

Modelling tools to predict potential distribution of forest species: using Pico Island and the Azores as a study case

Tese de Doutoramento

Lara Dutra Silva

Doutoramento em

Biologia



Modelling tools to predict potential distribution of forest species: using Pico Island and the Azores as a study case

Tese de Doutoramento

Lara Dutra Silva

Orientadores

Professor Doutor Luís Filipe Dias e Silva

Professor Doutor Rui Miguel Pires Bento da Silva Elias

Professor Doutor Mário Luís Gomes Ramalho Alves

Tese especialmente elaborada para obtenção do grau de Doutor em Biologia



Copyright © Lara Dutra Silva, 2018

All Rights Reserved

Title

Modelling tools to predict potential distribution of forest species: using Pico Island and the Azores as a study case

Author

Lara Dutra Silva¹

Advisors

Luís Silva¹, Rui Bento Elias² and Mário Alves³

¹InBIO, Rede de Investigação em Biodiversidade, Laboratório Associado, CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Polo-Açores, Universidade dos Açores, 9501-801 Ponta Delgada, Açores, Portugal, e-mail: laradutrasilva@gmail.com, luis.fd.silva@uac.pt

²CE3C – Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group, Universidade dos Açores, Faculdade de Ciências Agrárias e do Ambiente, 9700-042, Angra do Heroísmo, Açores, Portugal, e-mail: rui.mp.elias@uac.pt

³NATURALREASON, Lda, Caminho do Meio Velho, 5-B, 9760-114 Cabo da Praia, Açores, Portugal, e-mail: mario.alves@naturalreason.pt

Photographs

Pedro Cabral e Dutra (Lagoa das Furnas, São Miguel, Açores and Jardim Terra Nostra, São Miguel, Açores)

Welcomia (Scenic Forest Sunset)

Volodomir Kalina (Forest Trees)

Smileus (Bird's-eye view of a scenic sunset over the forest hills, with toned dramatic colors)

Vortex Magazine (Vinhas do Pico) and **Design** Pedro Cabral e Dutra



To my wonderful niece Anna Petra and
to my parents for raising me to believe that everything is possible.

Love, kindness, respect and dedication
are a source of inspiration to live.

PREFACE

This thesis is about the process of gaining knowledge about ecology using models and data of forest species. An overarching motivation is that all ecologists seek to solve research problems in field of biodiversity and how it affects ecosystems. Doing Ph.D. research is a challenge, but it also provided an opportunity for independent learning and creativity.

In the last decades, the interest in species distribution models (SDMs) has increased dramatically, as reflected by the number of publications. An SDM is a mathematical description of the species distribution in environmental space that can be used to predict the distribution of species in the geographic space. The models combine species occurrence records and environmental variables and the result can be projected to any geographic space to identify regions that are suitable for the target species. Those models allow us to potentially forecast anthropogenic effects on patterns of biodiversity at different spatial scales. Most modelling approaches developed are an invaluable tool for anticipating the potential range of invasive species. Understanding the relative importance of the variables that can control distributions is crucial, especially for invasive plant species that may pose significant challenges about managing indigenous biodiversity. The purpose of this thesis was to demonstrate the importance of SDMs, as modelling tools that related the distribution and occurrence of forest species to the environmental characteristics of the location from where it has been recorder (archipelago of Azores). The results of the thesis had satisfied the statistical performance measures; however, it was always subjected to doubts and questions by ecologists' community. It is my hope that at least the outcome of this study contributes to scientific knowledge and becomes a steppingstone in realize the importance of understanding the Azorean biodiversity.

Knowing more and knowing better is the healthiest ambition in the science fields.

Lara Dutra Silva

Ponta Delgada, June 2018

ABSTRACT

Species distribution models (SDMs) have been used in different areas within ecology, namely to model the potential spread of invasive species, to evaluate and manage priority species for conservation and to support forest management. An SDM is a mathematical description of the species distribution in the environmental space that can be used to predict the distribution of the species in the geographic space. The advances in computational capabilities have provided increasingly greater and more intensive statistical algorithms than was previously possible, as reflected by the increasing number of publications addressing SDMs and also the growing variety of modelling approaches. In the Azores, the growing abundance of the species distribution data, the diversity on island size and morphology, and the different spatial patterns that are possible among islands and species, make the archipelago a good model for the comparison of different modelling approaches and to test possible modelling constraints. The research questions addressed in this thesis were: (i) Did modelling approaches based on different theoretical backgrounds provide similar pictures of the potential distribution of forest species? (ii) Were there any relevant differences when calculating Generalized Linear Models (GLMs) using maximum likelihood methods or Bayesian methods? (iii) Was there a clear advantage in the use of a random field providing the spatial structure of the data in comparison with fixed effects models? (iv) Were different modelling approaches consistent when focusing on a limited number of available environmental variables? (v) Were modelling approaches profoundly affected by sample size, the type of species distribution and land use changes? To answer those questions, three modelling exercises were conducted: (i) A comparison of Ecological Niche Factor Analysis (ENFA) and Maximum Entropy Modelling (MaxEnt) using presence-only data for three species (*Pittosporum undulatum*, *Acacia melanoxylon* and *Morella faya*) in three islands (Pico,

Terceira and São Miguel) and including the effect of data reduction; (ii) A comparison of fixed effects and mixed effects models using an R platform for the calculation of GLMs and Integrated Nested Laplace Approximation (INLA), allowing the calculation of the spatial structure (Matérn covariance function), based on data from two islands (Pico and São Miguel) for two species (*P. undulatum* and *M. faya*), and including the effect of data reduction; and (iii) A comparison of GLMs and a selection of machine learning techniques used to model the potential change in the distribution range of *P. undulatum*, *A. melanoxydon* and *M. faya* in the three islands, according to the climate changes predicted to 2100. The results indicated that (i) both approaches provided similar results, particularly when the amount of information explained by ENFA was high, but models were clearly affected by reduction of sample size, models with better predictive accuracy included a specific set of environmental variables (combination of topographic, climatic and land use variables), and modelling approaches were somewhat sensitive to a transfer of the selected models to a different habitat (i.e. island); (ii) GLMs calculated using maximum likelihood methods or Bayesian methods provided very similar results, even with data reduction, and the addition of a random field increased model adjustment, particularly for the less abundant tree, *M. faya*, although the structure of the random field was clearly affected by sample size; and (iii) some limitations seem to exist when modelling the effect of climate change on species distributions, since the best models also included topographic variables, making modelling based on climate alone less reliable, with model fit varying among modelling approaches, and Random Forest often providing the best results. Overall, the results of this research can be expanded to support Azorean forestry management, and could be replicated in other island systems and forest regions, not only in projects addressing the ecology of particular forest species, but also when handling research questions related with the prediction of plant invader success and

expansion, the detection of areas potentially suited for restoration projects, modelling based on remote sense data, and modelling of the potential effect of climate change.

Keywords: Azores, climate change, ENFA, fixed effects models, INLA, MaxEnt, mixed effects models, modelling, SDMs.

RESUMO

Os modelos de distribuição de espécies (SDMs) têm sido aplicados em diferentes áreas da ecologia, nomeadamente para modelar a distribuição potencial de espécies invasoras, para avaliar espécies prioritárias no âmbito da conservação e para apoiar o planeamento florestal. Um SDM é uma descrição matemática da distribuição de uma espécie no espaço ambiental, a qual pode ser utilizada para prever a distribuição da espécie no espaço geográfico. O avanço ao nível da capacidade computacional disponibilizou uma diversidade de métodos estatísticos, que anteriormente não era possível utilizar. Esta diversidade de métodos reflete-se num número crescente de publicações direcionadas ao estudo e aplicação dos SDMs e também numa variedade crescente de métodos de modelação. Nos Açores, a abundância crescente de dados corológicos, a diversidade geomorfológica do arquipélago e os diferentes padrões espaciais que é possível encontrar em diferentes ilhas e em diferentes espécies, contribuem para que o arquipélago seja um bom modelo para a comparação de diferentes abordagens de modelação, bem como para testar possíveis constrangimentos inerentes ao processo de modelação. As perguntas de investigação a que pretendemos responder nesta tese foram as seguintes: (i) As abordagens de modelação, baseadas em diferentes fundamentos teóricos, originam resultados semelhantes, ao nível da distribuição potencial das espécies florestais estudadas? (ii) Existe alguma diferença relevante, entre o cálculo de Modelos Lineares Generalizados (GLMs) usando métodos de máxima verossimilhança ou métodos bayesianos? (iii) Existe alguma vantagem, no uso de um campo aleatório relativo à estrutura espacial dos dados, em comparação com os modelos que incluem apenas os efeitos fixos das variáveis ambientais? (iv) As diferentes abordagens de modelação originam resultados consistentes, em particular quando o número de variáveis ambientais

utilizadas na modelação é reduzido? (v) As diferentes técnicas de modelação são afetadas de um modo relevante pela dimensão da amostra, pelo tipo de distribuição da espécie e pelas alterações no uso do solo? Para responder a estas questões, foram desenvolvidos três exercícios de modelação: (i) Uma comparação da Análise Fatorial do Nicho Ecológico (ENFA) e da modelação baseada na Máxima Entropia (MaxEnt), utilizando dados relativos à presença de três espécies (*Pittosporum undulatum*, *Acacia melanoxylon* e *Morella faya*) em três ilhas (Pico, Terceira e São Miguel), e incluindo o efeito da redução da dimensão da amostra; (ii) A comparação de modelos com efeitos fixos ou mistos, utilizando a plataforma R para o cálculo de GLMs e da aproximação de Laplace (INLA), permitindo o cálculo da estrutura espacial dos dados (função de covariância de Matérn), baseada em dados de duas ilhas (Pico e São Miguel) para duas espécies (*P. undulatum* e *M. faya*), e incluindo o efeito da redução da dimensão da amostra; e (iii) A comparação de GLMs e de uma seleção de algoritmos de autoaprendizagem (*Machine Learning*), usados para modelar as possíveis alterações nas áreas de distribuição de *P. undulatum*, *A. melanoxylon* e *M. faya* nas três ilhas, resultantes das alterações climáticas previstas para 2100. Em relação ao primeiro exercício, ambas as abordagens originaram cenários semelhantes, particularmente quando a quantidade de informação explicada pela ENFA era elevada; os resultados da modelação foram afetados pela redução do tamanho da amostra; os modelos com melhor capacidade de previsão incluíam um conjunto variado de variáveis ambientais (topográficas, climáticas e de uso do solo); e os modelos eram afetados pela transferência para um novo habitat (i.e. ilha). Os resultados do segundo exercício de modelação indicaram que os GLMs, calculados através de métodos de máxima verossimilhança ou métodos bayesianos originaram resultados similares, mesmo nos casos em que a dimensão da amostra era reduzida; e que a adição de um campo aleatório aumentou o ajustamento dos modelos, particularmente para a árvore

menos abundante, *M. faya*, embora a estrutura do campo aleatório fosse claramente afetada pela dimensão da amostra. O terceiro exercício de modelação revelou que existem várias limitações quando se modela o efeito das alterações climáticas na distribuição das espécies, uma vez que os melhores modelos incluíram variáveis topográficas, demonstrando que a modelação baseada somente no clima poderá não ser fiável; verificou-se igualmente que o ajuste dos modelos variava de forma relevante entre as diferentes abordagens de modelação, e que o algoritmo *Random Forest* apresentou, em geral, os melhores resultados. De uma forma geral, os resultados desta investigação poderão ser aplicados como forma de apoio à gestão da floresta açoriana. Poderão ser replicados em outros sistemas insulares e noutras regiões florestais, não somente em projetos direcionados para a ecologia das espécies florestais, mas também em questões de investigação relacionadas com a previsão do sucesso e expansão das plantas invasoras, a deteção de áreas adequadas para projetos de restauro, a modelação baseada em dados de deteção remota e a modelação do efeito potencial das alterações climáticas.

Palavras-chave: Açores, alterações climáticas, Análise Fatorial do Nicho Ecológico, Aproximação de Laplace, Máxima Entropia, modelação, modelos com efeitos fixos, modelos com efeitos mistos, modelos de distribuição de espécies.

TABLE OF CONTENTS

PREFACE	VI
ABSTRACT.....	VII
RESUMO	X
ACKNOWLEDGEMENTS	XX
LIST OF ACRONYMS/NOMENCLATURE	XXI
GENERAL INTRODUCTION AND PURPOSE STATEMENT	1
INTRODUCTION.....	1
SPECIES DISTRIBUTION MODELS AND FOREST MANAGEMENT.....	4
THE IMPORTANCE OF MODELLING POTENTIAL EFFECTS OF CLIMATE CHANGE	9
PROPOSED OBJECTIVES.....	13
RESEARCH QUESTIONS	13
(I) FIRST MODELLING EXERCISE: A COMPARISON OF ENFA AND MAXENT FOR PRESENCE-ONLY DATA	14
(II) SECOND MODELLING EXERCISE: A COMPARISON OF FIXED AND MIXED EFFECTS MODELS USING INLA	15
(III) THIRD MODELLING EXERCISE: LIMITATIONS OF SPECIES DISTRIBUTION MODELS BASED ON AVAILABLE CLIMATE CHANGE VARIABLES	16
CHAPTER I	17
INTRODUCTION.....	17
1.1. ABOUT MODELS AND MODELLING	17
1.2. SPECIES DISTRIBUTION MODELS – AN ECOLOGICAL CONTEXT.....	18
1.3. TYPES OF BIOTIC DATA IN SMDs.....	21
1.4. SELECTION OF ENVIRONMENTAL VARIABLES.....	24
1.5. STATISTICAL ALGORITHMS AND FITTING PROCEDURES.....	27
1.5.1. GENERALIZED LINEAR MODELS	27
1.5.2. GENERALIZED ADDITIVE MODELS	28
1.5.3. ECOLOGICAL NICHE FACTOR ANALYSIS	29
1.5.4. SURFACE RANGE ENVELOPE	31
1.5.5. QUANTILE REGRESSION	32
1.5.6. STRUCTURAL EQUATION MODELLING.....	32
1.5.7. GEOGRAPHICALLY WEIGHTED REGRESSION	33
1.5.8. MULTIVARIATE ADAPTIVE REGRESSION SPLINES	34
1.5.9. ARTIFICIAL NEURAL NETWORKS.....	35
1.5.10. CLASSIFICATION AND REGRESSION TREES	36
1.5.11. CLASSIFICATION TREE ANALYSIS	37
1.5.12. BOOSTED REGRESSION TREES	38
1.5.13. RANDOM FOREST	39
1.5.14. BAYESIAN ANALYSIS AND APPROACH.....	39
1.5.15. GENERALIZED DISSIMILARITY MODELLING	43
1.5.16. ENTROPY MAXIMIZATION (MAXENT)	44

1.5.17. GENETIC ALGORITHM FOR RULE SET PRODUCTION	46
1.6. PLANT INVADERS AND MODELLING – BIBLIOMETRIC ANALYSIS AND REVIEW	47
1.7. ADVANTAGES AND LIMITATIONS OF MODELLING APPROACHES	53
CHAPTER 2	65
MATERIALS AND METHODS	65
2.1. STUDY AREA	65
2.1.1. PICO ISLAND	67
2.1.2. TERCEIRA ISLAND	68
2.1.3. SÃO MIGUEL ISLAND	68
2.2. TARGET SPECIES	69
2.2.1. <i>PITTOSPORUM UNDULATUM</i>	69
2.2.2. <i>ACACIA MELANOXYLON</i>	70
2.2.3. <i>MORELLA FAYA</i>	71
2.3. SPECIES DATA	73
2.4. ECOGEOGRAPHICAL VARIABLES	74
2.4.1. TOPOGRAPHICAL VARIABLES	74
2.4.2. CLIMATIC VARIABLES	76
2.4.3. LAND USE VARIABLES	76
2.5. FIRST MODELLING EXERCISE: A COMPARISON OF ENFA AND MAXENT FOR PRESENCE-ONLY DATA ...	77
2.5.1. SPECIES DATA	77
2.5.2. EGVs	78
2.5.3. MODELLING APPROACH	78
2.5.4. MODEL VALIDATION	79
2.6. SECOND MODELLING EXERCISE: A COMPARISON OF FIXED AND MIXED EFFECTS MODELS USING INLA	81
2.6.1. SPECIES DATA	81
2.6.2. EGVs	82
2.6.3. MODELLING APPROACH	82
2.6.4. MODEL VALIDATION	84
2.6.5. CALCULATION OF THE POSTERIOR DISTRIBUTION AND OF THE RANDOM FIELD	86
2.7. THIRD MODELLING EXERCISE: LIMITATIONS OF SPECIES DISTRIBUTION MODELS BASED ON AVAILABLE CLIMATE CHANGE VARIABLES	88
2.7.1. SPECIES DATA	88
2.7.2. EGVs	88
2.7.3. MODELLING APPROACH	90
2.7.4. MODEL VALIDATION	91
2.7.5. MODEL PROJECTIONS	92
CHAPTER 3	93
RESULTS	93
3.1. FIRST MODELLING EXERCISE: A COMPARISON OF ENFA AND MAXENT FOR PRESENCE-ONLY DATA ...	93
3.1.1. SELECTED EGVs	93

3.1.2. ENFA.....	93
3.1.3. MAXENT	94
3.1.4. COMPARATIVE PERFORMANCE OF ENFA AND MAXENT	99
3.1.5. TESTING THE EGV SETS FOR OTHER ISLANDS/SPECIES	105
3.2. SECOND MODELLING EXERCISE: A COMPARISON OF FIXED AND MIXED EFFECTS MODELS USING INLA	110
3.2.1. PRELIMINARY SELECTION OF MODELS	110
3.2.2. FIXED EFFECTS MODELS	110
3.2.3. MIXED EFFECTS MODELS	112
3.2.4. PICO ISLAND.....	115
3.2.5. SÃO MIGUEL ISLAND.....	115
3.2.6. PREDICTIONS OF THE RANDOM FIELD	119
3.2.6.1. PICO ISLAND.....	121
3.2.6.2. SÃO MIGUEL ISLAND.....	121
3.2.7. PREDICTION OF THE RESPONSE GRID.....	126
3.2.7.1. PICO ISLAND.....	126
3.2.7.2. SÃO MIGUEL ISLAND.....	129
3.3. THIRD MODELLING EXERCISE: LIMITATIONS OF SPECIES DISTRIBUTION MODELS BASED ON AVAILABLE CLIMATE CHANGE VARIABLES	132
3.3.1. SUMMARY OF THE BEST MODELS	132
3.3.2. RELATIVE IMPORTANCE OF EGVs	137
3.3.3. PRESENT AND FUTURE PROJECTIONS	137
CHAPTER 4.....	143
DISCUSSION	143
4.1. FIRST MODELLING EXERCISE: A COMPARISON OF ENFA AND MAXENT FOR PRESENCE-ONLY DATA .	143
4.2. SECOND MODELLING EXERCISE: A COMPARISON OF FIXED AND MIXED EFFECTS MODELS USING INLA	148
4.3. THIRD MODELLING EXERCISE: LIMITATIONS OF SPECIES DISTRIBUTION MODELS BASED ON AVAILABLE CLIMATE CHANGE VARIABLES	154
4.4. GLOBAL DISCUSSION	161
CHAPTER 5.....	166
CONCLUSIONS	166
REFERENCES	170
APPENDICES	
APPENDIX S1. GLOSSARY – DEFINITIONS OF SCIENTIFIC TERMS	1-A
REFERENCES	8-A
APPENDIX S2. GENERALIZED LINEAR MODELS.....	16-A
APPENDIX S3. THE INTEGRATED NESTED LAPLACE APPROXIMATION (INLA)	18-A
APPENDIX S4. STOCHASTIC PARTIAL DIFFERENTIAL EQUATION (SPDE)	21-A
REFERENCES	25-A
APPENDIX S5. LIST OF JOURNALS WITH ARTICLES ON SDMS	30-A
REFERENCES	32-A

APPENDIX S6. SECOND MODELLING EXERCISE: A COMPARISON OF FIXED AND MIXED EFFECTS MODELS USING INLA	62-A
APPENDIX S7. THIRD MODELLING EXERCISE: LIMITATIONS OF SPECIES DISTRIBUTION MODELS BASED ON AVAILABLE CLIMATE CHANGE VARIABLES	64-A
APPENDIX S8. COMPLEMENTARY LITERATURE	70-A
APPENDIX S9. PUBLICATION I	82-A
APPENDIX S10. PUBLICATION II.....	84-A
APPENDIX S11. ADDITIONAL INFORMATION	86-A

LIST OF FIGURES

CHAPTER 1

FIGURE 1.6.1. YEAR OF PUBLICATION PER NUMBER OF ARTICLES ON SDMS	50
FIGURE 1.6.2. PERCENTAGE OF ARTICLES (TOP) AND CITATIONS (BOTTOM) BY STUDY METHODS OF SDMS BETWEEN 1996 AND 2018	52

CHAPTER 2

FIGURE 2.1.1. THE ISLANDS OF THE AZORES AND THEIR LOCATION IN ATLANTIC OCEAN	66
FIGURE 2.3.1. DISTRIBUTION DATA OF <i>P. UNDULATUM</i> (PU), <i>A. MELANOXYLON</i> (AM) AND <i>M. FAYA</i> (MF) IN THREE AZORES ISLANDS	73

CHAPTER 3

FIGURE 3.1.1. MODELLING RESULTS FOR <i>PITTOSPORUM UNDULATUM</i> IN PICO ISLAND USING BIOMAPPER ..	95
FIGURE 3.1.2. MODELLING RESULTS FOR <i>PITTOSPORUM UNDULATUM</i> IN PICO ISLAND USING MAXENT	96
FIGURE 3.1.3. ANALYSIS OF VARIABLE CONTRIBUTION FOR MODEL EGV SET 4 USING MAXENT FOR <i>PITTOSPORUM UNDULATUM</i> IN PICO ISLAND	98
FIGURE 3.1.4. RESULTS OF CONTINUOUS BOYCE CURVES	99
FIGURE 3.1.5. RESPONSE CURVES OF <i>PITTOSPORUM UNDULATUM</i> TO THE THREE EGV CONSIDERED AS MORE IMPORTANT IN PICO ISLAND (EGV SET 4; Tr 100%) OBTAINED USING MAXENT.....	102
FIGURE 3.1.6. MODELLING RESULTS FOR <i>PITTOSPORUM UNDULATUM</i> IN PICO ISLAND USING ENFA (4 FACTORS) AND MAXENT	103
FIGURE 3.1.7. HABITAT SUITABILITY MAPS FOR <i>P. UNDULATUM</i> , <i>A. MELANOXYLON</i> AND <i>M. FAYA</i> BASED ON THE EGV SET 4, DERIVED WITH OCCURRENCE DATA FROM PICO ISLAND, AND APPLYING TWO MODELLING APPROACHES, ENFA AND MAXENT.....	106
FIGURE 3.1.8. HABITAT SUITABILITY MAPS FOR <i>P. UNDULATUM</i> , <i>A. MELANOXYLON</i> AND <i>M. FAYA</i> BASED ON THE EGV SET 4, DERIVED WITH OCCURRENCE DATA FROM TERCEIRA ISLAND, AND APPLYING TWO MODELLING APPROACHES, ENFA AND MAXENT	107
FIGURE 3.1.9. HABITAT SUITABILITY MAPS FOR <i>P. UNDULATUM</i> , <i>A. MELANOXYLON</i> AND <i>M. FAYA</i> BASED ON THE EGV SET 4, DERIVED WITH OCCURRENCE DATA FROM SÃO MIGUEL ISLAND, AND APPLYING TWO MODELLING APPROACHES, ENFA AND MAXENT	108

FIGURE 3.2.1. SUMMARY OF THE SPATIAL RANDOM EFFECT (GAUSSIAN RANDOM FIELD) FOR THE BEST MIXED EFFECTS MODELS AND THEIR REDUCTION FOR <i>P. UNDULATUM</i> IN PICO ISLAND	122
FIGURE 3.2.2. SUMMARY OF THE SPATIAL RANDOM EFFECT (GAUSSIAN RANDOM FIELD) FOR THE BEST MIXED EFFECTS MODELS AND THEIR REDUCTION FOR <i>M. FAYA</i> IN PICO ISLAND	123
FIGURE 3.2.3. SUMMARY OF THE SPATIAL RANDOM EFFECT (GAUSSIAN RANDOM FIELD) FOR THE BEST MIXED EFFECTS MODELS AND THEIR REDUCTION FOR <i>P. UNDULATUM</i> IN SÃO MIGUEL ISLAND	124
FIGURE 3.2.4. SUMMARY OF THE SPATIAL RANDOM EFFECT (GAUSSIAN RANDOM FIELD) FOR THE BEST MIXED EFFECTS MODELS AND THEIR REDUCTION FOR <i>M. FAYA</i> IN SÃO MIGUEL ISLAND	125
FIGURE 3.2.5. PREDICTION OF <i>P. UNDULATUM</i> OCCURRENCE IN PICO ISLAND BY POSTERIOR MEAN (A) AND STANDARD DEVIATION (B) OF THE LINEAR PREDICTOR	127
FIGURE 3.2.6. PREDICTION OF <i>M. FAYA</i> OCCURRENCE IN PICO ISLAND BY POSTERIOR MEAN (A) AND STANDARD DEVIATION (B) OF THE LINEAR PREDICTOR	128
FIGURE 3.2.7. PREDICTION OF <i>P. UNDULATUM</i> OCCURRENCE IN SÃO MIGUEL ISLAND BY POSTERIOR MEAN (A) AND STANDARD DEVIATION (B) OF THE LINEAR PREDICTOR.....	130
FIGURE 3.2.8. PREDICTION OF <i>M. FAYA</i> OCCURRENCE IN SÃO MIGUEL ISLAND BY POSTERIOR MEAN (A) AND STANDARD DEVIATION (B) OF THE LINEAR PREDICTOR	131
FIGURE 3.3.1. EVALUATION OF <i>P. UNDULATUM</i> (THREE DISTINCT ISLANDS) DISTRIBUTION MODELS/ALGORITHMS (MEAN VALUES) BY THE TRUE SKILL STATISTIC (TSS) AND THE RECEIVER OPERATING CHARACTERISTICS CURVE (AUC), USING TWO DIFFERENT EGV SETS.....	134
FIGURE 3.3.2. EVALUATION OF <i>A. MELANOXYLON</i> (THREE DISTINCT ISLANDS) DISTRIBUTION MODELS/ALGORITHMS (MEAN VALUES) BY THE TRUE SKILL STATISTIC (TSS) AND THE RECEIVER OPERATING CHARACTERISTICS CURVE (AUC), USING TWO DIFFERENT EGV SETS.....	135
FIGURE 3.3.3. EVALUATION OF <i>M. FAYA</i> (THREE DISTINCT ISLANDS) DISTRIBUTION MODELS/ALGORITHMS (MEAN VALUES) BY THE TRUE SKILL STATISTIC (TSS) AND THE RECEIVER OPERATING CHARACTERISTICS CURVE (AUC), USING TWO DIFFERENT EGV SETS.....	136
FIGURE 3.3.4. PRESENT AND FUTURE SPECIES DISTRIBUTION MAPS MODELLED FOR PICO ISLAND	138
FIGURE 3.3.5. PRESENT AND FUTURE SPECIES DISTRIBUTION MAPS MODELLED FOR TERCEIRA ISLAND	139
FIGURE 3.3.6. PRESENT AND FUTURE SPECIES DISTRIBUTION MAPS MODELLED FOR SÃO MIGUEL ISLAND	140
APPENDICES	
FIGURE S7.1. MEAN VALUES OF RELATIVE EXPLANATORY VARIABLES USED TO PREDICT THE DISTRIBUTION OF <i>P. UNDULATUM</i> (PU) IN PICO, TERCEIRA AND SÃO MIGUEL ISLANDS	67-A
FIGURE S7.2. MEAN VALUES OF RELATIVE EXPLANATORY VARIABLES USED TO PREDICT THE DISTRIBUTION OF <i>A. MELANOXYLON</i> (AM) IN PICO, TERCEIRA AND SÃO MIGUEL ISLANDS.....	68-A
FIGURE S7.3. MEAN VALUES OF RELATIVE EXPLANATORY VARIABLES USED TO PREDICT THE DISTRIBUTION OF <i>M. FAYA</i> (MF) IN PICO, TERCEIRA AND SÃO MIGUEL ISLANDS	69-A

LIST OF TABLES

CHAPTER 1

TABLE 1.6.1. DISTRIBUTION BY YEAR OF PUBLICATION, NUMBER OF ARTICLES AND NUMBER OF CITATIONS AND THEIR PERCENTAGE BETWEEN 1996 AND 2018	51
TABLE 1.6.2. JOURNALS WITH MORE ARTICLES ON SDMS AND THEIR IMPACT FACTOR	51
TABLE 1.6.3. DISTRIBUTION BY STUDY METHODS OF SDMS, NUMBER OF ARTICLES AND THEIR PERCENTAGE BETWEEN 1996 AND 2018	52

CHAPTER 2

TABLE 2.4.1. LIST OF THE 28 EGV USED TO MODEL INDIGENOUS AND NON-INDIGENOUS WOODY SPECIES .	75
TABLE 2.5.1. NUMBER OF SPECIES RECORDS USED IN MODELLING, PER ISLAND AND SPECIES	77
TABLE 2.6.1. NUMBER OF SPECIES RECORDS USED IN MODELLING, PER ISLAND AND SPECIES – INLA	81
TABLE 2.7.1. SUMMARY OF PRESENT AND FUTURE CLIMATIC VARIABLES FOR THREE DISTINCT ISLANDS ...	89

CHAPTER 3

TABLE 3.1.1. SETS OF EGV THAT PRODUCED MODELS WITH HIGHEST PREDICTIVE ABILITY.....	94
TABLE 3.1.2. RESULTS OF THE ENFA APPLIED TO <i>PITTOSPORUM UNDULATUM</i> IN PICO ISLAND, ACCORDING TO THE EGV SET	96
TABLE 3.1.3. SCORES OF THE EGV ON THE FIRST FOUR AXES OF THE ECOLOGICAL NICHE FACTOR ANALYSIS AND THE ESTIMATES OF THE CONTRIBUTIONS AND PERMUTATION IMPORTANCE OF EACH EGV FOR THE MODELS DERIVED USING MAXENT FOR <i>PITTOSPORUM UNDULATUM</i> IN PICO ISLAND.....	101
TABLE 3.1.4. RESULTS OF A MCNEMAR’S TEST COMPARING THE PREDICTED HABITAT SUITABILITY MAPS FOR <i>PITTOSPORUM UNDULATUM</i> , DERIVED USING ENFA AND MAXENT	104
TABLE 3.1.5. RESULTS OF MODELLING THE POTENTIAL DISTRIBUTION OF THREE WOODY SPECIES IN THREE AZOREAN ISLANDS USING FOUR DIFFERENT EGV SETS WITH THE ENFA (4 FACTORS) AND MAXENT.....	109
TABLE 3.2.1. SETS OF EGV THAT PRODUCED MODELS WITH HIGHEST FIT IN PICO AND SÃO MIGUEL ISLANDS USING FIXED EFFECTS GLMS.....	111
TABLE 3.2.2. RESULTS OF FIXED AND MIXED EFFECTS MODELS REGARDING DIFFERENT COMBINATIONS OF VARIABLES FOR <i>P. UNDULATUM</i> AND <i>M. FAYA</i> IN PICO AND SÃO MIGUEL ISLANDS.....	114
TABLE 3.2.3. RESULTS OF SAMPLING REDUCTION FOR FIXED AND MIXED EFFECTS MODELS REGARDING EGV SET 5 FOR <i>P. UNDULATUM</i> AND <i>M. FAYA</i> IN PICO ISLAND	116
TABLE 3.2.4. RESULTS OF SAMPLING REDUCTION FOR FIXED AND MIXED EFFECTS MODELS REGARDING EGV SET 5 FOR <i>P. UNDULATUM</i> AND <i>M. FAYA</i> IN SÃO MIGUEL ISLAND.....	117
TABLE 3.2.5. NUMERICAL SUMMARY OF THE POSTERIOR DISTRIBUTIONS OF THE PARAMETERS FOR THE BEST MIXED EFFECTS MODEL FOR THE TWO SPECIES STUDIED AND ISLANDS	118
TABLE 3.2.6. RANDOM FIELD HYPERPARAMETERS FOR THE BEST MODEL SELECTED.....	119
TABLE 3.2.7. NUMERICAL SUMMARY OF THE POSTERIOR DISTRIBUTIONS OF THE BEST MIXED EFFECTS MODELS AND THEIR REDUCTION FOR THE TWO SPECIES STUDIED IN PICO AND SÃO MIGUEL ISLANDS	120
TABLE 3.3.1. SETS OF EGV THAT PRODUCED MODELS WITH HIGHEST FIT IN PICO, TERCEIRA AND SÃO MIGUEL ISLANDS USING BIOMOD 2	133
TABLE 3.3.2. VALUES OF RESPECTIVE QUANTILES FOR HABITAT SUITABILITY MAPS	142

APPENDICES

TABLE S5.1. LIST OF JOURNALS WITH ARTICLES ON SDMS AND THEIR IMPACT FACTOR (IF – 2016) 30-A

TABLE S6.1. VARIABLES WITH SIGNIFICANCE (+) AND NON-SIGNIFICANCE (-) REGARDING EGV SET 5 FOR *P. UNDULATUM* AND *M. FAYA* IN PICO ISLAND 62-A

TABLE S6.2. VARIABLES WITH SIGNIFICANCE (+) AND NON-SIGNIFICANCE (-) REGARDING EGV SET 5 FOR *P. UNDULATUM* AND *M. FAYA* IN SÃO MIGUEL ISLAND 63-A

TABLE S7.1. EVALUATION OF DISTRIBUTION MODELS/ALGORITHMS (MEAN VALUES) IN PICO ISLAND BY THE TRUE SKILL STATISTIC (TSS) AND THE RECEIVER OPERATING CHARACTERISTICS CURVE (AUC), USING TWO DIFFERENT EGV SETS 64-A

TABLE S7.2. EVALUATION OF DISTRIBUTION MODELS/ALGORITHMS (MEAN VALUES) IN TERCEIRA ISLAND BY THE TRUE SKILL STATISTIC (TSS) AND THE RECEIVER OPERATING CHARACTERISTICS CURVE (AUC), USING TWO DIFFERENT EGV SETS 65-A

TABLE S7.3. EVALUATION OF DISTRIBUTION MODELS/ALGORITHMS (MEAN VALUES) IN SÃO MIGUEL ISLAND BY THE TRUE SKILL STATISTIC (TSS) AND THE RECEIVER OPERATING CHARACTERISTICS CURVE (AUC), USING TWO DIFFERENT EGV SETS 66-A

ACKNOWLEDGEMENTS

The long stretch of Ph.D. can be demanding both intellectually and emotionally. And, around three years and half, I am seeing the end of my project journey.

In general, the initial goals of this study were achieved, despite the normal setbacks inherent to a research work. This setback was gradually overcome by the tireless support of all. An achievement only made possible with the constant support of family members and friends, encouraging me and consistently insisting that everything will be fine while I was not very optimistic.

I would like to thank my advisors at Universidade dos Açores, Prof. Dr. Luís Silva, Prof. Dr. Rui Bento Elias and Prof. Dr. Mário Alves. But yet, I would like to express my deeply gratitude to Prof. Dr. Luís Silva, who gave me the opportunity to pursue this Ph.D. in the first place and offered your invaluable advice and support. You taught me the philosophy of science.

I also thanks: Direção Regional dos Recursos Florestais, Prof. Dr. Eduardo Brito de Azevedo and Hugo Costa for the provision of data; Hugo Costa, Artur Gil, Joana Vicente and Elias T. Krainski (The R-INLA project) for knowledge sharing.

LIST OF ACRONYMS/NOMENCLATURE

INSTITUTIONS

- ∅ **CE3C** – Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group
- ∅ **CIBIO** – Centro de Investigação em Biodiversidade e Recursos Genéticos
- ∅ **InBIO** – Rede de Investigação em Biodiversidade, Laboratório Associado

TECHNICAL TERMS

- ∅ **AIC** – Akaike Information Criterion
- ∅ **ANN** – Artificial Neural Networks
- ∅ **AUC** – Area Under the Curve
- ∅ **BIC** – Bayesian Information Criterion
- ∅ **BIOMOD** – BIOdiversity MODelling
- ∅ **BRT** – Boosted Regression Trees
- ∅ **BS** – Brier Score
- ∅ **CART** – Classification and Regression Trees
- ∅ **CIELO** – Clima Insular à Escala Local
- ∅ **CPO** – Conditional Predictive Ordinate
- ∅ **CRDT** – Constrained Refined Delaunay Triangulation
- ∅ **CSV** – Comma Separated Values
- ∅ **CTA** – Classification Tree Analysis
- ∅ **DIC** – Deviance Information Criterion
- ∅ **ENFA** – Ecological Niche Factor Analysis
- ∅ **GA** – Genetic Algorithms
- ∅ **GAM** – Generalized Additive Models
- ∅ **GARP** – Genetic Algorithm for Rule Set Production
- ∅ **GBA** – Network “guilt by association”
- ∅ **GBM** – Generalized Boosted Models
- ∅ **GDM** – Generalized Dissimilarity Modelling
- ∅ **GIS** – Geographic Information System
- ∅ **GLM** – Generalized Linear Models
- ∅ **GMRF** – Gaussian Markov Random Field
- ∅ **GRF** – Gaussian Random Field
- ∅ **GWR** – Geographically Weighted Regression

- ∅ **HS** – Habitat Suitability
- ∅ **IAP** – Invasive Alien Plants
- ∅ **IAS** – Invasive Alien Species
- ∅ **INLA** – Integrated Nested Laplace Approximation
- ∅ **IPCC/AR5** – Intergovernmental Panel on Climate Change/ Fifth Assessment Report
- ∅ **LCPO** – Logarithm of Conditional Predictive Ordinate
- ∅ **LM** – Linear Model
- ∅ **MAR** – Mid-Atlantic Ridge
- ∅ **MARS** – Multivariate Adaptive Regression Splines
- ∅ **MCMC** – Markov Chain Monte Carlo
- ∅ **MaxEnt** – Entropy Maximization
- ∅ **PCA** – Principal Component Analysis
- ∅ **PRAC** – Programa Regional para as Alterações Climáticas dos Açores
- ∅ **QGIS** – Quantum GIS
- ∅ **QR** – Quantile Regression
- ∅ **RCs** – Representative Concentration Pathways
- ∅ **RF** – Random Forest
- ∅ **ROC** – Receiver Operating Characteristic
- ∅ **SDMs** – Species Distribution Models
- ∅ **SEM** – Structural Equation Modelling
- ∅ **SPDE** – Stochastic Partial Differential Equation
- ∅ **SRE** – Surface Range Envelope
- ∅ **TSS** – True Skill Statistic
- ∅ **WAIC** – Watanabe-Akaike Information Criterion
- ∅ **WIP** – Woody Invasive Plant

ABBREVIATIONS

- ∅ **° C** – Degree Celsius
- ∅ **%** – Percentage
- ∅ **cm** – Centimetre
- ∅ **CO₂** – Carbon dioxide
- ∅ **km** – Kilometer
- ∅ **km²** – Kilometer square
- ∅ **m** – Meter

- ∅ **m.a.s.l** – Meter above sea level
- ∅ **Max.** – Maximum
- ∅ **Min.** – Minimum
- ∅ **mm** – Millimeter
- ∅ **mm/m²** – Millimeter per square meter
- ∅ **Myr** – Million years
- ∅ **W/m²** – Watt per square metre
- ∅ **AM** – *Acacia melanoxylon*
- ∅ **ASP** – Aspect
- ∅ **CI** – Credible Interval
- ∅ **CRV** – Curvature
- ∅ **DEM** – Digital elevation model
- ∅ **Dist.** – Distance from the nearest mainland
- ∅ **df** – Degree of freedom
- ∅ **DL 1** – Distance to forest
- ∅ **DL 2** – Distance to natural vegetation
- ∅ **DL 3** – Distance to pastureland
- ∅ **DL 4** – Distance to agriculture
- ∅ **DL 5** – Distance to barren/bare areas
- ∅ **DL 6** – Distance to urban/industrial areas
- ∅ **EGV** – Ecogeographical variables
- ∅ **e.g.** – “exempli gratia”
- ∅ **Expl. Inf.** – Total information explained
- ∅ **F1** – Factor 1
- ∅ **F2** – Factor 2
- ∅ **F3** – Factor 3
- ∅ **F4** – Factor 4
- ∅ **FLA** – Flow accumulation
- ∅ **i.e.** – “id est”
- ∅ **IF** – Impact factor
- ∅ **INF** – Infinite
- ∅ **MF** – *Morella faya*
- ∅ **ODL** – An ordinal variable that the highest score for habitats of forest type and lowest for urban/social areas
- ∅ **PM** – Annual mean precipitation
- ∅ **PMAX** – Annual maximum precipitation

- ∅ **PMIN** – Annual minimum precipitation
- ∅ **PMRA** – Annual mean precipitation range
- ∅ **PRA** – Annual precipitation range
- ∅ **PT** – Total annual precipitation
- ∅ **PU** – *Pittosporum undulatum*
- ∅ **Q1** – First quartile
- ∅ **Q2** – Second quartile
- ∅ **Q3** – Third quartile
- ∅ **Q4** – Fourth quartile
- ∅ **RHM** – Annual mean relative humidity
- ∅ **RHMAX** – Annual maximum relative humidity
- ∅ **RHMIN** – Annual minimum relative humidity
- ∅ **RHRA** – Annual relative humidity range
- ∅ **RTRG** – Jackknife of regularized training gain
- ∅ **sd** – Standard deviation
- ∅ **SHS** – Summer hill shade
- ∅ **SLP** – Slope
- ∅ **T** – Test
- ∅ **TG** – Jackknife of test gain
- ∅ **TM** – Annual mean temperature
- ∅ **TMAX** – Annual maximum temperature
- ∅ **TMIN** – Annual minimum temperature
- ∅ **TMRA** – Annual mean temperature range
- ∅ **Tr** – Training
- ∅ **TRA** – Annual temperature range
- ∅ **V** – Validation
- ∅ **WHS** – Winter hill shade
- ∅ **YP** – Year of publication



“Those who contemplate the beauty of the earth find reserves of strength that will endure as long as life lasts. There is something infinitely healing in the repeated refrains of nature – the assurance that dawn comes after night, and spring after winter.”

Rachel Carson, *Silent Spring*

“Conservation is a table of harmony between men and land.”

Aldo Leopold

GENERAL INTRODUCTION AND PURPOSE STATEMENT

INTRODUCTION

In the distant past, the Earth's mountains and oceans represented natural barriers to species dispersal and the ecosystems evolved in relative isolation (Lowe *et al.*, 2000). However, humans have broken down barriers to the long-dispersal of plants and animals, thus contributing to the increase of invasive alien species and changes in composition and functioning of the earth's ecosystems (Crawley, 1987; Roy 1990; Dukes and Mooney, 1999; Bois *et al.*, 2011). Are ecosystems flexible and able to cope with change? Can a new arrival have far-reaching repercussions or do permanent damage?

Invasive species are among the most important, least controlled, and least reversible of human impacts on the world's ecosystems, with negative consequences, affecting their sustainability, biodiversity, biogeochemistry and social/economic systems (Vitousek *et al.*, 1997; Cox, 1999; Carboni *et al.*, 2010; Miller *et al.*, 2011; Zimmermann *et al.*, 2011; Lockwood *et al.*, 2013; Zimmermann *et al.*, 2015). This invasive presence is known to influence the structure and function of invaded ecological communities and is recognized as a major component of global environmental change with implications in sustainable management and conservation (Higgins and Richardson, 1996; Vitousek *et al.*, 1997; Barthlott and Winiger, 1998; Chapin *et al.*, 2000; Mack *et al.*, 2000; Drake and Lodge, 2006; Ricciardi, 2007; Ricciardi and Cohen, 2007; Rodríguez *et al.*, 2007; van Kleunen and Richardson, 2007; Caley *et al.*, 2008; Crossman and Bass, 2008; Roura-Pascual *et al.*, 2009).

Biological invasions represent 40% of the economic losses in agricultural, grassland and forest ecosystems (Pimentel *et al.*, 2001; Hulme, 2012; Chen *et al.*, 2015a). In these terms, developing accurate risk assessment tools to predict potential invasive species is an important economy and policy goal since, once the invasion has been spread

in the wild, the possibility of eradication is limited (Alpert *et al.*, 2000; Kolar and Lodge, 2001, 2002; Regan *et al.*, 2006; Hulme, 2012).

From a theoretical perspective, biological invasions provide large-scale natural experiments which offer insights into issues fundamental to ecological theory (Higgins and Richardson, 1996). The evolutionary changes caused by invasive species are so profound that some ecologists have suggested that we are entering a new era, the Homogocene – a term apparently coined by Gordon Orians (1986), in which all the continents are connected into a “New Pangaea” through anthropogenic activities (Rosenzweig, 2001; Williams and Cameron, 2006; Janssen, 2007; Strayer, 2010; Winter *et al.*, 2010).

A main challenge in ecology is identifying and classifying determinants of invasiveness to provide suitable species habitats and to analyse the appropriate correlation of successful *versus* unsuccessful invasion, taking into account environmental variables and biotic interactions (Leibold *et al.*, 2004; Ebeling *et al.*, 2008; van Kleunen *et al.*, 2010a,b; Chang and Smith, 2012; Mainali *et al.*, 2015).

Oceanic islands have long been considered to be particularly vulnerable to biotic invasions, and much research has focused on invasive plants (Caujapé-Castells *et al.*, 2010; Kueffer *et al.*, 2010). The vulnerability can be explained for two main reasons. Firstly, these islands are isolated and vary broadly in size, geology and ecology (Denslow, 2003; Gillespie *et al.*, 2008; Denslow *et al.*, 2009; Kueffer *et al.*, 2010; Kaiser-Bunbury *et al.*, 2011). In comparison with less isolated settings, the biotas of oceanic islands are more sensitive to disturbance, because of the limited biotic interaction and adaptation through evolutionary time (Whittaker and Fernández-Palacios, 2007; Whittaker *et al.*, 2014). Thus, invasive plants have successfully spread outside their native range (Williamson, 1996; Richardson *et al.*, 2000; Richardson and Pyšek, 2006; Richardson and

Rejmánek, 2011; Bradley *et al.*, 2015). Determining spread patterns is challenging, particularly when invading plants become so abundant and dominant (“keystone species”), causing serious threat to biotic interactions (D’Antonio and Vitousek, 1992; Sax and Brown, 2000; D’Antonio and Meyerson, 2002; D’Antonio *et al.*, 2004; Sax and Gaines, 2008; Foxcroft *et al.*, 2011). These spread patterns depend on the scale of the study, the stage of invasion and on the spatio-temporal variation in the landscape structure (Collingham *et al.*, 2000; Davis, 2006; Melbourne *et al.*, 2007; Theoharides and Dukes, 2007; Gimona *et al.*, 2009).

The second explanation is linked to socioeconomic variables. The high level of islands invasions depends on human demography, agriculture, forestry, trade, industry and tourism (Preston *et al.*, 2004; Guézou *et al.*, 2010). Disconnecting the factors that determine the magnitude of the impacts of invasive species involves exploring the dependency of impacts on species features and ecosystem characteristics (Levine *et al.*, 2003; Gaertner *et al.*, 2009; Jarošík *et al.*, 2011; Pyšek *et al.*, 2012; Vilà *et al.*, 2015).

In the Azores islands, the land use changes were pronounced, following human colonization in the 15th century when native vegetation became gradually modified (Borges *et al.*, 2006). The increase in pastureland and spread of invasive plants can produce profound changes in the ecosystems, causing serious problems not only for biodiversity, but also in forestry farming and hydrologic cycles (Silva and Smith, 2006; Silva *et al.*, 2008; Costa *et al.*, 2012, 2013; Dutra Silva *et al.*, 2017a). Azorean native vegetation is threatened by several invaders. These invaders (e.g. *Pittosporum undulatum*) compete with species to be protected (e.g. *Morella faya*) and the management plans of nature reserves should incorporate this duality in long term of conservation plans. Pico Island is still one of the islands with a large forest cover. While important areas are

included inside the Island Natural Park and are primarily intended for conservation, many areas are presently invaded by *P. undulatum*.

Research and management of biological invasions require a specific view of invasions, as early as, possible so that impact risk can be assessed, and suitable long term/large-scale strategies can be formulated in appropriately time (Kriticos *et al.*, 2003, 2011; van Wilgen and Richardson, 2014). A variety of predictive models are currently in use to support management decisions and test key hypotheses regarding the patterns of spread of invasive plants (Guisan and Zimmermann, 2000; Scott *et al.*, 2001; Engler *et al.*, 2004; Andrew and Ustin, 2008, 2010; Petty *et al.*, 2012; Fukuda *et al.*, 2013). Species distribution models (SDMs) have become a fundamental tool in ecology, biodiversity, and have broad applications in assessing the relationships between species occurrence, the environment and the impact of ecological change (Brzeziecki *et al.*, 1993; Franklin *et al.*, 1998; Guisan and Thuiller, 2005; Guisan *et al.*, 2006a; Newbold, 2010; Catford *et al.*, 2011; Franklin, 2013; Guisan *et al.*, 2013; Guillera-Arroita *et al.*, 2015).

SPECIES DISTRIBUTION MODELS AND FOREST MANAGEMENT

Discoveries of pattern in nature provided both incentive and reward throughout the histories of science and art. Natural history is rich in the variety of biological and geographic information that is required in order to understand the landscape patterns that exist on earth (Burgess and Sharpe, 1981).

An appreciation of landscape is something we share with our earliest ancestors. For them, an understanding of the resources associated with the land meant the difference between the prosperity and collapse of their communities (Young *et al.*, 1993). For the current generations, particularly those in the developed world, our relationships with the

landscape are subtler, but just as important culturally and economically (Barbour *et al.*, 1987).

In the last two decades, the fascinating question of how plants are distributed in space and time has inspired many ecologists to find convincing explanations (Dutra Silva *et al.*, 2017a). To promote the sustainability of forests, one of the strategic issues for forest specialists is to recognize how site quality, species distribution, and eventually plant growth are affected by site conditions and how the variables may respond with changes in climate (Bourque *et al.*, 2008).

The advances in computational capabilities have provided increasingly greater and more intensive statistical computations than was previously possible (Zaniewski *et al.*, 2002). The development and use of numerous statistical techniques in ecology and forestry is extremely important to understand the relationship between species and their habitats (Huston, 2002). This relationship is fundamentally tied to the possibility of predicting species potential distribution by relating known species distributions to the spatial distributions of environmental variables (Franklin, 1995; Guisan and Zimmermann, 2000; Diggle and Ribeiro, 2007; Zimmermann *et al.*, 2015).

Modelling habitat requirements of species has been a particularly helpful tool in the domain of ecosystem management where identification and protection of areas containing high biological diversity has become crucial, but where species data sets are commonly limited or lacking (Austin and Meyers, 1996; Scott *et al.*, 1996; Corsi *et al.*, 1999; Jarvis and Robertson, 1999; Hijmans *et al.*, 2000; Williams and Hero, 2001; Storch, 2002; Ricklefs, 2004; Graham and Hijmans, 2006; Gottschalk *et al.*, 2007; Czado *et al.*, 2009; Jackson and Robertson, 2011). Numerous methods can be used to model species distributions (Guisan and Zimmermann, 2000; Franklin, 2010) that are most often allied to geographic information system (GIS) techniques, allowing the evaluation of their

predictive performance (Elith *et al.*, 2006; Tsoar *et al.*, 2007; Cawer *et al.*, 2014; Dutra Silva *et al.*, 2017a).

Species distribution models (SDMs) can serve as a tool to ensure consistency in ecological studies, while reducing the time and costs of large-scale studies of biodiversity involving large numbers of species (Stockwell, 1993, 1997, 1999; Pearce and Ferrier, 2000; Dutra Silva *et al.*, 2017a). For example, SDMs have been used to assess the potential distribution of invasive species (Chytrý *et al.*, 2009; Hallstan *et al.*, 2010; Chytrý *et al.*, 2012; Jones, 2012; Barbosa *et al.*, 2013) and to evaluate the possible impact of land use changes or climate change on species distribution (Schleupner and Link, 2008; Ferreira *et al.*, 2016; Dutra Silva *et al.*, 2017a). In forestry, these models were originally designed and used for research purposes but are presently being developed for use in practical forest management (Johnsen *et al.*, 2001; Buckley *et al.*, 2003; Falk and Mellert, 2011; Henderson *et al.*, 2014; Dutra Silva *et al.*, 2017a). Robust predictive models of forest biota distribution are important tools to understand the ecology theory and environmental processes which are connected to the control of species distribution (Austin and Meyers, 1996).

Human activities have either directly or indirectly influenced almost every part of our world (Liu, 2001). There are many activities which can affect landscapes in numerous ways, ranging from landscapes without any significant human impact to urban landscapes (Forman and Godron, 1986), and also vegetation dynamics (Pickett and Cadenasso, 2005; Essl *et al.*, 2011). Due to globalization, SDMs are being used to predict spatial patterns of biological invasions and prioritize locations for early detection and control of invasion outbreaks (Peterson and Vieglais, 2001; Fonseca *et al.*, 2006; Meentemeyer *et al.*, 2008; Strubbe and Mathysen, 2009; Jones *et al.*, 2010; Dutra Silva *et al.*, 2017a). Biological invasions by non-indigenous species are recognized to pose significant losses in the

economic value, biodiversity, and health of the invaded systems (Wittenberg and Cock, 2001; Hulme, 2009). Globally, invasive species are considered to be one of the most important causes of extinction and decline of indigenous species faced by island ecosystems (Magin *et al.*, 1994; Williamson, 1999; Olesen *et al.*, 2002; Williamson, 2006; Hortal *et al.*, 2010; Dutra Silva *et al.*, 2017a).

An important step for the construction of species distribution models is the determination of the statistical association between the data regarding the distribution of the species and the different independent variables that describe the topographic and climatic conditions of the study area, influencing the quality of the final result (Hirzel *et al.*, 2001; Le Lay *et al.*, 2001; Hirzel *et al.*, 2006; Hirzel and Le Lay, 2008; Mateo *et al.*, 2011; Lima *et al.*, 2013; Dutra Silva *et al.*, 2017a).

A major problem in the application of modelling and theory to field research and experimentation in ecology is that mathematical modelling in ecology requires simplifying assumptions, most of which are not compatible with the reality of ecology systems (DeAngelis and Rose, 1992). One of the most important of these assumptions is that individual members of populations can be aggregated into a single state variable representing population size (DeAngelis and Rose, 1992; Krebs, 1999).

The inference is a critical component in ecological modelling, although, the balance between accurate predictions and inference is the ultimate goal in ecological studies (Widrlechner *et al.*, 2004, 2009; Evans *et al.*, 2011). Practical applications of ecology in conservation planning, ecosystem assessment, and biodiversity are highly dependent on very accurate spatial predictions of ecological process and spatial patterns (Millar *et al.*, 2007; Evans *et al.*, 2011; Václavík *et al.*, 2012; Dutra Silva *et al.*, 2017a).

The complex nature of ecological systems hinders the ability to generate accurate models using the traditional frequentist data model (Evans *et al.*, 2011). Because of this,

we need to define a wide range of issues in ecological modelling, such as complex non-linear interactions, spatial autocorrelation, historic signal, anisotropy and scale, in order to overcome the problems that the frequentist data model has difficulty addressing.

A variety of statistical algorithms can be used for modelling species potential distributions (e.g. Guisan and Zimmermann, 2000; Dunlop *et al.*, 2006; Chejara *et al.*, 2010; Mukherjee *et al.*, 2011; Liang *et al.*, 2014). Costa *et al.* (2013) recently revised presence-only species distribution modelling approaches directed to exotic species. Several techniques used for modelling rely on presence-absence or abundance data. However, in the case of introduced species, absence data are of limited use because certain sites may be suitable, but not yet colonized by the species. Because of this fact, the use of alternative analyses relying on presence-only data is strongly advocated (Hirzel *et al.*, 2002; Hirzel and Le Lay, 2008; Elith and Leathwick, 2009).

Several presence-only methods of increasing complexity have been presented in recent years such as the maximum entropy modelling (MaxEnt; Phillips *et al.*, 2006). However, complex methods are more suitable for predicting realized species distributions (Phillips and Dudík, 2008), while the potential distribution might be more relevant in the case of exotic species (Dunlop *et al.*, 2006). Ecological Niche Factor Analysis (ENFA) is one of the commonly applied approaches for modelling that is suitable for predicting potential distributions. This is because the ENFA provides smooth responses to environmental factors (Václavík and Meentemeyer, 2012), which is desirable for modelling potential distributions, as models fitting complex responses may not accurately predict species not at equilibrium (Elith *et al.*, 2006). Good results have been reported on predicting species potential presence by using ENFA (e.g. Hirzel *et al.*, 2001, 2002; Martínez-Meyer and Peterson, 2006; Strubbe and Matthysen, 2009).

In forestry, species distribution models could be used to support the design of control strategies, including the definition of target areas and the potential replacement of invasive by a native species (Hortal *et al.*, 2005, 2010; Costa *et al.*, 2012, 2013; Dutra Silva *et al.*, 2017a). Finally, Bayesian methods are potentially applicable to many areas in biology and ecology (McCarthy, 2007; Kéry, 2010), and were already applied to the study of plant communities in the Azores (Marcelino *et al.*, 2013; Queiroz, 2013).

THE IMPORTANCE OF MODELLING POTENTIAL EFFECTS OF CLIMATE CHANGE

During the last several decades, biological invasions are gaining attention as a paradigm of ecological sustainability, historical variability and ecological integrity (Lackey, 1998; Dukes and Mooney, 1999; Landres *et al.*, 1999; Root *et al.*, 2003; Parmesan, 2006; Lenoir and Svenning, 2015). Biological invasions and climate change are two key drivers behind such paradigms (Walther *et al.*, 2009).

The planet Earth has entered an era of rapid environmental changes that has resulted in significant negative impacts on ecosystem resiliency (Orians, 1986; Walker and Vitousek, 1991; Peters and Lovejoy, 1992; Vitousek *et al.*, 1996; Millar *et al.*, 2007; Wenger *et al.*, 2013). However, scientific and societal unknowns make it difficult to predict how global environmental changes, such as climate change and biological invasions will affect species interactions and ecosystem processes (Walther *et al.*, 2002; Root *et al.*, 2003; Hellmann *et al.*, 2008; Beaumont *et al.*, 2014). In the long term, these changes may have interacting effects and compound the uncertainty associated with each individual driver (Zachos *et al.*, 2001; Hellmann *et al.*, 2008; Kerr and Dobrowski, 2013; Garcia *et al.*, 2014).

Global climate change is expected to further expand the risk of plant invasion through ecosystem disturbance and enhanced competitiveness due to elevated CO₂

(Dukes and Mooney, 1999; Mack *et al.*, 2000; Weltzin *et al.*, 2003; Thuiller, 2007; Thuiller *et al.*, 2008; Bradley *et al.*, 2009; Williams *et al.*, 2013). It is in combination with these threats that global warming becomes potentially problematic (Peters and Darling 1985; Dudley, 1998; Sala *et al.*, 2000; Noss, 2001; Brooks *et al.*, 2004, 2006; Bush *et al.*, 2014). Expanded risk from invasive plants due to climate changes has been identified for several species (Beerling, 1993; Sutherst, 1995; Zavaleta and Royval, 2002; Kriticos *et al.*, 2003; Qian and Ricklefs, 2006; Mika *et al.*, 2008; Bradley *et al.*, 2010a). Although, climate change also reduces invasive plant competitiveness if conditions become climatically unsuitable (Bradley *et al.*, 2009). Some previous publications suggest that climate change is likely to favour some invasive species (e.g. Dukes and Mooney, 1999; Thuiller *et al.*, 2008; Vilà *et al.*, 2007), but few authors have identified specific consequences of climate change for invasive species (Hellmann *et al.*, 2008).

It is useful to examine climate change responses of invasive species in order to (i) stimulate discussion about the distinctive consequences of climate change for invasive species, (ii) identify key hypotheses that should be tested to develop theories about invasive species and climate change, and finally (iii) inform planning efforts to species conservation (Hellmann *et al.*, 2008).

Hence, many questions regarding the behaviour of invasive alien species (IAS) cannot be answered directly by science using traditional hypothesis testing methods (Weinberg, 1972). The early view of an invasion can gain a relevant impact by using modelling tools that synthesize available information to outline potential management implications (Rastetter, 1996; Ehrenfeld, 2010). Change scenarios are commonly used with species distribution models (SDMs) to assess shifts in species range induced by climate change (Thuiller *et al.*, 2011; Lawler *et al.*, 2013; Casajus *et al.*, 2016). The relationships between species' climatic and habitat preferences thus need to be considered

in order to understand and predict changes in community composition (Jackson and Overpeck, 2000; Benito *et al.*, 2009; Clavero *et al.*, 2011; Barnagaud *et al.*, 2012, 2013; Regos *et al.*, 2017). If these relationships are understood, then local stressors could be proactively managed to improve tolerance and promote resilience to global climate change (Scheffer *et al.*, 2015; Regos *et al.*, 2017). In addition, innovations in predictive modelling techniques (e.g. Elith *et al.*, 2006) and modelling software allow applying and comparing several modelling methods, enabling the selection of best performing SDM for nature conservation. SDMs correlate the observed distribution of a species to a set of environmental predictors, including climate, and use this relationship to project the potential distribution into the future (Guisan and Thuiller, 2005; Elith and Leathwick, 2009; Bellard *et al.*, 2014).

In their nature state, oceanic islands typically support a substantial proportion of endemic species, many of which have been lost as a direct consequence of recent human habitation (Steadman, 2006; Whittaker and Fernández-Palacios, 2007; Triantis *et al.*, 2010).

The biodiversity “crisis” is a need of urgent action, once these remote islands are considered fragile ecosystems and highly vulnerable to biological invasion – introduction of non-native species (Paulay, 1994; Thompson *et al.*, 1995; Blackburn *et al.*, 2004; Steadman, 2006; Whittaker and Fernández-Palacios, 2007; Sax and Gaines, 2008). Thus, the biogeography and endemic biodiversity of oceanic islands makes them confined ecosystems that might serve as hot spots and laboratories for assessing changes in diverse “ecological stories” (Cowie and Holland, 2006). It is notable they are a home to unique forms of endemic plants and animals that have evolved in isolation over millions of years (Sadler, 1999; Cowie and Holland, 2006). Some of these ecosystems are the most sensitive and vulnerable in the world to climate change and may therefore provide

valuable monitoring sites for detecting the onset of human-induced global climate change (Markham, 1998; Randall *et al.*, 2007).

According to Ferreira *et al.* (2016), the Azorean archipelago provides a unique opportunity to study species – climate change relationships. Furthermore, the greatest estimated impact of global climate change the Azores may be the change in annual precipitation distribution, with wetter winters while the other seasons become drier (Santos *et al.*, 2004; Ferreira *et al.*, 2016). This change could have the potential effect on the islands' water resources (Ferreira *et al.*, 2016).

Particularly, forest ecosystems are under pressure, as climate change may threaten forest-dependent species across a wide range of species groups (Thomas *et al.*, 2004; Cherubini *et al.*, 2011; Thom *et al.*, 2017). The vulnerability of forest biodiversity alongside with the detail that the majority of terrestrial species depend on forest ecosystems highlights the lead role of forests in conservation management (Myers *et al.*, 2000; Levin *et al.*, 2003; Leishman *et al.*, 2007; Parrotta *et al.*, 2012; Thom *et al.*, 2017). To promote the sustainability of forests, one of the strategic issues for forest specialists is to recognize how site quality, species distribution, and eventually plant growth are affected by site conditions and how the variables may respond with changes in climate (Bourque *et al.*, 2008).