

MÓDULO

Modelagem de Áreas de Risco de Atropelamento (MARA)

(Parte integrante do PROMIRF)

CONTRATO: CT. 032/2022 - RDC 262/2021/SIE
ORDEM DE SERVIÇO: 066/2022

FEVEREIRO, 2023



Lista de figuras

Figura 1.- Diagrama do procedimento analítico empregado na obtenção do modelo de risco de atropelamento de fauna silvestre no Estado de Santa Catarina.	9
Figura 2.- Malha viária estadual e federal do Estado de Santa Catarina.	10
Figura 3.- Delimitação das áreas para desenvolvimento do modelo correspondente a cobertura do bioma Mata Atlântica (verde), expandido em 20 km e respectivamente representadas pelas áreas hachuradas.	17
Figura 4.- Variáveis ambientais em formato ASCII, após padronização. No exemplo abaixo são disponibilizadas as variáveis Bio05 e Solo.	18
Figura 5.- Dados utilizados para modelagem bio-ecológica, após processo de rarefação simples.	20
Figura 6.- Tela principal do MaxEnt e menus de configurações ilustrando o processo de configuração dos parâmetros para modelagem inicial da "sp_12" (<i>Coendou spinosus</i>) para o bioma Mata Atlântica. A partir desse cenário foram realizadas modificações com base nos resultados, neste caso específico foram desenvolvidos 10 ensaios (vide Figura 11).	22
Figura 7.- Gráfico das taxas de omissão da classificação produzida pelo MaxEnt para a sp_12 (<i>Coendou spinosus</i>).	23
Figura 8.- Curva ROC (<i>Receiver Operating Characteristics</i>) produzido pelo MaxEnt para a sp_12 (<i>Coendou spinosus</i>).	24
Figura 9.- Superfície de predição resultante do MaxEnt para o ensaio da sp_12 (<i>Coendou spinosus</i>).	24
Figura 10.- Análise <i>Jackknife</i> da importância das variáveis no modelo preditivo do MaxEnt para o ensaio da sp_12 (<i>Coendou spinosus</i>).	25
Figura 11.- Recorte ilustrativo da tabela com os resultados dos modelos aplicados à sp_12 (<i>Coendou spinosus</i>). Mesmo com resultados satisfatórios no primeiro ensaio, foram realizadas diversas tentativas de aprimorar o resultado do modelo, porém sem êxito. Neste caso específico, nenhuma das estratégias superou a qualidade do resultado do primeiro ensaio.	27
Figura 12.- Representação da BD Tipologia de rodovias para o Estado de Santa Catarina. A coluna id informa o identificador do segmento.	27
Figura 13.- Resultado da análise de K-Ripley encontrado para o <i>Didelphis aurita</i> no Trecho 1, no Estado de Santa Catarina.	28
Figura 14.- Resultado da análise de hotspot encontrado para o <i>Didelphis aurita</i> no Trecho 1, no Estado de Santa Catarina.	29
Figura 15.- Regressão entre os valores de AUC dos modelos de teste e do AUC dos modelos de treino.	37
Figura 16.- Regressão entre o valor do AUC de treino e o número de registros da respectiva espécie.	37
Figura 17.- Regressão entre o risco de atropelamento ponderado (Risco) e os hotspots somados e ponderados.	42
Figura 18.- Modelagem de risco de atropelamento de fauna silvestre para as estradas estaduais do Estado de Santa Catarina considerando os sete níveis de risco.	44
Figura 19.- Trechos da SC-435 que tangenciam o Parque Estadual da Serra do Tabuleiro e que possuem nível de risco 6 para o atropelamento de fauna silvestre.	46

Figura 20.- Imagem aérea e de uma seção da SC-414, em Penha, onde se identificou o único trecho de estrada com nível 5 de risco em área urbana.	47
Figura 21.- Variação do número de trechos em relação a altitude segundo o valor do índice de risco.....	48
Figura 22.- Área avaliada da rodovia SC-477 utilizando o modelo de risco de atropelamento de fauna silvestre atropelada, com indicação dos quatro trechos identificados como de maior prioridade para ações de mitigação (nível 6).	55

Lista de tabelas

Tabela 1.- Espécies e seus respectivos números de registros de atropelamento considerando o Bioma Mata Atlântica e o estado de Santa Catarina. Na coluna de código, substitui-se o "x" para o número da espécie, por exemplo sp_1 se refere à <i>Didelphis aurita</i> . CR=Criticamente em Perigo; DD=Dados insuficientes; EN=Em perigo; EW= Extinta na natureza; EX=Extinta; LC=Pouco preocupante; NE=Não avaliada; NT=Quase ameaçada; VU=Vulnerável.....	12
Tabela 2.- Layers de tipologia de rodovias e suas respectivas fontes de obtenção.....	14
Tabela 3.- Descrição das variáveis ambientais para os modelos bio-ecológicos.	15
Tabela 4.- Arquivos utilizados para definir os limites geográficos de processamento e suas respectivas fontes.....	16
Tabela 5.- Espécies e seus respectivos números de registros de atropelamento, após processo de rarefação simples, considerando o Bioma Mata Atlântica. Na coluna de código, substitui-se o "x" para o número da espécie, por exemplo sp_1 se refere à <i>Didelphis aurita</i>	19
Tabela 6.- Variáveis contínuas analisadas pela ferramenta 'Remove Highly Correlated Variables' e variáveis remanescentes após remoção das camadas com correlação superior a 70%.....	21
Tabela 7.- Parâmetros para análise dos resultados do modelo de cada uma das 30 espécies modeladas	26
Tabela 8.- Parâmetros obtidos dos modelos bio-ecológicos de atropelamento de cada uma das 30 espécies previamente definidas. (Continua)	35
Tabela 9.- Variáveis com percentual de contribuição ponderada superior a 70% para os modelos de risco de atropelamento das 30 espécies previamente definidas para o Estado de Santa Catarina.	39
Tabela 10.- Número e percentual de quilômetros com áreas de agregação para as 15 espécies analisadas para o Estado de Santa Catarina.....	40
Tabela 11.- Resultados de acurácias médias e desvio padrão na classificação.	41
Tabela 12.- Média (%) das acurácias e os <i>p</i> -valores dos conjuntos de resultados para o algoritmo kNN, baseados nas análises de Shapiro Wilk.	41
Tabela 13.- Resultados das médias de <i>mean absolute error</i> e desvio padrão na regressão.....	42
Tabela 14.- Número de trechos e suas respectivas extensões em quilômetros e percentuais para cada nível de risco.....	45
Tabela 15.- Distribuição dos trechos de risco de atropelamento em cada Coordenadoria Regional e considerando os diferentes níveis de risco de atropelamento de fauna silvestre.....	45
Tabela 16.- número de trechos, em seus respectivos níveis de risco, associados a unidades de conservação de diferentes categorias e esferas administrativas.	46
Tabela 17.- Número de trechos de diferentes níveis de risco, segundo a sua distribuição em áreas urbanas e não urbanas.	47
Tabela 18.- Variação da altitude dos trechos de risco segundo o valor do índice de risco.	47
Tabela 19.- Segmentos da rodovia SC-477 com nível 6 de risco de atropelamento de fauna silvestres e as espécies modeladas que obtiveram pesos de ocorrência superiores a 70.....	54
Tabela 20.- Comparação da extensão definida pelo MARA para cada nível de risco na SC-477 e o número de passagens de fauna instaladas em virtude dos estudos de licenciamento ambiental.	56

SUMÁRIO

SUMÁRIO EXECUTIVO	5
1 OBJETIVOS	7
2 MATERIAL E MÉTODOS	8
2.1 DADOS BRUTOS	10
2.1.1 DADOS DE ATROPELAMENTO DE FAUNA	10
2.1.2 DADOS DE RODOVIAS	13
2.1.3 DADOS DE VARIÁVEIS BIO-ECOLÓGICAS	14
2.2 ANÁLISES I - MODELAGEM ESPÉCIE A ESPÉCIE	15
2.2.1 MODELOS BIO-ECOLÓGICOS	15
2.2.2 ANÁLISES DE AGREGAÇÃO (HOTSPOTS)	28
2.3 BASES DE DADOS	29
2.4 ANÁLISES II - MACHINE LEARNING	30
2.4.1 PREPARAÇÃO DA BASE DE DADOS DE MODELO DE RISCO	30
2.4.2 TÉCNICAS DE FEATURE SELECTION	31
2.4.3 ALGORITMOS DE MACHINE LEARNING	32
2.4.4 DEFINIÇÃO DO MODELO DE EXTRAPOLAÇÃO	33
3 RESULTADOS	34
3.1 MODELOS BIO-ECOLÓGICOS	34
3.1 ANÁLISES DE AGREGAÇÃO	40
3.2 MACHINE LEARNING	41
3.2.1 ANÁLISE DE CLASSIFICAÇÃO	41
3.2.2 ANÁLISE DE REGRESSÃO	42
3.3 AJUSTES DOS MODELOS	42
3.4 MODELOS DE RISCO DE ATROPELAMENTO	43
3.5 OBSERVAÇÕES IMPORTANTES	48
3.6 USOS POTENCIAIS DO MARA	49
3.6.1 APERFEIÇOAR TERMOS DE REFERÊNCIA	50
3.6.2 APRIMORAR A TOMADA DE DECISÃO DE AÇÕES DE MITIGAÇÃO	50
3.6.3 PROGRAMAS DE CONCESSÕES DE RODOVIAS	51
3.6.4 PROTEÇÃO À VIDA	51
4 ESTUDO DE CASO	52
4.1 LICENCIAMENTO AMBIENTAL PRÉVIO	56
REFERÊNCIAS	58
ANEXOS	59

Sumário Executivo

O presente relatório se refere ao produto denominado de Modelagem de Áreas de Risco de Atropelamento (MARA), que integra o PROMIRF. O MARA é um produto tecnológico e inovador, que reúne conhecimento científico baseado em conceitos ecológicos, de engenharia e inteligência artificial para disponibilizar uma ferramenta de gestão territorial focada na proteção da vida, humana e animal, nas estradas do Estado de Santa Catarina.

A modelagem preditiva de áreas de risco se baseia no conceito ecológico de nicho. Parte do pressuposto que é possível se identificar condições (bio-ecológicas e/ou de tipologia de rodovias) que influenciam a ocorrência do atropelamento. Se esta premissa é confirmada, quanto mais dados de atropelamento georreferenciados utilizamos, melhores serão os modelos preditivos.

A modelagem do risco de atropelamento da fauna silvestre nas rodovias, está associada a dois fatores: [1] o potencial de ocorrência de uma determinada espécie na região em que se encontra o segmento de rodovia e [2] o grau de susceptibilidade do segmento de rodovia em questão ao risco de atropelamento.

O objetivo do MARA foi modelar a probabilidade de risco de atropelamento de fauna silvestre para a malha viária estadual de Santa Catarina. Como resultado principal, o MARA disponibiliza, quilômetro a quilômetro, um índice de risco de atropelamento que varia de 0 a 6, sendo 6 o maior risco.

A modelagem final foi capaz de definir riscos para 97,8% da malha estadual, com assertividades de até 86%. Se constatou que a região costeira esteve presente em todos os níveis de risco, provavelmente devido sua densidade populacional e viária, assim como seu elevado fluxo de veículos e sazonalidade das atividades econômicas. Também se constatou que os maiores níveis de risco de atropelamento se localizaram nas áreas de elevada altitude. Existe um forte desbalanceamento das áreas de risco conforme a coordenadoria regional considerada, sendo a Coordenadoria Regional do Meio Oeste a que possui mais trechos com níveis de risco 5 e 6.

As unidades de conservação do Estado de Santa Catarina, independente da esfera de administração, são pouco impactadas por estradas, com exceção da APA da Dona Francisca e um pequeno trecho do Parque Estadual da Serra do Tabuleiro. Importante enfatizar que as análises se restringiram às rodovias estaduais.

O MARA pode ser utilizado em diferentes fases do ciclo de vida de uma rodovia e propicia resultados que geram ganhos ambientais, sociais e econômicos para toda a sociedade do Estado

de Santa Catarina. Ao final deste documento são apresentados comentários e sugestões de uso aos modelos de risco de atropelamento de fauna silvestre.

1 Objetivos

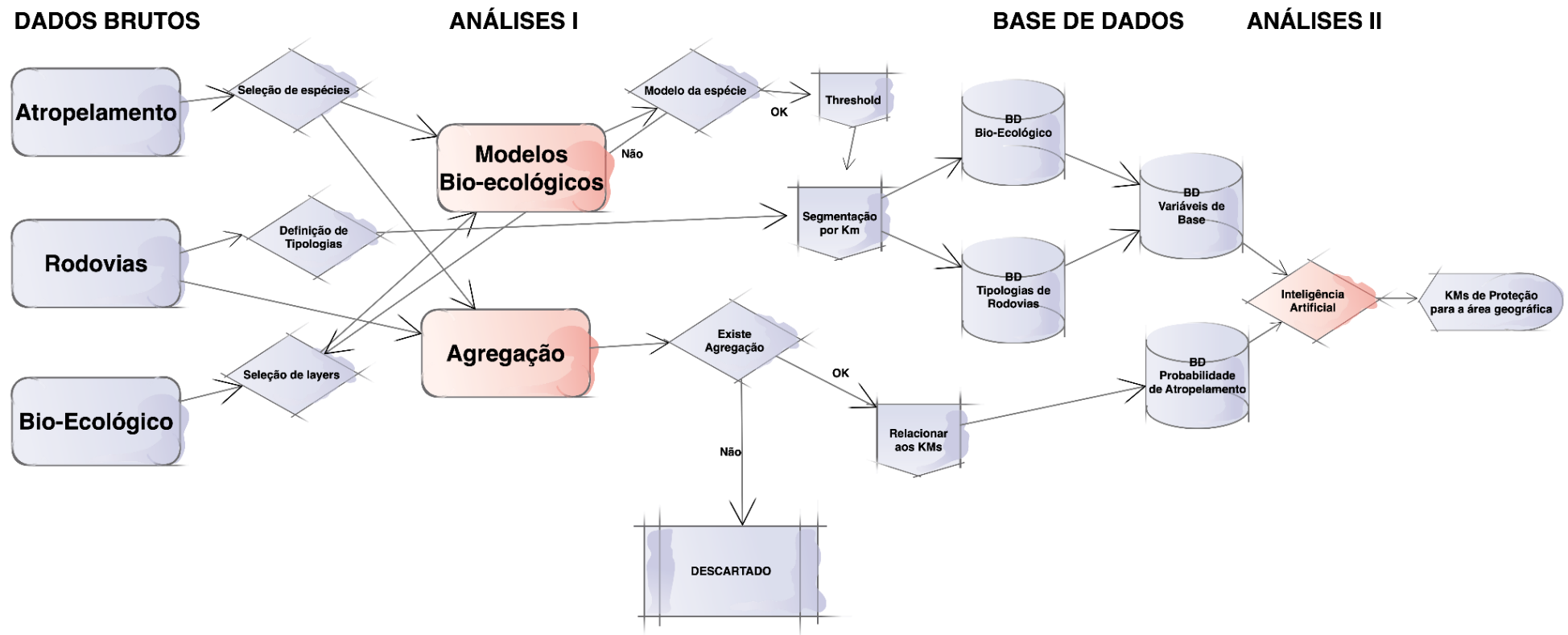
A Modelagem de Áreas de Risco de Atropelamento (MARA) foi desenvolvida com o objetivo principal de modelar a malha rodoviária estadual do Estado de Santa Catarina e identificar áreas de risco de atropelamento, tendo como base um conjunto de 30 espécies de fauna silvestre, sobretudo de mamíferos de médio e grande porte. Como resultado foi obtido um *ranking* de risco variando entre zero e seis, onde seis são os trechos com maior probabilidade de atropelamento. Toda a malha viária foi segmentada, quilômetro a quilômetro, tendo cada trecho seu respectivo valor de risco.

2 Material e métodos

Os procedimentos analíticos necessários à obtenção do modelo de risco de atropelamento de fauna silvestre de Santa Catarina podem ser observados na Figura 1. Essencialmente, os passos se dividem em:

- Obtenção e organização dos dados brutos: reunindo informações de (1) fauna silvestre atropelada; (2) malha viária e suas tipologias; (3) camadas bio-ecológicas;
 - Como passos intermediários, são definidas espécies para serem avaliadas; tipologias que serão utilizadas e seleção de camadas ambientais;
- Análises I (bio-ecológicas e de agregação): nos modelos bio-ecológicos são avaliadas, espécie a espécie, as variáveis ambientais com maior peso para a definição de suas áreas de risco de atropelamento. Nos modelos de agregação são identificados trechos de rodovias onde, sabidamente, existem agregações de atropelamento para uma ou mais espécies;
- Bases de dados: base de dados dos modelos bio-ecológicos de cada espécie, gerado na fase Análises I e as informações de diferentes tipologias de rodovias, são segmentadas, quilômetro a quilômetro. Da mesma forma, os pontos de agregação de atropelamento são alocados na malha viária segmentada.
- Análises II: Neste momento são estabelecidos dois grupos de dados, aqueles que possuem análise de agregação e o restante dos dados. Os trechos que possuem análise de agregação são analisados utilizando-se métodos de *machine learning* para se identificar quais as variáveis que se relacionam às áreas de agregação e seus respectivos pesos. Após a conclusão desta fase, o resultado é extrapolado para toda a malha viária do Estado, gerando o modelo de risco (KMs de Proteção).

Figura 1.- Diagrama do procedimento analítico empregado na obtenção do modelo de risco de atropelamento de fauna silvestre no Estado de Santa Catarina.



Cada um destes processos será detalhado nos tópicos abaixo.

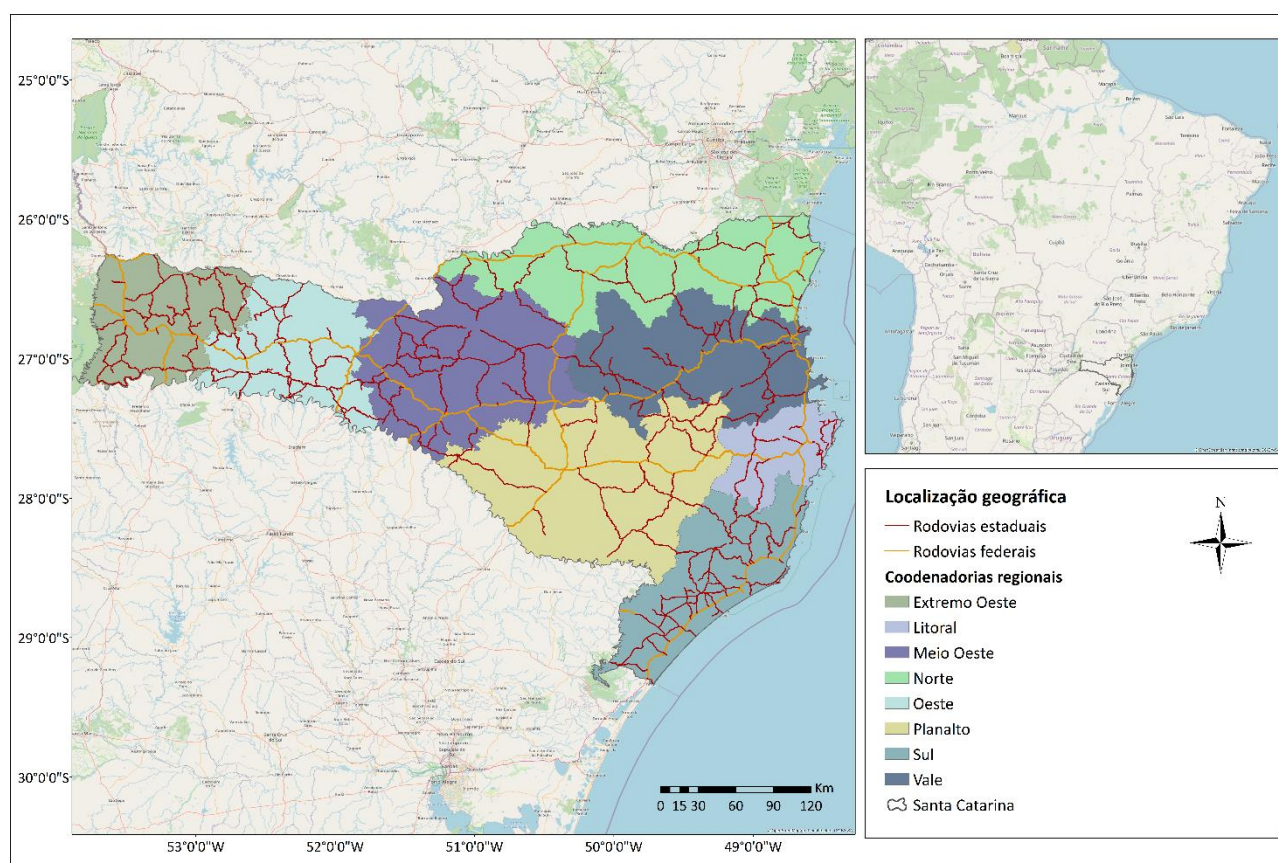
2.1 Dados Brutos

O desenvolvimento do modelo de risco requer o uso de três bases de dados distintas (Atropelamento; Tipologia de rodovias e Variáveis ambientais).

A última fase da definição dos modelos de risco utilizou uma série de métodos analíticos baseados nos bancos de dados de tipologia de rodovia (13 variáveis), distribuição bio-ecológica de atropelamento das 30 espécies previamente selecionadas (30 variáveis) e os pontos de hotspots conhecidos para a malha viária do Estado de Santa Catarina (15 variáveis).

Apesar de se buscar áreas de risco para fauna na malha viária estadual de Santa Catarina, as análises iniciais incluíram a malha viária federal, a qual posteriormente foi desconsiderada. Desta forma, as primeiras análises consideraram 8.666 trechos e 8.277,9 quilômetros de estradas (Figura 2).

Figura 2.- Malha viária estadual e federal do Estado de Santa Catarina.



2.1.1 Dados de atropelamento de fauna

O banco de dados utilizado para o desenvolvimento do modelo de risco de atropelamento de fauna silvestre nas rodovias estaduais de Santa Catarina foi constituído por 30 espécies, definidas no início do contrato pela equipe da EnvironBIT, com auxílio da SIE e do IMA (Tabela 1). A seleção destas espécies levou em consideração principalmente três fatores: (i)

número de registros disponíveis da espécie, para o bioma Mata Atlântica, que deveria ser superior a 30, para garantir bons resultados nas modelagens bio-ecológicas; (ii) a espécie possuir distribuição conhecida no Estado de Santa Catarina, porque apesar de o processo de modelagem ser realizado a nível de bioma (descrito no tópico Configuração geográfica), o resultado final do estudo se destina ao Estado de Santa Catarina e; (iii) ser uma espécie-foco, de acordo com a definição do IMA, a exemplo de espécies conhecidamente muito atropeladas, ameaçadas, que afetam diretamente a segurança viária, dentre outras.

Na Tabela 1, é apresentado o número de registros das espécies-foco do estudo, para o bioma Mata Atlântica e Estado de Santa Catarina. Para cada etapa do estudo se utilizou de recortes específicos desta base de dados de atropelamento de fauna:

- Geração de modelos bio-ecológicos: Para esta etapa considerou-se que o Estado de Santa Catarina está totalmente inserido no Bioma Mata Atlântica, sendo portanto, este bioma a base para a modelagem bio-ecológica de cada espécie, descritas no tópico Análises I - Modelagem espécie a espécie. Utilizou-se para esta etapa a base de dados completa, de 35.863 registros de fauna, para o bioma em questão (Tabela 1).
- Análises de hotspots: A análise de hotspots requer que existam monitoramentos sistemáticos de fauna atropelada e/ou um grande volume de dados eventuais em um trecho definido de rodovia. Para esta etapa, foram selecionados e utilizados registros das espécies-foco em trechos de rodovias pré-definidas do Estado de Santa Catarina. A base utilizada contou com 6.103 registros.

Tabela 1.- Espécies e seus respectivos números de registros de atropelamento considerando o Bioma Mata Atlântica e o estado de Santa Catarina. Na coluna de código, substitui-se o "x" para o número da espécie, por exemplo sp_1 se refere à *Didelphis aurita*. CR=Criticamente em Perigo; DD=Dados insuficientes; EN=Em perigo; EW= Extinta na natureza; EX=Extinta; LC=Pouco preocupante; NE=Não avaliada; NT=Quase ameaçada; VU=Vulnerável.

Código (sp_x)	Espécie	Nome Vulgar	Status de conservação (MMA, 2022)	Status de conservação (IMA SC, 2011)	Status de conservação (IUCN)	Registros Mata Atlântica	Registros SC
1	<i>Didelphis aurita</i>	Gambá-de-orelha-preta	LC	LC	LC	5727	673
2	<i>Dasypus novemcinctus</i>	Tatu-galinha	LC	LC	LC	5562	1202
3	<i>Didelphis albiventris</i>	Gambá-de-orelha-branca	LC	LC	LC	4461	2215
4	<i>Hydrochoerus hydrochaeris</i>	Capivara	LC	LC	LC	3724	694
5	<i>Cerdocyon thous</i>	Cachorro-do-mato	LC	LC	LC	3453	970
6	<i>Tamandua tetradactyla</i>	Tamanduá-mirim	LC	LC	LC	2561	616
7	<i>Euphractus sexcinctus</i>	Tatu-peba	LC	LC	LC	1768	141
8	<i>Salvator merianae</i>	Teiú-gigante	LC	LC	LC	1396	533
9	<i>Procyon cancrivorus</i>	Mão-pelada	LC	LC	LC	1183	427
10	<i>Alouatta guariba</i>	Bugio-ruivo	Sub-espécies: <i>clamitans</i> (VU) <i>guariba</i> (CR)*	Sub-espécie <i>clamitans</i> (VU)	VU	757	102
11	<i>Galictis cuja</i>	Furão-pequeno	LC	LC	LC	558	177
12	<i>Coendou spinosus</i>	Ouriço-cacheiro	LC	LC	LC	528	44
13	<i>Cabassous tatouay</i>	Tatu-de-rabo-mole-grande	LC	LC	LC	444	128
14	<i>Sphiggurus villosus</i>	Ouriço-cacheiro	LC	LC	LC	343	116
15	<i>Sapajus nigritus</i>	Macaco-prego	LC	LC	NT	335	36
16	<i>Leopardus pardalis</i>	Jaguarundi	LC	EN	LC	317	30
17	<i>Herpailurus yagouaroundi</i>	Jaguarundi	VU	LC	LC	302	90
18	<i>Dasypus septemcinctus</i>	Tatuí	LC	LC	LC	254	42
19	<i>Puma concolor</i>	Onça-parda	VU	VU	LC	247	27
20	<i>Leopardus guttulus</i>	Gato-do-mato-pequeno	VU	LC	VU	242	99
21	<i>Mazama gouazoubira</i>	Veado-catingueiro	LC	LC	LC	230	15
22	<i>Conepatus chinga</i>	Zorrilho	LC	LC	LC	220	138
23	<i>Eira barbara</i>	Irara	LC	LC	LC	214	47
24	<i>Lontra longicaudis</i>	Lontra	LC	LC	NT	208	61
25	<i>Cuniculus paca</i>	Paca	LC	VU	LC	206	12
26	<i>Leopardus tigrinus</i>	Gato-do-mato	EN	LC	VU	176	44
27	<i>Tapirus terrestris</i>	Anta	VU	EN	VU	134	2
28	<i>Leopardus wiedii</i>	Gato-maracajá	VU	LC	NT	128	41
29	<i>Pecari tajacu</i>	Cateto	LC	VU	LC	113	8
30	<i>Mazama americana</i>	Veado-mateiro	LC	EM	DD	72	3

2.1.2 Dados de rodovias

Conforme citado anteriormente, parte do processamento foi realizado considerando a malha viária federal e estadual de Santa Catarina.

Para geração das informações de caracterização e tipologia de rodovias, foram utilizadas algumas bases, especificadas abaixo:

- Duas bases principais (Federal e Estadual):
 - Para as Rodovias Estaduais, se utilizou os dados da SIE (Rodovias estaduais SC).
 - Para as Rodovias Federais, se utilizou os dados do SNV (DNIT) do ano de 2021:
<https://www.gov.br/dnit/pt-br/assuntos/atlas-e-mapas/pnv-e-snv>
- Para o trabalho com as elevações e cálculos de sinuosidade, foi utilizado um Modelo Digital de Elevação do Brasil do ano de 2011, disponível em:
<http://www.dsr.inpe.br/topodata/acesso.php>
- Para o VMDa, foram utilizadas três bases distintas:
 - SHP_GEMAM (SIE)
 - Modelagem 2020 SNV. Disponível em:
<http://servicos.dnit.gov.br/dadospnct/Modelagem>
 - PNCT

Algumas das informações foram apenas extraídas das bases supracitadas, outras por sua vez, foram calculadas pela equipe da EnvironBIT, como foi o caso da Sinuosidade Vertical, Sinuosidade Horizontal e Ângulo Vertical. Na Tabela 2 são listadas as informações de tipologia de rodovia utilizadas neste estudo e suas respectivas fontes.

Importante frisar que, a malha viária estadual utilizada foi aquela enviada no início de 2022, logo após a contratação do projeto. Não foi possível utilizar as informações modificadas em datas posteriores devido às análises já estarem em andamento, inviabilizando a substituição das mesmas.

Tabela 2.- Layers de tipologia de rodovias e suas respectivas fontes de obtenção.

Dado	Fonte
Altitude final (m)	MDE
Sinuosidade vertical - Quanto menor e mais próximo de 1, mais plano Fórmula: $\frac{\text{Distância na via do segmento}}{\text{Distância plana do segmento}}$	Calculado
Ângulo vertical (em graus)	Calculado
Sinuosidade horizontal - Quanto maior e mais próximo de 1, mais reto Fórmula: $\frac{((XA-XB)^2+(YA-YB)^2+((ZA-ZB)^2))}{\text{Distância AB medida na via}}$	Calculado
Material do pavimento CA - Concreto Asfáltico CC - Concreto Cimento LN - Leito Natural PF - Pré-Misturado a Frio RP - Revestimento Poliédrico TS - Tratamento Superficial	Bases Estadual e Federal
Estado da pavimentação DUP - Duplicada EOD - Em Obras de Duplicação EOP - Em Obras de Pavimentação IMP - Implantada LEN - Leito Natural PAV - Pavimentada TRA - Travessia	Bases Estadual e Federal
Esfera de gestão (Estadual ou Federal)	Bases Estadual e Federal
Número de pistas (faixas)	Bases Estadual e Federal
Volume médio diário	SHP_GEMAM, Modelagem SNV 2020 e PNCT
Índice de compatibilidade: - muito baixa compatibilidade - baixa compatibilidade - alta compatibilidade - muito alta compatibilidade	PRIM IVT, compatibilidade de rodovias

2.1.3 Dados de variáveis bio-ecológicas

As camadas de entrada do modelo são todas oficiais e públicas. A Tabela 3 traz, de forma organizada, as 21 variáveis bio-ecológicas e suas respectivas fontes. A lista inicial foi, posteriormente, reduzida para 14 camadas (5 categóricas e 9 contínuas) em decorrência da avaliação de informações redundantes. Todos os procedimentos para a seleção das variáveis finais podem ser vistos no tópico Correlação entre as variáveis.

Tabela 3.- Descrição das variáveis ambientais para os modelos bio-ecológicos.

Variável ambiental	Descrição	Fonte	Tipo
Vegetação	-	IBGE	Categórica
Relevo	-		
Uso da terra	Uso e cobertura do solo		
Densidade de drenagem	-	INPE	Contínua
Cobertura arbórea (%)	-		Contínua
Solo	-		Categórica
Altitude	-		Contínua
Declividade	-		Contínua
Clima	Climatologia de Koppen	IPEF	Categórica
Bio01	Temperatura média anual	WorldClim	Contínua
Bio02	Faixa Média Diurna (Média da média mensal (temperatura máxima - temperatura mínima))		
Bio03	Isotermalidade (BIO2/BIO7) (* 100)		
Bio04	Temperatura Sazonalidade (DP *100)		
Bio05	Temperatura Máxima do Mês mais quente		
Bio06	Temperatura Mínima do Mês mais Frio		
Bio07	Faixa Anual de Temperatura (BIO5-BIO6)		
Bio12	Precipitação anual		
Bio13	Precipitação do mês mais húmido		
Bio14	Precipitação do Mês Mais Seco		
Bio15	Sazonalidade de Precipitação (Coeficiente de variação)		
Sensibilidade PRIM IVT	Índice de sensibilidade (PRIM IVT) - Quanto maior, maior sensibilidade	PRIM-IVT	Contínua

2.2 Análises I - Modelagem espécie a espécie

2.2.1 Modelos Bio-Ecológicos

Os modelos bio-ecológicos são análises realizadas, espécie a espécie, onde o objetivo é identificar quais as variáveis ambientais são mais relevantes (maior peso) na ocorrência do atropelamento. A partir da definição de variáveis e suas respectivas contribuições, é possível a extrapolação para áreas potenciais em uma dada área geográfica previamente definida.

Modelos bio-ecológicos, normalmente denominados de modelos de distribuição potencial, são amplamente utilizados para avaliar áreas de vida de fauna e flora, risco de

introdução de espécies exóticas, evolução de efeitos climáticos na paisagem, entre outros. As modelagens realizadas para a obtenção do risco de atropelamento se diferem dos modelos tradicionais por utilizar, predominantemente, registros de fauna silvestre atropelada. Desta forma, as variáveis identificadas para comporem os modelos podem se diferenciar significativamente daquelas que definem a típica área de vida de uma dada espécie. A título de exemplo, uma espécie que tipicamente habita um ambiente florestal, pode sofrer o efeito do atropelamento em áreas mais abertas, de campo ou pasto.

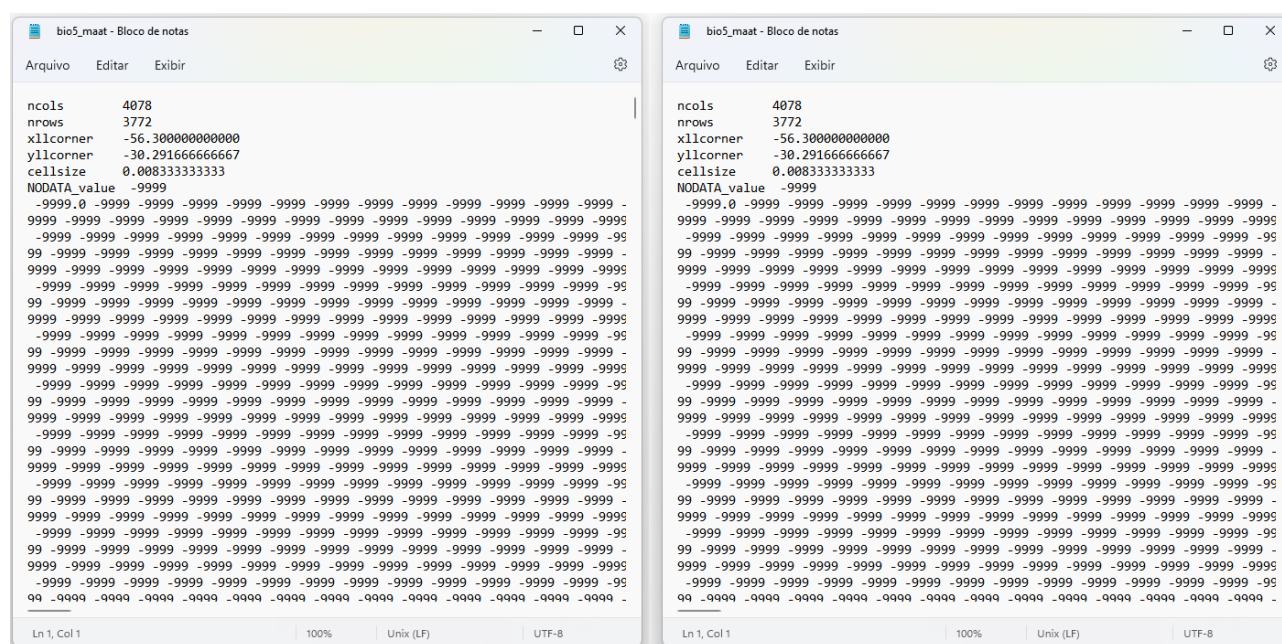
2.2.1.1 Configuração geográfica

A área de aplicação deste estudo compreende todo o território do Estado de Santa Catarina, o qual é composto pelo bioma Mata Atlântica (Tabela 4). Contudo, alguns pesquisadores sugerem o uso de dados que transcendam os limites geográficos da área de aplicação, de forma a permitir que o modelo seja sensível às características biofísicas locais das ocorrências das espécies. Isso permite que tais características sejam inseridas e ponderadas no processo de aprendizagem de máquina utilizado no modelo, que este possa analisar no contexto da área de estudo e a existência dessas condicionantes para calcular os mapas de potencial de atropelamento das espécies. Como máscara de recorte dos dados, variáveis ambientais e das características e tipologias das estradas fez-se uso do bioma Mata Atlântica, com extrapolação considerando um buffer de 20 km (Figura 3).

Tabela 4.- Arquivos utilizados para definir os limites geográficos de processamento e suas respectivas fontes.

Dado / arquivo	Formato	Fonte	Url
Bioma Mata Atlântica	Vetorial (polígono)	IBGE	https://www.ibge.gov.br/geociencias/downloads-geociencias.html
Limite Geográfico SC	Vetorial (polígono)		

Figura 4.- Variáveis ambientais em formato ASCII, após padronização. No exemplo abaixo são disponibilizadas as variáveis Bio05 e Solo.



O geoprocessamento dos dados de entrada gerou 21 arquivos em formato matricial, correspondentes às variáveis geográficas de entrada no modelo preditivo. Cabe ressaltar que os dados resultantes são devidamente armazenados em pastas para o histórico das operações, possibilitando futuras análises pontuais e reprocessamentos, se necessário.

○ *Dados redundantes*

Para a etapa de análise e eliminação dos dados redundantes foi utilizado o método de Rarefação simples, pela ferramenta SDM Toolbox (BROWN, 2017) - "Spatially rarefy occurrence data for SDM (Reduce Spatial Autocorrelation)". Este método é necessário para redução das amostras espacialmente auto correlacionadas.

Entende-se que amostras geograficamente próximas podem criar significativo viés na calibração dos parâmetros (alimentação) dos algoritmos de aprendizagem de máquina, posto que o mesmo contexto geográfico será super exposto, o que elevará seu respectivo peso de participação no processo. A rarefação consiste na remoção de amostras espacialmente auto correlacionadas, processo que na prática reduz múltiplas amostras espacialmente próximas para um único registro. A metodologia utiliza conceitos de geoestatística, mais especificamente o índice Moran-I para calcular a autocorrelação espacial com base na localização do fenômeno.

Segundo Veloz (2009), o ideal de um estudo, seria o uso de dados que tivessem sido coletados aleatoriamente na área de estudo, no entanto esta não é a realidade da maioria dos conjuntos de dados usados para modelagem. Ao fazer o método de rarefação, são removidas agregações de dados que poderiam influenciar para maiores valores de AUC.

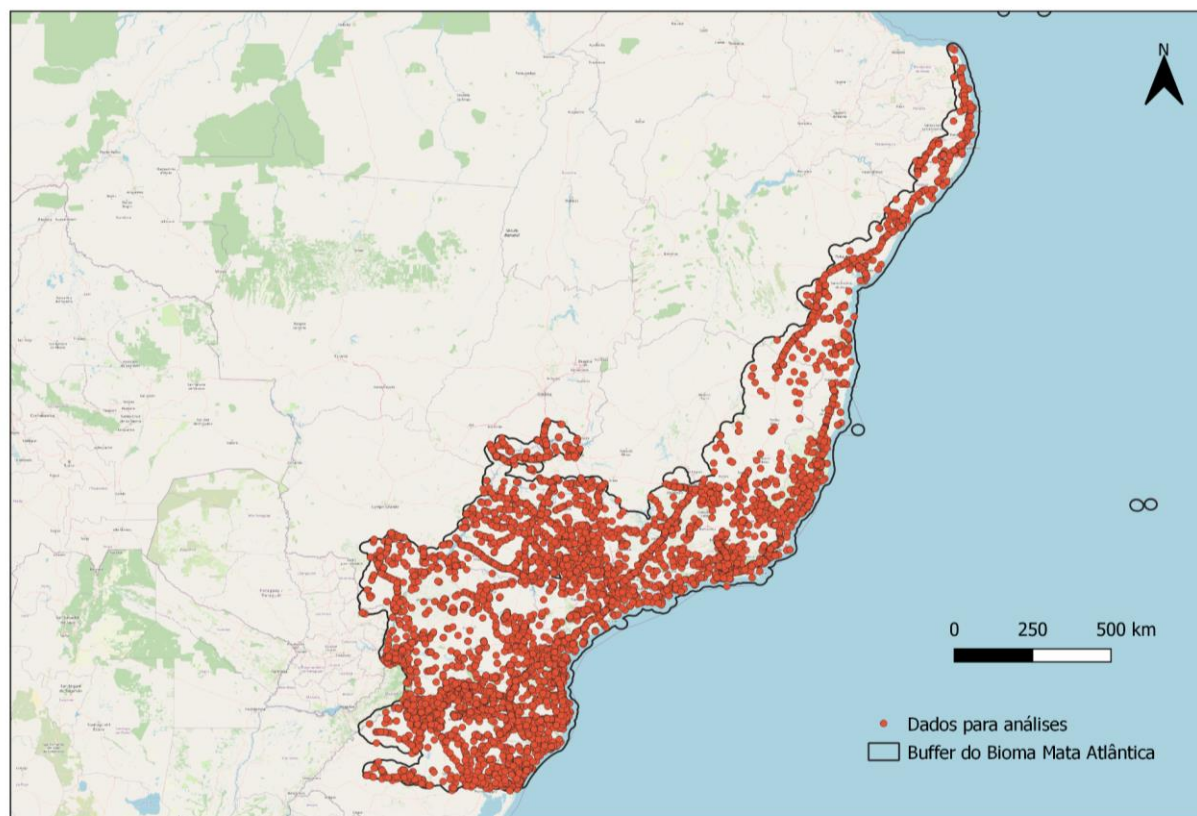
As análises realizadas no presente projeto empregaram a eliminação de duplicidade de dados e considerou um intervalo de 5km para a remoção de dados auto correlacionados. A definição do raio de 5km considerou a biologia e ecologia das espécies analisadas.

Importante ressaltar que, apenas os dados que não se tratavam de atropelamento passaram pelo processo de rarefação simples. Desta forma, os dados para modelagem compreenderam a totalidade dos dados de atropelamentos somados aos dados de ocorrência, após rarefação. A Tabela 5 e a Figura 5 apresentam os registros, após remoção dos dados redundantes.

Tabela 5.- Espécies e seus respectivos números de registros de atropelamento, após processo de rarefação simples, considerando o Bioma Mata Atlântica. Na coluna de código, substitui-se o "x" para o número da espécie, por exemplo sp_1 se refere à *Didelphis aurita*.

Código (sp_x)	Espécie	Registros da Mata Atlântica após rarefação
1	<i>Didelphis aurita</i>	2831
2	<i>Dasypus novemcinctus</i>	2314
3	<i>Didelphis albiventris</i>	2390
4	<i>Hydrochoerus hydrochaeris</i>	2262
5	<i>Cerdocyon thous</i>	2236
6	<i>Tamandua tetradactyla</i>	1348
7	<i>Euphractus sexcinctus</i>	948
8	<i>Salvator merianae</i>	868
9	<i>Procyon cancrivorus</i>	711
10	<i>Alouatta guariba</i>	504
11	<i>Galictis cuja</i>	404
12	<i>Coendou spinosus</i>	322
13	<i>Cabassous tatouay</i>	289
14	<i>Sphiggurus villosus</i>	65
15	<i>Sapajus nigritus</i>	261
16	<i>Leopardus pardalis</i>	227
17	<i>Herpailurus yagouaroundi</i>	287
18	<i>Dasypus septemcinctus</i>	195
19	<i>Puma concolor</i>	201
20	<i>Leopardus guttulus</i>	231
21	<i>Mazama gouazoubira</i>	130
22	<i>Conepatus chinga</i>	191
23	<i>Eira barbara</i>	171
24	<i>Lontra longicaudis</i>	173
25	<i>Cuniculus paca</i>	154
26	<i>Leopardus tigrinus</i>	153
27	<i>Tapirus terrestris</i>	56
28	<i>Leopardus wiedii</i>	115
29	<i>Pecari tajacu</i>	74
30	<i>Mazama americana</i>	63

Figura 5.- Dados utilizados para modelagem bio-ecológica, após processo de rarefação simples.



o *Correlação entre as variáveis*

A matriz de covariância contém valores de variâncias e covariâncias. A variância é uma medida estatística que mostra quanta variação existe da média. Para calcular essas métricas são calculados os quadrados das diferenças entre cada valor de célula e o valor médio de todas as células. As variâncias para cada camada podem ser lidas ao longo da diagonal da matriz de covariância, movendo-se da esquerda superior para a direita inferior. As variações são expressas em unidades de valor de célula ao quadrado.

A correlação entre duas variáveis é uma medida de dependência entre elas. Matematicamente, corresponde a razão da covariância entre as duas variáveis dividida pelo produto de seus desvios padrão. Por ser uma proporção, é um número sem unidade, variando entre -1 e +1. A correlação positiva indica uma relação direta entre duas variáveis, como quando os valores das células de uma camada aumentam os valores das células de outra camada também aumentam. Uma correlação negativa significa que uma variável muda inversamente à outra. Uma correlação de zero significa que duas variáveis são independentes.

O uso de variáveis com grande semelhança aumenta a generalidade, a instabilidade do modelo e os erros (WARREN et al., 2014; CRUZ-CÁRDENAS et al., 2014). Para remover as variáveis contínuas com alta correlação, foi utilizada a ferramenta "Remove Highly Correlated

Variables” - SDM Toolbox (BROWN, 2017). Esta ferramenta analisa as variáveis ambientais selecionadas, para cada bioma, e retorna um pool de variáveis adequadas para modelagem. Foram desconsideradas variáveis ambientais com autocorrelação igual ou superior a 70% (DALAPICOLLA, 2016). A Tabela 6 mostra o resultado da análise de correlação das variáveis contínuas. As variáveis remanescentes, juntamente com as variáveis categóricas (Tabela 3) compuseram o processo de modelagem no MaxEnt.

Tabela 6.- Variáveis contínuas analisadas pela ferramenta ‘Remove Highly Correlated Variables’ e variáveis remanescentes após remoção das camadas com correlação superior a 70%.

VARIÁVEIS ANALISADAS	Densidade de drenagem	VARIÁVEIS REMANESCENTES	Densidade de drenagem
	Cobertura arbórea (%)		Cobertura arbórea (%)
	Altitude		Declividade
	Declividade		Bio03
	Bio01		Bio05
	Bio02		Bio07
	Bio03		Bio12
	Bio04		Bio13
	Bio05		Sensibilidade PRIM IVT
	Bio06		
	Bio07		
	Bio12		
	Bio13		
	Bio14		
	Bio15		
	Sensibilidade PRIM IVT		

- Calibração do modelo
 - *MaxEnt*

O MaxEnt (Figura 6) é um programa desenvolvido para calcular os nichos ecológicos para compreensão e estimativas de distribuição de espécies utilizando variáveis ambientais (GAUCHEREL et. al 2016; HALVORSEN 2013; PHILIPS et al 2008). O programa utiliza uma arquitetura analítica baseada na formulação de máxima entropia (maximização do logaritmo da máxima verossimilhança estatística). Não obstante, sua robustez permite que a ferramenta possa também ser utilizada para modelagem geográfica preditiva de fenômenos geográficos que sejam dependentes de variáveis ambientais a partir de amostras pontuais para alimentar a base de dados. Desta forma, no presente trabalho são calculadas as superfícies de potencial atropelamento das espécies baseadas em evidências dadas pelas amostras.

Figura 6.- Tela principal do MaxEnt e menus de configurações ilustrando o processo de configuração dos parâmetros para modelagem inicial da "sp_12" (*Coendou spinosus*) para o bioma Mata Atlântica. A partir desse cenário foram realizadas modificações com base nos resultados, neste caso específico foram desenvolvidos 10 ensaios (vide Figura 11).

The image displays the MaxEnt software interface. The top window is the main configuration panel, and below it are three sub-windows showing detailed settings for Basic, Advanced, and Experimental options.

Main Configuration Panel:

- Samples:** File: ICA_Projeto_SC\12-Coendou spinosus\sp_12.csv. ☒ sp_12.
- Environmental layers:** Directory/File: _MAXENT\Variaveis_ambientais\MATA_ATLANTICA\ASCII. ☒ alt_maat, ☒ bio12_maat, ☒ bio13_maat, ☐ bio14_maat, ☐ bio15_maat, ☐ bio1_maat, ☐ bio2_maat, ☒ bio3_maat, ☐ bio4_maat, ☒ bio5_maat, ☐ bio6_maat. All are set to Continuous.
- Feature Selection:** ☒ Linear features, ☒ Quadratic features, ☒ Product features, ☐ Threshold features, ☒ Hinge features, ☒ Auto features.
- Advanced Options:** ☒ Create response curves, ☒ Make pictures of predictions, ☒ Do jackknife to measure variable importance.
- Output:** Output format: Logistic, Output file type: asc.
- Directories:** Output directory: am_feita\MATA_ATLANTICA\Projeto_SC\sp_12\teste 1, Projection layers directory/file: [empty].
- Buttons:** Run, Settings, Help.

Sub-windows:

- Basic:** ☒ Random seed, ☒ Give visual warnings, ☒ Show tooltips, ☒ Ask before overwriting, ☐ Skip if output exists, ☒ Remove duplicate presence records, ☒ Write clamp grid when projecting, ☒ Do MESS analysis when projecting. Random test percentage: 30, Regularization multiplier: 1, Max number of background points: 10000, Replicates: 10, Replicated run type: Bootstrap.
- Advanced:** ☒ Add samples to background, ☐ Add all samples to background, ☒ Write plot data, ☒ Extrapolate, ☒ Do clamping, ☒ Write output grids, ☒ Write plots, ☐ Append summary results to maxentResults.csv ..., ☒ Cache ascii files. Maximum iterations: 5000, Convergence threshold: 0.00001, Adjust sample radius: 0, Log file: maxent.log, Default prevalence: 0.5, Apply threshold rule: [dropdown], Bias file: [empty].
- Experimental:** ☒ Logscale raw/cumulative pictures, ☐ Per species results, ☐ Write background predictions, ☐ Show exponent in response curves, ☐ Fade by clamping, ☐ Verbose, ☐ Use samples with some missing data. Threads: 1, Lq to lqp threshold: 80, Linear to lq threshold: 10, Hinge threshold: 15, Beta threshold: -1, Beta categorical: -1, Beta lqp: -1, Beta hinge: -1, Default nodata value: -9999.

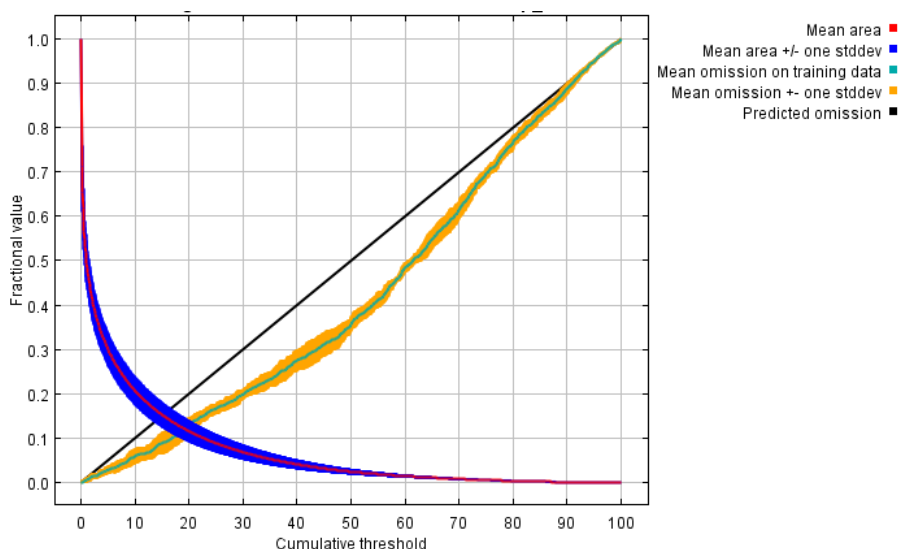
Em síntese, os modelos no MaxEnt foram desenvolvidos espécie a espécie, sendo que o número de amostras de cada espécie é o grande diferencial para a configuração do modelo, posto que o MaxEnt utiliza métodos de reamostragem que particionam as amostras em dados de treino e validação a fim de minimizar inconsistências e maximizar os resultados do modelo. Em alguns casos, o número reduzido de amostras pode impor limitações na modelagem, que por sua vez deve adotar métodos diferentes de amostragem. Embora o software opere com 3 técnicas de reamostragem (*Cross Validation*, *Bootstrap* e *Subsample*), adotou-se o *Bootstrap*, método indicado nos casos em que as espécies possuem mais de 14 amostras.

O *Bootstrap* é um método simples de reamostragem baseado na reposição das amostras no universo da população, ou seja, os dados usados no treinamento na primeira repetição retornam para o conjunto de dados para que seja feita uma nova rodada aleatória de amostragem, tanto quantas forem as repetições impostas na configuração do modelo. Por padrão, os parâmetros definidos para a modelagem corresponderam a outros estudos que se utilizaram do MaxEnt, sendo eles 10 repetições, 30% das amostras para teste e multiplicador de regularização 1.

○ *Análise dos resultados*

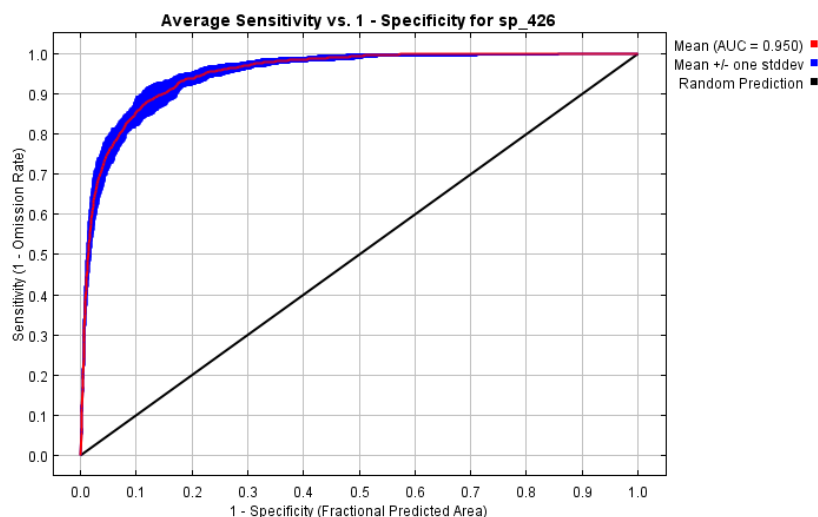
A etapa de análise dos resultados do MaxEnt é complexa e exige a observação de diversos indicadores. O gráfico de curvas ROC (*Receiver Operating Characteristics*) é uma das medidas mais utilizadas em aprendizagem de máquina para medir o desempenho de modelos de classificação. Ele equivale a uma curva de probabilidade criada traçando a taxa verdadeiro-positivo contra a taxa de falsos-positivos, ou seja, confronta graficamente os casos onde o classificador acertou a predição contra os casos onde a predição foi errada. Outra medida para avaliar o desempenho é o AUC (*Area Under The Curve*), que na prática se resume em um único índice adimensional. Quanto mais próximo o AUC estiver de 1, maior a performance do classificador em separar de forma correta as feições a partir dos dados de amostra. AUC próximo a 0,5 implica baixo desempenho do classificador (não há capacidade de separação de classes), e AUC próximo a 0 significa que o modelo está pobre e invertendo os valores de previsão. A Figura 7 ilustra o gráfico das taxas de omissão produzido pelo MaxEnt, para a qual se pode observar a aderência entre a taxa de omissão obtida com o treinamento (em verde), com desvio padrão (em amarelo) e a taxa de omissão prevista (em preto).

Figura 7.- Gráfico das taxas de omissão da classificação produzida pelo MaxEnt para a sp_12 (*Coendou spinosus*).



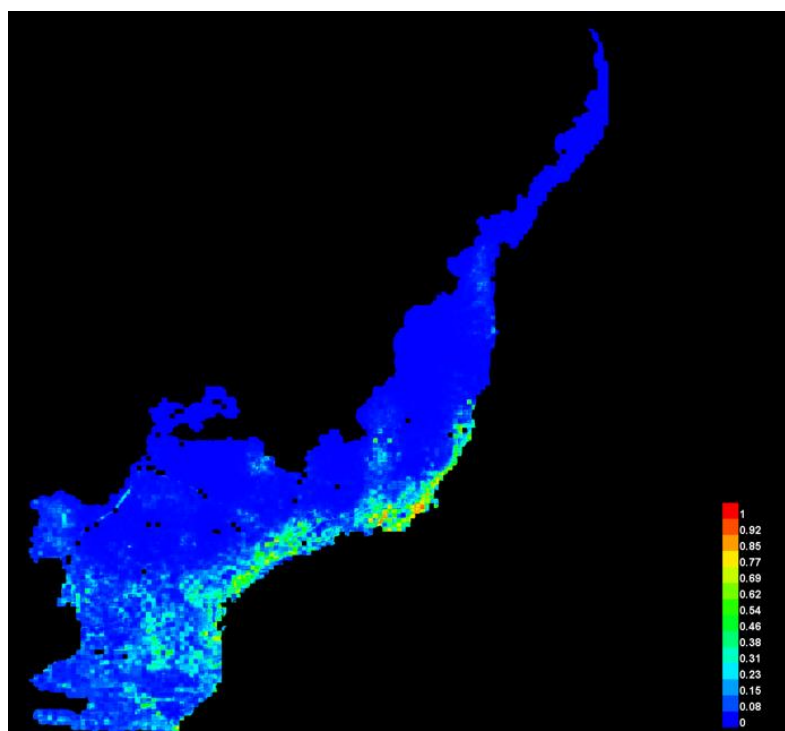
A Figura 8 ilustra a curva ROC com seus dois parâmetros: taxa de verdadeiros positivos (ou sensibilidade) e taxa de falsos positivos (especificidade ou alarme falso). Os cálculos indicam que a AUC máxima possível de teste seria 0,922 em vez de 1. Na prática, a AUC do treinamento pode exceder esse limite, e foi o que de fato ocorreu (AUC training = 0,950).

Figura 8.- Curva ROC (Receiver Operating Characteristics) produzido pelo MaxEnt para a sp_12 (*Coendou spinosus*).



Uma outra análise realizada é a inspeção visual da superfície de probabilidade resultante do modelo. O MaxEnt utiliza cores quentes para representar as áreas com melhor poder de predição, e consequentemente cores frias para representar as áreas onde a predição é deficiente. A análise visual serve para verificar a completitude espacial do resultado, posto que modelos deficientes tendem a apresentar lacunas (Figura 9).

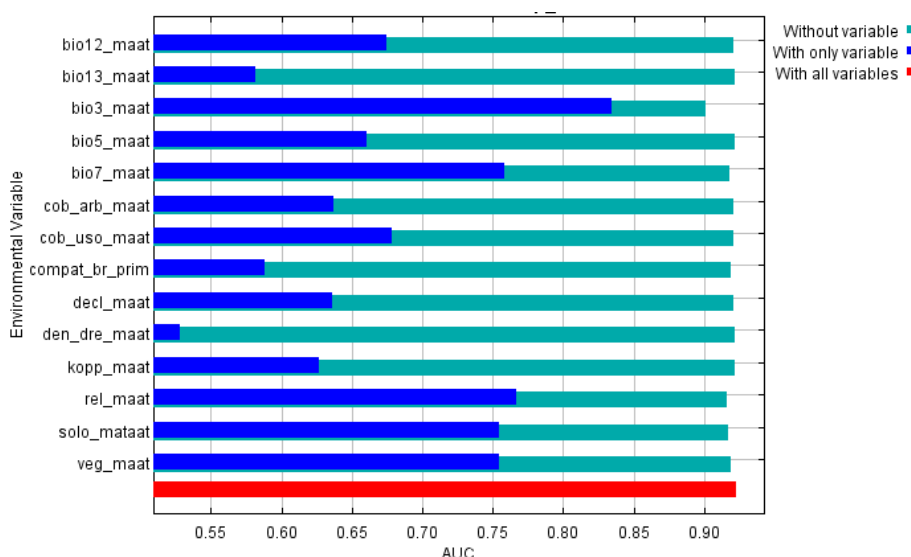
Figura 9.- Superfície de predição resultante do MaxEnt para o ensaio da sp_12 (*Coendou spinosus*).



Adicionalmente, o MaxEnt gera uma série de gráficos (curvas de predição) para cada uma das variáveis utilizadas, sejam estas de dados contínuos ou categóricos.

Por último, foi utilizado o gráfico de *Jackknife*. A ferramenta testa a importância da variável no poder explicativo do modelo de forma isolada (barras azuis), no conjunto de variáveis (barras vermelhas) e o modelo sem a variável em questão (barra verde). É esperado que o modelo com todas as variáveis tenha eficiência máxima para o caso em estudo (o que ocorreu aqui), e também é esperado que as variáveis com maior poder explicativo de forma isolada, impactem negativamente o modelo quando retiradas do conjunto de dados. Cabe observar que na Figura 10 a variável ambiental com maior ganho quando utilizada isoladamente é a Bio03, e também é a variável que mais impacta o modelo se desconsiderada do conjunto de dados.

Figura 10.- Análise *Jackknife* da importância das variáveis no modelo preditivo do MaxEnt para o ensaio da sp_12 (*Coendou spinosus*).



o Análise de priorização dos modelos

A análise de priorização dos modelos contou com um processo sistemático de verificação de métricas como o número de amostras de treinamento, número de amostras de teste, AUC (área abaixo da curva ROC) para o gráfico de treinamento e de teste, teste de omissão e de probabilidade binomial com treinamento mínimo, teste de omissão e de probabilidade binomial com treinamento a 10% e testes de sensibilidade máxima. Todos esses parâmetros foram elencados em planilhas que foram preenchidas a partir do relatório de saída do MaxEnt produzido em cada ensaio. A Tabela 7 sintetiza as principais métricas reportadas nos resultados no MaxEnt e que foram utilizadas no framework analítico.

Tabela 7.- Parâmetros para análise dos resultados do modelo de cada uma das 30 espécies modeladas

Parâmetro	Significado
<i>Training Samples</i>	Número de amostras utilizadas no modelo
<i>Test Samples</i>	Número de amostras utilizadas na validação do modelo
<i>Training AUC</i>	> 0,90 (Excelente); 0,80 – 0,90 (Bom); 0,70 – 0,80 (Aceitável); <0,70 (Não aceito).
<i>Test AUC</i>	Métrica da área abaixo da curva. > 0,90 (Excelente); 0,80 – 0,90 (Bom); 0,70 – 0,80 (Aceitável); <0,70 (Não aceito).
<i>AUC Standard Deviation</i>	Desvio padrão da AUC. Se alto pode indicar irregularidade na curva de resposta.
<i>Minimum training presence logistic threshold</i>	Valor mínimo do limiar
<i>Minimum training presence test omission</i>	Erro de omissão (falso negativo) do modelo com o limiar descrito acima. Aceitável até 0,2, sendo preferíveis valores menores que 0,15
<i>Minimum training presence binomial probability</i>	Estatística P-Valor do modelo. Descartar se superior a 5% e analisar os limiares de 10% seguintes.
<i>10 percentile training presence logistic threshold</i>	Valor do limiar com presença de 10% para treinamento do modelo
<i>10 percentile training presence test omission</i>	Erro de omissão com presença de 10% para treinamento do modelo
<i>10 percentile training presence binomial probability</i>	Estatística P-Valor do modelo com presença de 10% para treinamento do modelo
<i>Maximum test sensitivity plus specificity logistic threshold</i>	Valor máximo do limiar
<i>Maximum test sensitivity plus specificity test omission</i>	Erro de omissão do modelo para o limiar máximo
<i>Maximum test sensitivity plus specificity binomial probability</i>	Estatística P-Valor do modelo com limiar máximo

Todos os ensaios realizados para cada espécie foram transcritos para uma planilha (Figura 11), constituída por uma aba que resumizava todos os valores descritos na Tabela 7 e abas individuais para cada teste de modelo. Importante enfatizar que a prioridade foi dada ao 10 Percentil Training, o qual foi utilizado na elaboração dos mapas finais.

Esse processo analítico auxiliou na organização e entendimento do modelo, tanto em estatísticas globais como em métricas específicas que indicam o comportamento de cada uma das variáveis utilizadas no modelo. Desta forma, foi possível desenvolver progressivamente diferentes ensaios (denominados teste 1, teste 2, ...teste n) na tentativa de alcançar métricas mais consistentes. Para tanto, nos diferentes ensaios desenvolvidos foram alterados parâmetros, tais como a anulação ou substituição de variáveis, a depender do comportamento e do efeito que esta causou no modelo. Esse processo permitiu, com segurança, desenvolver o refinamento do modelo sem a preocupação quanto ao aspecto da imagem por ele produzida, posto que esta etapa contou integralmente com análises baseadas nos resultados numéricos dos dados e não no aspecto visual do mapa.

2.2.2 Análises de Agregação (hotspots)

Conhecer as áreas de agregação de atropelamento de fauna silvestre é fundamental no processo de validação dos modelos de risco. Os trechos analisados quanto à agregação são utilizados como variáveis descritoras das áreas de impacto e não impacto.

Foram analisados 1.089,78 quilômetros de rodovias no Estado de Santa Catarina, considerando dados de até 15 espécies passíveis de serem analisadas com o banco de dados existente para os trechos de estradas incluídos na análise. As espécies foram *Alouatta guariba*; *Cabassous tatouay*; *Cercopithecus thous*; *Conepatus chinga*; *Dasypus novemcinctus*; *Didelphis albiventris*; *Didelphis aurita*; *Euphractus sexcinctus*; *Galictis cuja*; *Hydrochoerus hydrochaeris*; *Leopardus spp.*; *Procyon cancrivorus*; *Herpailurus yagouaroundi*; *Sphiggurus villosus* e *Tamandua tetradactyla*. Estas espécies foram utilizadas sempre que, para um determinado trecho, existiam registros de atropelamentos suficientes para que fossem testadas a agregação das ocorrências.

Todas as análises foram realizadas através do software Siriema V.2. As análises ocorreram em duas fases:

- K de Ripley.- nesta fase foram avaliadas, espécie a espécie, e para cada trecho analisado, a existência de agregação de atropelamento e em qual raio as agregações ocorrem (Figura 13);
- Identificação de hotspots.- neste segundo momento se obteve a distribuição espacial das áreas de agregação e suas respectivas intensidades (Figura 14).

Figura 13.- Resultado da análise de K-Ripley encontrado para o *Didelphis aurita* no Trecho 1, no Estado de Santa Catarina.

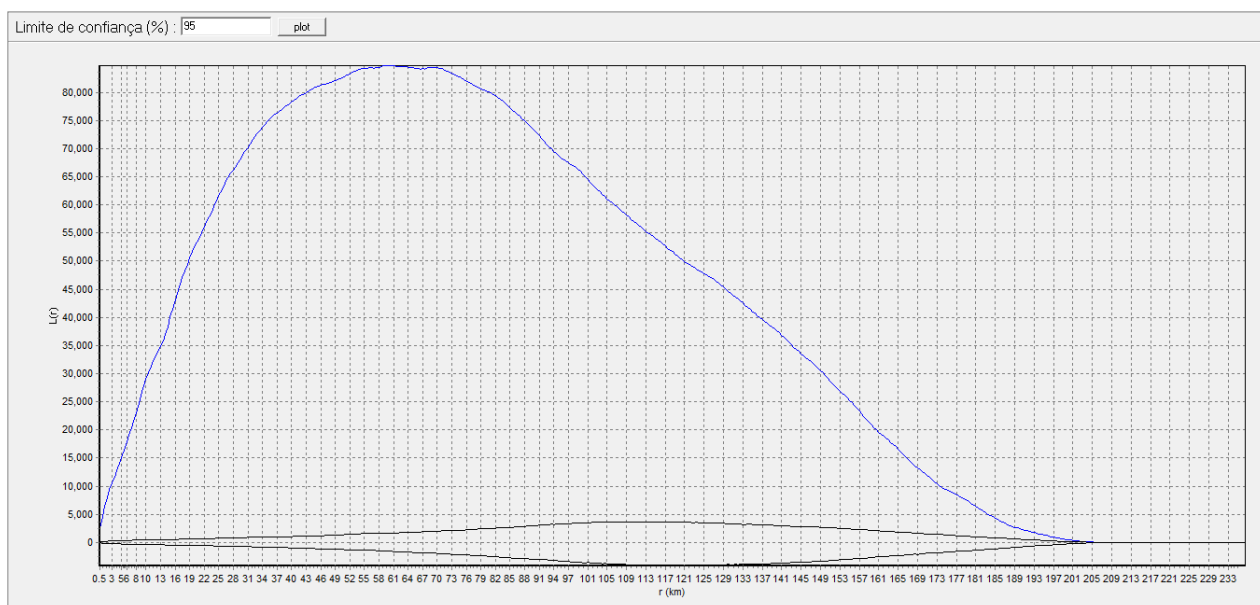
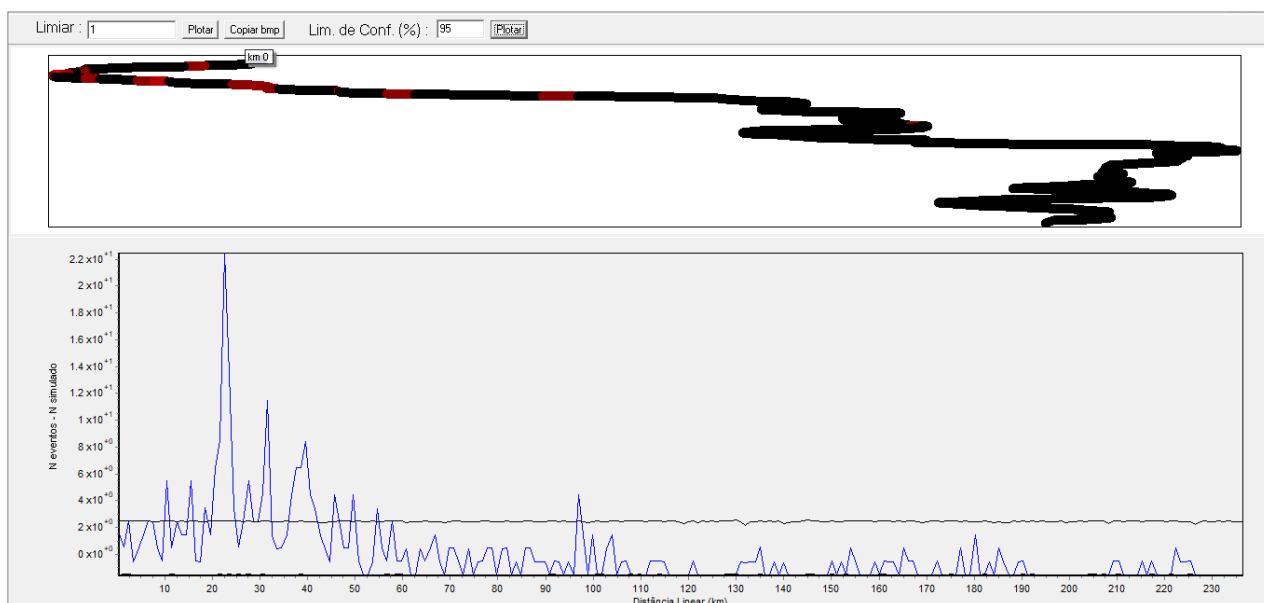


Figura 14.- Resultado da análise de hotspot encontrado para o *Didelphis aurita* no Trecho 1, no Estado de Santa Catarina.



Foi estabelecido um ranking de priorização dos 1.089,78 km, trecho a trecho. Este ranking considerou a riqueza de espécies identificadas com hotspots em um determinado segmento da estrada e o valor da intensidade deste hotspot. Espécies ameaçadas de extinção foram priorizadas, recebendo um coeficiente de correção que duplicou o valor de intensidade dos seus hotspots.

A partir desta base de dados foram geradas diferentes métricas para buscar o melhor ajuste dos modelos de risco. Algumas destas métricas foram:

- HS_Soma.- soma dos valores da análise de hotspot de cada espécie existente no trecho;
- HS_Soma_pond.- soma dos valores da análise de hotspot de cada espécie existente no trecho e ponderada para uma escala de 0 a 100;
- HS_Media_riq.- Soma dos valores da análise de hotspots de cada espécie existente no trecho, dividido pelo número de espécies existentes no trecho;
- HS_Media_riq_pond.- Soma dos valores da análise de hotspots de cada espécie existente no trecho, dividido pelo número de espécies existentes no trecho e ponderada para uma escala de 0 a 100;
- Entre outros.

2.3 Bases de dados

Após a Análise I, onde foram obtidas as variáveis condicionantes do atropelamento de cada espécie e seus respectivos pesos, assim como identificados os trechos de hotspots de atropelamento, foi possível a criação de um banco de dados constituído por quase 70 variáveis.

Como pode ser observado na Figura 1, a base de dados denominada de Variáveis de Base reúne todas as variáveis geradas pelos modelos bio-ecológicos das 30 espécies definidas como prioritárias e aquelas oriundas da malha viária. Já a base de dados denominada de Probabilidade de Atropelamento são aquelas geradas a partir das 15 espécies com registros suficientes para a geração dos modelos de hotspots.

É importante ressaltar que as variáveis iniciais sofreram processamentos de limpeza, normalização, análise de *outliers*, entre outros. Por outro lado, os pesos previamente definidos para cada espécie, tanto nos modelos bio-ecológicos, como para os hotspots, são afetados por padronizações de amplitude de efeito e são gerados subprodutos dos conjuntos destes dados, tais como somas, percentis, agrupamentos, entre outros, fato que pode multiplicar o total de variáveis consideradas nos modelos finais.

Ao final destes procedimentos, as duas bases de dados (Variáveis de base e Probabilidade de atropelamento) foram analisadas utilizando as técnicas de machine learning (ML), descritas abaixo.

2.4 Análises II - Machine learning

2.4.1 Preparação da base de dados de modelo de risco

Foi implementada a *Class Handle File* a qual é responsável pela manipulação dos dados, recebendo como entrada o caminho do arquivo da base de dados. Neste passo algumas das funções implementadas foram:

- `Prepare_files_controller`: controla todo o processo de manipulação e separação da base de dados;
- `read files`: realiza a leitura do arquivo da base de dados;
- `Separate_df`: separa o arquivo da base de dados em 2 (dois): `df_principal` que recebe os itens que possuem informações de hotspots e serão utilizados para a implementação do modelo baseado em Machine Learning e `df_extrapolate` que recebe os itens que não possuem tais informações que serão utilizados, no final, para a extrapolação das informações.
- `Prepare_hotspot`: separa dos dois arquivos criados na função `separate_df` as informações referentes aos hotspots, criando o arquivo `df_hotspot` que será utilizado para a implementação do modelo baseado em Machine Learning. Adicionalmente, esta realiza a chamada para a função `transform_hotspot_classifier`;
- `Transform_hotspot_classifier`: transforma as informações dos hotspots do tipo float para 1.0 ("YES") e 0.0 ("NO"). Isso foi realizado para a classificação para a presença de hotspots;

- Transform_df : transforma as informações do tipo string para tipo float. Foram atribuídos números floats entre -1.0 e -7.0. Além disso, para informações faltantes (NaN) foram atribuídos valores -10.0.

2.4.2 Técnicas de Feature Selection

Em machine learning (ML) e estatística, a seleção de recursos, também conhecida como feature selection, é o processo de seleção de um subconjunto de recursos relevantes (variáveis, preditores) para uso na construção do modelo de classificação/regressão. Resumidamente, algoritmos de feature selection são vistos como a combinação de uma técnica de busca para propor novos subconjuntos de características, juntamente com uma medida de avaliação que pontua os diferentes subconjuntos de características.

A abordagem mais simples é testar cada possível subconjunto de características, encontrando aquela que minimize a taxa de erro. No entanto, esta é uma busca exaustiva e computacionalmente intratável para conjunto de dados reais. Nesse sentido, a escolha da métrica de avaliação influencia fortemente a técnica de feature selection, e são essas métricas de avaliação que distinguem entre as três principais categorias de técnicas de feature selection, sendo elas:

- Filtro: utilizam uma medida de proxy em vez da taxa de erro para pontuar um subconjunto de recursos;
- Wrapper: utilizam um modelo preditivo para pontuar subconjuntos de features. Cada novo subconjunto é usado para treinar um modelo, que é testado em um conjunto de validação, e;
- Embedded: são um grupo abrangente de técnicas que realizam a seleção de recursos como parte do processo de construção do modelo.

Assim, no sentido de alcançar o objetivo do projeto, para verificar as variáveis de maior relevância na base de dados, selecionamos uma técnica para cada categorias de técnicas de feature selection, sendo elas: SelectKBest (filtro), Recursive Feature Elimination - RFE (wrapper) e Linear Regression (embedded). Adicionalmente as técnicas mencionadas, é aplicada a Pearson's Correlation como técnica de feature selection.

Nesse sentido, foi implementada a Class Selection no arquivo "feature_selection". Tal classe é responsável pela implementação das técnicas de feature selection citadas acima, recebendo como entrada o df_principal e o hotspot selecionado. Para melhor entendimento, a seguir são apresentadas as funções implementadas na Class Selection e uma breve descrição:

- feature_selection_controller : controla todo o processo de seleção de variáveis de acordo com as técnicas selecionadas;

- linear_algorithm: implementação da técnica Linear Regression;
- kbest_algorithm: implementação da técnica SelectKBest ;
- rfe_algorithm: implementação da técnica RFE;
- correlation_algorithm: implementação da técnica Pearson's Correlation;
- normalize: realiza a normalização dos dados para as técnicas de feature selection;
- select_features: realiza o tratamento dos resultados das técnicas Linear Regression e RFE, pois os algoritmos retornam uma lista (TRUE ou FALSE) das features selecionadas;
- get_features: realiza o tratamento dos resultados da técnica SelectKBest, pois esta é baseada em um escore de relevância da feature. Vale salientar, que a implementação realizada seleciona escores acima do threshold 0,05;
- save results: salva o arquivo de cada técnica de feature selection para cada hotspot na pasta "results/features".

2.4.3 Algoritmos de Machine Learning

Considerando o objetivo do contrato, de propor um modelo de áreas críticas de atropelamento para as rodovias estaduais do Estado de Santa Catarina, o projeto propõe a identificação de um ou mais modelos de ML que permitam atingir o objetivo. Buscando a identificação do melhor modelo, foram selecionados algoritmos que possuem paradigmas distintos, isto é, realizam a construção do modelo de formas diferentes entre eles. Foram definidos algoritmos baseados em Inteligência Artificial (classificação e regressão) aplicados para o experimento.

2.4.3.1 Classificação

A seguir segue uma breve descrição dos algoritmos de classificação selecionados para a implementação no projeto:

- k-Nearest Neighbors (kNN): Os exemplos de treinamento são vetores em um espaço de recursos multidimensional, cada um com um rótulo de classe. A fase de treinamento do algoritmo consiste em armazenar os vetores de características e rótulos de classe das amostras de treinamento. Na fase de classificação é atribuído o rótulo que é mais frequente entre as k amostras de treinamento mais próximas a esse ponto de consulta;
- Naive Bayes (Bayes): Assumem que o valor de um determinado recurso é independente do valor de qualquer outro recurso, dada a variável de classe. O classificador considera que cada características contribui independentemente para a probabilidade da classe, independentemente de qualquer possível correlações entre as suas características;
- Decision Tree (DT): se assemelha a um fluxograma, em que cada nó interno representa um "teste" em um atributo, cada ramo representa o resultado do teste e cada nó folha

representa um rótulo de classe (decisão tomada após computar todos os atributos). Os caminhos da raiz à folha representam regras de classificação.

- Support Vector Machine (SVM): são um dos métodos de previsão mais robustos. Dado um conjunto de exemplos de treinamento, cada um marcado como pertencente a suas categorias, um algoritmo de treinamento SVM constrói um modelo que atribui novos exemplos a uma categoria ou outra, tornando-o um classificador não probabilístico. O SVM mapeia exemplos de treinamento para pontos no espaço para maximizar a largura da lacuna entre as categorias. Novos exemplos são então mapeados nesse mesmo espaço e previstos para pertencer a uma categoria com base em qual lado da lacuna eles se enquadram.
- Random Forest (RF): é um método de aprendizado conjunto para classificação que opera construindo uma infinidade de DT em tempo de treinamento.

Nesse sentido, foi implementado a Class Classifiers no arquivo "evaluation_classifier". Tal classe é responsável pela implementação das técnicas de classificação citadas acima, recebendo como entrada o df_selection (arquivo com somente as variáveis selecionadas por cada técnica de feature selection), a técnica de feature selection e o hotspot selecionado. Vale salientar que na Class Classifiers é implementada a métrica de avaliação (acurácia) para avaliar cada abordagem aplicada na base de dados.

2.4.3.2 Regressão

Foram implementados 3 algoritmos de regressão. Esses são das mesmas classes de alguns utilizados em classificação (k-Nearest Neighbors - kNN; Decision Tree - DT; Support Vector Machine - SVM). Nesse sentido, foi implementado a Class Regression no arquivo "evaluation_regression". Tal classe é responsável pela implementação das técnicas de regressão citadas acima, recebendo como entrada o df_selection (arquivo com somente as variáveis selecionadas por cada técnica de feature selection), a técnica de feature selection e o hotspot selecionado. Vale salientar que na Class Regression é implementada a métrica de avaliação (mean absolute error) para avaliar cada abordagem aplicada na base de dados.

2.4.4 Definição do modelo de extrapolação

Apesar da base conter 8.666 linhas (trechos de estrada), para o treino/teste dos algoritmos foram utilizadas 1.304 linhas, as quais correspondem àquelas que possuem alguma informação de presença e ausência de hotspots (Base de dados de Probabilidade de Atropelamentos). Estes trechos são divididos em 70% para treino e 30% para teste e as análises foram repetidas 1.000 vezes. O modelo proposto é avaliado quanto a sua precisão, baseado nos registros de teste, sendo considerado adequado quando se obtém, no mínimo, 75% de

assertividade. Mesmo ao se obter um modelo com assertividade de 75%, novos modelos continuam sendo processados e testados, com o objetivo de se obter o melhor ajuste possível. Após se obter o modelo com melhor assertividade, o modelo é extrapolado para a totalidade dos trechos.

O valor de risco de um determinado trecho é o resultado das interações e pesos de todas as variáveis presentes no modelo final. Este valor é adimensional, podendo variar de zero ao infinito, e precisa ser normalizado para seu uso em planejamento territorial e sua aplicação na definição de áreas de risco nas rodovias. Para tanto, a partir dos valores originais, são estabelecidos sete níveis de risco, variando entre zero e seis. Cada nível de risco é atribuído considerando uma escala pré-estabelecida de percentual de contribuição. A definição dos limiares de risco para as estradas estaduais de Santa Catarina considerou os seguintes limiares:

- Risco 6.- trechos que possuem os valores de risco no limiar de 95% ou maior;
- Risco 5.- trechos que possuem os valores de risco entre os limiares de 78 e 95%;
- Risco 4.- trechos que possuem os valores de risco entre os limiares de 60 e 78%;
- Risco 3.- trechos que possuem os valores de risco entre os limiares de 40 e 60%;
- Risco 2.- trechos que possuem os valores de risco entre os limiares de 30 e 40%;
- Risco 1.- trechos que possuem os valores de risco entre os limiares de 18 e 60%;
- Risco 0.- trechos que possuem os valores de risco entre os limiares de 0 e 18%.

3 Resultados

3.1 Modelos bio-ecológicos

Os resultados dos modelos bio-ecológicos, analisados para cada uma das 30 espécies previamente definidas são apresentados na Tabela 8. Todos os valores de AUC superaram 0,7 e variaram de 0,71 a 0,98. O AUC de treino teve valores entre 0,73 e 0,98, com um valor médio de 0,86, enquanto que o AUC de teste apresentou valores entre 0,71 e 0,97, média de 0,81.

Para os modelos de treino, as espécies que tiveram os menores AUC (<0,8) foram *Dasypus novemcinctus*, *Cerdocyon thous*, *Tamandua tetradactyla* e *Euphractus sexcinctus*, enquanto que as espécies *Didelphis aurita*, *Leopardus guttulus*, *Coendou spinosus*, *Sphiggurus villosus* e *Conepatus chinga* tiveram AUC de 0,9 ou maiores. Já para os modelos de teste, as espécies *Dasypus novemcinctus*, *Tamandua tetradactyla*, *Euphractus sexcinctus*, *Cerdocyon thous*, *Puma concolor*, *Procyon cancrivorus*, *Leopardus pardalis*, *Herpailurus yagouaroundi* e *Dasypus septemcinctus* tiveram AUC menores que 0,8 e as espécies *Coendou spinosus*, *Sphiggurus villosus* e *Conepatus chinga* apresentaram ajustes maiores que 0,9.

Tabela 8.- Parâmetros obtidos dos modelos bio-ecológicos de atropelamento de cada uma das 30 espécies previamente definidas. (Continua)

	#Training samples	Training AUC	#Test samples	Test AUC	AUC Standard Deviation	Minimum training presence test omission	Minimum training presence binomial probability	10 percentile training presence test omission	10 percentile training presence binomial probability	Maximum test sensitivity plus specificity test omission	Maximum test sensitivity plus specificity binomial probability	Threshold médio
<i>Didelphis aurita</i>	672	0,90	288	0,89	0,009	0,004	0,000	0,120	0,000	0,173	0,000	0,178
<i>Dasypus novemcinctus</i>	1193	0,73	511	0,71	0,012	0,001	0,009	0,129	0,000	0,348	0,000	0,315
<i>Didelphis albiventris</i>	798	0,86	342	0,84	0,011	0,001	0,001	0,125	0,000	0,242	0,000	0,182
<i>Hydrochoerus hydrochaeris</i>	770	0,86	329	0,84	0,011	0,002	0,008	0,123	0,000	0,256	0,000	0,202
<i>Cerdocyon thous</i>	1148	0,74	491	0,72	0,011	0,001	0,096	0,122	0,000	0,323	0,000	0,311
<i>Tamandua tetradactyla</i>	707	0,76	303	0,71	0,015	0,002	0,016	0,144	0,000	0,328	0,000	0,308
<i>Euphractus sexcinctus</i>	488	0,77	209	0,71	0,017	0,008	0,004	0,144	0,000	0,301	0,000	0,295
<i>Salvator merianae</i>	488	0,88	172	0,86	0,016	0,006	0,000	0,137	0,000	0,274	0,000	0,169
<i>Procyon cancrivorus</i>	397	0,83	170	0,78	0,018	0,002	0,032	0,166	0,000	0,322	0,000	0,228
<i>Alouatta guariba</i>	290	0,86	123	0,80	0,018	0,009	0,000	0,166	0,000	0,218	0,000	0,230
<i>Galictis cuja</i>	231	0,88	98	0,82	0,022	0,004	0,021	0,180	0,000	0,262	0,000	0,173
<i>Coendou spinosus</i>	231	0,95	67	0,92	0,016	0,011	0,000	0,173	0,000	0,155	0,000	0,163
<i>Cabassous tatouay</i>	175	0,87	75	0,82	0,022	0,011	0,000	0,156	0,000	0,233	0,000	0,226
<i>Sphiggurus villosus</i>	42	0,96	17	0,94	0,029	0,047	0,000	0,171	0,000	0,129	0,000	0,158
<i>Sapajus nigritus</i>	169	0,84	72	0,81	0,022	0,011	0,000	0,125	0,000	0,188	0,000	0,266
<i>Leopardus pardalis</i>	135	0,86	57	0,78	0,028	0,007	0,004	0,198	0,000	0,261	0,000	0,251
<i>Herpailurus yagouaroundi</i>	141	0,86	60	0,78	0,031	0,022	0,060	0,180	0,000	0,242	0,000	0,233

Tabela 8.- Parâmetros obtidos dos modelos bio-ecológicos de atropelamento de cada uma das 30 espécies previamente definidas. (Conclusão)

	#Training samples	Training AUC	#Test samples	Test AUC	AUC Standard Deviation	Minimum training presence test omission	Minimum training presence binomial probability	10 percentile training presence test omission	10 percentile training presence binomial probability	Maximum test sensitivity plus specificity test omission	Maximum test sensitivity plus specificity binomial probability	Threshold médio
<i>Dasypus septemcinctus</i>	124	0,87	52	0,78	0,032	0,012	0,022	0,198	0,000	0,302	0,000	0,215
<i>Puma concolor</i>	128	0,84	54	0,77	0,031	0,022	0,001	0,185	0,000	0,298	0,000	0,230
<i>Leopardus guttulus</i>	112	0,93	47	0,89	0,021	0,030	0,000	0,179	0,000	0,128	0,000	0,212
<i>Mazama gouazoubira</i>	79	0,89	33	0,83	0,038	0,039	0,002	0,188	0,000	0,233	0,000	0,198
<i>Conepatus chinga</i>	110	0,98	46	0,97	0,009	0,013	0,000	0,163	0,000	0,072	0,000	0,199
<i>Eira barbara</i>	112	0,80	48	0,73	0,034	0,010	0,014	0,194	0,014	0,215	0,000	0,289
<i>Lontra longicaudis</i>	101	0,85	42	0,79	0,033	0,026	0,001	0,167	0,000	0,210	0,000	0,274
<i>Cuniculus paca</i>	101	0,79	42	0,77	0,035	0,007	0,044	0,124	0,002	0,314	0,000	0,244
<i>Leopardus tigrinus</i>	89	0,86	38	0,80	0,033	0,021	0,000	0,197	0,000	0,250	0,000	0,290
<i>Tapirus terrestris</i>	38	0,92	16	0,89	0,036	0,063	0,000	0,156	0,000	0,163	0,000	0,267
<i>Leopardus wiedii</i>	63	0,87	26	0,82	0,032	0,027	0,009	0,185	0,000	0,112	0,000	0,303
<i>Pecari tajacu</i>	51	0,79	21	0,74	0,046	0,024	0,058	0,124	0,013	0,176	0,001	0,295
<i>Mazama americana</i>	39	0,88	16	0,84	0,040	0,019	0,004	0,131	0,001	0,194	0,000	0,242

Como esperado, modelos de treino que possuem AUCs maiores, tendem a apresentar modelos de teste também com AUC elevadas (Figura 15). Entretanto, ao contrário do esperado, um grande número de registros gera modelos com AUCs menores, além de que este dado gera modelos com grande amplitude de AUCs, como demonstra o valor de R^2 (Figura 16). Esta segunda análise demonstra que, mais importante que o número de registros, é a definição de variáveis que compõem o modelo.

Figura 15.- Regressão entre os valores de AUC dos modelos de teste e do AUC dos modelos de treino.

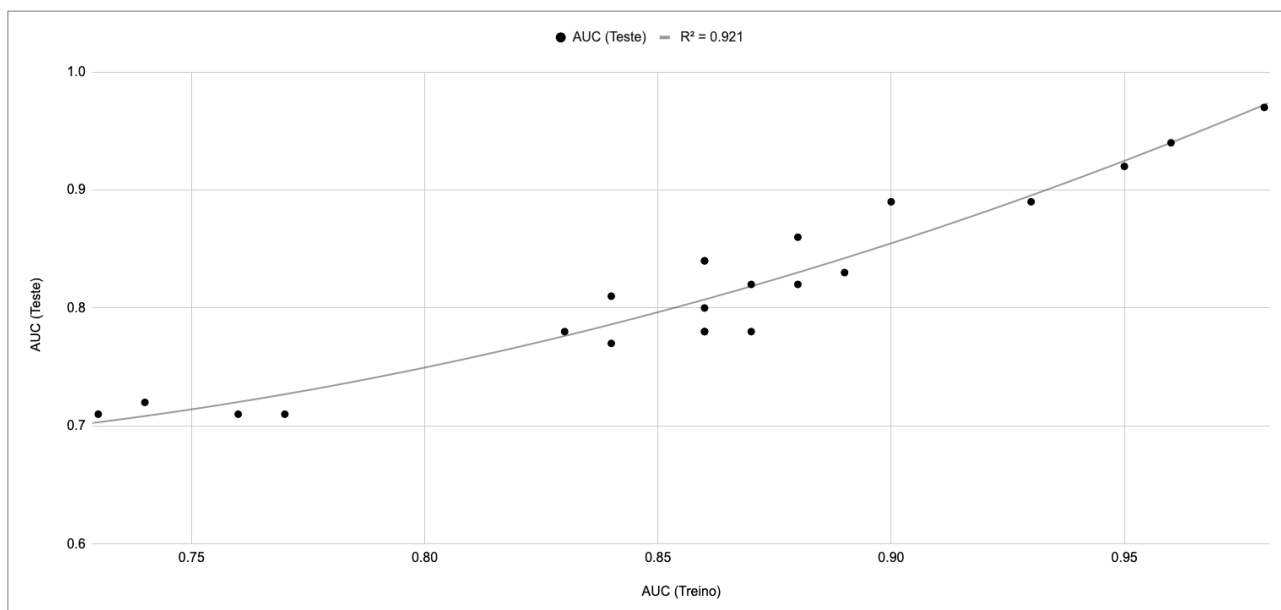
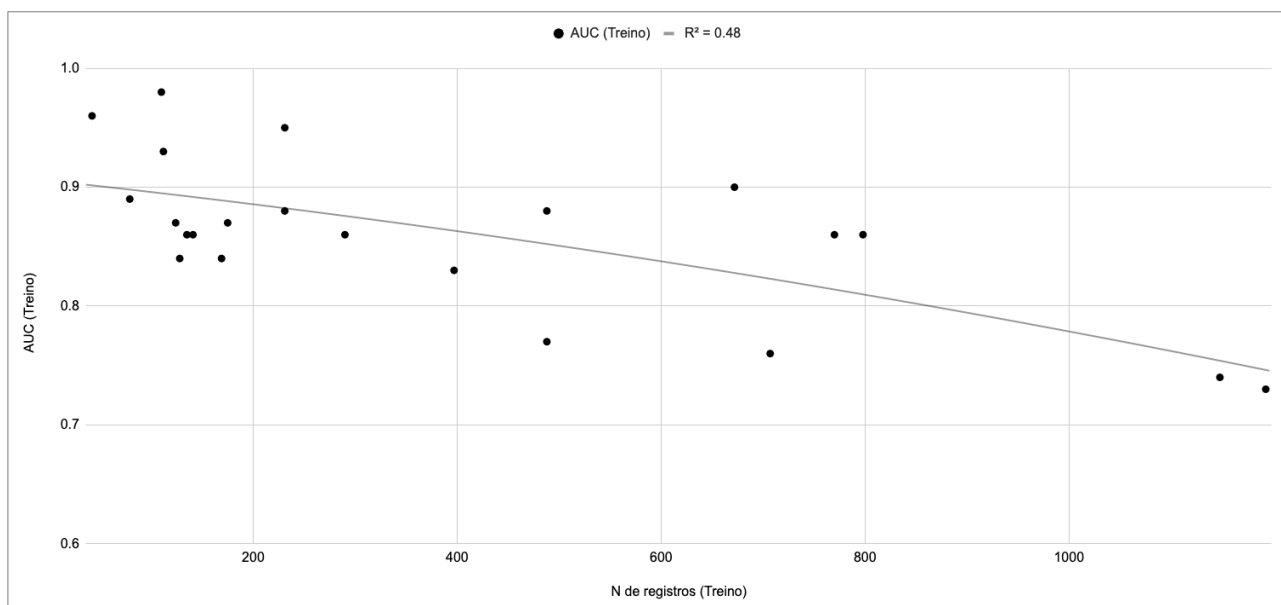


Figura 16.- Regressão entre o valor do AUC de treino e o número de registros da respectiva espécie.



Os valores da contribuição (peso) de cada variável bio-ecológica no modelo de risco da espécie são adimensionais e variam de espécie para espécie. Para facilitar a análise dos resultados e demonstrar quais variáveis são mais relevantes para cada espécie, foi realizada uma

padronização dos valores variando de 0 a 100. Desta forma, a variável com maior contribuição foi definida como o valor 100 e aos valores das demais variáveis foi aplicada uma regra de três para redistribuir sua importância. Para efeito de priorização, foram consideradas todas as variáveis que resultaram em uma importância igual ou superior a 70%, conforme visto na Tabela 9.

A variável com maior representatividade nos modelos foi Vegetação, tendo um peso superior a 70% em 43% dos modelos. Outras três variáveis têm elevada representatividade nos modelos das diferentes espécies, são elas: Bio03, com 37%; Clima, com 27% e Solo com 23%. Por outro lado, a variável Densidade de drenagem, que está presente em 15 modelos gerados, não apresentou elevada importância em nenhum deles.

É importante ressaltar que os modelos de risco ponderam a totalidade das variáveis e não apenas aquelas com maior peso. Assim, o fato da Tabela 9 mostrar que a anta teve a variável Declividade com peso maior que 70%, não implica que apenas ela seja importante na definição das áreas de risco. No caso específico da anta, as variáveis Bio12, Bio03, Bio05, Bio07, Cobertura arbórea, Declividade e PRIM participam da definição da modelagem de risco de atropelamento e possuem diferentes pesos, variando de 10 a 100.

Das 30 espécies modeladas, sete apresentaram três variáveis com peso superior a 70, 10 espécies tiveram duas variáveis e as demais (13) tiveram uma única variável prioritária (>70).

Tabela 9.- Variáveis com percentual de contribuição ponderada superior a 70% para os modelos de risco de atropelamento das 30 espécies previamente definidas para o Estado de Santa Catarina.

	Bio12	Bio13	Bio03	Bio05	Bio07	Cob. Arb.	Decliv.	Dens. Dren.	Uso. Solo	Clima	Relevo	Solo	Veget.	PRIM
<i>Didelphis aurita</i>			X											
<i>Dasypus novemcinctus</i>	X											X	X	
<i>Didelphis albiventris</i>			X											
<i>Hydrochoerus hydrochaeris</i>					X					X			X	
<i>Cerdocyon thous</i>										X		X		
<i>Tamandua tetradactyla</i>													X	
<i>Euphractus sexcinctus</i>											X		X	
<i>Salvator merianae</i>			X									X	X	
<i>Procyon cancrivorus</i>										X				
<i>Alouatta guariba</i>										X		X	X	
<i>Galictis cuja</i>										X		X	X	
<i>Coendou spinosus</i>			X											
<i>Cabassous tatouay</i>			X										X	
<i>Sphiggurus villosus</i>			X										X	
<i>Sapajus nigratus</i>					X				X		X			
<i>Leopardus pardalis</i>												X	X	
<i>Herpailurus yagouaroundi</i>			X										X	
<i>Dasypus septemcinctus</i>													X	
<i>Puma concolor</i>													X	
<i>Leopardus guttulus</i>			X											
<i>Mazama gouazoubira</i>										X	X			
<i>Conepatus chinga</i>										X		X		
<i>Eira barbara</i>						X								
<i>Lontra longicaudis</i>			X											
<i>Cuniculus paca</i>			X	X										
<i>Leopardus tigrinus</i>				X										
<i>Tapirus terrestris</i>							X							
<i>Leopardus wiedii</i>			X											
<i>Pecari tajacu</i>	X													X
<i>Mazama americana</i>		X								X	X			

3.1 Análises de agregação

Dos 1.089,78 quilômetros avaliados na busca de pontos de agregação de atropelamento, foram identificados 463 hotspots em 320 quilômetros (29,4% dos quilômetros avaliados). Deste total, 223 quilômetros (70%) possuem hotspots de uma única espécie, 64 quilômetros (20%) possuem duas espécies; 23 quilômetros (7%) possuem três espécies; sete quilômetros (2%) possuem quatro espécies e três quilômetros (1%) reúnem áreas de agregação conjuntas de cinco espécies.

O número de áreas de agregação por espécie variou de 5 (*Cabassous tatouay*) a 83 (*Didelphis albiventris*), conforme a Tabela 10. As cinco espécies com maiores números de hotspots (*Didelphis albiventris*, *Dasypus novemcinctus*, *Tamandua tetradactyla*, *Hydrochoerus hydrochaeris* e *Cerdocyon thous*) totalizam 63% das áreas de agregação. Estas espécies se caracterizam por serem generalistas e de grande ocorrência no Estado de Santa Catarina. Das cinco espécies, a capivara (*Hydrochoerus hydrochaeris*) é a de maior porte e agrega um elevado risco à segurança viária de motoristas. O tamanduá-mirim (*Tamandua tetradactyla*) e o cachorro-do-mato (*Cerdocyon thous*) possuem porte médio e, normalmente, geram prejuízos financeiros quando atropelados, mas também acidentes em decorrência da resposta comportamental de escape de motoristas.

Tabela 10.- Número e percentual de quilômetros com áreas de agregação para as 15 espécies analisadas para o Estado de Santa Catarina.

Espécie	Nº Hotspots	% Hotspots
<i>Didelphis albiventris</i>	83	0,18
<i>Dasypus novemcinctus</i>	65	0,14
<i>Tamandua tetradactyla</i>	55	0,12
<i>Hydrochoerus hydrochaeris</i>	46	0,10
<i>Cerdocyon thous</i>	41	0,09
<i>Leopardus spp.</i>	33	0,07
<i>Didelphis aurita</i>	32	0,07
<i>Procyon cancrivorus</i>	25	0,05
<i>Conepatus chinga</i>	17	0,04
<i>Galictis cuja</i>	16	0,03
<i>Herpailurus yagouaroundi</i>	13	0,03
<i>Euphractus sexcinctus</i>	12	0,03
<i>Sphiggurus villosus</i>	11	0,02
<i>Alouatta guariba</i>	9	0,02
<i>Cabassous tatouay</i>	5	0,01
Total	463	

3.2 Machine learning

As análises de machine learning, quando empregadas para buscar o ajuste das variáveis de base e o banco de dados de probabilidade de atropelamento obtiveram os seguintes resultados:

3.2.1 Análise de Classificação

O algoritmo kNN apresenta melhor acurácia e menor desvio padrão se comparado aos outros algoritmos utilizados (Tabela 11). No entanto, como a diferença entre as técnicas de feature selection do algoritmo kNN foram semelhantes e foi utilizado o método de *Shapiro-Wilk* para verificar a adequação à hipótese de normalidade e, portanto, indicar para testes paramétricos ou não paramétricos. Como os p-valores obtidos foram menores que 0,05 (Tabela 12), rejeitamos a hipótese de normalidade com 95% de confiabilidade e se define o uso de testes não paramétricos (*Wilcoxon Rank Sum*), no intuito de verificar a diferença estatisticamente significativa entre os grupos de resultados.

Os resultados da análise de *Wilcoxon Rank Sum*, realizado par a par, demonstram que para o algoritmo kNN, todas as técnicas apresentam diferença estatística se comparado entre elas, com exceção da RFE. Nesse sentido, é possível se definir o uso do algoritmo kNN e da técnica "Linear", por requererem o menor número de variáveis para definição do modelo, possuir uma taxa de acurácia de 76,6% e um desvio padrão de 0,023.

Tabela 11.- Resultados de acurácias médias e desvio padrão na classificação.

Algoritmos	Completo	Linear	KBest	RFE	Correlação
kNN	0,758 (0,024)	0,766 (0,023)	0,757 (0,023)	0,758 (0,022)	0,761 (0,023)
Bayes	0,687 (0,076)	0,759 (0,025)	0,706 (0,058)	0,688 (0,076)	0,682 (0,084)
DT	0,698 (0,065)	0,746 (0,032)	0,710 (0,050)	0,699 (0,066)	0,694 (0,072)
SVM	0,717 (0,067)	0,754 (0,033)	0,727 (0,0530)	0,718 (0,067)	0,714 (0,073)
RF	0,726 (0,063)	0,756 (0,032)	0,733 (0,050)	0,726 (0,063)	0,723 (0,068)

Tabela 12.- Média (%) das acurácias e os p-valores dos conjuntos de resultados para o algoritmo kNN, baseados nas análises de Shapiro Wilk.

Feature Selection	Acurácia (%)	p-valores (Shapiro Wilk)
Complete	0,75825	0,00822
Linear	0,76630	0,00446
KBest	0,75798	0,03308
RFE	0,75874	0,01087
Correlation	0,76144	0,04647

3.2.2 Análise de Regressão

O algoritmo kNN apresenta menor *mean absolute error* e desvio padrão se comparado aos outros algoritmos utilizados. No entanto, não há diferença entre as técnicas de *feature selection* do algoritmo kNN, fato que reforça o uso da técnica "Linear", devido ao menor número de variáveis necessárias para o modelo (Tabela 13).

Tabela 13.- Resultados das médias de *mean absolute error* e desvio padrão na regressão.

Algoritmos	Completo	Linear	KBest	RFE	Correlação
kNN	0,027 (0,002)	0,027 (0,002)	0,027 (0,002)	0,027 (0,002)	0,027 (0,002)
DT	0,030 (0,004)	0,030 (0,004)	0,030 (0,005)	0,030 (0,005)	0,030 (0,005)
SVM	0,046 (0,022)	0,050 (0,028)	0,046 (0,022)	0,046 (0,022)	0,046 (0,023)

3.3 Ajustes dos modelos

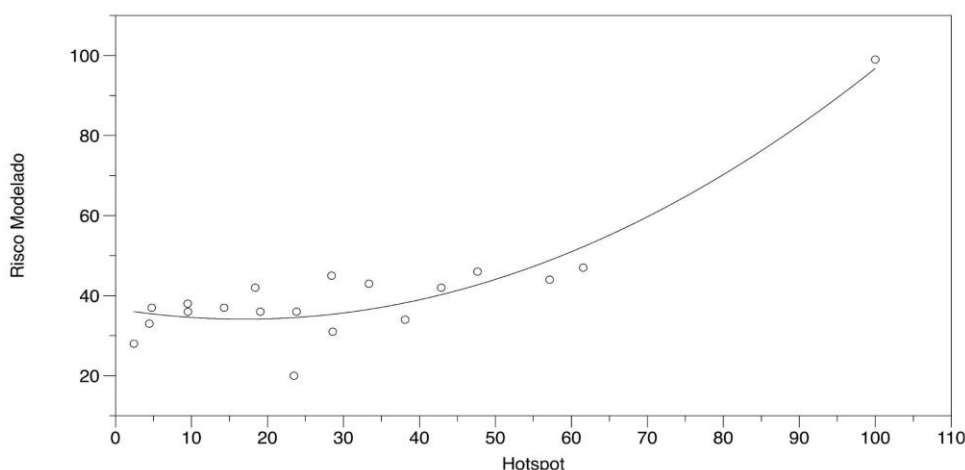
A partir das definições obtidas nas análises de classificação e de regressão, foram realizados os testes de ajustes dos modelos obtidos nas análises de *machine learning* (ML) e os dados das análises de hotspots.

Considerando os resultados das análises de agregação, dos 1.089,78 quilômetros analisados, um total de 320 possuem hotspots e 769 não indicaram pontos de agregação. Baseado nestes dados e se aplicando os modelos gerados com as técnicas de ML, se obteve uma precisão de 0,97 na identificação de áreas de não agregação. Já para as áreas com agregações o ajuste da curva (R^2) foi de 0,86.

O melhor ajuste foi obtido com uma equação polinomial de segunda ordem, tendo a variável dependente o risco de atropelamento ponderado (Risco) e a variável preditora o valor dos hotspots somados e ponderados (HS) (Figura 17):

$$\text{Risco} = 36,7 - 0,307 \cdot \text{HS} + 0,00907 \cdot \text{HS}^2$$

Figura 17.- Regressão entre o risco de atropelamento ponderado (Risco) e os hotspots somados e ponderados.



3.4 Modelos de risco de atropelamento

Os modelos de risco foram obtidos para um total de 5.922,54 quilômetros de rodovias estaduais do Estado de Santa Catarina, subdivididos em 6.212 segmentos com extensões que variaram de 0,1 a 1,1 km. Deste total, 91% tiveram extensões de 1 a 1,1 km (Figura 18).

Para o restante da malha viária estadual (~2,2%) os modelos foram inconclusivos. Dois fatores podem ter determinado a impossibilidade da modelagem nestes trechos: (1) o conjunto de espécies de animais definidas para modelagem foram pouco representativas para as localidades e/ou (2) as variáveis bio-ecológicas e de tipologia de estradas foram incapazes de obter modelos com a efetividade almejada.

O número de trechos e suas respectivas extensões variaram conforme os índices de risco (Tabela 14). Para efeito de planejamento territorial, definição de políticas públicas de conservação da biodiversidade e direcionamento para estudos aprimorados em licenciamento ambiental de rodovias, os níveis 5 e 6 são aqueles que devem ser considerados prioritários. Os trechos de nível 4, em geral se caracterizam como áreas complementares, onde monitoramentos podem ser propostos e/ou na definição de segmentos de rodovias onde estratégias complementares de mitigação são implementadas. Os demais níveis, 0 a 3, representam pouca relevância para ações de mitigação e planejamento territorial.

Como esperado, os trechos de estradas estaduais de Santa Catarina, definidos com risco 6 foram os menos representativos, tendo 326 segmentos e 319,4 quilômetros de estradas estaduais, totalizando 5,4% da malha viária modelada. Já os trechos com nível 5 de risco, os quais ainda apresentam significativo risco à fauna silvestre, foram definidos para 1.014 segmentos de estradas estaduais, representando 986,5 quilômetros e 16,7% da malha modelada.

Figura 18.- Modelagem de risco de atropelamento de fauna silvestre para as estradas estaduais do Estado de Santa Catarina considerando os sete níveis de risco.

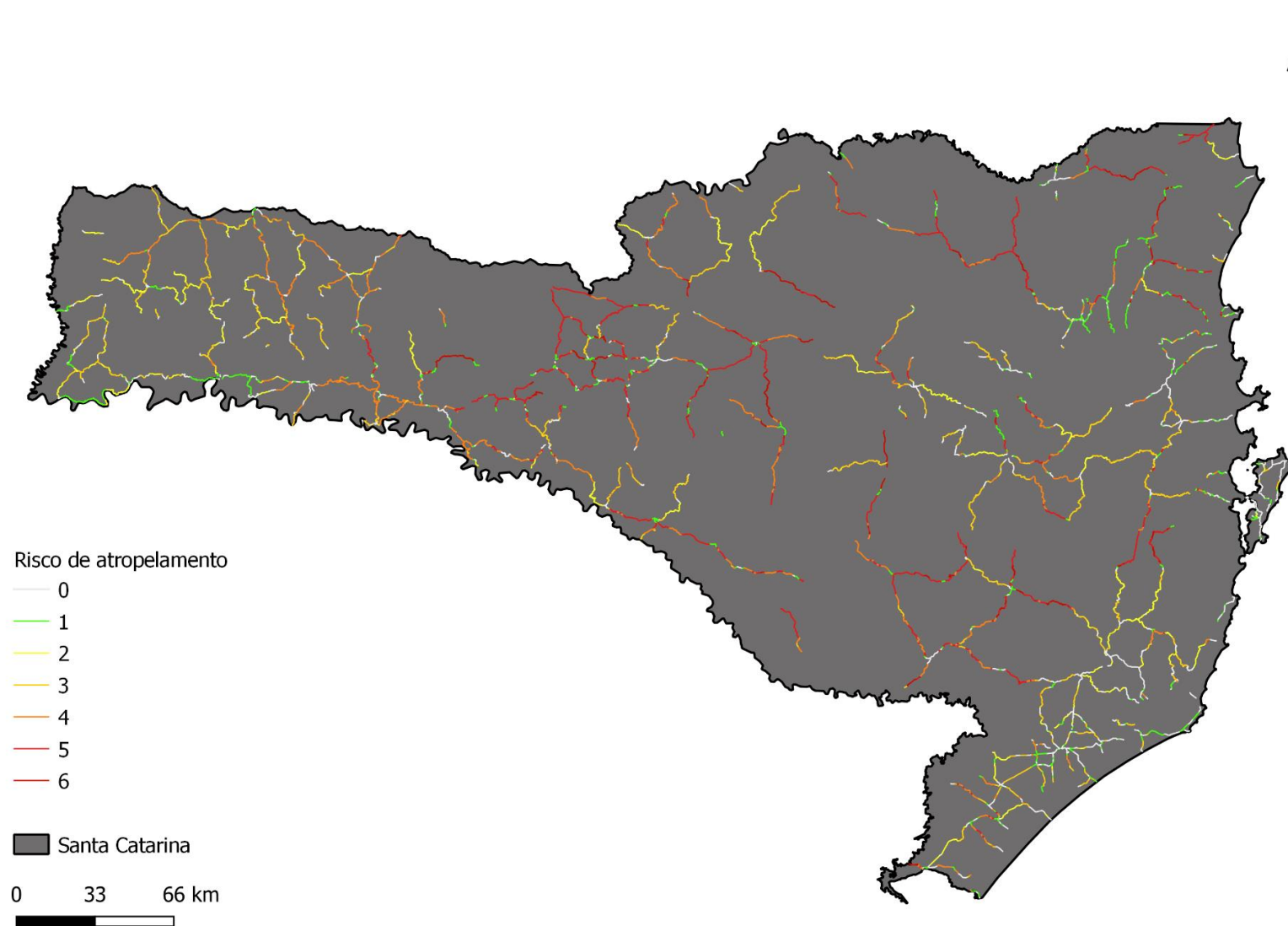


Tabela 14.- Número de trechos e suas respectivas extensões em quilômetros e percentuais para cada nível de risco.

Risco	Trechos	Extensão (Km)	Extensão (%)
0	1.122	1.000,08	16,9
1	700	643,64	10,9
2	772	752,17	12,7
3	1.207	1.177,12	19,9
4	1.071	1.043,68	17,6
5	1.014	986,47	16,7
6	326	319,38	5,4
Total	6.212	5.922,54	100

A Coordenadoria Regional (CR) com maior número de trechos foi a Meio Oeste, com 1.103, totalizando 17,8% dos trechos analisados para estradas estaduais de Santa Catarina (Tabela 15). Esta CR também teve os maiores percentuais de trechos com nível de risco 5 e 6, respectivamente 37% e 41% do total em cada nível. No extremo oposto, a CR Extremo Oeste teve apenas dois trechos identificados no nível 5, ambos na SC-157.

Tabela 15.- Distribuição dos trechos de risco de atropelamento em cada Coordenadoria Regional e considerando os diferentes níveis de risco de atropelamento de fauna silvestre.

	Nível							Total
	6	5	4	3	2	1	0	
Oeste	23	46	294	114	61	65	125	728
Litoral	15	53	12	53	40	38	107	318
Norte	59	194	89	42	72	82	56	594
Meio Oeste	133	377	182	160	70	94	87	1103
Planalto	71	245	192	167	64	61	60	860
Sul	10	24	91	202	188	118	352	985
Extremo Oeste	0	2	125	269	214	94	124	828
Vale	15	73	86	200	63	148	211	796
Total	326	1014	1071	1207	772	700	1122	6212

Foram avaliados os trechos de risco associados às unidades de conservação (UCs) no Estado de Santa Catarina considerando um buffer de 500 m no entorno de cada unidade. As estradas estaduais de Santa Catarina possuem 152 trechos afetando unidades de conservação, o que representa 2,4% da malha estadual (Tabela 16). Deste total, 63 estão em UCs de proteção integral e 89 em UCs de uso sustentável. No que se refere a esfera de administração, 59 trechos afetam UCs federais, 30 afetam UCs estaduais e 63 as UCs municipais. Um total de 62 trechos (41% dos trechos em UCs) possuem níveis 5 e 6 de risco e requerem ações de mitigação.

Duas UCs sob gestão federal são afetadas por trechos de estradas com riscos 5 ou 6, sendo elas o Parque Nacional de São Joaquim e a Estação Ecológica da Mata Escura. O Parque

Nacional de São Joaquim tem cinco trechos, todos de nível 6, localizados na estrada SC-370, no trecho denominado de Serra do Corvo Branco. Já a ESEC da Mata Escura, possui um único trecho, de nível 5, na SC-155 (-26,464569, -52,221762), localizado próximo à divisa entre o Estado de Santa Catarina e Paraná. A unidade de conservação estadual com maior número de quilômetros com trechos em níveis 5 e 6 é o Parque Estadual da Serra do Tabuleiro, em um trecho de 28 quilômetros entre as coordenadas (-27,876653, -48,941808) e (-27,731117, -48,907999), associados às rodovias SC-435 (Figura 19). Na gestão municipal, a APA da Dona Francisca possui a totalidade dos trechos de risco com níveis 5 e 6 para essas UCs nesta esfera de gestão. São 34,2 quilômetros de extensão, na SC-418, entre as coordenadas (-26,181145, -49,177784) e (-26,197205, -48,921764).

Tabela 16.- número de trechos, em seus respectivos níveis de risco, associados a unidades de conservação de diferentes categorias e esferas administrativas.

		Nível							Total
		6	5	4	3	2	1	0	
PROTEÇÃO INTEGRAL	Federal	5	1	5	2	0	0	0	13
	Estadual	7	8	0	4	0	2	5	26
	Municipal	0	0	0	0	0	5	19	24
USO SUSTENTÁVEL	Federal	0	0	2	6	0	19	19	46
	Estadual	1	2	1	0	0	0	0	4
	Municipal	21	17	0	0	0	1	0	39
Total		34	28	8	12	0	27	43	152

Figura 19.- Trechos da SC-435 que tangenciam o Parque Estadual da Serra do Tabuleiro e que possuem nível de risco 6 para o atropelamento de fauna silvestre.

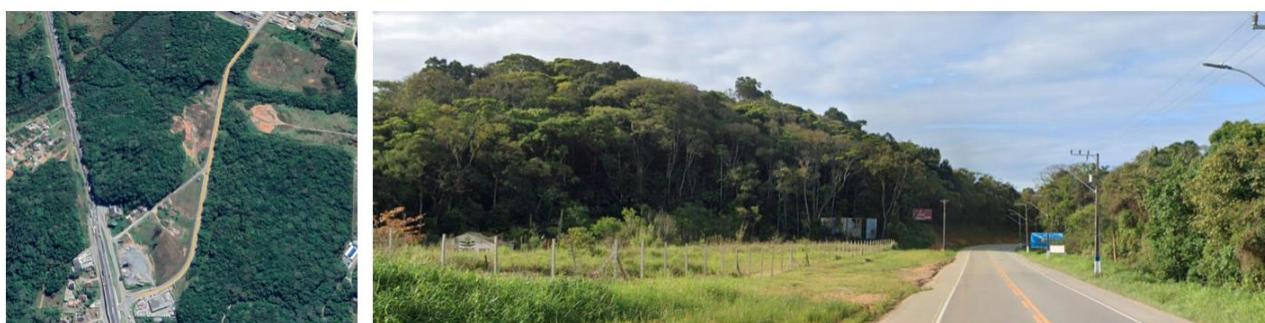


Do total de trechos, 4.546 foram fora das áreas urbanas e 1.666 em áreas urbanas. Como esperado, pouco mais de 50% dos trechos urbanos foram definidos como áreas de risco zero e outros 33% foram modelados como no nível 1 (Tabela 17). O único quilômetro com nível 5 de risco, localizado em um trecho urbano, se encontra no acesso da SC-414 a BR-101, no município de Penha (Figura 20).

Tabela 17.- Número de trechos de diferentes níveis de risco, segundo a sua distribuição em áreas urbanas e não urbanas.

Risco	Não Urbano	Urbano	Total
0	10	1.112	1.122
1	149	551	700
2	772	0	772
3	1.206	1	1.207
4	1.070	1	1.071
5	1.013	1	1.014
6	326	0	326
Total	4.546	1.666	6.212

Figura 20.- Imagem aérea e de uma seção da SC-414, em Penha, onde se identificou o único trecho de estrada com nível 5 de risco em área urbana.



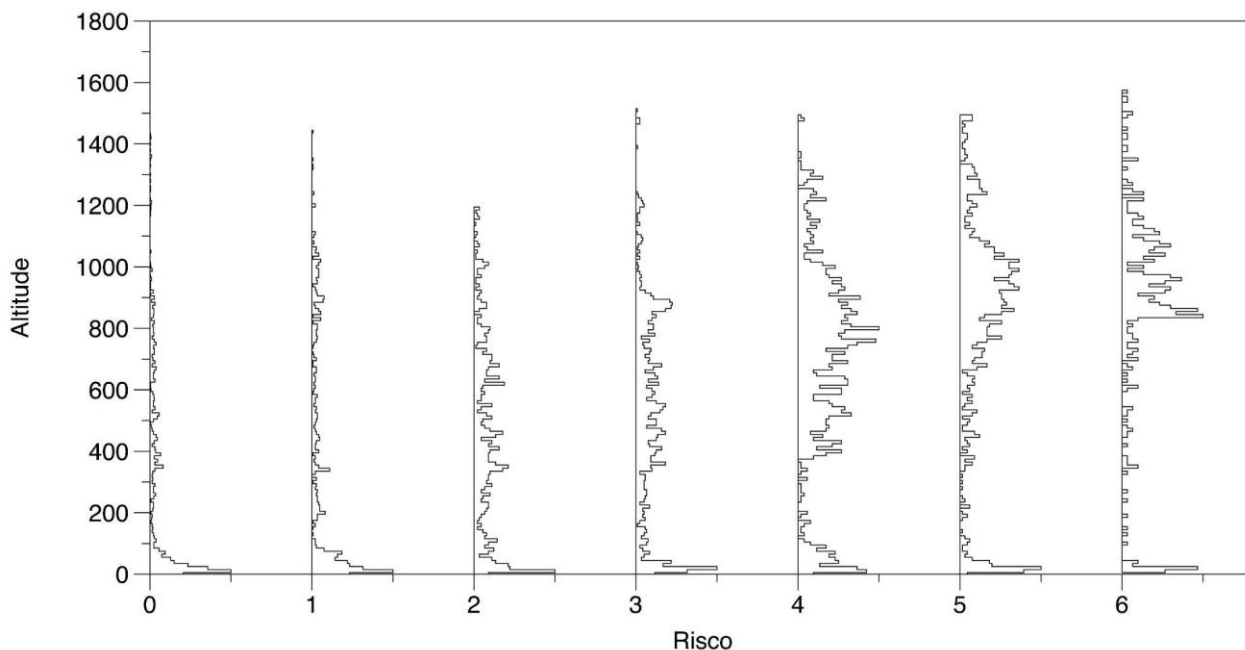
A Tabela 18 e a Figura 21 mostram que os trechos com diferentes índices de risco estão presentes em todas as altitudes existentes no Estado de Santa Catarina. Pode-se notar que os valores variam de áreas menores que 10 metros a mais de 1.400m, exceto para o risco 2, que atinge 1.191 m. Entretanto, tanto a mediana quanto a média mostram um aumento contínuo da altitude dos trechos conforme aumenta o nível de risco. A Figura 21 evidencia que todos os níveis de risco possuem ocorrências em áreas de baixa altitude e se diferenciam no restante da distribuição, com maiores altitudes nos índices maiores.

O fato de todos os níveis de risco apresentarem trechos nas áreas de baixa altitude pode estar relacionado ao fato da maior ocupação e densidade da malha viária nesta região. Tradicionalmente, em todo o mundo, as zonas costeiras possuem maior densidade populacional e são afetadas por diferentes atividades econômicas, sazonais ou não, que amplificam os riscos à fauna.

Tabela 18.- Variação da altitude dos trechos de risco segundo o valor do índice de risco.

Risco	Mínimo	Mediana	Média	Máximo
0	0,0	147,2	305,4	1.433,3
1	1,0	223,4	364,6	1.441,0
2	0,7	411,6	440,6	1.191,7
3	1,9	527,9	520,7	1.505,3
4	2,1	741,7	685,7	1.486,5
5	2,2	895,5	815,9	1.494,3
6	9,5	925,8	859,6	1.570,3

Figura 21.- Variação do número de trechos em relação a altitude segundo o valor do índice de risco.



3.5 Observações importantes

Algumas considerações são importantes de serem apresentadas sobre os modelos de risco e os resultados obtidos para o Estado de Santa Catarina para que o uso dos dados e a tomada de decisão seja assertiva.

Os modelos foram concebidos para um conjunto de espécies altamente representativas de diferentes grupos funcionais da classe dos mamíferos, sobretudo para espécies de médio e grande porte. Este fato permite o uso dos modelos para a tomada de decisão para esta classe, independente da espécie considerada, mas não deve ser utilizado para tomada de decisão para outras classes, como aves, répteis e anfíbios.

A modelagem integra informações bio-ecológicas das 30 espécies, reunindo a participação de cada espécie e sua importância na definição dos riscos de cada trecho. Por serem uma integração de espécies, os valores de risco não podem ser analisados para uma ou poucas espécies separadamente.

Os modelos bio-ecológicos, gerados para cada uma das 30 espécies, como mostrado no diagrama disponibilizado no material e métodos, são uma parte do processo de análise e se baseiam exclusivamente em variáveis bio-ecológicas. Estes modelos não são conclusivos e devem ser ponderados com outras variáveis, podendo ser utilizados como indicadores preliminares de risco, mas seus valores não são definitivos e devem ser utilizados com cautela.

O modelo de risco é ponderado para todo o Estado de Santa Catarina e indica quais os segmentos de rodovias têm maior probabilidade de atropelamento em comparação a todos os

demaís. O valor do nível, em um trecho específico, foi comparado com os outros 6.211 trechos e recebeu sua probabilidade de risco. Todos os trechos de estradas possuem alguma probabilidade de atropelamento, mesmo aqueles definidos com o risco zero na modelagem aqui realizada.

A análise de uma rodovia específica deve ser ponderada para os valores nela existentes. Por exemplo, se uma rodovia de 50 quilômetros possui apenas trechos com níveis de risco variando de 0 a 4, a tomada de decisão para o trecho deve considerar os trechos de nível 4 como de prioridade máxima. Contudo, se a tomada de decisão estiver buscando respostas de priorização entre duas rodovias e uma delas possuir trechos com níveis 5 ou 6, estes devem ser priorizados para ações de mitigação.

Os modelos de risco aqui modelados são um retrato instantâneo da realidade conhecida dos impactos potenciais. Sua assertividade se reduzirá com o passar do tempo, devendo ser refeito. Mudanças nas características ambientais, como a mudança da paisagem, ou das características da estrada, como alteração de pavimento e fluxo de veículos, alteram substancialmente a precisão dos modelos. É possível uma reanálise exclusiva destes trechos. Por ser um instantâneo, existe a possibilidade de uma ou mais espécies serem afetadas com diferentes intensidades no decorrer do tempo. Um bom exemplo é a mudança de comportamento de capivaras conforme a precipitação e a variação do nível dos ambientes aquáticos. Modelos sazonais podem ser desenvolvidos no futuro. Da mesma forma, os modelos gerados consideram a ocorrência de espécies independente do horário do dia, uma vez que o objetivo do projeto foi identificar os trechos com maior risco de atropelamento. É possível modelar risco para os períodos diurnos, noturnos e crepusculares, mas estas análises somente se justificam para a implantação de ações de mitigação vinculadas ao aplicativo U-Safe.

A modelagem preditiva, como demonstrado anteriormente, obteve um nível de assertividade em Santa Catarina que chegou a 86%. Entretanto, modelos são falhos em algum nível e, existindo dados reais, coletados e analisados de forma adequada, sempre devem prevalecer, mesmo que contraponham aos resultados dos modelos.

3.6 Usos potenciais do MARA

O MARA é uma tecnologia inovadora, que aprimora a tomada de decisões sobre diferentes aspectos da gestão ambiental de estradas em grandes áreas geográficas. No caso específico da malha viária de Santa Catarina, o MARA poderá contribuir nos seguintes aspectos:

3.6.1 Aperfeiçoar termos de referência

O PROMIRF como um programa de mitigação aos efeitos de estradas na fauna integra um grande conjunto de ações que aperfeiçoam a proposição de estudos e termos de referência para estradas no Estado de Santa Catarina. Os protocolos de monitoramento de atropelamento e medidas de mitigação são fundamentais para estabelecer um padrão de coleta de dados comparáveis no tempo e no espaço. Entretanto, o MARA é o maior diferencial do PROMIRF, uma vez que permite ao empreendedor otimizar recursos, aplicando-os em áreas previamente definidas como de maior impacto à fauna silvestre.

Em termos operacionais, se sugere os seguintes passos na elaboração de termos de referência baseados no MARA:

- 1 Acessar a camada "Modelos de Risco de Atropelamento", disponibilizada em formato .shp e .kml, e identificar o segmento de estrada (início e fim) que se pretende avaliar;
- 2 Identificar os trechos com maior nível de risco, priorizando aqueles com valores 5 e 6;
- 3 Acessar as camadas dos modelos bio-ecológicos de cada uma das 30 espécies e verificar aquelas com valores superiores a 70;
- 4 O passo anterior é subjetivo e dependendo do objetivo e número de espécies o valor do limiar pode ser remanejado para mais ou para menos. Sempre lembrar que os modelos bio-ecológicos são indicativos e não devem ser utilizados sozinhos na tomada de decisão;
- 5 De posse da lista de espécies focais, avaliar a possibilidade de priorização de áreas de amostragem. Importante lembrar que todas as informações se baseiam em dados de mamíferos de médio e grande porte.

3.6.2 Aprimorar a tomada de decisão de ações de mitigação

Neste tópico, o MARA pode contribuir na tomada de decisão de ações de mitigação em todas as fases de implantação de uma estrada.

Na fase de projeto, o MARA pode ser utilizado conforme descrito no tópico anterior e auxiliar na proposição prévia de medidas de mitigação, qualificando o projeto executivo e permitindo uma melhor valoração dos custos de implantação da rodovia.

Na fase de planejamento e implantação, além das oportunidades já descritas, é possível se disponibilizar os trechos de riscos para a população, seja através de placas de sinalização ou diretamente em seus celulares, utilizando a tecnologia embarcada de proteção. A vantagem deste último método é sua rapidez para implantação e sua elevada taxa de sucesso, podendo reduzir em até 100% os atropelamentos nos trechos indicados.

3.6.3 Programas de concessões de rodovias

Um dos maiores entraves aos programas de concessões de rodovias são os custos que surgem após a submissão das propostas de valores. Um destes custos são as ações de mitigação, as quais normalmente são desconhecidas na fase pré-contrato e posteriormente acarretam longos processos de readequação orçamentária. O uso dos modelos de risco, sobretudo os de níveis 5 e 6, no planejamento de longo prazo de concessões, viabiliza estudos de viabilidade mais precisos e a qualificação das propostas frente às demandas de sustentabilidade e ESG.

3.6.4 Proteção à vida

Os modelos de risco do MARA são a principal informação necessária para se disponibilizar proteção aos usuários de estradas de todo o Estado de Santa Catarina, através do aplicativo U-Safe.

Os dados do MARA podem ser liberados para toda a sociedade, gratuitamente, desde que sejam adaptados para o modelo de dados do aplicativo U-Safe, gerando redução de acidentes na totalidade da malha viária do Estado e, por consequência, salvando vidas humanas e de milhares de animais. Importante ressaltar que os dados podem ser priorizados para áreas como unidades de conservação e outras regiões de elevada sensibilidade.

4 Estudo de caso

O presente estudo de caso será utilizado para exemplificar uma das formas de uso dos modelos de risco na tomada de decisão para ações de mitigação. De forma secundária, esta análise demonstrará a efetividade do processo de modelagem comparando os resultados aqui obtidos e um processo de licenciamento já implementado. Vale ressaltar que toda a comparação está sendo realizada após a modelagem estar concluída, não havendo uso de informações do processo de licenciamento para a tomada de decisão.

Para efeito deste estudo de caso será considerado o trecho de 55 quilômetros entre Dr. Pedrinho e Moema, na rodovia SC-477, entre as coordenadas (-26,56650491, -49,67399881) e (-26,5624441, -49,80019327).

A análise do referido trecho, considerando os resultados do modelo de risco de atropelamento de fauna silvestre, mostra que este segmento é uma área de elevado risco à fauna, requerendo ações de mitigação para minimizar seus efeitos. Se observa que todos os trechos possuem risco de atropelamento de nível 4 ou maior. Dos 55 quilômetros, 12 possuem nível 4; 31 têm nível 5 e 12 têm nível 6. Como afirmado anteriormente, trechos de nível 5 e 6 devem ser priorizados na tomada de decisão para ações de mitigação, entretanto, devido o trecho demonstrar uma elevada complexidade, a estruturação de uma estratégia de mitigação deve se basear nos trechos de nível 6.

Os 12 quilômetros com nível 6 estão distribuídos em quatro trechos que variam de um a seis quilômetros de extensão (Tabela 19 e Figura 22). Considerando estes trechos, é possível se utilizar as camadas de cada espécie (modelos bio-ecológicos) e avaliar aquelas que possuem pesos superiores a 70.

Constata-se que, das 30 espécies modeladas, nove possuem pesos igual ou superior a 70 (Tabela 19). Agora, como afirmado de forma recorrente no texto, a análise com foco nas espécies requer extrema cautela e deve ser conduzida por profissional que possua bom conhecimento da biologia e ecologia das espécies analisadas. Também é necessário lembrar que os modelos bio-ecológicos são uma fase preliminar do processo total da geração dos modelos de risco, sendo posteriormente refinados e aprimorados por análises complementares, como o *machine learning*.

A título de exemplo, considere que os modelos bio-ecológicos analisam as condições de um determinado ponto e informam se neste local existem características que ampliam a probabilidade de risco de atropelamento caso a espécie habite aquele local. Entretanto, o modelo bio-ecológico não avalia se a espécie realmente vive no ponto. Esta discussão é necessária devido a presença do caititu (*Pecari tajacu*) na lista de espécies de alguns dos trechos

analisados. O banco de dados de Santa Catarina possui apenas oito registros da espécie, o que indica uma baixa probabilidade da sua ocorrência atual na região. Os modelos bio-ecológicos nos informam que, havendo a presença de caititu na região, estes trechos devem ser avaliados com foco na espécie.

Por outro lado, havendo algum requerimento específico que deva ser analisado, como o potencial risco de atropelamento de uma determinada espécie (p.ex. *Puma concolor*), é possível utilizar os modelos bio-ecológico como norteadores das áreas com maior potencial de ocorrência de atropelamentos. Nos 55 quilômetros avaliados os valores dos modelos bio-ecológicos da onça-parda variaram de 28 a 57. O trecho com os maiores valores, variando entre 55 e 57, se encontra em um segmento contínuo de 11 quilômetros localizados entre as coordenadas (-26,65282277, -49,51846228) e (-26,5765638, -49,54514432) e seis destes quilômetros se sobrepõem ao trecho 3 da Tabela 19.

Tabela 19.- Segmentos da rodovia SC-477 com nível 6 de risco de atropelamento de fauna silvestres e as espécies modeladas que obtiveram pesos de ocorrência superiores a 70.

Trecho	Lat. Inicial	Long. Inicial	Lat. Final	Long. Final	Extensão (Km)	Espécies	Observação
1	-26.5636199	-49.65642389	-26.55494044	-49.65201495	1	<i>Cabassous tatouay</i> , <i>Leopardus pardalis</i> , <i>Dasypus novemcinctus</i> , <i>Mazama gouazoubira</i> , <i>Cerdocyon thous</i> , <i>Tamandua tetradactyla</i> , <i>Procyon cancrivorus</i>	
2	-26.54707287	-49.56564493	-26.56107068	-49.55337799	3	<i>Cabassous tatouay</i> , <i>Leopardus pardalis</i> , <i>Dasypus novemcinctus</i> , <i>Mazama gouazoubira</i> , <i>Cerdocyon thous</i> , <i>Tamandua tetradactyla</i> , <i>Procyon cancrivorus</i>	<i>Mazama gouazoubira</i> esteve presente em apenas um dos três quilômetros. As demais espécies possuem risco em todo o trecho
3	-26.5765638	-49.54514432	-26.61905934	-49.52083789	6	<i>Cabassous tatouay</i> , <i>Leopardus pardalis</i> , <i>Dasypus novemcinctus</i> , <i>Mazama gouazoubira</i> , <i>Leopardus wiedii</i> , <i>Pecari tajacu</i> , <i>Cerdocyon thous</i> , <i>Tamandua tetradactyla</i>	<i>Cabassous tatouay</i> , <i>Cerdocyon thous</i> e <i>Tamandua tetradactyla</i> estiveram presentes em apenas um dos seis quilômetros deste trecho. As demais espécies possuem risco em todo o trecho
4	-26.69644677	-49.50306151	-26.7106696	-49.49192286	2	<i>Leopardus pardalis</i> , <i>Dasypus novemcinctus</i> , <i>Mazama gouazoubira</i> , <i>Leopardus wiedii</i> , <i>Pecari tajacu</i>	<i>Dasypus novemcinctus</i> e <i>Mazama gouazoubira</i> estiveram presentes em apenas um dos dois quilômetros deste trecho. As demais espécies possuem risco em todo o trecho

Figura 22.- Área avaliada da rodovia SC-477 utilizando o modelo de risco de atropelamento de fauna silvestre atropelada, com indicação dos quatro trechos identificados como de maior prioridade para ações de mitigação (nível 6).



4.1 Licenciamento ambiental prévio

O levantamento de medidas de mitigação instaladas na rodovia SC-477, no trecho em análise, indica a presença de, pelo menos, 27 passagens de fauna de diferentes dimensões. Correlacionando-se a posição destas passagens e cada trecho analisado pelo MARA foi possível verificar que dos 12 quilômetros indicados como de risco 6, foram instaladas nove passagens inferiores, implicando em uma precisão de 75% na seleção de áreas críticas de atropelamento (Tabela 20). Considerando que os trechos indicados com nível 5 devem ser menos representativos que aqueles de nível 6, identificamos uma convergência de 52%, tendo 16 passagens nos trechos indicados com o risco 5. Finalmente, os 12 quilômetros indicados como de nível 4, receberam apenas duas passagens com base nos estudos de licenciamento prévio, mais uma vez reforçando a precisão do MARA em indicar áreas que realmente são significativas para a mitigação de impactos.

Tabela 20.-Comparação da extensão definida pelo MARA para cada nível de risco na SC-477 e o número de passagens de fauna instaladas em virtude dos estudos de licenciamento ambiental.

Risco	Extensão (KM)	N passagens	% de cobertura
4	12	2	17
5	31	16	52
6	12	9	75

Se considerarmos que os modelos do MARA foram gerados sem a necessidade de um único dia de campo, utilizando bancos de dados pré-existent e baseando sua tomada de decisão em conhecimento científico e robustos processos analíticos, os resultados obtidos demonstram uma elevada assertividade. O MARA foi capaz de prever a necessidade da instalação de passagens em 75% dos trechos indicados como de maior risco. Em termos ambientais, este resultado demonstra que o uso do MARA em diferentes fases do ciclo de vida de uma rodovia pode qualificar a localização de passagens de fauna e a definição de outras estratégias de mitigação. Agora, em termos econômicos, o MARA pode gerar milhões em economia com estudos ambientais prévios, permitindo que recursos sejam melhor alocados em ações de monitoramento e aprimoramento de ações de mitigação.

Outro ponto que reforça a assertividade do MARA foi sua capacidade de identificar a principal área de risco de atropelamento de onça-parda com uma precisão inferior a 2 quilômetros. O licenciamento ambiental da SC 477 teve como base um estudo realizado por Mazzolli (2013), o qual não tivemos acesso, mas que relatórios de monitoramentos de fauna na região sugerem ter sido a base para o posicionamento das passagens de fauna. Se desconhece os métodos, período de estudo ou os custos envolvidos neste estudo, mas o MARA, dos 55

quilômetros considerados, indicou 11 como prioritários para a onça-parda. Destes 11 quilômetros, os dois trechos com valores máximos (57) para a onça-parda foram alocados pelo MARA a menos de 2.000 metros da maior passagem existente no trecho (7x4m) e que foi instalada com foco em mitigar potenciais riscos à onça-parda.

Espera-se que este estudo de caso e sua comparação com os resultados obtidos pelos estudos destinados ao licenciamento ambiental da SC-477 demonstrem o elevado grau de precisão e assertividade que o MARA pode propiciar. Não é esperado que o MARA substitua totalmente estudos em campo ou monitoramentos de atropelamento e medidas de mitigação, mas que seus resultados sejam empregados como uma ferramenta complementar e de elevada confiabilidade em diferentes fases do ciclo de vida de qualquer estrada estadual do Estado de Santa Catarina.

Referências

- BROWN et al. SDMtoolbox 2.0: the next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *PeerJ* 5:e4095. 2017; DOI 10.7717/peerj.4095
- CRUZ-CÁRDENAS, G. et al. Selección de predictores ambientales para el modelado de la distribución de especies en MaxEnt. *Revista Chapingo. Serie ciencias forestales y del ambiente*, v. 20, n. 2, p. 187-201, 2014.
- DALAPICOLLA, J. Tutorial de modelos de distribuição de espécies: Guia prático usando o MaxEnt e o ArcGIS 10. Laboratório de Mastozoologia e Biogeografia, Universidade Federal do Espírito Santo, Vitória. 2016. Disponível em: <http://blog.ufes.br/lamab/tutoriais>
- HALVORSEN, R. A strict maximum likelihood explanation of MaxEnt, and some implications for distribution modeling. *Sommerfeltia*, 36, n. 1, 2013. 1-132.
- GAUCHEREL, C. et al. Spatial analysis of endemism to redefine conservation areas in Western Ghats (India). *Journal for Nature Conservation*, 34, 2016. 33-41.
- PHILLIPS, S. J.; DUDÍK, M.. Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation. *Ecography* 31: 161-175, 2008.
- WARREN, D. L. et al. Incorporating model complexity and spatial sampling bias into ecological niche models of climate change risks faced by 90 California vertebrate species of concern. *Diversity and distributions*, v. 20, n. 3, p. 334-343, 2014.
- VELOZ, S. D. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of biogeography*, v. 36, n. 12, p. 2290-2299, 2009.

Anexos

Anexo ao MARA são encaminhados os arquivos listados abaixo:

- Pasta MODELOS DE RISCO DE ATROPELAMENTO: Arquivos/Layers referentes ao resultado deste contrato. Onde poderão ser consultados, km a km, o risco das rodovias estaduais do Estado de Santa Catarina. Nestes arquivos, tanto os destinados aos sistemas de informações geográficas, quanto às planilhas de dados, o(a) usuário(a) poderá obter o valor final, denominado de nível de risco de atropelamento, para cada trecho da malha viária estadual de Santa Catarina, com valores que variam entre 0 e 6.
 - Shapefile “Modelos de risco de atropelamento”, em SIRGAS 2000
 - Arquivo de Excel “Modelos de risco de atropelamento”, com o risco final para cada segmento de um quilômetro. O arquivo contém uma aba explicativa das colunas da tabela.
 - Arquivo KML “Modelos de risco de atropelamento”
- Pasta MODELOS BIO-ECOLÓGICOS PONDERADOS: Arquivos/Layers referentes à Base de Dados dos Modelos Bio-ecológicos, utilizada para geração do Modelo de Risco de Atropelamento. Os dados originais foram normalizados para que os pesos calculados para cada espécie, em cada segmento da malha viária estadual de Santa Catarina variassem em uma amplitude entre 0 e 100. Estão sendo disponibilizados os dados individualizados das 30 espécies analisadas no MARA. Nestes arquivos, tanto os destinados aos sistemas de informações geográficas, quanto às planilhas de dados, o(a) usuário(a) poderá obter o valor final, denominado de peso de risco de atropelamento, para cada trecho da malha viária estadual de Santa Catarina, com valores que variam entre 0 e 100.
 - Shapefile “Modelos Bio-Ecológicos Ponderados - Ponderada total”, em SIRGAS 2000, com os riscos ponderados de todas as 30 espécies;
 - Arquivo de Excel “Modelos Bio-Ecológicos Ponderados - Ponderada total”, com os riscos ponderados de todas as 30 espécies para cada segmento de um quilômetro. O arquivo contém uma aba explicativa das colunas da tabela.
 - Shapefiles com os nomes das espécies, em SIRGAS 2000, com o risco ponderado da respectiva espécie.