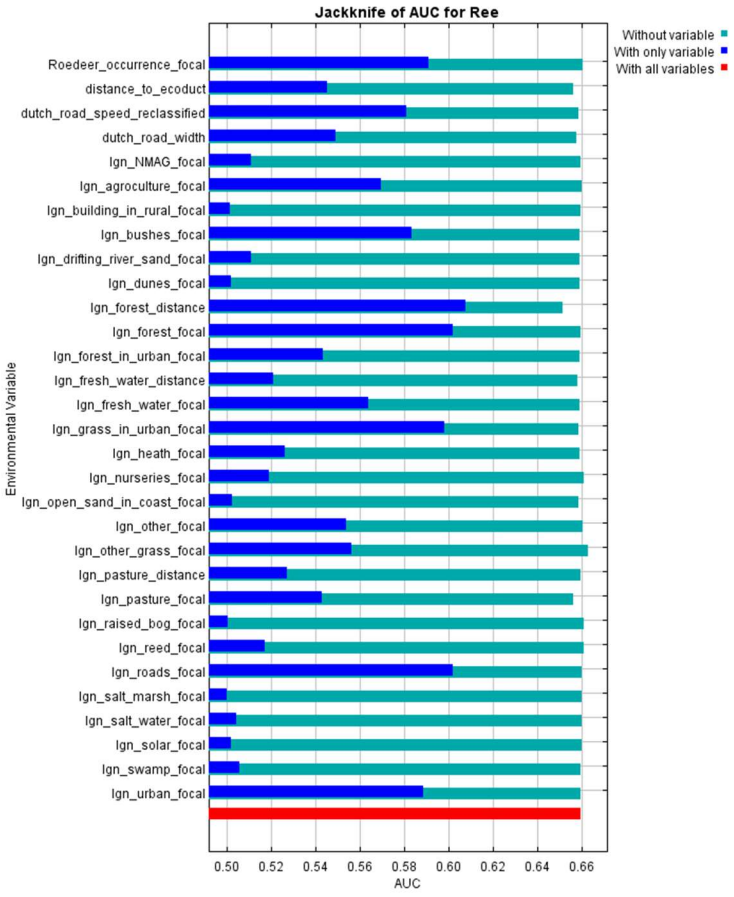

Figure 10. Jackknife test results for environmental variable contributions in the MaxEnt model. The bars represent model gain with each variable used in isolation (dark blue), and with that variable excluded (light blue). Longer dark blue bars indicate variables with high predictive power on their own. therefore a higher dark blue bar suggests that the variable contains information not present in other variables.

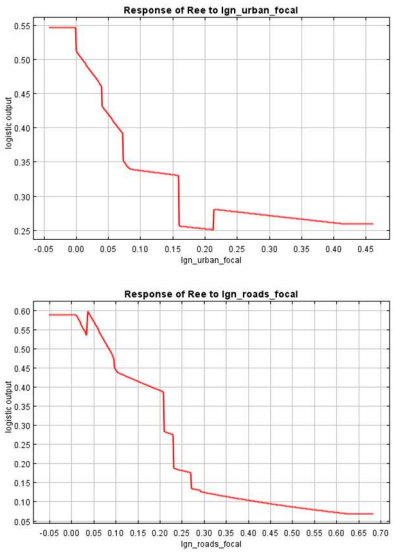

*Figure 11. Trajectories of both Urban focal as well as road focal. On the Y axes the effect on roe deer (*Capreolus capreolus*) vehicle collision suitability and on the X axes the number of the urban/road focal value.*

Model Performance

For the collision model, the AUC of the training data is 0.730 (*Figure 12*). The collision model with road width as a predictive factor used 1,433 points out of the total 11,430 points provided by FBE. As with the binomial test for the occurrence, 30% was used as testing data for the binomial test. When looking at the binomial test we see a success change of 65.2% (280 out of 429 trials), significantly (p= <2.2 e-16) more than expected based on the average collision probability (41.6%). Something that is also confirmed by the 95% confidence interval [0.605,0.698]. This binomial test was also conducted based on the "Maximum training sensitivity plus specificity", which is 0.431 in this case. The procedure is the same as in the occurrence model described above.

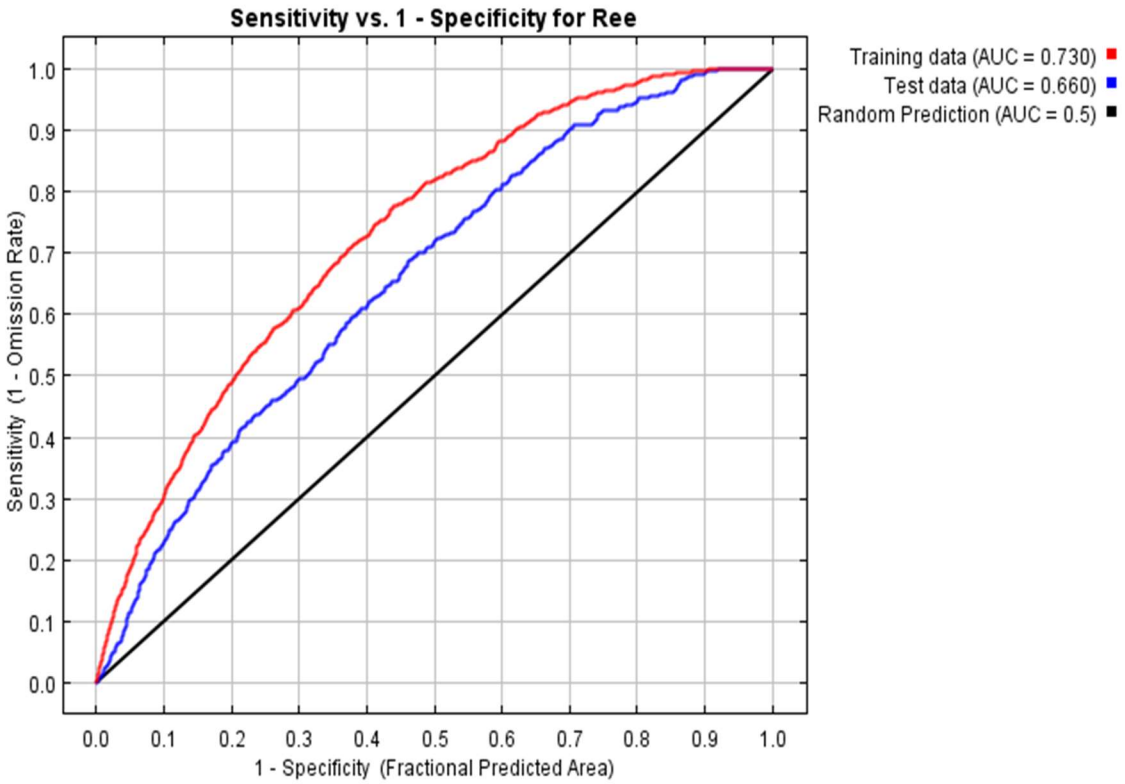

*Figure 12. The AUC of the roe deer (*Capreolus capreolus*) vehicle collision suitability model training data, AUC = 0.730.*

Model without Road Width

Seeing that the road width map excludes a lot of the road network of the Netherlands, and therefore a lot of recorded collisions and their respective conditions are lost. Thus, when removing road width from the environmental factors we end up with the following map, seen in *Figure 13*. The model is based on a total of 6,940 collisions. This map also shows a higher collision suitability in the east and south of the Netherlands, with areas further away from the larger cities as the highest suitability. Since this map has more road data, it can be seen that the cities have the lowest collision suitability. The same can be seen as in the model with road width, that the east and the south of the Netherlands show the highest collision suitability.

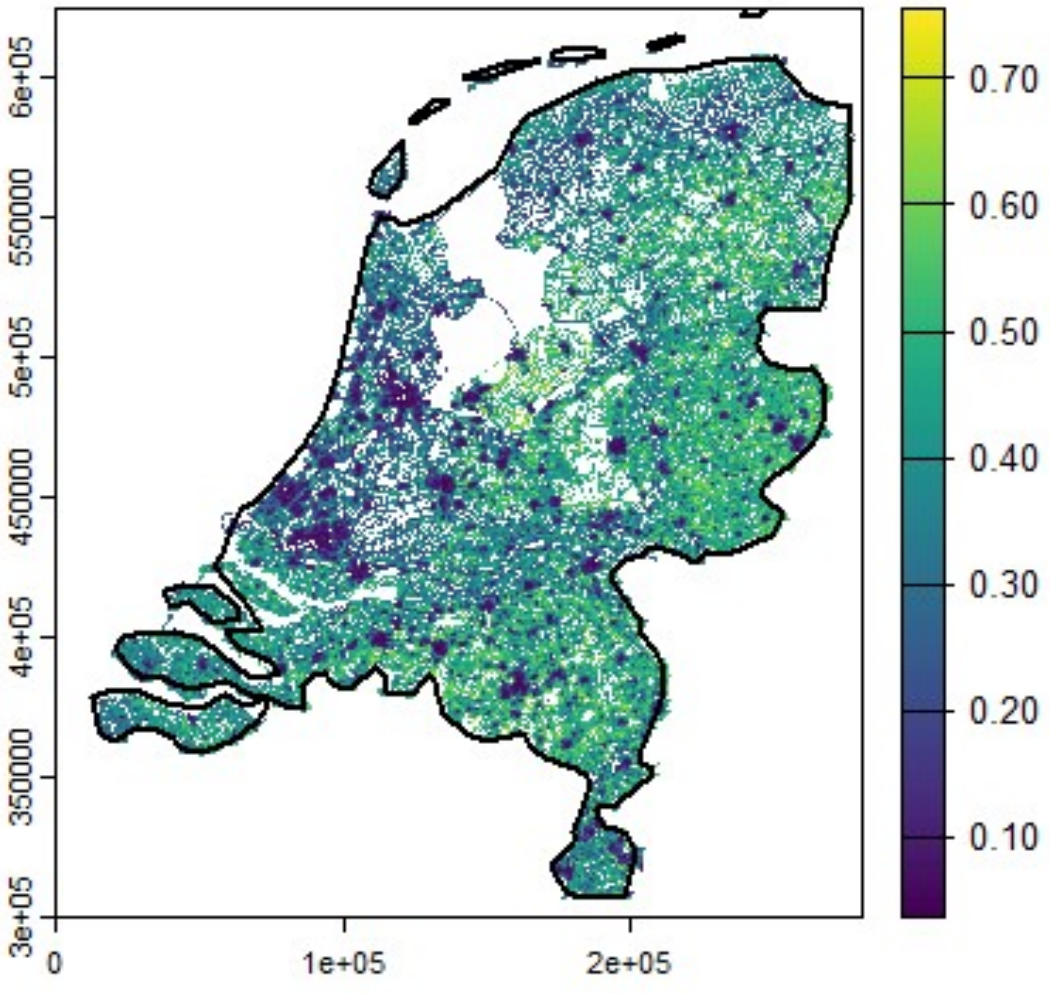

*Figure 13. A wildlife-vehicle collision suitability map for roe deer (*Capreolus capreolus*) generated using MaxEnt. In which 0 is not suitable and 1 is most suitable*, based on 6,940 collision points. With the coordinates on the X and Y axis.

Predictive Factors

MaxEnt provides us with the following top predictors (>5%): Road speed (25.3%), grass in urban areas (22.9%), forest distance (18.2%), bushes (9.1%), roe deer occurrence (5.6%), and freshwater distance (5.3%) (*Table 4*). The trajectory of these six predictive factors in three groups of trajectories can be described as follows (*Figure 14*). The first one being a positive relationship between predictive factors and collision suitability. This trajectory can be seen in both the road speed, which shows an increasing collision proneness with increasing speeds up to 80km/h (*Figure 14a*), as well as the bushes (*Figure 14d*) and roe deer occurrence focal (*Figure 14e*). The second trajectory that can be seen is a negative relation between predictive factor and collision suitability, this is the case for both the grass in urban focal (*Figure 14b*) as well as the forest distance (*Figure 14c*). In which forest distance shows high proneness to collision up until 1km from the road. Lastly, the freshwater distance model (*Figure 14f*) shows a hump shape with an early peak, indicating relatively high collision proneness close to freshwater.

Variable	Percent contribution	Permutation importance
dutch_road_speed_reclassified	25.3	18.9
lgn_grass_in_urban_focal	22.9	13.5
lgn_forest_distance	18.2	9.7
lgn_bushes_focal	9.1	5.7
Roedeer_occurrence_focal	5.6	0.9
lgn_fresh_water_distance	5.3	14.3
lgn_urban_focal	2.6	0.9
lgn_fresh_water_focal	2.3	8
lgn_forest_focal	1.9	5.3
lgn_pasture_focal	1.6	6.3
lgn_other_grass_focal	1.3	0.8
lgn_other_focal	0.9	3.3
lgn_forest_in_urban_focal	0.8	2.1
distance_to_ecoduct	0.7	2.5
lgn_pasture_distance	0.7	4.5
lgn_agroculture_focal	0.5	1.9
lgn_building_in_rural_focal	0.1	0.3
lgn_swamp_focal	0.1	0.4
lgn_NMAG_focal	0.1	0.1
lgn_solar_focal	0	0.1
lgn_dunes_focal	0	0
lgn_reed_focal	0	0.1
lgn_drifting_river_sand_focal	0	0.1
lgn_nurseries_focal	0	0.1
lgn_heath_focal	0	0
lgn_salt_marsh_focal	0	0.2
lgn_salt_water_focal	0	0
lgn_open_sand_in_coast_focal	0	0
lgn_raised_bog_focal	0	0
lgn_roads_focal	0	0

Table 4. List of the environmental variables and their contribution to the model.

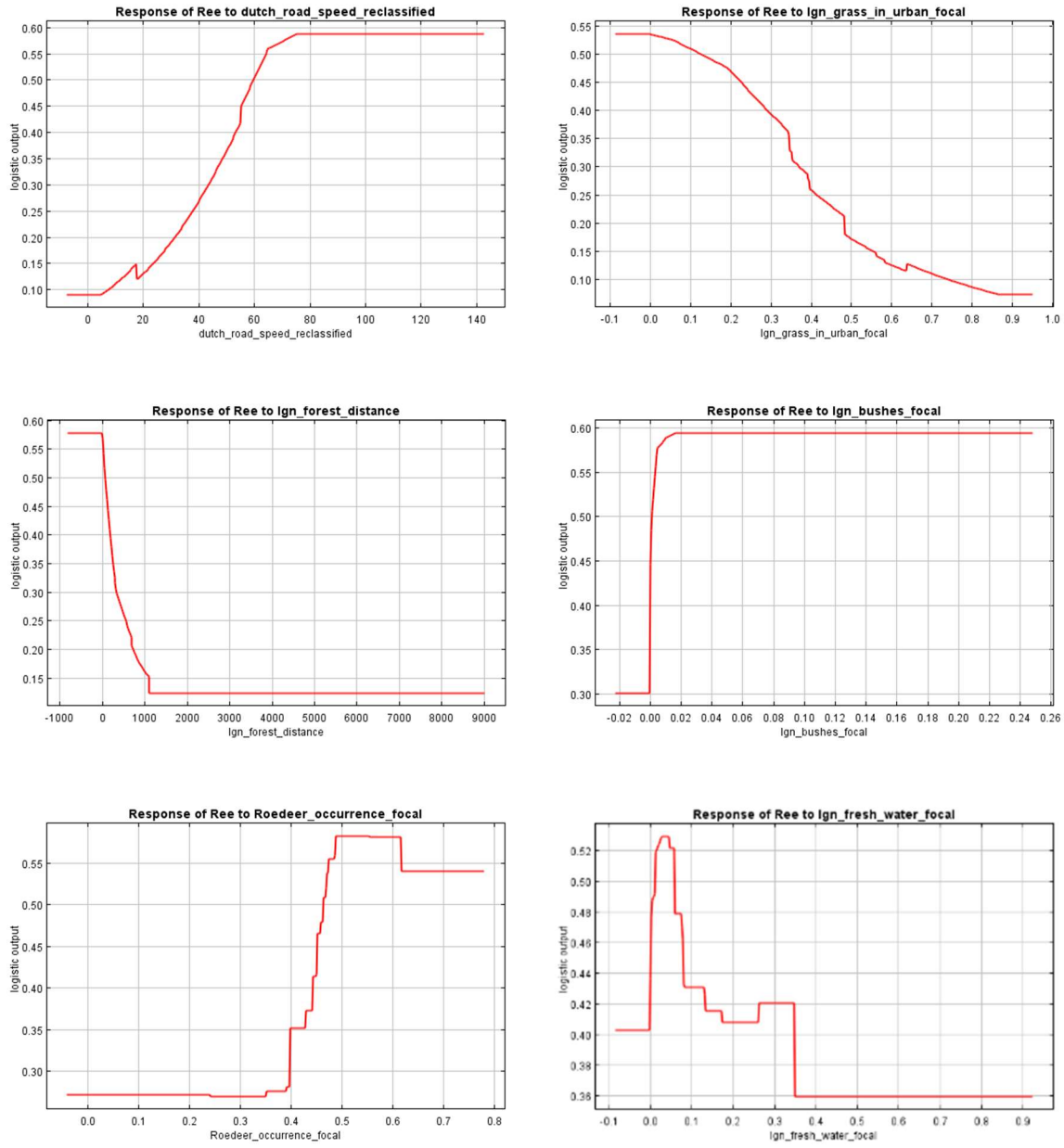

*Figure 14. The trajectory of each of the top predicters. On the Y axes the effect on roe deer (*Capreolus capreolus*) vehicle collision suitability and on the X axes the number of the environmental variable (focal of landcover for grass in urban, bushes, roe deer occurrence and fresh water, distance in meters for distance to forest and km/h for road speed).*

A: Response of roe deer collision risk to maximum road speed. B: Response of roe deer collision risk to grass in urban. C: Response of roe deer collision risk to forest distance. D: Response of roe deer to bushes. E: Response of roe deer collision risk to occurrence. F: Response of roe deer collision risk to fresh water.

Within this model, we also have a few predictive factors that show up in the jackknife test whilst not showing a high contribution percentage (*Figure 15*). In this model, the urban focal in addition to both the focal and the distance measure of pasture seem to be of predictive value. In which urban shows a negative relationship with collision suitability, meaning a higher percentage of urban ground cover the lower collision suitability. Whilst pasture shows a hump shape in which between 40% and 50% ground cover, and distance to these pastures show a negative correlation with shorter distances being more suitable for collisions (*Figure 16*).

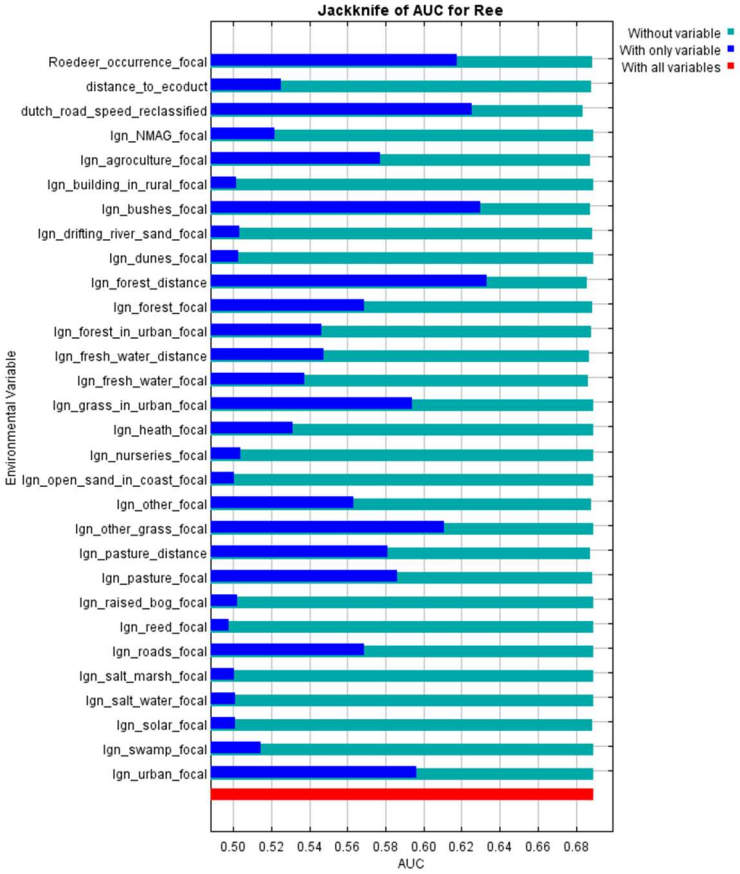

Figure 15. Jackknife test results for environmental variable contributions in the MaxEnt model. The bars represent model gain with each variable used in isolation (dark blue), and with that variable excluded (light blue). Longer dark blue bars indicate variables with high predictive power on their own. therefore a higher dark blue bar suggests that the variable contains information not present in other variables.

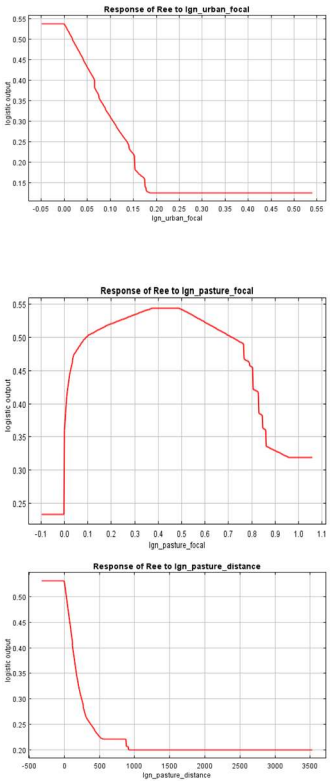

*Figure 16. The trajectory of each of the top predicters. On the Y axes the effect on roe deer (*Capreolus capreolus*) vehicle collision suitability and on the X axes the focal number of the environmental variable(urban, pasture, and distance in meters for pasture distance).*

Model Performance

The collision model without road width, has an AUC value of 0.692 (*Figure 17*). For the collision model, 6,940 collisions out of the same 11,430 points that were used. In this case of these 6,940 collisions, 30% is used as test data for the binomial test. This binomial test gives us a probability of success of 73.2% (1,525 of the 2,082 trials). Significantly better (p = <2,2e-16) than expected based on the average collision suitability. Which is also confirmed by the 95% confidence interval (0.721, 0.751). Both indicate that the model performs better than is expected when it would be based on random change, in which case the correctly predictive value would be around 32%.

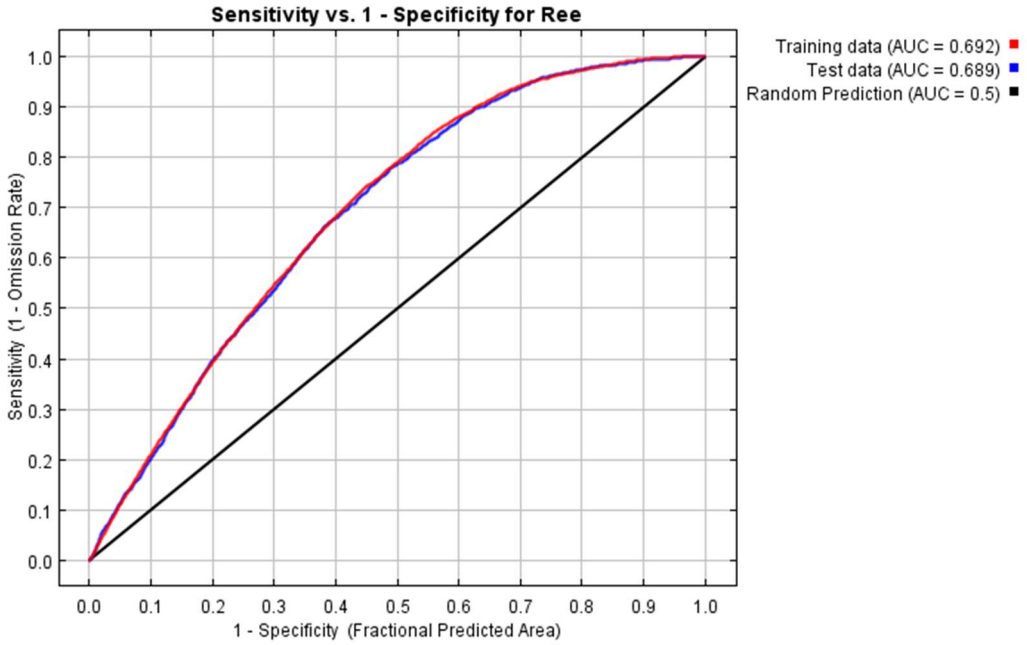

*Figure 17. The AUC of the roe deer (*Capreolus capreolus*) vehicle collision suitability model. Training data AUC = 0.692.*

Comparing the Two Collision Models

Considering the performance of both models, the model with the road width has a higher AUC but a lower probability of success. What does this mean and how does this affect the interpretation of the models? The AUC of a model tells us how well the model can differentiate between the different classes of predictive factors used to create the model. The binomial test, however, gives us a percentage of correct predictions, based on the calculated maximum training sensitivity plus specificity. Therefore, if we compare the two models, the model using road width is better in discriminating between the different factors, while the model without road width has a higher predictive factor.

Secondly, when assessing the predictive factors between the two different models, a difference between them can be seen both in the contribution of each factor as well as in the relevance of each factor. Obviously, road width is removed from the second model, but this does not enhance the importance of the other factors. This is most likely because the road width model only includes the main Dutch road network (highways and provincial roads). As visible by the number of collisions that are not used in the model with road width (1,433) in comparison with the model without (6,940), a lot of roads and therefore the collisions occurring on these roads are not used in the road width model.

As mentioned apart from differences between the percentage of contribution of predictive factors, the factors themselves also show slightly different trajectories between the two different models. When we compare the four response curves that are within the top predictors of both models, only one shows a, to some degree, different trajectory. The bushes focal (*Figure 18b*) shows a hump shape in the road width model suggesting a peak in collision suitability at this hump, whereas the model without a road shows a quickly rising asymptote. Thus, while both models see an increase in collision suitability at a bush cover of around 1%, this drops again around 5% and even further at 10% for the road width model, while the model without road width does not see this suitability drop. Whereas roe deer occurrence shows the opposite, where the model width road width does not see a drop in collision suitability at higher occurrence values, while the model without road width does see this drop(*Figure 18c*). The suitability value is however similar between all response curves at around 0.6 at its peak (*Figure 18*).

Response curve
Road width model

Response curve no
Road width model

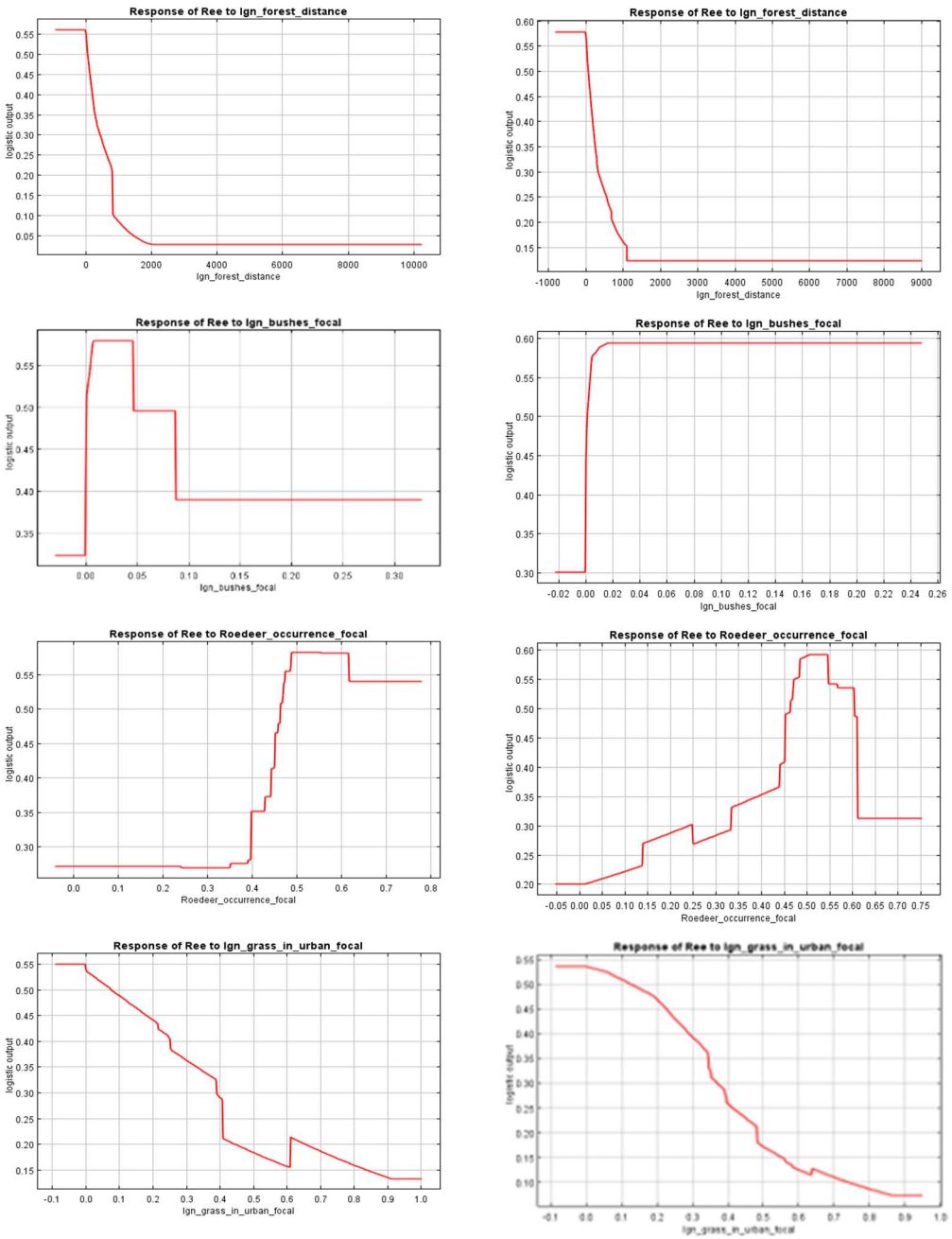

*Figure 18. The trajectory of each of the top predictors. On the Y axes the effect on roe deer (*Capreolus capreolus*) vehicle collision suitability and on the X axes the focal number of the environmental variable (bushes, grass in urban), suitability of roe deer for roe deer occurrence, and distance in meters for pasture distance).*
A: Response of roe deer collision risk to forest distance, with and without road width.
B: Response of roe deer collision risk to bushes, with and without road width.
C: Response of roe deer collision risk to occurrence, with and without road width.
D: Response of roe deer collision risk to urban, with and without road width.

Lastly, there is one predictive factor that even though it is not high in the percental contribution, it did have a larger bar within both jackknifes, namely Urban. Both models show that with an increasing percentage of urban landscape, the collision suitability goes down (*Figure 19*).

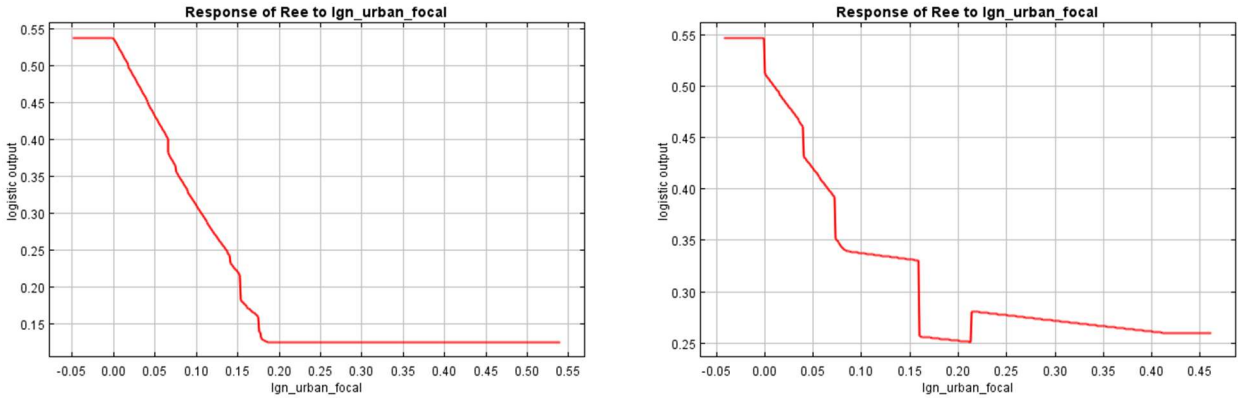

*Figure 19.The trajectory of each of the top predicters. On the Y axes the effect on roe deer (*Capreolus capreolus*) vehicle collision suitability and on the X axes the focal number of the environmental variable (urban, pasture).*

Discussion & Conclusion

This study aimed to identify and map the predictive factors of wildlife-vehicle collisions in the Netherlands, with a specific focus on roe deer. The analysis showed that several key factors influence the risk of collision, including factors such as land use, road characterization and roe deer occurrence. These findings provide useful insights into the interactions between human infrastructure, naturally occurring landcover and the collision risk for wildlife within the highly fragmented landscape of the Netherlands. Each of these categories will be discussed in more detail following the same structure as in the result section of the paper.

The Occurrence Model

Although not the primary focus of this paper, discussing the results of the occurrence model could be valuable as it serves as a predictive factor for both wildlife-vehicle collision models (with and without road width). The distribution map (*Figure 3*) created by MaxEnt indicates the highest occurrence of roe deer around natural areas within the Netherlands, such as the Veluwe, Utrechtse Heuvelrug, Oostvaardersplassen and the coastal dunes. Something that aligns with prior research that identified the highest percentage of roe deer occurrences within Gelderland and Utrecht (Hoekman, 2013).

Predictive Factors

The model identifies bushes, dunes, forest cover and forest distance to be the top predictors of roe deer occurrence aligning with the species' ecology and prior research (Coenen, 2022; Zorzi et al., 2022) and the hypothesis at the beginning of this research. Dunes however are not something that is typically indicative of roe deer populations and is presumably influenced by the coastal roe deer population. Furthermore, the low contribution of pasture is also something that is contradictory to earlier research, which found that roe deer prefer grasslands near the forest edge (Coenen, 2022; Zorzi et al., 2022). While the forest edge is represented in the form of distance-to-forest, in which lower distances are preferred, the grasslands in the Netherlands, mainly pasture, are only contributing 3% to the model. However, the model did show by means of the jackknife that the distance to pasture is a relevant factor to the occurrence of roe deer, with lower distances being more suitable and peaking at a distance of around 500 meters, similar to earlier research (Coenen, 2022).

Model Performance

The model shows an AUC of 0.610, indicating a performance slightly above random. This could explain why some of the expected land uses are not showing up as an important predictive factor. This AUC is also lower than the AUC (0.713) of an earlier master thesis by a WUR student(Coenen, 2022). This thesis also used data from waarnemingen.nl as well as the LGN database (*LGN*, n.d.; *Welkom - Waarneming.Nl*,

n.d.). The binomial test also suggests that the occurrence prediction is not the strongest with only 44% correctly predicted (compared to the 37% for a null model). Overall, this indicates that predicting roe deer occurrence within a highly fragmented and human-modified landscape can be very challenging. Something that is most likely increased by the aggregation of the cells during the data preparation, which may have affected the model's accuracy. As the Netherlands is a highly fragmented landscape, therefore by aggregating this landscape into larger cells, each with a percentage of land use occurring within the boundaries of this cell, results in a reduction of the details. This could explain why pasture would not show up as an important predictive factor. Since within the Netherlands a large percentage consists of pasture, each grid cell of 100 meters by 100 meters now contains at least some pasture, diluting its potential effect on roe deer occurrence. It could therefore be that factors such as interspecific competition, predation pressure or fine-scale habitat quality are of high significance when it comes to predicting roe deer occurrence.

Nevertheless, since the model still performed better than chance and improved the performance of both collision models, included in the final models.

Collision Models

This study identified multiple key predictive factors across the Netherlands, using two different models, one with and one without road width. Finding several factors that are indeed potentially linked to collision risk.

First, the performance of the models will be discussed and compared to prior research indicating similarities and differences. Whilst secondly, the initial hypothesis will be re-examined, and examined based on the findings of this and relevant prior research. These will be discussed in the same order as the predictions at the beginning of the paper.

Model Performance

Both collision models show a reasonable performance, with an AUC of 0.730 for the model with road width and 0.692 for the model without road width. These results are considered fair (Çorbacıoğlu & Aksel, 2023), and thus indicate the ability to distinguish between each of the different factors used. The success rates of 65% and 73%, respectively, indicate that both models are able to correctly predict the collision suitability to some extent. These predictive values are lower than those in an earlier study conducted in Spain in which their model predicted 87% of the collisions correctly, although they used a different modelling approach (Malo et al., 2004). Within this Spanish research low, medium and high collision areas were compared among each other and compared among each other.

Nevertheless, it should be noted that both models still lost a lot of recorded collisions. In which the model with road width lost 9,997 (11,340 – 1,433), while the model without road width lost 4,490 collision points (11,340 – 6,940). Something that is most likely due

to the road network data being the limiting factor. Specifically, the model with road width only includes data of the main roads within the Netherlands, excluding a lot of the collisions that occurred on smaller roads. Likewise, the data used for the maximum road speed likely does not perfectly depict the road network of the Netherlands, resulting in a loss of usable collision incidents. As mentioned by the paper of Gunson, the composition of the road network data can influence the quality of the model (Gunson et al., 2011). Furthermore, no collinearity test was performed among the different predictive factors. This is something that is highly recommended in further research to improve the model by reducing strongly correlated factors.

Land Use

It was predicted that a higher percentage of urban landscape decreased the collision suitability. Earlier research found an absence of foxes and badgers in areas with greater urbanization, and therefore an absence of collisions involving these animals (Grilo et al., 2009). Although this research was based on the risk to carnivore-vehicle collisions, the impact of urban areas could as well hold up for roe deer, as roe deer are similarly shy animals, that react with higher stress levels to human stimuli (Cahler & Marsten, 2012). This could be explained by the deterring characteristics of urban areas, as shown by this research. On the contrary, the distance to urban areas would have an effect on the number of collisions recorded, as suggested by prior research (Tajchman et al., 2020). This is at odds with both the results of this research as well as the research of Grilo(et al., 2009).

Other predictive factors in both models are the amount of, and distance to, forest(both a negative correlation) as well as the amount of bushes(positive correlation), and grassland in urban areas(negative correlation). Forest has been suggested to be of influence in the number of wildlife-vehicle collisions by prior research using roe deer, which showed the relative probability of a collision to increase with an increasing forest cover (Mayer et al., 2021). Similarly, a paper by Madsen advised to reduce the amount of vegetation near the road to prevent wildlife-vehicle collisions (Madsen et al., 2001). This also relates to other types of vegetation, such as bushes, and is something that also was shown to be of importance within this paper's models.

Road Composition

In the road width model, road width is predictive of the suitability of wildlife-vehicle collisions. The response curve shows a peak at a road width between 5 to 9 meters, and a lower plateau at 9 to 15 meters. This meets my hypothesis in which I expected medium size roads to have a higher collision suitability. Something that a South Korean and a Danish study showed as well, however, the Korean research did not use roe deer as one of the animals within its study, but it did use water deer (*Hydropotes inermis*) ,another small ungulate deer, among a long list of other animals (Hwang & Jo, 2024; Mayer et al., 2021). This same study also linked road speed to wildlife-vehicle collisions, something that could also be seen in this studies model. However, the importance of road speed can only be found in the second collision model, where road width is removed as one of the predictive factors. This has likely to do with the limiting data scope of the road width dataset. Nevertheless, this studies model suggests that with increasing speed the collision suitability rises. Something that is confirmed by prior studies in Europe (Langbein et al., 2011).

Mitigation Measures

As previously hypothesized, mitigation measures would decrease the suitability of wildlife-vehicle collisions. However, neither of the models in this study showed any predictive power for the distance to wildlife crossings. Something that earlier research showed to have an effect on the number of collisions (Sugiarto, 2022). However, this does not mean that wildlife crossings are shown to be not effective within this research, since I only checked whether the distance to the nearest wildlife crossing has a predictive effect on the collision suitability. To check whether or not these structures are effective another study should be designed were data before and after the construction of a crossing, or two similar roads, one with and one without a wildlife crossing, is compared. Other mitigation measures such as fences or similar deterring measures, were not included in either model due to the lack of reliable data. Something which could be improved in future research, because early research shows that these mitigation structures do have an effect on the number of wildlife-vehicle collisions and therefore the suitability of roads that have these mitigation measures (Hwang & Jo, 2024; Langbein et al., 2011; Seiler, 2005).

Roe Deer Occurrence

Seeing that in both models roe deer occurrence had a positive relationship with the collision suitability, it may be concluded that areas with a higher population of roe deer also are more suitable for collisions. This can also be seen in the maps made in which we see drops in both occurrence suitability as well as in collision suitability, around the larger cities of the Netherlands for example. Likewise, areas around the larger natural parks see both an increase in occurrence suitability and collision suitability. These findings are in line with the prediction made in the beginning of the research.

Limitations and Future Directions

While this research provides new insights on the predictive factors of wildlife-vehicle collisions, this research is not without its own limitations. These weaknesses are both caused by limitations in the models design as well as limitations in the data, the latter of which will now be discussed.

Quality of the data

The difference in the number of data points used (1,433 in the model with road width and 6,940 for the model without road width) and data points collected (11,430) highlights potential improvement in the comprehensiveness of the road-network data. Using a more comprehensive data set would result in more available collision points and therefore an expected improvement of the overall model. This is again highlighted by the low AUC values of both models. Another potential improvement is trying to limit the aggregation performed on all raster-data. Whether this can be done by using more powerful hardware or other software than MaxEnt which allows larger datafiles to be uploaded, remains open for discussion.

Missed Potential Predictive Factors

Another possible improvement could be made by presenting the model with more potential predictive factors, for instance animal related factors, driver related factors and weather conditions. These are some of the factors that were not taken into account but could provide extra information and therefore increase the predictability of the model. Each of these potential missed factors will be discussed below.

Animal characteristics, as sex, age and activity pattern, could well influence the probability of collision (Madsen et al., 2001). Therefore, knowing the age and sex of the roe deer in each collision case could help explain some of the still not explained patterns, something that now was not precent for all collisions. Adult roe deer make up the largest group of collision victims, however, a peak in yearlings can be seen during the spring season (Mayer et al., 2021).

As already hinted at in the previous paragraph, season, weather, and time of day could all be presumed other predictive factors. Each of these factors will affect both the animals' movement patterns as well as the behaviour of motorists. Roe deer have a diurnal activity pattern (Steiner et al., 2021), thus during these times they are active and have the largest chance of being on the road making a collision possible (Mayer et al., 2021). At the same time, the weather could influence the visibility of the driver making the animals harder to see (Su et al., 2023). Additionally, think of effects like black ice making the roads more slippery (Su et al., 2023). Even seasons could play a role with research by Seo suggesting that during both spring and autumn see a higher mammal collision rate (Seo et al., 2015). However, knowing the exact time of a collision and therefore the exact weather conditions at the time of collision. Nevertheless, seasonal effects should be able to be shown if included in a model as shown by prior research (Su et al., 2023). Lastly, drivers may not only be affected by the weather, but their behaviour perception and driving style can have an influence on collision risk, as well as the overall traffic composition (Mayer et al., 2021).

With all these missing variables, a new conceptual model could be made, and is depicted in *Figure 21*.

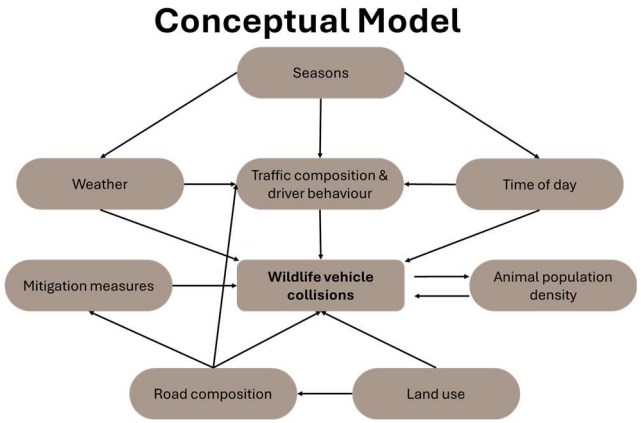

Figure 21. An updated conceptual model of the expected mechanisms contributing to wildlife-vehicle collisions. The model illustrates the interplay between animal factors, as the population density, mitigation measures as wildlife crossings, road composition as road width and maximum speed and landscape features as the different land-use cover classes. As well as the newly added climatical factors such as weather time of day and season. Each arrow indicates a relationship between the components of the full model.

Management Implications

Both the created suitability maps, as well as the analysis of the predictive factors, can be used to highlight areas of higher risk. The importance of forest, urban, bushes and urban grass suggests focusing on edge habitats, within 1km of the forest edge as shown by the trajectory, and areas with a higher bush cover, to be the best strategy for reducing the number of wildlife-vehicle collisions. This is in line with prior research suggesting the management of edge habitats for wildlife-vehicle collision management (Putman, 2004). Seeing that road speed also was one of the predictive factors (the largest in the model without road width) suggests that lowering the maximum speed could reduce the number of wildlife-vehicle collisions. This, however, has not yet been proven effective by prior research (Bénard et al., 2023).

Furthermore, the collision suitability maps (*Figures 8 and 13*) highlight areas of potential high collision risk, therefore conducting further research in these areas could be beneficial to identifying/perfecting the new and already found predictive factors. Which can be done in addition to conducting more specific research in order to mitigate these collisions as best as possible. Furthermore, the roads that are shown to have a higher proneness to collisions should be checked for mitigation measures, and if not present these should be installed.

Conclusion

In conclusion, this study provided information about the occurrence of roe deer as well as the factors and potential locations of wildlife-vehicle collisions with roe deer as a model species. The occurrence model showed that the cover of: forest, bushes and grasses are all positively correlated with the roe deer habitat suitability, while an increasing distance to forest and pasture both are negatively correlated to the habitat suitability. The analysis of the collision model showed land use patterns, road composition and animal occurrence all to have an effect on wildlife-vehicle collision risk. The creation of a MaxEnt model allowed for the identification of the effects of the predictive factors. The results indicated that higher urbanized areas (areas with a higher disturbance) saw a decrease in wildlife-vehicle collision suitability, whilst areas closer to forest increased the wildlife-vehicle collision suitability. Furthermore, an increase in bushes increased the suitability of a potential wildlife-vehicle collision, likewise, an increase in roe deer population size also saw an increased suitability for wildlife-vehicle collisions. Other effects were both road width and road speed, in which roads with a width between 5 to 9 meters saw the highest wildlife-vehicle collision suitability. As with road speed, the higher the maximum speed limit on a road, the higher the wildlife-vehicle collision suitability. The distance to wildlife crossings, however, did not seem to hold predictive value regarding the collision risk. Using these factors and the roads that were indicated to have a higher risk could aid in future mitigation of wildlife-vehicle collisions. This research, therefore, contributed to the ongoing efforts to balance urban development with wildlife conservation in a densely populated and fragmented area such as the Netherlands. With hopes of reducing the number of wildlife-vehicle collisions saving both animals as well as human lives.

References

Bénard, A., Lengagne, T., & Bonenfant, C. (2023). *A biologically realistic model to predict wildlife-vehicle collision risks*. https://doi.org/10.1101/2023.02.15.528614

Cahler, A. A. ., & Marsten, J. P. . (2012). *Deer : habitat, behavior and conservation*. Nova Science Publishers, Inc.

Coenen, I. (2022). *Predicting population densities and the population-habitat relationship of European Roe deer (Capreolus capreolus) in the Netherlands*.

Coffin, A. W. (2007). From roadkill to road ecology: A review of the ecological effects of roads. *Journal of Transport Geography*, *15*(5), 396–406. https://doi.org/10.1016/j.jtrangeo.2006.11.006

Çorbacıoğlu, Ş. K., & Aksel, G. (2023). Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. In *Turkish Journal of Emergency Medicine* (Vol. 23, Issue 4, pp. 195–198). Wolters Kluwer Medknow Publications. https://doi.org/10.4103/tjem.tjem_182_23

Dataregister Rijkswaterstaat. (n.d.). Retrieved January 25, 2025, from https://maps.rijkswaterstaat.nl/dataregister/srv/dut/catalog.search#/home

De Molenaar, J. G., & Henkens, R. J. H. (1998). *Effectiviteit van wiLdspiegels: een literatuurevaluatie BID 0 C ibn-dlo*.

De Vries, M. (2015). *Road kills of roe deer (Capreolus capreolus) in the Netherlands Assessment of impacts and mitigation measures*. www.natuurmonumenten.nl

Fahrig, L., & Rytwinski, T. (2009). *Effects of Roads on Animal Abundance: an Empirical Review and Synthesis* (Vol. 14, Issue 1). http://www.ecologyandsociety.org/vol14/iss1/art21/

GBIF. (n.d.). Retrieved January 25, 2025, from https://www.gbif.org/

Gibbs, J. P. (2001). *Demography versus habitat fragmentation as determinants of genetic variation in wild populations*. www.elsevier.com/locate/biocon

Grilo, C., Bissonette, J. A., & Santos-Reis, M. (2009). Spatial-temporal patterns in Mediterranean carnivore road casualties: Consequences for mitigation. *Biological Conservation*, *142*(2), 301–313. https://doi.org/10.1016/j.biocon.2008.10.026

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. In *Journal of Environmental Management* (Vol. 92, Issue 4, pp. 1074–1082). Academic Press. https://doi.org/10.1016/j.jenvman.2010.11.027

Haddad, N. M., Brudvig, L. A., Clobert, J., Davies, K. F., Gonzalez, A., Holt, R. D., Lovejoy, T. E., Sexton, J. O., Austin, M. P., Collins, C. D., Cook, W. M., Damschen, E. I., Ewers, R. M., Foster, B. L., Jenkins, C. N., King, A. J., Laurance, W. F., Levey, D. J., Margules, C. R., … Townshend, J. R. (2015). Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances*, *1*(2). https://doi.org/10.1126/sciadv.1500052

Hoekman, E. D. (2013). *Dutch roe deer (Capreolus capreolus), review of cases presented at the Dutch Wildlife Health Centre*.

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* (Vol. 14, Issue 2). https://about.jstor.org/terms

Hwang, J. W., & Jo, Y. S. (2024). Wildlife–Vehicle Collisions and Mitigation: Current Status and Factor Analysis in South Korea. *Animals*, *14*(20). https://doi.org/10.3390/ani14203012

Langbein, J., Putman, R., & Pokorny, B. (2011). Traffic collisions involving deer and other ungulates in Europe and available measures for mitigation. In *Ungulate Management in Europe* (pp. 215–259). Cambridge University Press. https://doi.org/10.1017/cbo9780511974137.009

LGN. (n.d.). Retrieved February 16, 2025, from https://lgn.nl/basiskaart

Madsen, A. B., Strandgaard, H., & Prang, A. (2001). *factors_causing_traffic_killings_of_roe_deer_in_denmark*.

Malo, J. E., Suárez, F., & Díez, A. (2004). Can we mitigate animal-vehicle accidents using predictive models? *Journal of Applied Ecology*, *41*(4), 701–710. https://doi.org/10.1111/j.0021-8901.2004.00929.x

Mayer, M., Coleman Nielsen, J., Elmeros, M., & Sunde, P. (2021). Understanding spatio-temporal patterns of deer-vehicle collisions to improve roadkill mitigation. *Journal of Environmental Management*, *295*. https://doi.org/10.1016/j.jenvman.2021.113148

Milner, G. R., & Boldsen, J. L. (2022). *Population trends and the transition to agriculture: Global processes as seen from North America*. https://doi.org/10.1073/pnas

Pagany, R. (2020). Wildlife-vehicle collisions - Influencing factors, data collection and research methods. In *Biological Conservation* (Vol. 251). Elsevier Ltd. https://doi.org/10.1016/j.biocon.2020.108758

Putman, R. J. (2004). *DEER AND ROAD TRAFFIC ACCIDENTS: A REVIEW OF MITIGATION MEASURES: COSTS AND COST-EFFECTIVENESS*. www.deercollisions.co.uk/ftp/mit_review.doc

Rytwinski, T., & Fahrig, L. (2015). The Impacts of Roads and Traffic on Terrestrial Animal Populations. In *Handbook of Road Ecology* (pp. 237–246). wiley. https://doi.org/10.1002/9781118568170.ch28

Rytwinski, T., Soanes, K., Jaeger, J. A. G., Fahrig, L., Findlay, C. S., Houlahan, J., Van Ree, R. Der, & Van Der Grift, E. A. (2016). How effective is road mitigation at reducing road-kill? A meta-analysis. *PLoS ONE*, *11*(11). https://doi.org/10.1371/journal.pone.0166941

Seiler, A. (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

Seo, C., Thorne, J. H., Choi, T., Kwon, H., & Park, C. H. (2015). Disentangling roadkill: the influence of landscape and season on cumulative vertebrate mortality in South Korea. *Landscape and Ecological Engineering*, *11*(1), 87–99. https://doi.org/10.1007/s11355-013-0239-2

Steiner, W., Scholl, E. M., Leisch, F., & Hacklander, K. (2021). Temporal patterns of roe deer traffic accidents: Effects of season, daytime and lunar phase. *PLoS ONE*, *16*(3 March). https://doi.org/10.1371/journal.pone.0249082

Su, H., Wang, Y., Yang, Y., Tao, S., & Kong, Y. (2023). An Analytical Framework of the Factors Affecting Wildlife–Vehicle Collisions and Barriers to Movement. In *Sustainability (Switzerland)* (Vol. 15, Issue 14). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/su151411181

Sugiarto, W. (2022). *The Impact of Wildlife Crossing Structures on Wildlife-Vehicle Collisions*. https://ssrn.com/abstract=4025079

Tajchman, K., Gawryluk, A., & Fonseca, C. (2020). Predicting wildlife–vehicle collisions in an urban area by the example of Lublin in Poland. *Applied Ecology and Environmental Research*, *18*(1), 1981–1997. https://doi.org/10.15666/aeer/1801_19811997

United Nations. (2024). *World Population Prospects 2024 Summary of Results Ten key messages*.

van der Grift, E. A., van der Ree, R., Fahrig, L., Findlay, S., Houlahan, J., Jaeger, J. A. G., Klar, N., Madriñan, L. F., & Olson, L. (2013). Evaluating the effectiveness of road mitigation measures. *Biodiversity and Conservation*, *22*(2), 425–448. https://doi.org/10.1007/s10531-012-0421-0

Welkom - Waarneming.nl. (n.d.). Retrieved January 25, 2025, from https://waarneming.nl/

WorldData.info. (2024). *WorldData.info The Netherlands*. https://www.worlddata.info/europe/netherlands/index.php#google_vignette

Zheng, Y. M. (2020). *Human Population Growth Since the Agricultural and Industrial Revolution*.

Zorzi, P., Nardotto, A., Bottazzo, M., & Zotto, M. D. (2022). Habitat selection of the roe deer Capreolus capreolus (Artiodactyla: Cervidae) in an agroforestry system. *Natural History Sciences*, *9*(2), 3–6. https://doi.org/10.4081/nhs.2022.550

Appendix 1

Code LGN data of the occurrence model:

```
library(terra)

lgn_raster <- rast("LGN_2023/LGN2023.tif")

chunksize <- 1e8


#road mask

lgn_roads <- as.int(lgn_raster %in% c(251:253))

lgn_roads <- terra::aggregate(lgn_roads, fact=20, fun="mean")


#forest mask

lgn_forest <- as.int(lgn_raster %in% c(11:12,40,43))

lgn_forest <- terra::aggregate(lgn_forest, fact=20, fun="mean")


#pasture mask

lgn_pasture <- as.int(lgn_raster %in% c(1))

lgn_pasture <- terra::aggregate(lgn_pasture, fact=20, fun="mean")


#agroculture mask

lgn_agroculture <- as.int(lgn_raster %in% c(2:10))

lgn_agroculture <- terra::aggregate(lgn_agroculture, fact=20, fun="mean")


#fresh_water mask

lgn_fresh_water <- as.int(lgn_raster %in% c(16))

lgn_fresh_water <- terra::aggregate(lgn_fresh_water, fact=20, fun="mean")


#salt_water mask

lgn_salt_water <- as.int(lgn_raster %in% c(17))

lgn_salt_water <- terra::aggregate(lgn_salt_water, fact=20, fun="mean")


#Urban mask

lgn_urban <- as.int(lgn_raster %in% c(18:19))

lgn_urban <- terra::aggregate(lgn_urban, fact=20, fun="mean")


#Forest_in_urban mask

lgn_forest_in_urban <- as.int(lgn_raster %in% c(20,22))

lgn_forest_in_urban <- terra::aggregate(lgn_forest_in_urban, fact=20, fun="mean")
```

```
#Grass_in_urban mask
lgn_grass_in_urban <- as.int(lgn_raster %in% c(23,28))
lgn_grass_in_urban <- terra::aggregate(lgn_grass_in_urban, fact=20, fun="mean")

#Building_in_rural mask
lgn_building_in_rural <- as.int(lgn_raster %in% c(9))
lgn_building_in_rural <- terra::aggregate(lgn_building_in_rural, fact=20, fun="mean")

#Other_landuse mask
lgn_other <- as.int(lgn_raster %in% c(27))
lgn_other <- terra::aggregate(lgn_other, fact=20, fun="mean")

#Solar_park mask
lgn_solar <- as.int(lgn_raster %in% c(29))
lgn_solar <- terra::aggregate(lgn_solar, fact=20, fun="mean")

#Salt_marsh mask
lgn_salt_marsh <- as.int(lgn_raster %in% c(30))
lgn_salt_marsh <- terra::aggregate(lgn_salt_marsh, fact=20, fun="mean")

#Open_sand_in_coast mask
lgn_open_sand_in_coast <- as.int(lgn_raster %in% c(31))
lgn_open_sand_in_coast <- terra::aggregate(lgn_open_sand_in_coast, fact=20, fun="mean")

#Dunes mask
lgn_dunes <- as.int(lgn_raster %in% c(32:34))
lgn_dunes <- terra::aggregate(lgn_dunes, fact=20, fun="mean")

#Drifting/river_sand mask
lgn_drifting_river_sand <- as.int(lgn_raster %in% c(35))
lgn_drifting_river_sand <- terra::aggregate(lgn_drifting_river_sand, fact=20, fun="mean")

#heath mask
lgn_heath <- as.int(lgn_raster %in% c(36:38))
lgn_heath <- terra::aggregate(lgn_heath, fact=20, fun="mean")

#Raised_bog mask
lgn_raised_bog <- as.int(lgn_raster %in% c(39))
lgn_raised_bog <- terra::aggregate(lgn_raised_bog, fact=20, fun="mean")
```

```
#Swamp mask
lgn_swamp <- as.int(lgn_raster %in% c(41))
lgn_swamp <- terra::aggregate(lgn_swamp, fact=20, fun="mean")

#Reed mask
lgn_reed <- as.int(lgn_raster %in% c(42))
lgn_reed <- terra::aggregate(lgn_reed, fact=20, fun="mean")

#Naturally managed agricultural grassland mask
lgn_NMAG <- as.int(lgn_raster %in% c(45))
lgn_NMAG <- terra::aggregate(lgn_NMAG, fact=20, fun="mean")

#Grass_in_coast mask
lgn_grass_in_coast <- as.int(lgn_raster %in% c(46))
lgn_grass_in_coast <- terra::aggregate(lgn_grass_in_coast, fact=20, fun="mean")

#Other_grass mask
lgn_other_grass <- as.int(lgn_raster %in% c(47))
lgn_other_grass <- terra::aggregate(lgn_other_grass, fact=20, fun="mean")

#Nurseries mask
lgn_nurseries <- as.int(lgn_raster %in% c(61:62))
lgn_nurseries <- terra::aggregate(lgn_nurseries, fact=20, fun="mean")

#Bushes mask
lgn_bushes <- as.int(lgn_raster %in% c(321:323, 331:333))
lgn_bushes <- terra::aggregate(lgn_bushes, fact=20, fun="mean")

raster_list <- list(lgn_roads, lgn_forest, lgn_pasture, lgn_agroculture,
                    lgn_fresh_water, lgn_salt_water, lgn_urban,
                    lgn_forest_in_urban, lgn_grass_in_urban,
                    lgn_building_in_rural, lgn_other, lgn_solar,
                    lgn_salt_marsh, lgn_open_sand_in_coast, lgn_dunes,
                    lgn_drifting_river_sand, lgn_heath, lgn_raised_bog,
                    lgn_swamp, lgn_reed, lgn_NMAG, lgn_grass_in_coast,
                    lgn_other_grass, lgn_nurseries, lgn_bushes)

# Combine all rasters into a single multi-layer raster
```

```
multi_layer_LGN <- do.call(c, raster_list)

# Optionally, you can name the layers
names(multi_layer_LGN) <- c("roads", "forest", "pasture", "agroculture",
                            "fresh_water", "salt_water", "urban",
                            "forest_in_urban", "grass_in_urban",
                            "building_in_rural", "other", "solar",
                            "salt_marsh", "open_sand_in_coast", "dunes",
                            "drifting_river_sand", "heath", "raised_bog",
                            "swamp", "reed", "NMAG", "grass_in_coast",
                            "other_grass", "nurseries", "bushes")



plot(multi_layer_LGN)

##create the mask files
lgn_water_mask <- as.int(lgn_raster %in% c(16, 17))
lgn_water_mask <- terra::aggregate(lgn_water_mask, fact=20, fun="mean")
lgn_water_mask[lgn_water_mask >= 0.5] <- NA
lgn_water_mask[lgn_water_mask < 0.5] <- 1



lgn_urbancenter_mask <- as.int(lgn_raster %in% c(18))
lgn_urbancenter_mask <- terra::aggregate(lgn_urbancenter_mask, fact=20, fun="mean")
lgn_urbancenter_mask[lgn_urbancenter_mask >= 0.3] <- NA
lgn_urbancenter_mask[lgn_urbancenter_mask < 0.3] <- 1



lgn_fullmask <- lgn_raster
lgn_fullmask[lgn_fullmask > 0] <- 1
lgn_fullmask <- terra::aggregate(lgn_fullmask, fact=20, fun="mean")

multi_layer_LGN_masked <- multi_layer_LGN * lgn_water_mask * lgn_urbancenter_mask * lgn_fullmask

plot(multi_layer_LGN_masked)

writeRaster(multi_layer_LGN_masked, "multi_layer_LGN_masked.tif")

predictor_names <- c("roads.asc", "forest.asc", "pasture.asc", "agroculture.asc",
```

```
            "fresh_water.asc", "salt_water.asc", "urban.asc",
            "forest_in_urban.asc", "grass_in_urban.asc",
            "building_in_rural.asc", "other.asc", "solar.asc",
            "salt_marsh.asc", "open_sand_in_coast.asc", "dunes.asc",
            "drifting_river_sand.asc", "heath.asc", "raised_bog.asc",
            "swamp.asc", "reed.asc", "NMAG.asc", "grass_in_coast.asc",
            "other_grass.asc", "nurseries.asc", "bushes.asc")

writeRaster(multi_layer_LGN_masked, predictor_names, NAflag=-999999, overwrite=TRUE)

## mask the distance map
freshwater_distance <- rast("lgn_fresh_water_distance.asc")
forest_distance <- rast("lgn_forest_distance.asc")
pasture_distance <- rast("lgn_pasture_distance.asc")
roads_distance <- rast("lgn_roads_distance.asc")

freshwater_distance_masked <- freshwater_distance * lgn_water_mask * lgn_urbancenter_mask * lgn_fullmask
forest_distance_masked <- forest_distance * lgn_water_mask * lgn_urbancenter_mask * lgn_fullmask
pasture_distance_masked <- pasture_distance * lgn_water_mask * lgn_urbancenter_mask * lgn_fullmask
roads_distance_masked <- roads_distance * lgn_water_mask * lgn_urbancenter_mask * lgn_fullmask

plot(freshwater_distance_masked)
plot(forest_distance_masked)

writeRaster(freshwater_distance_masked, "freshwater_distance_masked.asc", NAflag=-999999, overwrite=TRUE)
writeRaster(forest_distance_masked, "forest_distance_masked.asc", NAflag=-999999, overwrite=TRUE)
writeRaster(pasture_distance_masked, "pasture_distance_masked.asc", NAflag=-999999, overwrite=TRUE)
writeRaster(roads_distance_masked, "roads_distance_masked.asc", NAflag=-999999, overwrite=TRUE)


##mask coast and waddeneilanden
library(sf)
library(geodata)

lgn_agroculture <- rast("agroculture.asc")
plot(lgn_agroculture)

polygon(x=c(240000,90000,0,20000,115000,120000,180000),
    y=c(620000,610000,350000,350000,500000,560000,600000))
```

```
polygon <- vect(cbind(c(240000,90000,0,20000,115000,120000,180000),
                      c(620000,610000,350000,350000,500000,560000,600000)), type="polygon")

lgn_multilayer <- rast("multi_layer_LGN_masked.tif")
plot(lgn_multilayer)

crs(polygon) <- crs(lgn_agroculture)
polygon <- project(polygon, crs(lgn_agroculture))

nocoast_lgn_agroculture <- mask(lgn_agroculture, polygon, inverse = TRUE)

plot(nocoast_lgn_agroculture)

writeRaster(nocoast_lgn_agroculture, "agroculture_nocoast.asc", NAflag=-999999, overwrite=TRUE)
```

Code LGN data script for the Collision model:

```
## LGN data

library(terra)
library(predicts)
library(geodata)
library(dplyr)
library(rJava)
library(ncdf4)

## set WD

lgn_raster <- rast("LGN_2023/LGN2023.tif")
plot(lgn_raster, main = 'lgn_raster' )



# Set chunk size (adjust as needed)
chunksize <- 1e8

#road mask
lgn_roads <- as.int(lgn_raster %in% c(251:253))
lgn_roads <- terra::aggregate(lgn_roads, fact=20, fun="mean")
plot(lgn_roads, main = 'lgn_roads')
```

```
#create the focal circle
fw <- focalMat(lgn_roads, d = 300, type = "circle")
# Process roads focal in chunks
lgn_roads_f <- focal(lgn_roads, fw, fun = "sum")
plot(lgn_roads_f, main = 'lgn_roads_f')
##only use the roadcells
lgn_roads_f[lgn_roads == 0] <- NA
plot(lgn_roads_f, main = 'lgn_roads_focal_road_only')
# Calculate distance in chunks
lgn_roads_dist <- distance(lgn_roads, target = 0)
plot(lgn_roads_dist, main = 'lgn_roads_dist')

#forest mask
lgn_forest <- as.int(lgn_raster %in% c(11:12,40,43))
lgn_forest <- terra::aggregate(lgn_forest, fact=20, fun="mean")
plot(lgn_forest, main = 'lgn_forest')
lgn_forest_f <- focal(lgn_forest, fw, fun = "sum")
plot(lgn_forest_f, main = 'lgn_forest_f')
##only use the roadcells
lgn_forest_f[lgn_roads == 0] <- NA
plot(lgn_forest_f, main = 'lgn_forest_f_road_only')

###create distance to forest map
lgn_forest_dist <- distance(lgn_forest, target = 0)
plot(lgn_forest_dist, main = 'lgn_forest_dist')

#pasture mask
lgn_pasture <- as.int(lgn_raster %in% c(1))
lgn_pasture <- terra::aggregate(lgn_pasture, fact=20, fun="mean")
plot(lgn_pasture, main = 'lgn_pasture')
lgn_pasture_f <- focal(lgn_pasture, fw, fun = "sum")
plot(lgn_pasture_f, main = 'lgn_pasture_f')
##only use the roadcells
lgn_pasture_f[lgn_roads == 0] <- NA
plot(lgn_pasture_f, main = 'lgn_pasture_f_road_only')
###create distance to pasture map
lgn_pasture_dist <- distance(lgn_pasture, target = 0)
plot(lgn_pasture_dist, main = 'lgn_pasture_dist')
```

```
#agroculture mask
lgn_agroculture <- as.int(lgn_raster %in% c(2:10))
lgn_agroculture <- terra::aggregate(lgn_agroculture, fact=20, fun="mean")
plot(lgn_agroculture, main = 'lgn_agroculture')
lgn_agroculture_f <- focal(lgn_agroculture, fw, fun = "sum")
plot(lgn_agroculture_f, main = 'lgn_agroculture_f')
##only use the roadcells
lgn_agroculture_f[lgn_roads == 0] <- NA
plot(lgn_agroculture_f, main = 'lgn_agroculture_f_road_only')

#fresh_water mask
lgn_fresh_water <- as.int(lgn_raster %in% c(16))
lgn_fresh_water <- terra::aggregate(lgn_fresh_water, fact=20, fun="mean")
plot(lgn_fresh_water, main = 'lgn_fresh_water')
lgn_fresh_water_f <- focal(lgn_fresh_water, fw, fun = "sum")
plot(lgn_fresh_water_f, main = 'lgn_fresh_water_f')

#distance to fresh water
lgn_fresh_water_dist <- distance(lgn_fresh_water, target = 0)
plot(lgn_fresh_water_dist, main = 'lgn_fresh_water_dist')

#salt_water mask
lgn_salt_water <- as.int(lgn_raster %in% c(17))
lgn_salt_water <- terra::aggregate(lgn_salt_water, fact=20, fun="mean")
plot(lgn_salt_water, main = 'lgn_salt_water')
lgn_salt_water_f <- focal(lgn_salt_water, fw, fun = "sum")
plot(lgn_salt_water_f, main = 'lgn_salt_water_f')

#Urban mask
lgn_urban <- as.int(lgn_raster %in% c(18:19))
lgn_urban <- terra::aggregate(lgn_urban, fact=20, fun="mean")
plot(lgn_urban, main = 'lgn_urban')
lgn_urban_f <- focal(lgn_urban, fw, fun = "sum")
plot(lgn_urban_f, main = 'lgn_urban_f')

#Forest_in_urban mask
lgn_forest_in_urban <- as.int(lgn_raster %in% c(20,22))
lgn_forest_in_urban <- terra::aggregate(lgn_forest_in_urban, fact=20, fun="mean")
```

```
plot(lgn_forest_in_urban, main = 'lgn_forest_in_urban')
lgn_forest_in_urban_f <- focal(lgn_forest_in_urban, fw, fun = "sum")
plot(lgn_forest_in_urban_f, main = 'lgn_forest_in_urban_f')

#Grass_in_urban mask
lgn_grass_in_urban <- as.int(lgn_raster %in% c(23,28))
lgn_grass_in_urban <- terra::aggregate(lgn_grass_in_urban, fact=20, fun="mean")
plot(lgn_grass_in_urban, main = 'lgn_grass_in_urban')
lgn_grass_in_urban_f <- focal(lgn_grass_in_urban, fw, fun = "sum")
plot(lgn_grass_in_urban_f, main = 'lgn_grass_in_urban_f')

#Building_in_rural mask
lgn_building_in_rural <- as.int(lgn_raster %in% c(9))
lgn_building_in_rural <- terra::aggregate(lgn_building_in_rural, fact=20, fun="mean")
plot(lgn_building_in_rural, main = 'lgn_building_in_rural')
lgn_building_in_rural_f <- focal(lgn_building_in_rural, fw, fun = "sum")
plot(lgn_building_in_rural_f, main = 'lgn_building_in_rural_f')

#Other_landuse mask
lgn_other <- as.int(lgn_raster %in% c(27))
lgn_other <- terra::aggregate(lgn_other, fact=20, fun="mean")
plot(lgn_other, main = 'lgn_other')
lgn_other_f <- focal(lgn_other, fw, fun = "sum")
plot(lgn_other_f, main = 'lgn_other_f')

#Solar_park mask
lgn_solar <- as.int(lgn_raster %in% c(29))
lgn_solar <- terra::aggregate(lgn_solar, fact=20, fun="mean")
plot(lgn_solar, main = 'lgn_solar')
lgn_solar_f <- focal(lgn_solar, fw, fun = "sum")
plot(lgn_solar_f, main = 'lgn_solar_f')
#Salt_marsh mask
lgn_salt_marsh <- as.int(lgn_raster %in% c(30))
lgn_salt_marsh <- terra::aggregate(lgn_salt_marsh, fact=20, fun="mean")
plot(lgn_salt_marsh, main = 'lgn_salt_marsh')
lgn_salt_marsh_f <- focal(lgn_salt_marsh, fw, fun = "sum")
plot(lgn_salt_marsh_f, main = 'lgn_salt_marsh_f')

#Open_sand_in_coast mask
```

```
lgn_open_sand_in_coast <- as.int(lgn_raster %in% c(31))
lgn_open_sand_in_coast <- terra::aggregate(lgn_open_sand_in_coast, fact=20, fun="mean")
plot(lgn_open_sand_in_coast, main = 'lgn_open_sand_in_coast')
lgn_open_sand_in_coast_f <- focal(lgn_open_sand_in_coast, fw, fun = "sum")
plot(lgn_open_sand_in_coast_f, main = 'lgn_open_sand_in_coast_f')

#Dunes mask
lgn_dunes <- as.int(lgn_raster %in% c(32:34))
lgn_dunes <- terra::aggregate(lgn_dunes, fact=20, fun="mean")
plot(lgn_dunes, main = 'lgn_dunes')
lgn_dunes_f <- focal(lgn_dunes, fw, fun = "sum")
plot(lgn_dunes_f, main = 'lgn_dunes_f')

#Drifting/river_sand mask
lgn_drifting_river_sand <- as.int(lgn_raster %in% c(35))
lgn_drifting_river_sand <- terra::aggregate(lgn_drifting_river_sand, fact=20, fun="mean")
plot(lgn_drifting_river_sand, main = 'lgn_drifting_river_sand')
lgn_drifting_river_sand_f <- focal(lgn_drifting_river_sand, fw, fun = "sum")
plot(lgn_drifting_river_sand_f, main = 'lgn_drifting_river_sand_f')

#heath mask
lgn_heath <- as.int(lgn_raster %in% c(36:38))
lgn_heath <- terra::aggregate(lgn_heath, fact=20, fun="mean")
plot(lgn_heath, main = 'lgn_heath')
lgn_heath_f <- focal(lgn_heath, fw, fun = "sum")
plot(lgn_heath_f, main = 'lgn_heath_f')

#Raised_bog mask
lgn_raised_bog <- as.int(lgn_raster %in% c(39))
lgn_raised_bog <- terra::aggregate(lgn_raised_bog, fact=20, fun="mean")
plot(lgn_raised_bog, main = 'lgn_raised_bog')
lgn_raised_bog_f <- focal(lgn_raised_bog, fw, fun = "sum")
plot(lgn_raised_bog_f, main = 'lgn_raised_bog_f')

#Swamp mask
lgn_swamp <- as.int(lgn_raster %in% c(41))
lgn_swamp <- terra::aggregate(lgn_swamp, fact=20, fun="mean")
plot(lgn_swamp, main = 'lgn_swamp')
lgn_swamp_f <- focal(lgn_swamp, fw, fun = "sum")
```

```
plot(lgn_swamp_f, main = 'lgn_swamp_f')

#Reed mask
lgn_reed <- as.int(lgn_raster %in% c(42))
lgn_reed <- terra::aggregate(lgn_reed, fact=20, fun="mean")
plot(lgn_reed, main = 'lgn_reed')
lgn_reed_f <- focal(lgn_reed, fw, fun = "sum")
plot(lgn_reed_f, main = 'lgn_reed_f')

#Naturally managed agricultural grassland mask
lgn_NMAG <- as.int(lgn_raster %in% c(45))
lgn_NMAG <- terra::aggregate(lgn_NMAG, fact=20, fun="mean")
plot(lgn_NMAG, main = 'lgn_NMAG')
lgn_NMAG_f <- focal(lgn_NMAG, fw, fun = "sum")
plot(lgn_NMAG_f, main = 'lgn_NMAG_f')

#Grass_in_coast mask
lgn_grass_in_coast <- as.int(lgn_raster %in% c(46))
lgn_grass_in_coast <- terra::aggregate(lgn_grass_in_coast, fact=20, fun="mean")
plot(lgn_grass_in_coast, main = 'lgn_grass_in_coast')
lgn_grass_in_coast_f <- focal(lgn_grass_in_coast, fw, fun = "sum")
plot(lgn_grass_in_coast_f, main = 'lgn_grass_in_coast_f')

#Other_grass mask
lgn_other_grass <- as.int(lgn_raster %in% c(47))
lgn_other_grass <- terra::aggregate(lgn_other_grass, fact=20, fun="mean")
plot(lgn_other_grass, main = 'lgn_other_grass')
lgn_other_grass_f <- focal(lgn_other_grass, fw, fun = "sum")
plot(lgn_other_grass_f, main = 'lgn_other_grass_f')

#Nurseries mask
lgn_nurseries <- as.int(lgn_raster %in% c(61:62))
lgn_nurseries <- terra::aggregate(lgn_nurseries, fact=20, fun="mean")
plot(lgn_nurseries, main = 'lgn_nurseries')
lgn_nurseries_f <- focal(lgn_nurseries, fw, fun = "sum")
plot(lgn_nurseries_f, main = 'lgn_nurseries_f')

#Bushes mask
lgn_bushes <- as.int(lgn_raster %in% c(321:323, 331:333))
```

```
lgn_bushes <- terra::aggregate(lgn_bushes, fact=20, fun="mean")
plot(lgn_bushes, main = 'lgn_bushes')
lgn_bushes_f <- focal(lgn_bushes, fw, fun = "sum")
plot(lgn_bushes_f, main = 'lgn_bushes_f')

##create the mask files
lgn_water <- as.int(lgn_raster %in% c(16, 17))
lgn_water <- terra::aggregate(lgn_water, fact=20, fun="mean")
plot(lgn_water >= 0.5)
lgn_water[lgn_water >= 0.5] <- NA
lgn_water[lgn_water < 0.5] <- 1
plot(lgn_water)

lgn_urbancenter <- as.int(lgn_raster %in% c(18))
lgn_urbancenter <- terra::aggregate(lgn_urbancenter, fact=20, fun="mean")
plot(lgn_urbancenter >= 0.3)
lgn_urbancenter[lgn_urbancenter >= 0.3] <- NA
lgn_urbancenter[lgn_urbancenter < 0.3] <- 1
plot(lgn_urbancenter)

lgn_fullmask <- lgn_raster
lgn_fullmask[lgn_fullmask > 0] <- 1
lgn_fullmask <- terra::aggregate(lgn_fullmask, fact=20, fun="mean")
plot(lgn_fullmask, main = 'lgnfull_mask')

## apply mask over focalmaps
lgn_forest_focal <- mask(lgn_forest_focal, lgn_water)
#roads
lgn_roads_focal <- lgn_roads_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_roads_focal, main = 'lgn_roads_focal')
writeRaster(lgn_roads_focal, filename = "lgn_roads_focal.asc", NAflag=-9999, overwrite = TRUE)

lgn_roads_distance <- lgn_roads_dist * lgn_water* lgn_urbancenter * lgn_fullmask
plot(lgn_roads_distance, main = "lgn_roads_distance")
writeRaster(lgn_roads_distance, filename = "lgn_roads_distance.asc", NAflag=-9999, overwrite = TRUE)

#forest
lgn_forest_focal <- lgn_forest_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_forest_focal, main = 'lgn_forest_focal')
```

```
writeRaster(lgn_roads_focal, filename = "lgn_forest_focal.asc", NAflag=-9999, overwrite = TRUE)

lgn_forest_distance <- lgn_forest_dist * lgn_water* lgn_urbancenter * lgn_fullmask
plot(lgn_forest_distance, main = "lgn_forest_distance")
writeRaster(lgn_forest_distance, filename = "lgn_forest_distance.asc", NAflag=-9999, overwrite = TRUE)

#pasture
lgn_pasture_focal <- lgn_pasture_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_pasture_focal, main = 'lgn_pasture_focal')
writeRaster(lgn_pasture_focal, filename = "lgn_pasture_focal.asc", NAflag=-9999, overwrite = TRUE)

lgn_pasture_distance <- lgn_pasture_dist * lgn_water* lgn_urbancenter * lgn_fullmask
plot(lgn_pasture_distance, main = "lgn_pasture_distance")
writeRaster(lgn_pasture_distance, filename = "lgn_pasture_distance.asc", NAflag=-9999, overwrite = TRUE)

#agroculture
lgn_agroculture_focal <- lgn_agroculture_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_agroculture_focal, main = 'lgn_agroculture_focal')
writeRaster(lgn_agroculture_focal, filename = "lgn_agroculture_focal.asc", NAflag=-9999, overwrite = TRUE)

#fresh water
lgn_fresh_water_focal <- lgn_fresh_water_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_fresh_water_focal, main = 'lgn_fresh_water_focal')
writeRaster(lgn_fresh_water_focal, filename = "lgn_fresh_water_focal.asc", NAflag=-9999, overwrite = TRUE)

lgn_fresh_water_distance <- lgn_fresh_water_dist * lgn_water* lgn_urbancenter * lgn_fullmask
plot(lgn_fresh_water_distance, main = "lgn_fresh_water_distance")
writeRaster(lgn_fresh_water_distance, filename = "lgn_fresh_water_distance.asc", NAflag=-9999, overwrite = TRUE)

#salt water
lgn_salt_water_focal <- lgn_salt_water_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_salt_water_focal, main = 'lgn_salt_water_focal')
writeRaster(lgn_salt_water_focal, filename = "lgn_salt_water_focal.asc", NAflag=-9999, overwrite = TRUE)

#salt water
lgn_salt_water_focal <- lgn_salt_water_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_salt_water_focal, main = 'lgn_salt_water_focal')
writeRaster(lgn_salt_water_focal, filename = "lgn_salt_water_focal.asc", NAflag=-9999, overwrite = TRUE)
```

```
#urban
lgn_urban_focal <- lgn_urban_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_urban_focal, main = 'lgn_urban_focal')
writeRaster(lgn_urban_focal, filename = "lgn_urban_focal.asc", NAflag=-9999, overwrite = TRUE)

#forest in urban
lgn_forest_in_urban_focal <- lgn_forest_in_urban_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_forest_in_urban_focal, main = 'lgn_forest_in_urban_focal')
writeRaster(lgn_forest_in_urban_focal, filename = "lgn_forest_in_urban_focal.asc", NAflag=-9999, overwrite = TRUE)

#grass in urban
lgn_grass_in_urban_focal <- lgn_grass_in_urban_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_grass_in_urban_focal, main = 'lgn_grass_in_urban_focal')
writeRaster(lgn_grass_in_urban_focal, filename = "lgn_grass_in_urban_focal.asc", NAflag=-9999, overwrite = TRUE)

#building in rural
lgn_building_in_rural_focal <- lgn_building_in_rural_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_building_in_rural_focal, main = 'lgn_building_in_rural_focal')
writeRaster(lgn_building_in_rural_focal, filename = "lgn_building_in_rural_focal.asc", NAflag=-9999, overwrite = TRUE)

#other
lgn_other_focal <- lgn_other_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_other_focal, main = 'lgn_other_focal')
writeRaster(lgn_other_focal, filename = "lgn_other_focal.asc", NAflag=-9999, overwrite = TRUE)

#solar
lgn_solar_focal <- lgn_solar_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_solar_focal, main = 'lgn_solar_focal')
writeRaster(lgn_solar_focal, filename = "lgn_solar_focal.asc", NAflag=-9999, overwrite = TRUE)

#salt marsh
lgn_salt_marsh_focal <- lgn_salt_marsh_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_salt_marsh_focal, main = 'lgn_salt_marsh_focal')
writeRaster(lgn_salt_marsh_focal, filename = "lgn_salt_marsh_focal.asc", NAflag=-9999, overwrite = TRUE)

#open sand in coast
lgn_open_sand_in_coast_focal <- lgn_open_sand_in_coast_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_open_sand_in_coast_focal, main = 'lgn_open_sand_in_coast_focal')
writeRaster(lgn_open_sand_in_coast_focal, filename = "lgn_open_sand_in_coast_focal.asc", NAflag=-9999, overwrite
= TRUE)
```

```
#dunes
lgn_dunes_focal <- lgn_dunes_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_dunes_focal, main = 'lgn_dunes_focal')
writeRaster(lgn_dunes_focal, filename = "lgn_dunes_focal.asc", NAflag=-9999, overwrite = TRUE)

#drifting river sand
lgn_drifting_river_sand_focal <- lgn_drifting_river_sand_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_drifting_river_sand_focal, main = 'lgn_drifting_river_sand_focal')
writeRaster(lgn_drifting_river_sand_focal, filename = "lgn_drifting_river_sand_focal.asc", NAflag=-9999, overwrite =
TRUE)

#heath
lgn_heath_focal <- lgn_heath_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_heath_focal, main = 'lgn_heath_focal')
writeRaster(lgn_heath_focal, filename = "lgn_heath_focal.asc", NAflag=-9999, overwrite = TRUE)

#raised bog
lgn_raised_bog_focal <- lgn_raised_bog_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_raised_bog_focal, main = 'lgn_raised_bog_focal')
writeRaster(lgn_raised_bog_focal, filename = "lgn_raised_bog_focal.asc", NAflag=-9999, overwrite = TRUE)

#swamp
lgn_swamp_focal <- lgn_swamp_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_swamp_focal, main = 'lgn_swamp_focal')
writeRaster(lgn_swamp_focal, filename = "lgn_swamp_focal.asc", NAflag=-9999, overwrite = TRUE)

#reed
lgn_reed_focal <- lgn_reed_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_reed_focal, main = 'lgn_reed_focal')
writeRaster(lgn_reed_focal, filename = "lgn_reed_focal.asc", NAflag=-9999, overwrite = TRUE)

#NMAG
lgn_NMAG_focal <- lgn_NMAG_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_NMAG_focal, main = 'lgn_NMAG_focal')
writeRaster(lgn_NMAG_focal, filename = "lgn_NMAG_focal.asc", NAflag=-9999, overwrite = TRUE)

#other grass
lgn_other_grass_focal <- lgn_other_grass_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_other_grass_focal, main = 'lgn_other_grass_focal')
```

```
writeRaster(lgn_other_grass_focal, filename = "lgn_other_grass_focal.asc", NAflag=-9999, overwrite = TRUE)

#nurseries
lgn_nurseries_focal <- lgn_nurseries_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_nurseries_focal, main = 'lgn_nurseries_focal')
writeRaster(lgn_nurseries_focal, filename = "lgn_nurseries_focal.asc", NAflag=-9999, overwrite = TRUE)

#bushes
lgn_bushes_focal <- lgn_bushes_f * lgn_water * lgn_urbancenter * lgn_fullmask
plot(lgn_bushes_focal, main = 'lgn_bushes_focal')
writeRaster(lgn_bushes_focal, filename = "lgn_bushes_focal.asc", NAflag=-9999, overwrite = TRUE)
```

Code Road characteristics:

```
library(terra)

## set WD

## ROAD SPEED

Road_speed_vct <- terra::vect("Weg_eigenschappen/Snelheid_weg/Snelheden.shp")

Road_speed_vct <- terra::project(Road_speed_vct, "EPSG:28992") # but it was already in RDNew

r <- terra::rast(ext = ext(Road_speed_vct),

        res = 100, # 100m resolution (since EPSG:28992 is in meters, yes)

        crs = crs(Road_speed_vct))


road_speed_raster <- terra::rasterize(Road_speed_vct, r, field = "MAXSHD")

plot(road_speed_raster)

writeRaster(road_speed_raster, filename = "Rdata/Weg_eigenschappen/dutch_road_speeds.asc", NAflag=-9999,
overwrite = TRUE)


## Road width
library(sf)
```

```
library(dplyr)
library(tidyverse)
library(terra)

## set WD

# Read the Rijkswaterstaat file (assuming it's a CSV)
rws_data <- read.csv("Weg_eigenschappen/wkd_036-WEGBRDTV2.csv", sep = ";")

# Read the NWB GPKG file
nwb_data <- st_read("Weg_eigenschappen/nwb_wegen.gpkg")



# Check column names in both datasets
names(nwb_data)
names(rws_data)

## rework rws_data for With
rws_data_clean <- rws_data %>%

 select(WVK_ID, BREEDTE) %>%

 mutate(BREEDTE = ifelse(BREEDTE == "onbekend", NA, BREEDTE),

      BREEDTE = as.numeric(BREEDTE)) %>%

 rename(wvk_id = WVK_ID)

## check for duplicates

rws_data_clean

length(rws_data_clean$wvk_id)

length(unique(rws_data_clean$wvk_id))

# there are some duplicates

rws_data_clean <- rws_data_clean %>%
```

```
 group_by(wvk_id) %>%

 summarise(BREEDTE = mean(BREEDTE, na.rm = TRUE)) %>%

 ungroup()

length(rws_data_clean$wvk_id)

## join the data

rws_nwb_join <- nwb_data %>%

 inner_join(rws_data_clean, by = "wvk_id")

rws_nwb_join

rws_nwb_join$BREEDTE



rws_nwb_join_vect <- vect(rws_nwb_join)

rws_nwb_join_vect

ext(rws_nwb_join_vect)

crs(rws_nwb_join_vect, proj = TRUE)

r <- rast(ext = ext(rws_nwb_join_vect),

      res = 100, # 100m resolution

      crs = crs(rws_nwb_join_vect))
```

```
rws_nwb_join_rast <- rasterize(rws_nwb_join_vect, r, field = "BREEDTE", fun = "max")


rws_nwb_join_rast


plot(rws_nwb_join_rast)


## save file


 writeRaster(rws_nwb_join_rast, filename = "Rdata/Weg_eigenschappen/dutch_road_width.asc", NAflag=-9999,
overwrite = TRUE)
```

Code relocate the collisions to road:

```
# Load required libraries

library(raster)

library(sf)

library(terra)

library(data.table)

library(pbapply)


# Read the landuse raster

landuse <- rast("LGN_2023/LGN2023.tif")


# Plot the cropped landuse raster

plot(landuse, main = 'lgn_raster')


# Extract road cells from the landuse raster

road_cells <- landuse %in% c(

 251, # Main road

 252, # Slow road

 253  # Small road

)


road_mask <- road_cells * 1

road_mask[road_mask == 0] <- NA


# Plot the road mask

plot(road_mask, col = 'red', main = 'lgn_roadmap')


# Read the animal collision points

collision_points <- st_read("Rdata/Aanrijdingen/Ree/Ree_coll_Total_new.shp")
```

```
collision_points <- st_transform(collision_points, crs = 28992)

# Function to process large rasters
process_large_raster <- function(raster, chunksize = 1e8) {
 result <- data.table()

 for(i in seq(1, ncell(raster), by = chunksize)) {
  end <- min(i + chunksize - 1, ncell(raster))
  cells <- i:end

  # Extract values and coordinates for this chunk
  vals <- terra::extract(raster, cells)
  coords <- terra::xyFromCell(raster, cells)

  # Combine into a data.table
  chunk_dt <- data.table(x = coords[,1], y = coords[,2], value = vals[,1])

  # Filter for road cells
  road_cells <- chunk_dt[value == 1]

  if(nrow(road_cells) > 0) {
   result <- rbindlist(list(result, road_cells[, .(x, y)]))
  }

  # Optional: Print progress
  cat("Processed cells", i, "to", end, "out of", ncell(raster), "\n")
 }

 return(result)
}

# Process the road_mask raster
road_coords <- process_large_raster(road_mask)

# Create a spatial points object for road cells
road_points <- st_as_sf(road_coords, coords = c("x", "y"), crs = st_crs(road_mask))

# Read the animal collision points
collision_points <- st_read("Rdata/Aanrijdingen/Ree/Ree_coll_Total_new.shp")
```

```
collision_points <- st_transform(collision_points, crs = st_crs(road_mask))

# Function to process chunks of collision points
process_chunk <- function(chunk) {
 nearest_indices <- st_nearest_feature(chunk, road_points)
 distances <- st_distance(chunk, road_points[nearest_indices,], by_element = TRUE)
 nearest_coords <- st_coordinates(road_points[nearest_indices,])

 data.table(
  original_x = st_coordinates(chunk)[,1],
  original_y = st_coordinates(chunk)[,2],
  nearest_road_x = nearest_coords[,1],
  nearest_road_y = nearest_coords[,2],
  distance = as.numeric(distances)
 )
}

# Process collision points in chunks
chunk_size <- 1e8  # Adjust based on your system's capacity
n_chunks <- ceiling(nrow(collision_points) / chunk_size)

result_list <- pblapply(1:n_chunks, function(i) {
 start_idx <- (i-1) * chunk_size + 1
 end_idx <- min(i * chunk_size, nrow(collision_points))

 chunk <- collision_points[start_idx:end_idx,]
 process_chunk(chunk)
})

# Combine results from all chunks
result <- rbindlist(result_list)

# Filter points to keep only those with distance <= 25 meters
filtered_points <- result[distance <= 25]

# Move the remaining points to the nearest road coordinates
filtered_points[, c("final_x", "final_y") := .(nearest_road_x, nearest_road_y)]

# Create a new spatial object with the filtered and moved points
```

```
final_points <- st_as_sf(filtered_points, coords = c("final_x", "final_y"),
                         crs = st_crs(collision_points))

# Plot the final points
plot(landuse, main = 'Final Collision Points')
plot(st_geometry(final_points), add = TRUE, col = 'red', pch = 20)

zoom_extent <- extent(175000,200000,450000,475000)
plot(landuse, ext=zoom_extent, main = 'Final collision Points zoom')
plot(st_geometry(final_points), add = TRUE, col = 'red', pch = 20)

# Write the final points to a new shapefile
st_write(final_points, "Rdata/Aanrijdingen/Ree/final_collision_points_ree.shp", append = FALSE)

#save shapefile as a CSV

final_collisionpoints <- st_read("Rdata/Aanrijdingen/Ree/final_collision_points_ree.shp")

final_collisionpoints_new <- final_collisionpoints[, 1:2]
final_collisionpoints_new$species <- 'Ree'
final_collisionpoints_new$geometry <- NULL
final_collisionpoints_new <- final_collisionpoints_new[c(3, 1, 2)]

write.csv(final_collisionpoints_new, "Rdata/Aanrijdingen/Ree/final_collision_points_ree.csv", row.names = FALSE)
```