ASSESSING THE GENETIC DIVERSITY OF CATFACE ROCKCOD *EPINEPHELUS ANDERSONI* IN THE SUBTROPICAL WESTERN INDIAN OCEAN AND MODELLING THE EFFECTS OF CLIMATE CHANGE ON THEIR DISTRIBUTION

Thesis submitted in fulfilment of the requirements for the degree of

MASTER OF SCIENCE

of

RHODES UNIVERSITY

by

CHRISTINE ROSE COPPINGER

November 2013

Abstract

The catface rockcod *Epinephelus andersoni* is a range-restricted species that is endemic to the south-east coast of Africa from Quissico in Mozambique (subtropical) to Knysna in South Africa (warm-temperate). Its complex life-history, long-lived nature and high residency make *E. andersoni* potentially vulnerable to over-exploitation. *Epinephelus andersoni* is an important fishery species and has shown signs of depletion. Due to inadequate information necessary for management and conservation, further research is vital, particularly in the face of potentially significant climatic changes which could put further pressure on *E. andersoni*. The aim of the study was to provide information for the management of *E. andersoni*, with considerations for the possibly detrimental effects of future climate change. The objectives of this study were to describe the genetic structure and diversity of *E. andersoni* and to determine possible range shifts of *E. andersoni* with future changes in sea surface temperature.

Genetic samples were collected throughout the distribution of *E. andersoni*. Standard DNA extraction and PCR using universal primers were conducted and nuclear (RPS7-1) and mitochondrial (cytochrome *b*) data were analysed to determine genetic diversity. A combination of nuclear and mitochondrial markers was used to ensure that the results were robust. RPS7-1 haplotype diversity was high (0.801) and an AMOVA on the RPS7-1 data showed significantly high among group variation (Φ_{CT} = 0.204, $p < 0.05$) between five groups: 1. Quissico to Inhaca; 2. Cape Vidal to Port Edward; 3 Port St Johns to Coffee Bay; 4. Mbashe; 5. Port Alfred. This geographic structuring could be attributed to low gene flow across barriers such as the Port Alfred upwelling cell, the Mozambique Channel eddies and smaller more localised upwelling cells such as the Port St Johns cell. The cytochrome *b* results contrastingly indicate low haplotype diversity (0.309) and no differentiation (Φ_{CT} = 0.265, p = 0.074) between groups and support the hypothesis of a historical population bottleneck. This may be due to an unusually slower mutation rate of the cytochrome *b* region than the RPS7-1 region, resulting in the RPS7-1 data showing a more recent picture of diversification.

To complement the genetic results, niche modelling techniques were used to determine range shifts of *E. andersoni* with future temperature trends using species distribution and climatic data. The model illustrated a contraction of the *E. andersoni* distribution as well as future intensification of various upwelling cells along the south-east African coast including the Port Alfred upwelling cell. Due to the low gene flow across these barriers this intensification could decrease the resilience of *E. andersoni*, as its range becomes more limited with global change. The genetic data and modelling results combined provide useful information on which to base future fisheries management.

Acknowledgements

I would like to thank both my supervisors, Dr Nicola James and Dr Monica Mwale, for their unwavering patience and support for the duration of the project. Their guidance has been invaluable and without them, the completion of my MSc would not have been possible. Thanks go to them for editing several drafts of this thesis. I have learnt so much in the last few years and I value the experience I have gained very highly.

Without the funding for the project that was provided by the Western Indian Ocean Marine Science Association (WIOMSA) and the personal funding provided to me by the National Research Foundation of South Africa (NRF), this project would not have been possible. Thanks are extended to these funders. The Technology and Human Resources for Industry Programme (THRIP) is also thanked for the personal funding that they provided. This project was only achieved with the extensive collaboration and support of José Halafo from the Instituto de Investigação Pesqueira (IIP) and of Almeida Guissamulo from the Universidade Eduardo Mondlane in Mozambique. The data and genetic samples collected by these collaborators made the project possible and they are thanked for their support.

I would like to thank The Oceanographic Research Institute (ORI) for providing many of the genetic samples I needed through the ORI/WWF Tagging Project. I thank ORI as well, for the provision of tagging data on the catface rockcod which has helped me unravel their story. Thanks also to countless fishermen and spearfishermen who have helped in collecting the genetic material I needed: Rod Hastier in Port St Johns, Lisa Powell and the Blue Dolphin team in Xai Xai, Justin Lindsay, Nick Riddin, Terence Bellingan and many others.

The last few years have been an emotional rollercoaster and without the support of my family and my dearest partner Bruce Ellender, I doubt I would have got very far. Thank you so much for your encouragement and patience during some of the toughest times. I'm sure you will all be relieved that the stress will soon be over and your daughter/sister/partner will be back to normal! Thanks also to Bruce and my sister Jen for their help with many fishing trips in the Eastern Cape to collect genetic samples - those spiky, slippery little catface rockcods were elusive! To Bruce especially - thank you so much for everything you've done for me in the past few years. Your help and many invaluable comments on countless drafts of my thesis have meant so much and have been instrumental in getting me where I am.

Thank you also to my colleagues, Murray Duncan and Devin Isemonger. Your help in the field - from the fish markets in Mozambique to the deep sea fishing in Coffee Bay - and your support during the analysis and writing up phase have been vital... I may just forgive you for abandoning me in Xai Xai, to catch a minibus taxi back to Maputo! My other colleagues at the Department of Ichthyology and Fisheries Science at Rhodes University have also supported me during the course of this project, especially Geraldine Taylor, Chenelle de Beer and some of the lecturers of the department, notably Warren Potts. Your help gave me the positive boost I needed at just the right time.

Thank you all so much for everything!

THE world is too much with us; late and soon,
Getting and spending, we lay waste our powers:
Little we see in Nature that is ours;
We have given our hearts away, a sordid boon!
The Sea that bares her bosom to the moon;
The winds that will be howling at all hours,
And are up-gathered now like sleeping flowers;
For this, for everything, we are out of tune;
It moves us not.--Great God! I'd rather be
A Pagan suckled in a creed outworn;
So might I, standing on this pleasant lea,
Have glimpses that would make me less forlorn;
Have sight of Proteus rising from the sea;
Or hear old Triton blow his wreathed horn.

William Wordsworth 1888

Table of Contents

Abstract ii

Acknowledgements iii

CHAPTER 1: General Introduction 1

1.1 *Epinephelus andersoni* life-history characteristics and threats 1

1.2 South African and Mozambican fisheries 4

1.3 Climate change and impacts on genetic diversity 7

1.4 The role of southern African oceanographic features in shaping genetic diversity 9

1.5 Rationale for the study 12

1.6 Aims and hypotheses 12

1.7 Thesis outline 13

CHAPTER 2: The genetic stock structure of *Epinephelus andersoni* 14

2.1 Introduction 14

2.2 Materials and Methods 17

2.3 Results 28

2.4 Discussion 57

2.5 Conclusions 64

CHAPTER 3: Modelling the distribution of *Epinephelus andersoni* with future sea surface temperature changes 66

3.1 Introduction 66

3.2 Materials & Methods 69

3.3 Results 81

3.4 Discussion 89

3.5 Conclusions 94

CHAPTER 4: General Discussion 96

4.1 Will genetic diversity be lost with future climate change? 96

4.2 Management and implications for fisheries 100

4.3 Further study 106

4.4 Conclusions 108

References 110

Appendix I 131

5.1 Model run 1: Surface Range Envelope (SRE) projected pseudo-absences 131

5.2 Model run 2: Randomly projected pseudo-absences 139

List of Figures

Figure 1.1: Map of the distribution of *Epinephelus andersoni* on the south-east African coast. Bioregions are shown. 1

Figure 1.2: A juvenile *Epinephelus andersoni* (approximately 140 mm TL), caught in the Kowie Estuary, Eastern Cape, South Africa 2011. 2

Figure 1.3: Observations of the recent warming in the Agulhas Current system, taken from Rouault *et al*. (2009). Linear trend in Advanced Very High Resolution Radiometer sea surface temperature from 1985 to 2006 (°C/decade) showing a warming of up to 0.7 °C/decade. The black contours represent mean surface currents and arrows indicate the direction of the flow. The blue and black lines and red box are not relevant here. 8

Figure 1.4: General oceanographic currents of South Africa, Mozambique and Madagascar (taken from Lutjeharms 2006). 10

Figure 1.5: Oceanographic currents off the south-east African mainland. Cream colour indicates the continental shelf; large southward pointing arrows represent the Agulhas Current; blue areas indicate upwelling; circular arrows indicate eddies or irregularities in flow. The distributional limits of *Epinephelus andersoni* are indicated by black dots and other localities are indicated by white dots. The Transkei coast extends southward from the KZN border to the Mbashe R. Mouth. 11

Figure 2.1: Map of south-eastern Africa showing the sampling localities for *Epinephelus andersoni*. Total sample numbers are indicated in parentheses after the name of the locality followed by the numbers of RPS7-1 and mitochondrial DNA cytochrome *b* sequences respectively. 30

Figure 2.2: Minimum spanning haplotype network of the 269 sequences from 12 localities distributed amongst 12 unique cytochrome *b* haplotypes. Circle sizes are proportional to haplotype frequencies: the smallest circle represents one sample. Connecting branches between haplotypes are one mutational step while circle segment colours represent the geographical origin of the samples. 32

Figure 2.3: Observed (bars) and expected (line) mismatch distributions of the frequency of pairwise nucleotide differences between all pairs of haplotypes in the cytochrome *b* data, under the spatial expansion (a) and demographic expansion (b) models. 33

Figure 2.4: Scatterplots of pairwise F_{ST} comparisons versus geographic distance (r = -0.015, p = 0.691) (a) and log-transformed pairwise F_{ST} versus geographic distances (r = 0.003, p = 0.420) (b) between all pairs of samples for the mitochondrial cytochrome *b* region. 38

Figure 2.5: Spatial autocorrelation plot for the mitochondrial cytochrome *b* data using 14 distance classes of 95 km throughout the sampling range of *E. andersoni*. The dotted line indicates overall mean genetic distance and the solid line indicates genetic distance (*Ay*) within each distance class. Significant correlation is indicated by a star. 39

Figure 2.6: Cytochrome *b* genetic seascape interpolation with distance weighting parameter a = 1 and 100 × 100 gridded latitude and longitude axes. Surface plot heights (on the D axis) represent residual genetic distances (range: -0.00013 - 0.00010). Compass directions are indicated on the X and Y axes as S or N and E or W. Views from the southern (a) and northern (b) edges of the plot are shown. Localities are indicated on the plot: QU = Quissico; XX = Xai Xai; MP = Maputo; IN = Inhaca; CV = Cape Vidal; RB = Richards Bay; PEd = Port Edward; MS = Msikaba; PSJ = Port St Johns; CB = Coffee Bay; MB = Mbashe; PA = Port Alfred. 41

Figure 2.7: Minimum spanning haplotype network of 39 alleles represented from the phased 480 nuclear RPS7-1 sequences from the 12 localities. Nine of the shared alleles are numbered and circle sizes are proportional to the haplotype frequencies: the smallest circle represents one sample and the largest allele (allele 1) represents 183 samples. Colours of the circle segments represent the geographical origin of the samples. Connecting branches between the centres of haplotypes

represent one mutational step except when transverse bars are present, each of which represent an additional step. The network representing the most likely genealogy (a), as well as the full allele network with all seven possible genealogies (b), are shown. 45

Figure 2.8: Observed (bars) and expected (line) mismatch distributions of the frequency of pairwise nucleotide differences between all pairs of haplotypes in the nuclear RPS7-1 data, under the spatial (a) and demographic expansion models (b). 46

Figure 2.9: Scatterplots of pairwise F_{ST} comparisons versus geographic distance (r = 0.032, p = 0.029) (a) and log-transformed pairwise F_{ST} versus geographic distances (r = 0.092, p = 0.001) (b) between all pairs of samples for the nuclear RPS7-1 region. 51

Figure 2.10: Spatial autocorrelation analysis for the RPS7-1 data using 14 distance classes of 95 km throughout the sampling range of *Epinephelus andersoni*. The dotted line indicates the overall mean genetic distance and the solid line indicates genetic distance (*Ay*) within each distance class. Significant correlation is indicated by a star. 52

Figure 2.11: Nuclear RPS7-1 genetic seascape interpolation with distance weighting parameter a = 1 and 100 × 100 gridded latitude and longitude axes. Surface plot heights (on the D axis) represent residual genetic distances (range: -0.00085 - 0.00157). Compass directions are indicated on the X and Y axes as S or N and E or W. Views from the southern (a) and northern (b) edges of the plot are shown. Localities are indicated on the plot: QU = Quissico; XX = Xai Xai; MP = Maputo; IN = Inhaca; CV = Cape Vidal; RB = Richards Bay; PEd = Port Edward; MS = Msikaba; PSJ = Port St Johns; CB = Coffee Bay; MB = Mbashe; PA = Port Alfred. 56

Figure 3.1: Map of the study area showing the 210 *Epinephelus andersoni* presence points used for modelling. Blue dots show previously recorded distributional limits. 71

Figure 3.2: Decadal trends in NOAA OI SST between 1982-2010 projected for southern Africa for the four seasons: A) summer (January, February, March); B) autumn (April, May, June); C) winter (July, August, September); D) spring (October, November, December). The accompanying key represents decadal SST change (i.e. cooling is negative and warming is positive). 82

Figure 3.3: Mean variable importances for each of the seven environmental variables with standard error whiskers. The solid horizontal line indicates the overall mean variable importance (0.306). 84

Figure 3.4: Probability mean weighted ensemble probability distributions of *E. andersoni* projected for the present (A), 2020 (B), and 2030 (C). Occurrences are plotted against the current projection. 87

Figure 3.5: Binary probability mean weighted ensemble distributions of *E. andersoni* projected for the present (A), 2020 (B), and 2030 (C). Occurrences are plotted against the current projection. 88

Figure 3.6: Bathymetry (scale in metres below sea level) of the study area with overlaid occurrences. 94

Figure 4.1: Binary current (top) and 2030 (bottom) projected distributions for *Epinephelus andersoni*. Occurrence records are shown (black dots) and localities that were sampled for genetic data are indicated (blue-filled dots, bolded names). 99

Figure 4.2: South African MPAs are shown in red. Knysna, the accepted southern *E. andersoni* distribution limit is shown. MPAs shown include: 1 Maputaland sanctuary/restricted areas; 2 St Lucia sanctuary/restricted areas; 3 Aliwal Shoal controlled area; 4 Trafalgar MPA; 5 Pondoland controlled/restricted areas; 6 Hluleka MPA; 7 Dwesa-Cwebe MPA; 8 Nyara River to Great Kei River; 9 Nahoon Point to Gonubie Point; 10 Christmas Rock to Gxulu River; 11 Bird Island MPA; 12 Sardinia Bay MPA; 13 Tsitsikamma MPA; 14 Robberg MPA; 15 Goukamma MPA; 16 De Hoop MPA; 17 Betty's Bay MPA; 18 Helderberg MPA; 19 Table Mountain National Park MPA; 20 Sixteen Mile Beach MPA;

21 Langebaan Lagoon MPA; 22 Jutten Island MPA; 23 Malgas Island MPA; 24 Marcus Island MPA; 25 Namaqualand MPA. 104

Figure 4.3: MPAs are shown in red. Quissico, the accepted northern *E. andersoni* distribution limit is shown. Mozambique MPAs shown include: A) Maputo Special; B) Ilhas da Inhaca e dos Portuguesas; C) Pomene; D) Bazaruto; E) Marromeu Complex; F) Primeiras and Segundas; G) Quirimbas. 105

Figure 5.1: Individual model projections for the current *E. andersoni* distribution using 2100 pseudo-absences projected using the SRE approach. The probability of occurrence distributions are shown on the left (all figures labelled 1) and the binary transformed distributions are shown on the right (all figures labelled 2). The results of the following models are shown: A) CTA, B) GAM, C) GBM, D) GLM, E) MARS, F) MAXENT, G) RF. 133

Figure 5.2: Individual model projections for the 2020 *E. andersoni* distribution using 2100 pseudo-absences projected using the SRE approach. The probability of occurrence distributions are shown on the left (all figures labelled 1) and the binary transformed distributions are shown on the right (all figures labelled 2). The results of the following models are shown: A) CTA, B) GAM, C) GBM, D) GLM, E) MARS, F) MAXENT, G) RF. 135

Figure 5.3: Individual model projections for the 2030 *E. andersoni* distribution using 2100 pseudo-absences projected using the SRE approach. The probability of occurrence distributions are shown on the left (all figures labelled 1) and the binary transformed distributions are shown on the right (all figures labelled 2). The results of the following models are shown: A) CTA, B) GAM, C) GBM, D) GLM, E) MARS, F) MAXENT, G) RF. 137

Figure 5.4: Mean variable importances for each of the seven environmental variables with standard error whiskers. The solid horizontal line indicates the overall mean variable importance (0.281). .. 140

Figure 5.5: Probability mean weighted ensemble probability distributions of *E. andersoni* projected for the present (A), 2020 (B), and 2030 (C). Occurrences are plotted against the current projection. 142

Figure 5.6: Binary probability mean weighted ensemble distributions of *E. andersoni* projected for the present (A), 2020 (B), and 2030 (C). Occurrences are plotted against the current projection. ... 143

Figure 5.7: Individual model projections for the current *E. andersoni* distribution using 2100 pseudo-absences projected randomly. The probability of occurrence distributions are shown on the left (all figures labelled 1) and the binary transformed distributions are shown on the right (all figures labelled 2). The results of the following models are shown: A) CTA, B) GAM, C) GBM, D) GLM, E) MARS, F) MAXENT, G) RF. 146

Figure 5.8: Individual model projections for the 2020 *E. andersoni* distribution using 2100 pseudo-absences projected randomly. The probability of occurrence distributions are shown on the left (all figures labelled 1) and the binary transformed distributions are shown on the right (all figures labelled 2). The results of the following models are shown: A) CTA, B) GAM, C) GBM, D) GLM, E) MARS, F) MAXENT, G) RF. 148

Figure 5.9: Individual model projections for the 2030 *E. andersoni* distribution using 2100 pseudo-absences projected randomly. The probability of occurrence distributions are shown on the left (all figures labelled 1) and the binary transformed distributions are shown on the right (all figures labelled 2). The results of the following models are shown: A) CTA, B) GAM, C) GBM, D) GLM, E) MARS, F) MAXENT, G) RF. 150

List of tables

Table 1.1: Summary of Mozambique catch statistics between 1994 and 2009, calculated from mainly semi-industrial vessel catches (modified from Fennessy *et al.* 2011 and van der Elst *et al*. 2003). Figures indicate percentage contribution to the total catch. Other fish including sciaenids, lethrinids, carangids and scombrids were not included in this summary. Blanks indicate no data. 6

Table 2.1: Forward (F) and reverse (R) primers tested for the PCR amplification of the *E. andersoni* cytochrome *b* region (cyt *b*) and S7 ribosomal protein genes. Primers selected for the study are indicated in bold. 19

Table 2.2: The number of *Epinephelus andersoni* samples (N) per locality and per sub-site within localities; GPS co-ordinates of localities; and the number of sequences for the RPS7-1 after phasing and cytochrome *b* regions. 29

Table 2.3: Cytochrome *b* (711 bp) diversity indices and the test statistics and corresponding *p*-values for the tests for selective neutrality. Diversity indices represented include haplotype (*Hd*) and nucleotide (π) diversities as well as the numbers of sequences (*N*), polymorphic sites (*S*), haplotypes (*H*), private haplotypes (*Hp*) and mean number of nucleotide differences between sequences (*k*) for each of the 12 sampling localities. Significant values ($p < 0.05$) are indicated in bold. 31

Table 2.4: Pairwise F_{ST} values (above diagonal) and exact tests of genetic differentiation *p*-values (below diagonal) between localities for 771 bp cytochrome *b* data and the AMOVA grouping of localities that maximised F_{CT}.(Groups 1 to 6) are defined. Significant *p*-values ($\alpha = 0.05$) are indicated in bold. 35

Table 2.5: AMOVA results for 269 cytochrome *b* sequences of *Epinephelus andersoni*. Genetic variance among geographic groups (6); among localities within groups (12 sampling localities); and among individuals within localities ($\alpha = 0.05$). 36

Table 2.6: Summary of SAMOVA results showing the level of cytochrome *b* genetic variation among geographic groups with varying numbers of groups (*K*). The percent variation explained by the groups, F_{CT} and corresponding *p*-values are indicated for all *K*. Significant *p*-values ($\alpha = 0.05$) are indicated in bold. 37

Table 2.7: RPS7-1 (706 bp) diversity indices and the test statistics and corresponding *p*-values for the tests for selective neutrality. Diversity indices represented include haplotype (*Hd*) and nucleotide (π) diversities as well as the numbers of sequences (*N*), polymorphic sites (*S*), haplotypes (*H*), private haplotypes (*Hp*) and mean number of nucleotide differences between sequences (*k*) for each of the 12 sampling localities. Significant values ($p < 0.05$) are indicated in bold..... 43

Table 2.8: Pairwise F_{ST} values (above diagonal) and exact tests of genetic differentiation *p*-values (below diagonal) between localities for 706 bp RPS7-1 data and the AMOVA grouping of localities that maximised F_{CT}.(Groups 1 to 5) are defined. Significant *p*-values ($\alpha = 0.05$) are indicated in bold. 48

Table 2.9: AMOVA results for 480 nuclear RPS7-1 sequences of for *Epinephelus andersoni*. Genetic variance among geographic groups (5); among localities within groups (12 sampling localities); and among individuals within localities were all significant ($\alpha = 0.05$). 49

Table 2.10: Summary of SAMOVA results showing the level of RPS7-1 genetic variation among geographic groups with varying numbers of groups (*K*). The percent variation explained by the groups, F_{CT} and corresponding *p*-values are shown for all *K* values. Significant *p*-values ($\alpha = 0.05$) are indicated in bold. 50

Table 2.11: Matrix of pairwise geographic distances (km) between localities with significantly negative spatial autocorrelation between RPS7-1 genetic and geographic distances ($\alpha = 0.05$) indicated by an asterisk (*****). Significantly positive autocorrelations ($\alpha = 0.05$) are indicated by two asterisks (******). 53

Table 2.12: The groups of localities resulting from each of the barriers generated from the nine iterations of Monmonier's algorithm using RPS7-1 data are shown. 54

Table 2.13: Haplotype diversity (*H*) and nucleotide diversity (π) indices for population studies on selected grouper species, based on mitochondrial DNA sequencing. The gene region studied is shown, as well as whether genetic structure or isolation by distance (IBD) was found and the proposed reason for genetic structure. All species presented have long-lived planktonic larvae and are either generally reef-dependent or completely reliant upon rocky reefs, similar to *Epinephelus andersoni*. 58

Table 3.1: Strengths and weaknesses of all models used in the ensemble and general comments on the class of each model and how it functions. 76

Table 3.2: Measures of model's predictive accuracy calculated from a two × two confusion matrix, modified from McPherson *et al*. (2004) and Allouche *et al*. (2006). 79

Table 3.3: Variable importances for each of the seven environmental variables (columns) when the *Epinephelus andersoni* distribution was modelled using seven different modelling algorithms (rows). Mean importance, standard deviation and standard error for each variable across modelling techniques are shown. 83

Table 3.4: Three measures of model evaluation, True Skills Statistic (TSS), area under the Receiver Operating Curve (ROC), and Cohen's KAPPA are shown for each of the seven modelling algorithms used to model the distribution of *Epinephelus andersoni* into the future. 85

Table 3.5: Summary of the range changes between the Probability Weighted Mean ensemble current and the 2020 and 2030 projections. Figures represent number of pixels except where percentages are indicated. "Stable0" indicates the number of pixels that are unoccupied in the current projection and that remain unoccupied in the future projections. "Stable1" indicates the number of pixels that are occupied in the current projection and are predicted to remain occupied in the future. 89

Table 4.1: The private haplotypes and alleles (*Hp*), and haplotype diversities (*Hd*) for the cytochrome *b* and RPS7-1 regions respectively, for each locality. The total number of haplotypes (*H*) for each gene region (bottom of *Hp* columns) and overall *Hd* is also shown. 98

Table 5.1: Range changes between the current and the 2020, and the current and the 2030 binary transformed distributions for each model. Figures represent number of pixels except where percentages are indicated. Stable0 indicates the number of pixels that are unoccupied in the current projection and that remain unoccupied in the future projections. Stable1 indicates the number of pixels that are occupied in the current projection and are predicted to remain occupied in the future. 138

Table 5.2: Variable importances for each of the seven environmental variables (columns) when the *Epinephelus andersoni* distribution was modelled using seven different modelling algorithms (rows). Mean importance, standard deviation and standard error for each variable across modelling techniques are shown. 139

Table 5.3: Three measures of model evaluation, True Skills Statistic (TSS), area under the Receiver Operating Curve (ROC), and Cohen's KAPPA are shown for each of the seven modelling algorithms used to model the distribution of *Epinephelus andersoni* into the future. 141

Table 5.4: Summary of the range changes between the Probability Weighted Mean ensemble current and the 2020 and 2030 projections. Figures represent number of pixels except where percentages are indicated. Stable0 indicates the number pixels that are unoccupied in the current projection and that remain unoccupied in the future projections. Stable1 indicates the number of pixels that are occupied in the current projection and are predicted to remain occupied in the future. 144

Table 5.5: Range changes between the current and the 2020, and the current and the 2030 binary transformed distributions for each model. Figures represent number of pixels except where percentages are indicated. Stable0 indicates the number of pixels that are unoccupied in the current projection and that remain unoccupied in the future projections. Stable1 indicates the number of pixels that are occupied in the current projection and are predicted to remain occupied in the future. 151

CHAPTER 1: General Introduction

1.1 *Epinephelus andersoni* life-history characteristics and threats

1.1.1 Life-history characteristics

The catface rockcod *Epinephelus andersoni* Boulenger, 1993 belongs to the family Serranidae and the subfamily Epinephelinae (rockcods or groupers), which is the largest of the four subfamilies of serranids (Fennessy and Sadovy 2002). *Epinephelus andersoni* is endemic to the south-east African coast, its known distribution extending from Quissico in Mozambique to Knysna in South Africa (Heemstra and Randall 1993; Fennessy and Mann 2013). Although it has also been reported to occur in southern Madagascar (Fourmanoir 1957; Craig *et al.* 2011) recent work suggests that it is either rare or absent from this region (S. Fennessy, pers. comm.). Its south-east African distribution spans three Indian Ocean bioregions namely: the tropical Delagoa bioregion in Mozambique; the subtropical bioregion between Cape Vidal and Coffee Bay; and the warm-temperate bioregion south of Coffee Bay (Sink *et al.* 2004) (Figure 1.1).

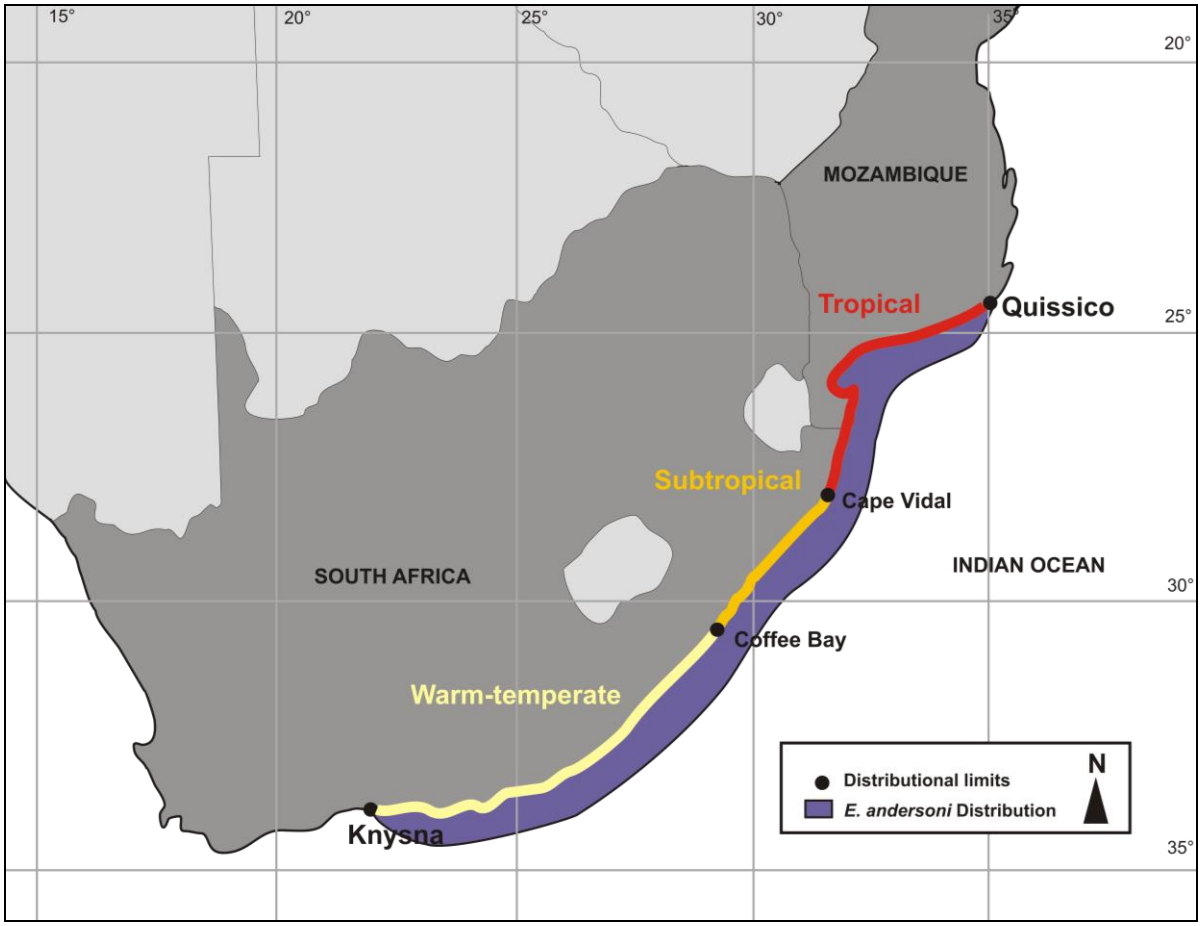


Figure 1.1: Map of the distribution of *Epinephelus andersoni* on the south-east African coast. Bioregions are shown.

Epinephelus andersoni (Figure 1.2) is a sedentary, ambush predator which subsists largely on crustaceans and other fishes (Heemstra and Randall 1993). The species is relatively slow-growing and long-lived (Fennessy 2000), reaching a maximum length of approximately 870 mm total length (TL) (Fennessy and Sadovy 2002), weight of 9.4 kg (S. Fennessy, ORI, unpubl. data), and age of 11 years (Fennessy 1998). Male age at first maturity and minimum length at maturity are 2.7 years and 430 mm TL respectively, whereas female age at first maturity and length at 50% maturity are 3.9 years and 492 mm TL respectively (Fennessy and Mann 2013). Like most other epinepheline serranids, *E. andersoni* are generally found in shelf waters (Fennessy and Sadovy 2002), and inhabit rocky (non-coral) reefs in the surf zone down to approximately 70 m depth (Fennessy *et al*. 2000). Juveniles inhabit tidal pools, shallow subtidal reefs and estuaries (Heemstra and Randall 1993; Heemstra and Heemstra 2004). *Epinephelus andersoni* have planktonic larvae and the species has a long-lived larval phase (Connell 2012), making long distance dispersal from spawning sites via ocean currents possible (Heemstra and Randall 1993).

Figure 1.2: A juvenile *Epinephelus andersoni* (approximately 140 mm TL), caught in the Kowie Estuary, Eastern Cape, South Africa 2011.

During a study on the reproductive biology of *E. andersoni*, Fennessy and Sadovy (2002) found a complete overlap in male and female sizes in sexually mature fish and significantly more males among sampled fish. Males can develop either from juveniles or from mature females making *E. andersoni* a diandric protogynous hermaphrodite, a sexual pattern which is unusual within the genus

Epinephelus (Fennessy and Sadovy 2002). Fennessy and Sadovy (2002) found that spawning occurs in the relatively warm waters off KwaZulu-Natal (KZN) during austral summer (November-January), suggesting that spawning may be triggered by warmer temperatures that occur in the region at this time. Anecdotal evidence suggests that *E. andersoni* spawn close inshore (Connell 2012) near the surface (Heemstra and Heemstra 2004). Since spawning has not been observed elsewhere and Fennessy and Sadovy (2002) only found reproductively active fish in KZN during the spawning season, it is possible that KZN fish are important contributors to genetic diversity of the species (Fennessy and Mann 2013). There is insufficient evidence, however, to rule out the possibility that infrequent spawning events may occur elsewhere.

Epinephelus andersoni adults have been found to display pioneering behaviour and to undertake small migrations, supposedly in search of available habitat that has been vacated by other predatory fishes (Maggs 2011). Despite this, available tagging data suggests that *E. andersoni* are relatively sedentary and do not undergo substantial migrations as adults (Heemstra and Randall 1993; Maggs 2011; Dunlop and Mann 2012a). This leads to the assumption that the majority of the dispersal of the species occurs through larval dispersal. Little is known about the eggs, larvae, spawning or migratory habits of *E. andersoni*, and therefore hypotheses concerning possible dispersal patterns are speculative.

1.1.2 Threats to *Epinephelus andersoni* and rockcods

Epinephelus andersoni is one of a number of endemics that inhabit the southern African coastline (van der Elst *et al*. 2005). The high level of endemism found in the marine environment off South Africa is attributed to the unique environments available in this region (van der Elst *et al*. 2005). Furthermore, favourable rocky reef habitat for *E. andersoni* is patchy, which further restricts the range of the species (Fennessy 2004). Although *E. andersoni* is endemic to the south-east coast of Africa (Heemstra and Randall 1993), the species is relatively widespread and common within this region (Fennessy 2000). Sadovy de Mitcheson *et al.* (2013) found that most groupers that are considered threatened or near threatened, generally have wide geographic ranges, indicating that a wide distribution does not necessarily infer resilience of groupers to fishing or other pressures.

All groupers, including *E. andersoni*, have a high market value due to their good tasting flesh and are therefore among the most valuable species on the reef fish market, often being the first reef species to be exploited (Sadovy de Mitcheson *et al.* 2013). They are also popular in the aquarium and live reef fish food trades, and wild groupers are commonly harvested for ranching purposes (Sadovy de Mitcheson *et al.* 2013). This has increased their value enormously and contributed to further intense

fishing pressure on this family of fishes (Fennessy 2000). *Epinephelus andersoni*, like many groupers, is vulnerable to over-exploitation due to its complex life-history characteristics, being a diandric protogynous hermaphrodite (Fennessy and Sadovy 2002). Furthermore, many grouper species aggregate to spawn and these aggregations have been frequently targeted by fishers, which severely impacts grouper populations (Sadovy de Mitcheson *et al.* 2013). A combination of these factors has put a lot of pressure on grouper stocks worldwide and although their popularity has stimulated developments in grouper aquaculture in Asia and America (Sadovy de Mitcheson *et al.* 2013), these developments often rely heavily upon wild-caught juveniles putting further pressure on wild populations (Sadovy de Mitcheson *et al.* 2013). Therefore, over-exploitation is to blame for the majority of groupers that are currently considered at risk of extinction (Sadovy de Mitcheson *et al.* 2013).

Fishing pressure had reduced the spawner biomass-per-recruit of *E. andersoni* to 42% of pristine levels by the mid-1990s (Fennessy 2004). The stock status of *E. andersoni* was said to be at approximately 40% of pristine levels in 2000 (optimally exploited) (Fennessy 2000), although it is now thought that the stock may be exploited beyond optimal levels (Fennessy and Mann 2013). Due to extensive exploitation of the stock by fisheries and the restricted range of the species, *E. andersoni* is currently listed as near-threatened on the IUCN Red List of Threatened Species (Fennessy 2004). Furthermore, groupers such as *E. andersoni*, are generally poorly managed due to identification difficulties and are often lumped together in catch records where multiple species are targeted, making them even more susceptible to over-exploitation (Fennessy and Sadovy 2002; Sadovy de Mitcheson *et al.* 2013). Although fishing pressure is largely to blame for the decline of groupers, the effect of other environmental changes such as climate change on this group of fishes, have not been quantified (Sadovy de Mitcheson *et al.* 2013).

1.2 South African and Mozambican fisheries

The high levels of ichthyodiversity and ichthyological endemism in the south west Indian Ocean (SWIO), and particularly the subtropical region, is coupled with a high consumer demand for marine resources (van der Elst *et al.* 2005). The SWIO region also incorporates several of the worlds' developing countries - including Mozambique and South Africa - that rely heavily on marine resources (van der Elst *et al.* 2005). This is concerning in terms of management and conservation since many regional endemics are over-exploited due to the generally poor fisheries resource management in the region (van der Elst *et al.* 2005). Very few of the informal fisheries in the Western Indian Ocean are subject to effective management measures that ensure sustainable

exploitation while maintaining the rich biodiversity within the region (van der Elst *et al*. 2005). Existing management of shared stocks is also inconsistent between neighbouring countries, and the classification of different fisheries sectors (subsistence, artisanal and industrial) varies between countries making proper management difficult (van der Elst *et al*. 2005). Capacity and political will for developing appropriate management structures that are consistent is also low, and co-management and regional collaboration with regards to fisheries is very limited (van der Elst *et al*. 2005).

The Mozambique fishery

The linefishery in Mozambique is currently divided into the following official sectors: artisanal, recreational, semi-industrial and industrial (Fennessy *et al.* 2011). These sectors are then divided further, according to vessel type, crew numbers and gear used.

Catches in the southern Mozambique linefishery are dominated by sparids (Dengo and David 1993; Fennessy *et al*. 2011), although serranids are relatively well represented (van der Elst *et al.* 2003) (see Table 1.1). Between 2007 and 2009 serranids comprised approximately 2% (Fennessy *et al.* 2011) of the catches of semi-industrial linefish vessels in southern Mozambique, which is a decrease from 1998 and 2000 levels when they comprised about 18.49% and 8.9% respectively (van der Elst *et al*. 2003) (Table 1.1). Data collection methods were, however, not consistent between these periods, with 1998 data based on port landings and 2000 and 2007-2009 based on on-board observer data (Fennessy *et al.* 2011). Data collected between 2008 and 2010 from 430 recreational skiboat outings, mostly from Maputo, indicated that a total of 342 serranids were caught, contributing 9% of the total catch by numbers and 6% by weight (Fennessy *et al.* 2011). Pelagics and other game fish contributed a more significant portion of the recreational catch (Fennessy *et al.* 2011). *Epinephelus andersoni* are also common in artisanal and subsistence landings in and around Maputo (pers. obs.).

The management of the Mozambique fishery falls under the Ministry of Fisheries and the legislation that governs management are the Fisheries Master Plan and the Fishery Act (Lei No 3/90 of September 1990) (van der Elst *et al*. 2003). Current management is not effective at controlling fishing pressure, as participation in the Mozambique linefishery is controlled by permit issue with minimal restrictions placed on effort and so the fishery can essentially be classified as open access (van der Elst *et al*. 2003).

Table 1.1: Summary of Mozambique catch statistics between 1994 and 2009, calculated from mainly semi-industrial vessel catches (modified from Fennessy *et al.* 2011 and van der Elst *et al*. 2003). Figures indicate percentage contribution to the total catch. Other fish including sciaenids, lethrinids, carangids and scombrids were not included in this summary. Blanks indicate no data.

	1994	1996	1997	1998	2000	2007-2009
Serranidae	3.04	9.31	15.78	18.41	8.9	2
E. andersoni		0.9	1.6	0.94	0.57	
Sparidae	34.5	61.6	22	36.49	24.4	63
Lutjanidae	45.26	5.4	8	7.69	7.67	

The South African fishery

In South Africa, *E. andersoni* is caught in the commercial and recreational boat-based linefishery in KZN and the Transkei, as well as the shore-based fishery (Fennessy and Sadovy 2002; Dunlop and Mann 2013). It also contributes to spearfishing and charter boat catches (Mann *et al*. 1997; Dunlop and Mann 2013). *Epinephelus andersoni* is well represented in boat-based catches in KZN, constituting approximately 4.32% of recreational boat catches and 3.12% of commercial boat catches, by mass, between 2008 and 2009 (Dunlop and Mann 2013). Although the species composition of the recreational boat-based fishery has changed considerably in the past two decades, *E. andersoni* is still the sixth most important species caught (Dunlop and Mann 2013). In the Transkei, *Epinephelus* spp. comprised between 4.4% and 12.5% by mass of commercial boat-based catches between 1997 and 1999 (Fennessy *et al*. 2003). *Epinephelus andersoni* contributed 1.1%, by number, to the total catch of recreational shore competition anglers in the Transkei between 1977 and 2000 (Pradervand 2004). *Epinephelus andersoni* is not caught in significant numbers in the Eastern Cape (Stilbaai to Kei Mouth) shore or boat-based fishery (Brouwer and Buxton 2002).

The South African recreational linefishery is managed on species-specific bag and minimum size limits while the commercial fishery is currently managed mainly through total allowable effort (TAE) allocation, together with species specific size limits and bag limits for some species (including *E. andersoni*) (DAFF 2012). The bag limit for *E. andersoni* is five per day for both recreational and commercial sectors and a size limit of 500 mm (larger than male size at first maturity and female size at 50% maturity) has been implemented for all fishery sectors (Fennessy 2004; Dunlop 2011). However, under-size *E. andersoni* are commonly kept by shore anglers in KZN (Dunlop 2011). This could have a significant effect on the *E. andersoni* population since it is one of the few species targeted by both the shore and skiboat fishermen (Dunlop 2011).

1.3 Climate change and impacts on genetic diversity

Rijnsdorp *et al.* (2009) hypothesised that fish species with narrow dietary preferences; that are spatially restricted with narrow habitat requirements during part of their life-history; and that are under intense exploitation will be more vulnerable to climate change than generalists and under-exploited species. Being a range-restricted endemic (Heemstra and Randall 1993) that is probably over-exploited (Fennessy and Mann 2013), it is therefore likely that *E. andersoni* may be susceptible to the environmental changes brought on by climatic change (Rijnsdorp *et al.* 2009). The interactive effects of climatic changes and fishing pressure can be illustrated by the cod fishery off western Greenland. Cod stocks collapsed in the 1990s due to a combination of increased fishing pressure and alternating periods of warming and cooling which affected recruitment (Hamilton *et al.* 2000). Genetic diversity of wild populations is also reduced by exploitation, resulting in populations that are less resilient in dealing with environmental changes (Allendorf *et al*. 2013). The vulnerability of wild marine stocks to climatic changes is concerning given the high value of marine produce, and the reliance of many of the worlds' poorest populations on fisheries for the majority of their protein intake (van der Elst *et al*. 2005, Brander 2007).

General ocean warming is an expected result of a warmer climate brought about by climate change (Parmesan and Yohe 2003; Perry *et al.* 2005; Sabatés *et al.* 2006; Hiddink and Hofstede 2008; Cheung *et al.* 2009), and since many fish species live close to their thermal tolerance limits (Harley *et al.* 2006; Sunday *et al.* 2012), their distributions are predicted to change in response to warming (Perry *et al.* 2005). Coastal oceanic climate change and variability around southern Africa has recently been studied (Rouault *et al.* 2009, 2010) (Figure 1.3). Rouault *et al*. (2009) found that since the 1980s the sea surface temperature (SST) of the Agulhas Current has increased significantly as a result of changing wind patterns and subsequent intensification of the Agulhas Current. In coastal areas, warming has been recorded along the Transkei and KZN coastline (Rouault *et al.* 2010). However, changing wind patterns have also resulted in seasonal cooling in areas of strengthened upwelling (Rouault *et al.* 2010). Ocean circulation patterns such as coastal upwelling can act as barriers to movement of marine organisms: for example they have been used to explain observed patterns of genetic diversity in the Hawaiian grouper *Epinephelus quernus* (Rivera *et al.* 2004). Intensification of upwelling cells causing regional cooling in the future may fortify these barriers, and could result in distributional changes and possibly fragmentation of marine populations. This will be detrimental for gene flow between fragmented populations and the loss of diversity may subsequently limit the adaptability of species to new environmental conditions (Ayre and Hughes 2004).

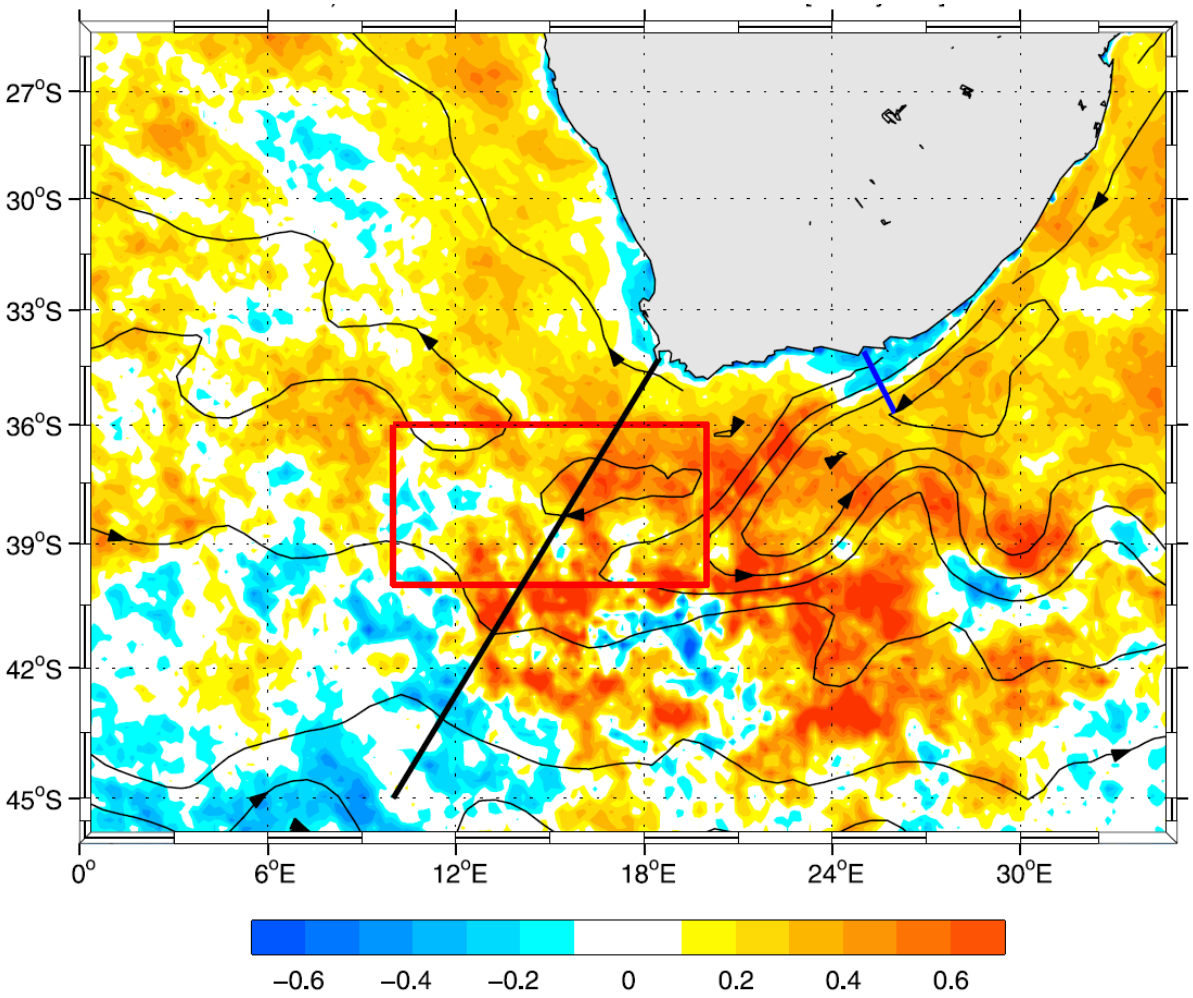


Figure 1.3: Observations of the recent warming in the Agulhas Current system, taken from Rouault *et al*. (2009). Linear trend in Advanced Very High Resolution Radiometer sea surface temperature from 1985 to 2006 (°C/decade) showing a warming of up to 0.7 °C/decade. The black contours represent mean surface currents and arrows indicate the direction of the flow. The blue and black lines and red box are not relevant here.

The genetic diversity present in populations provides the raw materials for adaptation to new climatic conditions (Ayre and Hughes 2004; Reusch *et al*. 2005; Hampe and Petit 2005; Ehlers *et al*. 2008). Although the predicted rate of climate change is thought to outpace the rate of evolution (Opdam and Wascher 2004), the resilience of populations is still highly dependent on the levels of genetic diversity present in a population at any given time (Ayre and Hughes 2004). An assessment of the overall genetic diversity of species can provide information on the capability of a species or population to deal with environmental change and should provide useful information for managers about the specific levels of population diversity and can therefore indicate areas that are conservation priorities (Ayre and Hughes 2004).

Despite their importance, little research has been done on the quantitative effects of climate change on marine fauna in south-east Africa (although see: Potts *et al*. in press; Clark *et al*. 2000; Roessig *et al*. 2004; Clark 2006; James *et al*. 2008; Lloyd *et al*. 2012). Globally, conservation organisations are becoming more aware of the effects of environmental variability on biodiversity and resources under exploitation, and are developing appropriate precautionary management strategies (Hannah *et al.* 2002; Willis and Birks 2006). Considering the depleted status of many of southern Africas' fisheries (van der Elst *et al.* 2005; Dunlop 2011), it is clear that if southern African fish stocks are to be managed sustainably, current and future climate change and associated environmental changes will have to be taken into account. A precautionary approach to management is vital in preventing the confounding effects of environmental change on over-exploited fish stocks (Rijnsdorp *et al.* 2009).

1.4 The role of southern African oceanographic features in shaping genetic diversity

The oceanographic features dominating the northern part of the *E. andersoni* distribution in Mozambique are the Delagoa Bight and the associated cyclonic and anticyclonic Mozambique Channel eddies. Mozambique waters are thought not to contribute much water volume to the Agulhas Current due to the slower meandering eddies in the Mozambique Channel (Lutjeharms 2006), possibly creating a barrier to gene flow between these two current systems. South of about 28°S in KZN, the *E. andersoni* distribution is dominated by the warm western boundary Agulhas Current, which sweeps close inshore due to the predominantly narrow and steep continental shelf in this region (Hutchings *et al.* 2002; Lutjeharms 2006) (Figure 1.4 and Figure 1.5). However, the continental shelf widens in some areas such as the Natal Bight (from St Lucia to just south of Durban), Port St Johns and the Algoa Bay area (Port Alfred upwelling cell), disrupting the predominantly unidirectional Agulhas flow and causing variable flow regimes, eddies or upwelling cells (Lutjeharms *et al.* 2000; Hutchings *et al.* 2002; Lutjeharms 2006; Roberts *et al.* 2010) (see Figure 1.5). The number of fish species occurring along the South African south-east coast declines significantly from Algoa Bay, and this has been attributed to a decrease in temperatures in this region as the continental shelf widens (Maree *et al*. 2000). Similarly, the Mbashe Estuary marks the distribution limits of some tropical marine organisms and represents the biogeographic boundary between the subtropical and warm-temperate bioregions (Mann *et al.* 2006). Biogeographic breaks and diverse oceanographic features could therefore be partly responsible for any patterns of genetic diversity observed in regional marine species that have planktonic larvae.

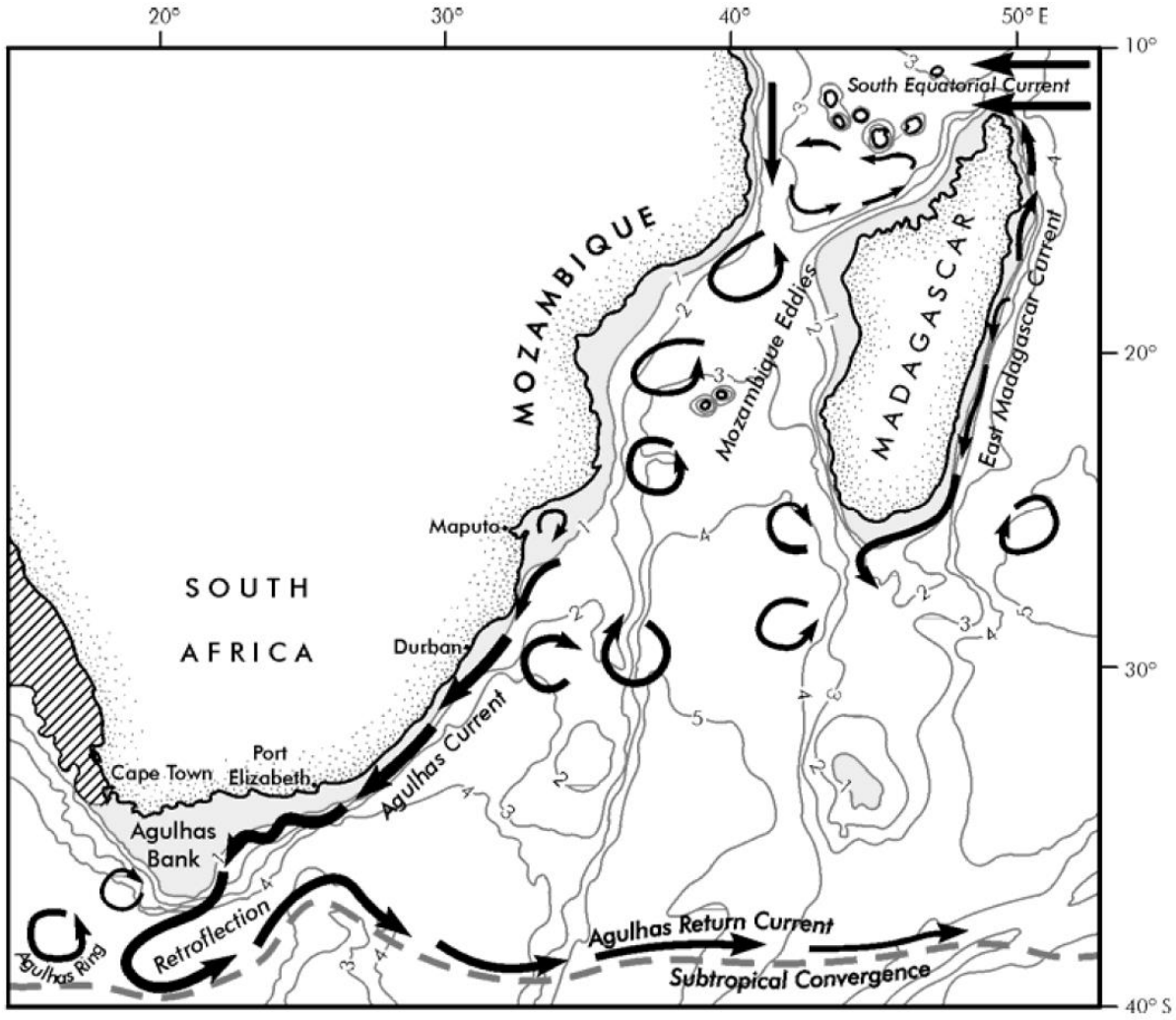


Figure 1.4: General oceanographic currents of South Africa, Mozambique and Madagascar (taken from Lutjeharms 2006).

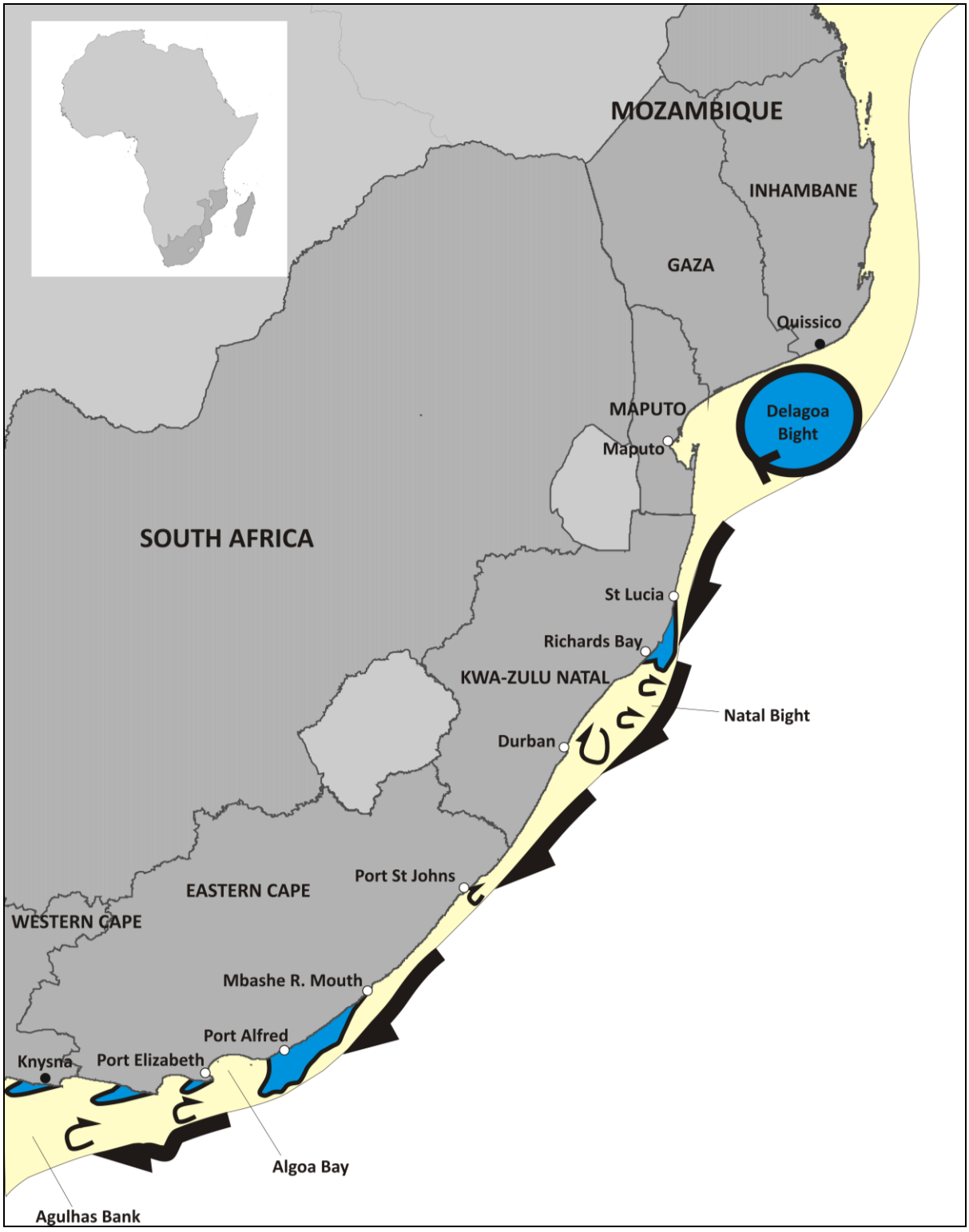


Figure 1.5: Oceanographic currents off the south-east African mainland. Cream colour indicates the continental shelf; large southward pointing arrows represent the Agulhas Current; blue areas indicate upwelling; circular arrows indicate eddies or irregularities in flow. The distributional limits of *Epinephelus andersoni* are indicated by black dots and other localities are indicated by white dots. The Transkei coast extends southward from the KZN border to the Mbashe R. Mouth.

1.5 Rationale for the study

Epinephelus andersoni occurs within a region of high endemism due to unique environmental conditions (van der Elst *et al.* 2005) and these unique habitats are predicted to shrink rapidly with future climatic changes, making them and their adapted species a conservation concern (Ohlemüller *et al.* 2008). Coupled with the fact that range-restricted endemics are a research priority, endemics are conveniently suitable for inclusion in species distribution modelling (SDM) studies such as the current study. This is because their narrower distributions can be fully captured by SDMs as the full realised niche (Broennimann *et al.* 2006) enabling SDMs to more accurately predict their distribution. Model errors are generally greater for more widespread species with larger environmental niches (Coetzee *et al.* 2009). *Epinephelus andersoni* is therefore an ideal candidate for inclusion in a modelling study since it is both endemic and a valuable resource (Fennessy 2000, 2004; Fennessy and Sadovy 2002), making it a research priority.

Studying the population genetics of *E. andersoni* should provide vital information on the levels of structuring of diversity and could therefore determine the adaptability of the stock to climate change effects (Ayre and Hughes 2004; Reusch *et al*. 2005; Hampe and Petit 2005; Ehlers *et al*. 2008; Allendorf *et al*. 2013). Determining levels and patterns of gene flow between subpopulations also provides important information on the potential impacts that climate change- induced fragmentation will have on the species (Ayre and Hughes 2004). Identifying key areas of high genetic diversity that are associated with range areas that may be lost due to environmental change, will also assist managers in determining where best to focus conservation and management efforts in the long-term. In this way, effective conservation measures can be implemented that will be more flexible and that will allow for the effects of climate change on the species (Hannah *et al*. 2002).

1.6 Aims and hypotheses

The aims of this study were to describe the genetic structure and diversity, and determine possible range shifts of *E. andersoni* with projected future changes in sea surface temperature. The first hypothesis is that there is genetic structure within the *E. andersoni* stock, and the second is that there will be a shift or a change in the distributional range of the species with future modelled temperature changes. It is hoped that the findings of this study will provide information to managers and promote a precautionary approach to the management of similarly over-exploited and range-restricted endemics in the region, especially in light of current climatic changes.

1.7 Thesis outline

This thesis is comprised of four chapters as follows:

Chapter 1 is a general introduction to the study, which includes background information on the study species, *E. andersoni*, its habitat and the current threats that affect its population. Some background information on the trends of exploitation in the South African and Mozambique fisheries is also provided. The impacts of climate change on fisheries are briefly mentioned as well as the need for a precautionary approach to fisheries management.

Chapter 2 focuses on the genetic stock structure of *Epinephelus andersoni* and describes the current levels of genetic diversity, with possible explanations for the observed sub-structuring. These analyses were done using two gene regions: the mitochondrial DNA cytochrome *b* and the nuclear ribosomal protein S7-1.

Chapter 3 concentrates on modelling the distribution of *Epinephelus andersoni* with future projected sea surface temperature changes. The possible changes in the distributional range of *E. andersoni* are determined using a number of SDMs. Future projections of sea surface temperature are also discussed.

Chapter 4 consists of a general discussion of the entire study. The key findings of the analysed data are summarised. The likelihood and possible extent of *E. andersoni* loss in genetic diversity as a result of range contractions caused by climate change are discussed as well as the implications this would have on the resilience and effective management of the species. Further study that is needed to clarify the findings of the current study is also detailed.

CHAPTER 2: The genetic stock structure of *Epinephelus andersoni*

2.1 Introduction

The delineation of separate stocks within a species of fish is a core aspect in the development of effective management plans aimed at maintaining the productivity and sustainability of marine fisheries and conserving genetic diversity (Waldman 1999; Palsbøll *et al.* 2007). Although stocks have been defined in a number of ways, it is currently accepted that the definition of stocks should be consistent with the concept of "populations" that are characterised by some degree of genetic integrity (Waldman 1999). This has stimulated the study of the genetics of populations within species (population genetics) so that management units can be identified and incorporated into stock management (Palsbøll *et al.* 2007). The development of the polymerase chain reaction (PCR) in 1983 (Mullis 1990) as well as coalescent theory by Kingman (1982a, b) combined with an increase in computational power for modelling and statistics, provided essential foundations for the practice of population genetics (Hudson 1990; Ohta 1992; Allendorf *et al.* 2013).

The relatively recent application of studying the population genetics of marine populations has resulted in a paradigm shift away from the generally accepted notion that marine species were panmictic (or genetically homogeneous), with limited local adaptation and speciation. It has been increasingly observed that many marine species exhibit significant genetic structuring, so as to warrant separate management strategies for each genetically distinct stock (Waldman 1999; Cowen *et al.* 2007; Palsbøll *et al.* 2007). This is because gene flow within marine fish populations may be more restricted and complex than was initially assumed (Hauser and Carvalho 2008). For example, although genetic studies of some endemic marine fishes inhabiting the south-eastern coast of Africa, such as slinger *Chrysoblephus puniceus* (M. Duncan, unpubl. data), red roman *Chrysoblephus laticeps* (Teske *et al.* 2010), and the southern African goby *Caffrogobius caffer* (Neethling *et al.* 2008) have revealed panmictic populations, there is also evidence of genetic structuring in other species within the region. These include species inhabiting the rocky intertidal environments (Neethling *et al.* 2008), such as the South African endemic bluntnose klipfish *Clinus cottoides* (von der Heyden *et al*. 2008); South African perlemoen *Haliotis midae* (Evans *et al.* 2004); the caridean shrimp *Palaemon peringueyi* (Teske *et al.* 2007); indigenous *Perna perna* and invasive *Mytilus galloprovincialis* mussels (Zardi *et al.* 2007).

Very little is known about the genetic structure of serranid populations of the genus *Epinephelus* despite these fishes being common along the eastern coast of southern Africa. It has been suggested that some species such as the catface rockcod *Epinephelus andersoni* are relatively sedentary and do

not undergo substantial migrations as adults (Heemstra and Randall 1993; Maggs 2011). *Epinephelus andersoni* is therefore theoretically likely to possess genetic structure (Antoro *et al.* 2006) similar to other sedentary marine organisms. The patterns of genetic structure observed for such species would be primarily influenced by the dispersal of planktonic eggs and larvae and would therefore be subject to larval duration, behaviour and swimming abilities and oceanic current circulation (Antoro *et al.* 2006; Cowen *et al.* 2007; Banks *et al.* 2007; Hogan *et al.* 2010). Factors precluding larval dispersal that create barriers to connectivity include biogeographic breaks (Turpie *et al.* 2000; Sink *et al.* 2004) such as river outflow (Carlin *et al.* 2003; Antoro *et al.* 2006), and upwelling cells (Lutjeharms *et al.* 2000; Hutchings *et al.* 2002; Roberts *et al.* 2010), among others. All these factors affect larval survival and successful recruitment (*sensu* Antoro et al. 2006).

According to Heemstra and Randall (1993), groupers have relatively long-lived planktonic larvae which may allow widespread dispersal from spawning sites. However, a long larval phase could also mean that there is more time for the adverse effects of oceanic currents to act on developing larvae, for example larvae could be swept offshore making them unable to recruit into favourable habitats successfully (Cowen *et al.* 2007). This highly variable survival rate of offspring is common among marine fishes that have a planktonic larval phase and has been termed the "sweepstakes-chance matching hypothesis" by Hedgecock (1994). Very little is known about the eggs, larvae, spawning or migratory habits of *E. andersoni* such that most of the current assumptions concerning possible patterns of gene flow are speculative.

Although oceanographic features and associated biogeographic breaks are known to play a role in shaping the genetic diversity of marine stocks, other influencing factors could include pre and post-settlement selection of larvae as well as variable sources of larvae where adults spawn in various discrete locations (Hogan *et al.* 2010). Appropriate rocky reef habitat for *E. andersoni* along the southern African coast is irregularly distributed among the different biogeographic regions, where there are varying levels of exploitation. The influence of several such factors as contributors to genetic structure of species, can sometimes result in highly complex patterns of genetic differentiation, referred to in the literature as "chaotic genetic patchiness" (Johnson and Black 1982, 1984).

A number of genetic studies on *Epinephelus* species worldwide have found genetic structure that has been attributed to a range of possible causes. Although these species may have slightly different habitat requirements and migratory habits when compared to *E. andersoni*, the majority of groupers occupy similar reef habitats that are inherently patchy (Heemstra and Randall 1993; Craig *et al.*

2011). Most groupers also have long pelagic larval phases (Heemstra and Randall 1993) that have been found to range between 25 and 75 days (Hateley 2005), making larval dispersal a major determinant of the genetic structure of this group of fishes (Cowen *et al.* 2007; Hogan *et al.* 2010). For example, Wang *et al*. (2011) found genetic differentiation in widely distributed orange-spotted grouper *E. coioides* populations in the South China Sea and Southeast Asia that could have been caused by historical dispersal or environmental influences such as currents. Antoro *et al*. (2006) suggested that the differentiation within *E. coioides* samples collected in Thailand and Indonesia could be due to small, fragmented, localised populations being separated by expanses of inappropriate habitat. Buchholz-sørensen and Vella (2010) attributed the genetic differentiation observed in the Maltese dusky grouper *E. marginatus* in the Maltese archipelago in the Mediterranean, to isolation due to depth range and larval retention patterns. De Innocentiis *et al*. (2001) found that genetic structure within Mediterranean *E. marginatus* could either be due to restricted gene flow or could reflect chaotic variance caused by variability in reproductive success. Genetic structure in the rock hind *E. adscensionis* in the tropical Atlantic was attributed to some form of isolation (Carlin *et al.* 2003) and to habitat fixation or self-recruitment in the red spotted grouper *E. akaara* in the South and East China Sea (Chen *et al.* 2008). Rivera *et al*. (2004) used ocean currents to explain the observed genetic structure in the Hawaiian grouper *E. quernus*. In the case of the Goliath grouper the observed genetic differentiation was enough to warrant the separation of what was previously treated as a single stock into two separate species due to the physical historical barrier of the Isthmus of Panama (Craig *et al.* 2008). Goliath grouper populations in northern Brazil also showed structuring and it was speculated that this could be explained by habitat preferences and fidelity to reproductive shoaling sites (Silva-oliveira *et al.* 2008).

Other studies on *Epinephelus* species, however, found little or no genetic structuring. For example the red grouper *E. morio* had little genetic structure in the south-eastern U.S. Atlantic coast and the Gulf of Mexico (Zatcoff *et al.* 2004) and the Nassau grouper *E. striatus* in the Caribbean had none (Hateley 2005). Zatcoff *et al*. (2004) speculated that the lack of genetic structure in *E. morio* could be due to theoretically widespread dispersal of long-lived larvae by swift currents of the Caribbean and the Gulf of Mexico. Similarly, Hateley (2005) attributed the apparent genetic homogeneity in the *E. striatus* population to widespread dispersal of larvae and also to the capability of adults and juveniles to migrate, although the results were preliminary since not all *E. striatus* populations were included in the analysis.

A number of genetic studies have found that the genetic structure detected within wild populations is dependent on the methods used to analyse genetic data (Maruyama 1983; Bohonak 1999; Palsbøll

et al. 2007; Allendorf *et al.* 2013), the gene regions studied (Palumbi and Baker 1994; Gaffney 2000; Ward 2000; De Innocentiis *et al.* 2001; Hellberg 2006; Tollefsrud *et al.* 2009), and the extent and design of population sampling (Waples 1998; Bohonak 1999; Ward 2000; Schwartz and McKelvey 2008). The sampling and analysis of genetic material therefore needs to be carefully planned and gene regions must be selected according to the research question of interest. For example, faster-mutating regions are generally of interest to population geneticists since they give a more recent picture of the patterns of genetic diversity in the population (Allendorf *et al.* 2013). Although advancements have been made in the analysis of genetic data (Allendorf *et al.* 2013), these are still based on simplified models of actually complex populations (Wakeley 2005; Allendorf *et al.* 2013). In order to analyse genetic data using these simple models, a number of assumptions need to be made about the population, some of which may sometimes be unrealistic (Wakeley 2005; Allendorf *et al.* 2013). For example, the assumption that the loci used in the analysis are selectively neutral is required for the analysis of genetic data (Allendorf *et al.* 2013). In light of the limitations of population genetics, results obtained using these simplistic models need to be interpreted with caution and with reference to all available information including the life-history of the species of interest (Allendorf *et al.* 2013).

The aim of this chapter was to determine the genetic stock structure of the *E. andersoni* population in southern Africa. The specific objectives were to determine possible avenues of gene flow and barriers to gene flow that would explain patterns of genetic differentiation. Another objective was to establish the historical demography of the species and to determine whether the population has experienced recent collapses or expansions. Determination of the genetic stock structure of *E. andersoni* should assist in determining whether any management units exist within the population, enabling more informed management decisions and conservation actions. The null hypothesis is that there is high gene flow within the *E. andersoni* stock and therefore no genetic structuring within the population.

2.2 Materials and Methods

2.2.1 Study sites and sample collection

Genetic samples were collected from the entire range of the *E. andersoni* distribution, including the northern and southern distribution limits (Quissico and Port Alfred respectively). Samples were collected using rod and line, by spearfishermen in southern Transkei, and researchers and taggers from several institutions, namely the Oceanographic Research Institute (ORI), the South African Institute for Aquatic Biodiversity (SAIAB) and the Department of Ichthyology and Fisheries Science

(DIFS) at Rhodes University. Fish collected by the ORI taggers and other researchers using rod and line were returned after taking a fin clipping (generally from the pectoral fin) as a tissue sample. Genetic samples from Mozambique were collected from fish markets, local fishing landing sites and from semi-industrial fishing boats which landed their catch in Maputo harbour.

All genetic samples consisted of tissue clips (the majority of which were fin clips) that were preserved in labelled Eppendorfs containing 99% ethanol that were stored at -80˚C until analyses. All available information about the fish from which the samples were taken was recorded. This included details such as length or weight; fishing method; the date of capture or collection of the fish or sample; and latitude and longitude co-ordinates of the capture locality (so that depth and habitat details could be determined). Where only locality names were available, co-ordinates were obtained using Google Earth®.

2.2.2 Laboratory methods

DNA extraction

A small sub-sample of tissue (approximately 4 mm^2) was used for DNA extractions which were conducted using the commercially available Wizard® Genomic DNA Purification Kit (Promega, USA). Extraction success was verified by electrophoresis, conducted by running 5 µl of the extracted product through a 1% agarose gel that contained ethidium bromide with 1 × TBE buffer. The gel products were then viewed and photographed on an ultraviolet (UV) trans-illuminator. The concentration of nucleic acid was measured from 1 µl of the extracted product using a Nano Drop 2000 spectrophotometer (Thermo Scientific, USA). The extraction was considered successful if nucleic acid concentrations were higher than 5 mM. Extractions were stored at 4˚C until further analysis.

DNA amplification and sequencing

The selected mitochondrial and nuclear gene regions were amplified, using standard polymerase chain reaction (PCR) amplification (Saiki *et al.* 1988). A number of universal primers of the cytochrome *b* region were tested to obtain successful amplifications (Table 2.1). Primer pairs L15424 and H16458; and L15803 and H16458 were tested with a number of conditions but did not produce clear bands. The first intron of the ribosomal protein S7 gene of the nuclear DNA was amplified using universal primers which were developed by Chow and Hazama (1998). Two primers combinations were tested with S7RPEX1F as the forward primer combined with two reverse primers S7RPEX2R and S7RPEX3R (Table 2.1).

Table 2.1: Forward (F) and reverse (R) primers tested for the PCR amplification of the *E. andersoni* cytochrome *b* region (cyt *b*) and S7 ribosomal protein genes. Primers selected for the study are indicated in bold.

Primer Name	F/R	Region	Sequence	Source
L15242	F	cyt *b*	5' - CGA ACG TTG ATA TGA AAA AAC ATC GTT G - 3'	(Kocher and White 1989)
H15910	R	cyt *b*	5' - ATC TTC GGT TTA CAA GAC CGG TG - 3'	(Cantatore *et al.* 1994)
L14725	**R**	**cyt *b***	**5' - GTG ACT TGA AAA ACC ACC GTT G - 3'**	(Song *et al.* 1998)
H15573	**F**	**cyt *b***	**5' - AAT AGG AAG TAT CAT TCG GGT TTG ATG - 3'**	(Taberlet *et al.* 1992)
S7RPEX1F	**F**	**RPS7-1**	**5' -TGGCCTCTTCCTTGGCCGTC- 3'**	(Chow and Hazama 1998)
S7RPEX2R	R	RPS7-1	5' - AACTCGTCTGGCTTTTCGCC - 3'	(Chow and Hazama 1998)
S7RPEX3R	**R**	**RPS7-1**	**5' - GCC TTC AGGTCAGAGTTC AT - 3'**	(Chow and Hazama 1998)

PCR reactions for the cytochrome *b* region were conducted in 25 µl of a solution containing 2.5 µl of 1 × Buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTPs (KAPPA Biosystems, South Africa), 0.2 mM of each primer (Integrated DNA Technologies Inc., USA), one unit of Super-ThermTaq Polymerase (JMR Holdings, USA), and 1 µl (or 50 – 100 ng) of extracted genomic DNA template and the solution was topped up with DNA free water. Cytochrome *b* thermal cycling conditions consisted of an initial denaturation step at 94°C for 3 minutes, followed by 35 cycles of denaturation at 94°C for 50 seconds, annealing at 58°C for 50 seconds and extension at 72°C for 90 seconds. There was a final extension of 10 seconds at 72°C and followed by a hold step at 4°C.

PCR reactions for the RPS7-1 region were carried out following a protocol modified from Chow and Hazama (1998). The PCR mixture contained 2.5 µl of 1 × Buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 mM of each primer, one unit of Taq polymerase, and 1 µl (or 50 - 100 ng) of extracted genomic DNA template and was topped up to 25 µl with DNA free water. Thermal cycling conditions consisted of an initial denaturation step at 95°C for 60 seconds, followed by 30 cycles of denaturation at 95°C for 30 seconds, annealing at 60°C for 60 seconds and extension via DNA polymerization at 72°C for 90 seconds; followed by a final extension of 10 minutes at 72°C and hold at 4°C.

Amplification for both genes was conducted on one of three machines: an Eppendorf Mastercycler thermal cycler (Eppendorf, Germany); a Hybaid Multiblock system PCR thermal cycler (Thermoscientific); or an Applied Biosystems Veriti® 96-well thermal cycler (0.2 mL VeriFlex™ Block). Amplifications were verified using gel electrophoresis of the PCR products. Successfully amplified PCR products were sent to a commercial sequencing facility (Macrogen Inc., Korea) where they were purified and cycle sequenced on an ABI 3730xl DNA analyser (Applied Biosystems Inc.). The numbers of DNA sequences that were subsequently obtained for each of the gene regions were different due to problems in amplifying the DNA. This could have been caused by poor quality or contaminated DNA and primers or a number of other problems during the DNA amplification process.

2.2.3 Analysis of molecular data

DNA sequence chromatograms were checked in Chromas Lite version 2.01 (Technelysium Pty Ltd) for quality and length. Unusable sequences were removed from the analysis and re-sequenced where possible. Good quality sequences were checked for mutations, polymorphisms and edited. All verified sequences were then aligned manually in Seqman Pro™ version 9 (DNASTAR® Inc., Madison, WI, USA). The alignment was then imported into Bioedit Sequence Alignment Editor, version 7.0.9.0 (Hall 1999) where a Clustal W multiple alignment (Thompson *et al.* 1994) was run to realign all sequences with a profile-based progressive alignment procedure. All gaps were checked carefully for

both gene regions with only the RPS7-1 data having gaps that were treated as the fifth state (Librado and Rozas 2009). The final aligned fasta file was imported into DnaSP version 5.10.01 (Librado and Rozas 2009) for analyses and conversion to other file formats. The RPS7-1 nuclear data consisted of polymorphic DNA sequence data from which the haplotype phases had to be reconstructed. This reconstruction was conducted using the algorithm provided by PHASE (Stephens *et al.* 2001; Stephens and Donnelly 2003) in DnaSP. PHASE uses a coalescent-based Bayesian method (Stephens and Donnelly 2003) to infer the haplotypes.

Genetic diversity

Several indices of genetic diversity were calculated using DnaSP and Arlequin version 3.5.1.2 (Excoffier and Lischer 2010, 2011) software, for each locality and the entire sample: haplotype diversity (*h*), nucleotide diversity (π), number of polymorphic or segregating sites (*S*), and the average number of pairwise nucleotide differences (*k*). The numbers of segregating sites and pairwise nucleotide differences are direct measures of diversity that affect the calculations of haplotype and nucleotide diversity. Haplotype diversity (Allendorf *et al*. 2013) indicates the relative frequencies of haplotypes in a population and is a gauge of the degree of variation in haplotypes found in a given population. It is measured as the probability of two individuals drawn randomly from a population having different haplotypes. Values of *h* close to zero indicate low levels of diversity whereas values closer to one indicate that the population (or subpopulation) is highly diverse where every sample is different. Nucleotide diversity was introduced by Kimura (1968b) and is defined as the average number of differences per site between two sequences from different haplotypes, drawn at random from a sample (Felsenstein 2007). It is calculated as the total number of nucleotide differences between haplotypes, divided by the total number of nucleotide bases and represents the probability that homologous nucleotides from randomly chosen haplotypes will be different (Nei and Tajima 1981).

Haplotype network

Median-joining (MJ) haplotype networks were created using Network software, version 4.6.1.0 (Fluxus Technology Ltd) to portray all possible relationships and nodes between haplotypes as well as the most likely genealogy. The minimum spanning network connects all haplotypes but does not infer additional nodes or cycles when there are ambiguous relationships between haplotypes (Bandelt *et al.* 1999). Phylogenetic trees are primarily used for interspecific comparisons (Posada and Crandall 2001) which require large sample sizes and generally involve small genetic distances, resulting in a large number of plausible trees which are best expressed as a network (Bandelt *et al.* 1999). Haplotype networks are a more appropriate method of portraying intraspecific genetic

patterns than phylogenetic trees and they are a useful way to visualise the haplotypes (Posada and Crandall 2001). When haplotypes are arranged according to their similarity to each other, taking into consideration their geographic source, genetic structure within populations becomes evident.

Genetic neutrality and demographic history

Tests for selective neutrality

Two neutrality tests were carried out in Arlequin to test for selective neutrality of the cytochrome *b* and RPS7-1 regions since several demographic events and natural influences can affect the evolution of a population. These different tests were used in conjunction as each test is most efficient at identifying a single category of deviations from the neutral model (Fu 1997). Many of the statistical models used to test for genetic structure within populations require the assumption of selective neutrality to be met (Simonsen *et al.* 1995; Ford 2002; Allendorf *et al.* 2013). The neutral mutation-random drift theory of molecular evolution (Kimura 1968a; Ohta and Kimura 1973; Kimura and Ohta 1974; Ohta 1992) hypothesises that "most of the variability within species is not caused by Darwinian selection but by random drift of mutant alleles that are selectively neutral or nearly neutral...In other words, the selection intensity involved in the process is so weak that mutation pressure and random drift prevail in molecular evolution" (modified from Kimura 1983, pg 34). It is not currently believed that all genetic variation is selectively neutral, although it is necessary to assume selective neutrality if population genetics theory is to be applied to interpret genetic data and make predictions (Ford 2002; Allendorf *et al.* 2013). Furthermore, neutral loci are more informative than adaptive loci when determining population sizes and patterns of migration from allele frequency distributions (Allendorf *et al.* 2013). Selection acting on the gene being studied can also incorrectly signify the presence of genetic structure (*sensu* Bennett 2012).

The first neutrality test used was Tajima's *D* (Tajima 1989a,b) which tests for selective neutrality of a random sample of DNA sequences under the infinite-site model without recombination (Excoffier and Lischer 2011). The infinite-site model proposed by Kimura (1969) assumes that there are a nearly infinite number of sites available for mutation relative to the number of already segregating sites and that the mutation rate at each site is so small that whenever a mutant appears it occurs at a site in which a previous mutation is not still segregating (Kimura 1969; Tajima 1996). A positive *D* value could indicate balancing selection, or a mixture of distinct, isolated populations (Simonsen *et al.* 1995; Rand 1996). If the *D*-statistic is close to zero (the two population mutation rate parameters should be in close agreement in this case), this could be consistent with a less recent selective sweep and weaker selection or population equilibrium and selective neutrality (Simonsen *et al.* 1995). If it is

negative it can signify either: a selective sweep; a less recent bottleneck (Simonsen *et al.* 1995; Rand 1996; Fu 1997); or population growth (Simonsen *et al.* 1995).

The second test was Fu's *Fs* test (Fu 1997) which is also based on the infinite-site model without recombination and is thought to be one of the most powerful neutrality tests for detecting the presence of genetic-hitchhiking (Fu 1997). It is based on the haplotype frequency distribution and evaluates the probability of observing a random neutral sample with a number of alleles similar or smaller than the observed value, given the number of pairwise differences (π) which is taken as an estimator of the population mutation rate (ϑ) (Ewens 1972; Fu 1997; Excoffier and Lischer 2011). The random F_S statistics are generated using a coalescent simulation algorithm adapted from Hudson (1990). The F_S statistic is very sensitive to population demographic expansion (causing an excess of recent mutations and subsequently rare alleles), which often results in large negative F_S values. The F_S test *P*-values should be considered significant at the 5% level when $P < 0.02$, since the critical point at the 5% level of significance corresponds to the lower 2nd percentile of the distribution (Fu 1997; Excoffier and Lischer 2011).

Population size history by mismatch distribution

Mismatch distributions of the observed number of pairwise differences between haplotypes were plotted for each gene region in Arlequin to determine whether the population was experiencing demographic increase or geographic range increase. Harpending's (1994) raggedness index (*r*) was then calculated to determine the smoothness of the distributions and the goodness-of-fit to the demographic and spatial expansion models. It is determined by the maximum number of observed differences between haplotypes and the observed relative frequencies of the mismatch classes (Excoffier and Lischer 2011).

A mismatch distribution is based on quantitative measures of variation between all pairs of haplotypes. Inferences about historical demographic or spatial population expansion and decline can be made by looking at the shape of the distribution of current nucleotide site differences between pairs of individuals (Rogers and Harpending 1992; Excoffier and Lischer 2011). A multimodal distribution is expected in samples originating from a population at demographic equilibrium. Conversely, unimodal distributions indicate that the population has experienced a recent demographic expansion (Slatkin and Hudson 1991; Rogers and Harpending 1992) or has experienced expansion with high levels of migration between adjacent demes (e.g. Bennet 2012). Under the assumption of a stationary haploid population at equilibrium suddenly having experienced pure demographic expansion, it is possible to estimate the mutation parameters ϑ_0 (before expansion)

and ϑ_1 (after expansion) as well as the time since the expansion event (τ) (Excoffier and Lischer 2011). This can be done using a generalized non-linear least-square approach proposed by Schneider and Excoffier (1999). The significance of the mismatch test is determined by the sum of squared deviations (*SSD*) between the parameters of the observed and expected distributions, assuming that estimated parameters are true. The *P*-value is calculated as the proportion of *SSD* values from the number of simulated sets of parameters (1,000 simulations were used for the current study) that are larger or equal to those observed (Excoffier and Lischer 2011). The mismatch test was also run under the assumption of spatial demographic expansion.

Population structure

Exact tests of population differentiation

Exact tests of population differentiation were done in Arlequin to test the null hypothesis of panmixia (Raymond and Rousset 1995) in *E. andersoni*. The exact test is a qualitative test of pairwise differences which tests for non-random distribution on individuals between pairs of groups using a contingency table of groups (*r*) × haplotypes (*k*). All possible states of the contingency table are explored using a Markov chain (Raymond and Rousset 1995) which alters values within each cell of the table slightly but keeping the row and column totals constant. The significance of the test is determined by estimating the probability of observing a table less or equally likely than the observed sample configuration (Raymond and Rousset 1995; Excoffier and Lischer 2011). The significance criterion (α) was set to 0.05, and a total of 100,000 steps of the Markov chain were conducted.

Pairwise population comparisons (F_{ST})

Pairwise fixation indices (F_{ST}) (Wright 1951) were calculated as a quantitative measure of genetic differentiation between localities by pairwise comparisons of haplotype frequencies using Arlequin. Pairwise fixation indices are defined as the variance of haplotype frequencies within each pair of samples, relative to the maximum possible value based on the mean haplotype frequency across all localities/populations (Beaumont and Hoare 2003), and are calculated under the infinite-alleles model (Wright 1931; Kimura and Crow 1964). F_{ST} values that are close to zero indicate little genetic differentiation between localities due to gene flow dominating over random genetic drift. Contrarily, F_{ST} values that are close to one indicate that genetic drift is the predominant force rather than gene flow, resulting in some differentiation (Beaumont and Hoare 2003). The significance of the test for each F_{ST} comparison was calculated by permuting (10,000 permutations) haplotypes between groups and is determined as the proportion of permutations resulting in an F_{ST} value larger than or equal to the observed value (Excoffier and Lischer 2011). Pairwise F_{ST} values indicate little (0 to 0.05),

moderate (0.05 to 0.15), high (0.15 to 0.25) and very high (>0.25) genetic differentiation (Wright 1978).

Analysis of molecular variance

Analyses of molecular variance (AMOVA) (Excoffier *et al*. 1992) were conducted in Arlequin and used to estimate indices of genetic structure at a number of hierarchical levels, based on the analyses of variance of gene frequencies but taking into account the number of mutations between molecular haplotypes. The hierarchical levels of assessment used in this analysis were: among geographic groups (F_{CT}); among localities within groups (F_{SC}); and inter-individual differences within localities (F_{ST}) (Excoffier and Lischer 2011). For the among geographic groups level, sampling localities were grouped with other adjacent localities with reference to the pairwise population comparisons (F_{ST}) of each gene region and to possible regional barriers to gene flow such as oceanographic features and biogeographic breaks. A number of groupings were tested until the grouping with the highest and most significant F_{CT} variation was found. A matrix of squared distances between haplotypes was created from nucleotide sequence differences. The covariance of the levels was then assessed at the different hierarchical levels mentioned above using a non-parametric permutation procedure (Excoffier *et al.* 1992; Excoffier and Lischer 2011). For each level of the hierarchy, 10000 permutations were conducted.

Spatial analyses of molecular variance

Spatial genetic variation was also investigated among localities using SAMOVA version 1.0 (Spatial Analysis of Molecular Variance) (Dupanloup *et al*. 2002). SAMOVA is a variation of the AMOVA (Excoffier *et al.* 1992) which incorporates geographic coordinate positions of localities. The method uses a simulated annealing procedure that aims to maximise F_{CT}. The advantage of this test is that groups are chosen randomly, unlike for the AMOVA, which may make it a more subjective test of genetic structure. The available localities were allocated into a number of groups (*K*) which were defined before each analysis, and variance components (F_{CT}, F_{SC} and F_{ST}) were estimated. About 100 simulated annealing steps were conducted to ensure that the final configuration of *K* groups was not affected by the initial grouping. The analysis requires *K* to be set at a minimum of two with at least one of the groups having more than one locality.

Spatial genetic patterns

Isolation by distance

A Mantel test (Mantel 1967) was done to determine whether there was isolation by distance among sampling localities using Alleles In Space version 1.0 (AIS) (Miller 2005). The mantel test checks for a significant correlation between any existing isolation and geographic distance that may occur when migratory movements or dispersal distances are a fraction of the entire distributional range of the species (Dupanloup *et al.* 2002) or when there is declining genetic connectivity between increasingly distant populations (Slatkin 1993). A matrix of pairwise F_{ST} values was regressed against a matrix of the geographic distances (km) between localities using geographic coordinates. The significance of the test was determined by the proportion of permuted correlation coefficients (*R*) equal to or greater than the *R* observed. The analysis was done with untransformed geographical and genetic distances and with distances log-transformed prior to analysis.

Spatial autocorrelation

A spatial autocorrelation analysis (SAA) was also done in AIS to assess whether the genetic and geographic distances were interrelated and whether any correlation had a spatial pattern (Manni *et al.* 2004). A number of different distance classes were tested with equal-sized distance classes being selected for the final analysis. The distance classes were set to be equal rather than unequal (i.e. classes contain an equal number of samples) since the numbers of sequence data among localities were not always equal.

The SAA is similar to the Mantel test as it also establishes whether there is a positive relationship between genetic and geographic distance. However, this test also permits the detection of genetic structure and allows for inferences to be made about the spatial scales over which the genetic structure occurs (Manel *et al.* 2003; Miller 2005). The number and therefore the size of distance classes analysed can be set by the user (Miller 2005), making this especially useful for species inhabiting patches of suitable habitat that are not necessarily equidistant from each other. The AIS software, unlike most analyses, estimates SAA by calculating the simple statistic *Ay*, which can be interpreted as the average genetic distance between pairs of individuals that fall within distance class *y*, instead of examining allele-specific patterns using Moran's *I* (Moran 2012). The value of *Ay* will be zero when all individuals within distance class *y* are genetically identical, and one if they are completely dissimilar. Bonferroni corrections were applied to each of the non-independent tests performed on the correlations in each of the distance classes. The quantity *V*, which is an overall assessment of the variability of *Ay* among the distance classes, was then calculated. The significance of *V* for each distance class is evaluated via random allocation of individuals and genotypes among

the set of sampled points on the landscape and calculating V^{RND} for each randomization replicate with 10,000 permutations. The proportion of randomization replicates where $V^{RND} = V$ provides an estimate of the *p*-value for each distance class (Miller 2005). A confidence interval was calculated for the observed correlation coefficient at each distance class and a *p*-value was returned for the probability of a correlation significantly lower than the lower confidence interval or significantly higher than the higher confidence interval, resulting in significantly negative or positive spatial autocorrelation respectively.

Location of barriers using Monmonier's algorithm

To further determine correlations between genetic and spatial patterns, Monmonier's Maximum Difference (MMD) algorithm (Monmonier 1973) was also conducted in AIS, to locate barriers to gene flow between the sampled localities. This approach is useful to visualise data contained in a genetic matrix on a geographical map in order to identify areas of low gene flow or potential genetic boundaries (Monmonier 1973; Manel *et al.* 2003; Ivetić *et al.* 2010). Monmonier's algorithm is a useful method to putatively identify genetic boundaries rather than inferring them *a priori*. The Delaunay triangulation (Brassel and Reif 1979; Watson 1992; Brouns *et al.* 2003) procedure was used to connect adjacent geographical positions of mapped sample localities resulting in a network that connects all samples. Genetic distances between neighbouring samples are then calculated and associated to each edge of the network (i.e. each of the lines connecting adjacent localities) using the MMD algorithm. Three forms of genetic distance namely: raw data, pseudoslopes (distance corrected genetic distances), and residual genetic distances can be used for the calculation of genetic boundaries in AIS. Residual genetic distances were chosen since it has been found that they perform better than pseudoslopes in AIS analyses (Miller 2005) and in cases where correlations exist between genetic and geographical distances, raw genetic distances would affect the results (Manni *et al.* 2004; Miller 2005).

Landscape or "seascape" shape

The genetic landscape or in this case seascape shape, which is a three-dimensional interpolated genetic surface plot represented within the geographic space of the study (Manni *et al.* 2004), was plotted in AIS as a visual representation of genetic variation relative to the geographic area. This approach has been used in landscape genetics studies (Miller *et al.* 2006; Funk *et al.* 2008; Ivetić *et al.* 2008; Kenchington *et al.* 2009; Tollefsrud *et al.* 2009; Shimono *et al.* 2010; Malaney *et al.* 2012) which combine molecular population genetics and landscape ecology (Turner *et al.* 2001) to provide information on the interaction between landscape features acting as barriers to gene flow and micro-evolutionary processes (Manel *et al.* 2003). Unlike more traditional population genetics

methods, individuals do not need to be grouped prior to analysis which minimizes potential bias in any resulting genetic structure (Manel *et al.* 2003; Miller 2005). The surface plot has three axes: X and Y axes correspond to geographical locations and surface heights on the Z-axis measure average genetic distances, indicated between locality points on the plot. Due to some correlation between genetic and geographic distances for some sampling localities, residual genetic distances were used as the heights for the surface plot as opposed to raw genetic distances, as recommended by Manni *et al*. (2004). Residual genetic distances were derived from the linear regression of all pairwise genetic distances on geographical distances (Manni *et al.* 2004; Miller 2005). Large genetic distances are reflected as peaks on the surface plot whereas relative genetic similarities are represented as troughs (Miller 2005). Although the grid dimensions do not impact the surface plot as much as the distance weighting parameter, some substantial detail may be lost if the grids are too large (Miller 2005). Therefore two different grid dimensions (with X and Y axes split into 50 × 50; and with X and Y axes split into 100 × 100) and three different weighting values (a = 0.5, 1 and 2) were tested in order to ensure that interpretations were not sensitive to these parameters, as has been done in other studies (Miller *et al.* 2006; Funk *et al.* 2008; Kenchington *et al.* 2009; Shimono *et al.* 2010).

2.3 Results

2.3.1 Sample localities and numbers

Samples were collected from a total of 22 sites throughout the distribution of *E. andersoni* (Table 2.2). These sampling sites were a minimum of 12 km and a maximum of 300 km away from each other. DNA samples that were collected in small numbers and from sites within a small distance of each other were pooled into one locality (see Table 2.2 and Figure 2.1). For example, the Msikaba locality includes samples that were collected along an 11 km stretch of coastline (from Mnyameni to Msikaba) with sampling sites 5.7 to 5.9 km away from each other. The Coffee Bay samples were collected along a 20.5 km stretch of coastline from Presleys Bay to Hole in the Wall with sampling sites being 4.8 to 8.5 km apart.

Table 2.2: The number of *Epinephelus andersoni* samples (N) per locality and per sub-site within localities; GPS co-ordinates of localities; and the number of sequences for the RPS7-1 after phasing and cytochrome *b* regions.

Locality region	Locality name	Abbr.	Sub-sampling site	GPS Co-ordinate		Samples (N)	Sequences	
				Lat	Long		S7	Cyt
Mozambique	1. Quissico	QU		-24.66	35.05	18	12	15
	2. Xai Xai	XX		-25.12	33.74	52	31	44
	3. Maputo	MP		-25.86	32.72	26	18	24
	4. Inhaca	IN		-26.10	32.86	40	33	28
KwaZulu-Natal (KZN)	5. Cape Vidal	CV		-28.12	32.56	37	32	34
	6. Richards Bay	RB		-28.77	32.13	35	32	34
Transkei	7. Port Edward	PEd		-31.06	30.22	11	10	11
	8. Msikaba	MS		-31.32	29.97	17	15	17
			Mnyameni			9	8	9
			Mtentu			5	4	5
			Mkambati			2	2	2
			Msikaba			1	1	1
	9. Port St Johns	PSJ		-31.63	29.55	21	20	21
	10. Coffee Bay	CB		-31.98	29.15	10	10	9
			Presleys Bay			4	4	3
			Mdumbi			2	2	2
			Coffee Bay			3	3	3
			Hole in the Wall			1	1	1
Southern Cape	11. Mbashe	MB		-32.24	28.92	23	18	21
	12. Port Alfred	PA		-33.60	26.90	13	9	11
			Kowie estuary			11	8	10
			Kenton			2	1	1
Total						*303*	*240*	*269*

Note: The shading in the table indicates marine bioregions: white, light grey and dark grey indicate the tropical Delagoa, subtropical, and warm-temperate bioregions respectively. The bioregions used were taken from the South African National Biodiversity Assessment 2004 (Sink *et al.* 2004).

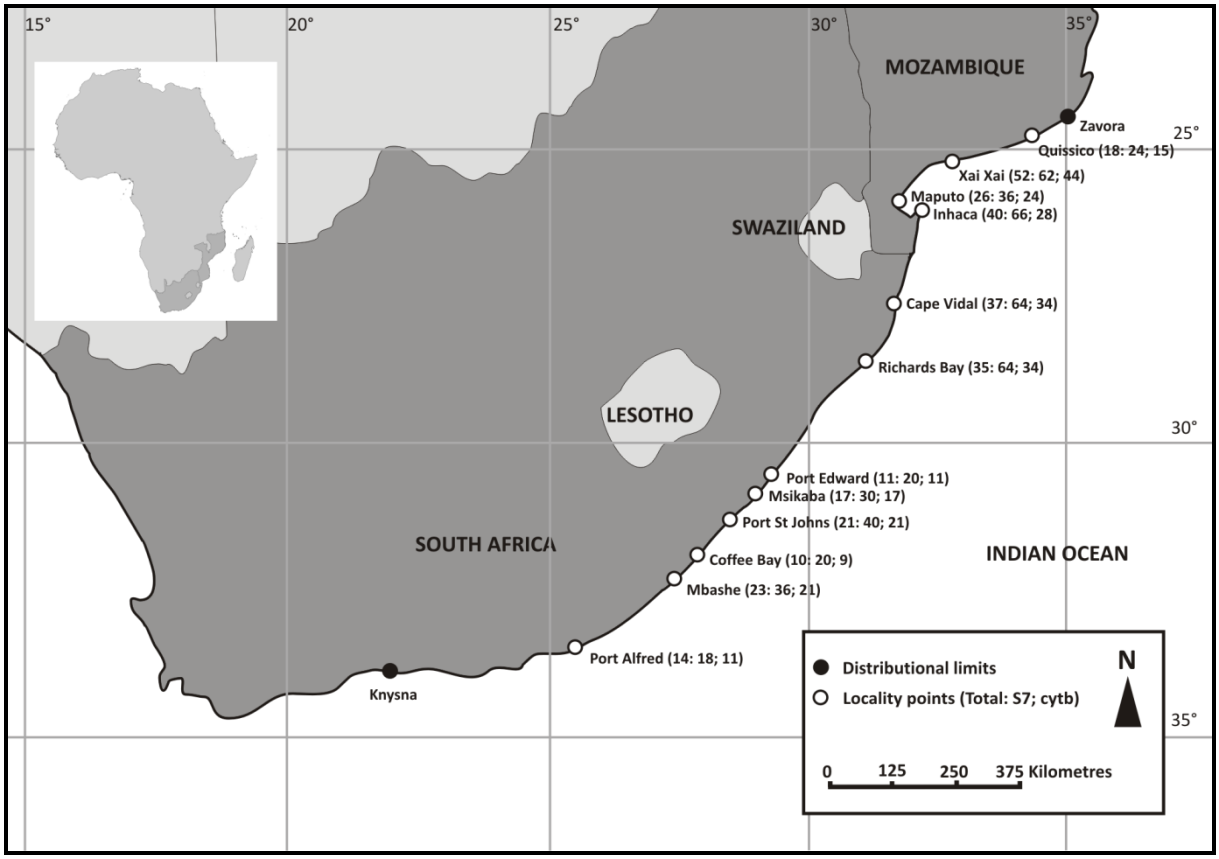


Figure 2.1: Map of south-eastern Africa showing the sampling localities for *Epinephelus andersoni*. Total sample numbers are indicated in parentheses after the name of the locality followed by the numbers of RPS7-1 and mitochondrial DNA cytochrome *b* sequences respectively.

2.3.2 Mitochondrial DNA - Cytochrome *b*

Genetic diversity

The primer pair H15573 and L14725 was the most successful as it produced a clean PCR product for the last 771 base pairs (bp) of the cytochrome *b* gene (base pair position 370 to 1141) and was therefore selected for the study. These sequences were successfully generated for 269 individuals of the 303 genetic samples that were collected. All sequences were checked against other *Epinephelus* cytochrome *b* sequences downloaded from GenBank. There were a total of 12 unique haplotypes represented in the 771 bp data resulting from a total of nine segregating sites (Table 2.3). The number of segregating sites (*S*) was relatively consistent among localities, with Richards Bay having the highest number of haplotypes (i.e. six). There were a very few private haplotypes (four in total) with two being restricted to Cape Vidal. Haplotype diversities were low to moderate (0.174 - 0.507) across all localities and nucleotide diversities were low (0 - 0.001).

Table 2.3: Cytochrome *b* (711 bp) diversity indices and the test statistics and corresponding *p*-values for the tests for selective neutrality. Diversity indices represented include haplotype (*Hd*) and nucleotide (π) diversities as well as the numbers of sequences (*N*), polymorphic sites (*S*), haplotypes (*H*), private haplotypes (*Hp*) and mean number of nucleotide differences between sequences (*k*) for each of the 12 sampling localities. Significant values ($p < 0.05$) are indicated in bold.

								Tajima's		Fu's	
Locality	*N*	*S*	*H*	*Hp*	*k*	*Hd*	π	*D*	*p*	*Fs*	*p*
Quissico	15	2	3	0	0.381	0.362	0.000	-1.002	0.195	-0.918	0.079
Xai Xai	44	3	4	0	0.180	0.174	0.000	-1.574	**0.017**	-3.353	**0.005**
Maputo	24	3	4	0	0.326	0.308	0.000	-1.494	**0.049**	-2.383	**0.006**
Inhaca	28	3	4	0	0.347	0.267	0.000	-1.321	0.076	-2.109	**0.024**
Cape Vidal	34	4	4	2	0.446	0.271	0.001	-1.353	0.092	-1.374	0.099
Richards Bay	34	5	6	1	0.504	0.414	0.001	-1.553	**0.031**	-3.724	**0.002**
Port Edward	11	2	3	0	0.364	0.345	0.000	-1.430	**0.018**	-1.246	**0.021**
Msikaba	17	4	5	0	0.574	0.507	0.001	-1.577	**0.038**	-2.826	**0.002**
Port St Johns	21	4	5	1	0.381	0.352	0.000	-1.873	**0.008**	-3.736	**0.001**
Coffee Bay	9	2	3	0	0.444	0.417	0.001	-1.362	0.098	-1.081	**0.048**
Mbashe	21	3	4	0	0.286	0.271	0.000	-1.727	**0.017**	-2.820	**0.001**
Port Alfred	11	2	3	0	0.364	0.345	0.000	-1.430	**0.032**	-1.246	**0.036**
Overall	*269*	*9*	*12*		*0.367*	*0.309*	*0.000*	*-1.475*	*0.056*	*-2.235*	***0.027***

Haplotype network

The minimum spanning haplotype network of the cytochrome *b* samples (Figure 2.2) has a typical star-like topology, with a small number of haplotypes representing a relatively small number of individuals and branching off the central common haplotype. Eight of the haplotypes were shared with the central haplotype (haplotype 1) being present in all 12 geographic localities and representing 223 individuals which was more than 80% of the sample. Three of the shared haplotypes (2, 3 and 4) were also common and distributed among most of the localities with eight, six and seven of the 12 localities being represented in each respectively. Haplotypes 5 to 8 were also shared but to a lesser degree. The four rare haplotypes consisted of individuals from Cape Vidal (two haplotypes), Richards Bay and Port St Johns. These unique haplotypes could indicate some degree of isolation for these localities.

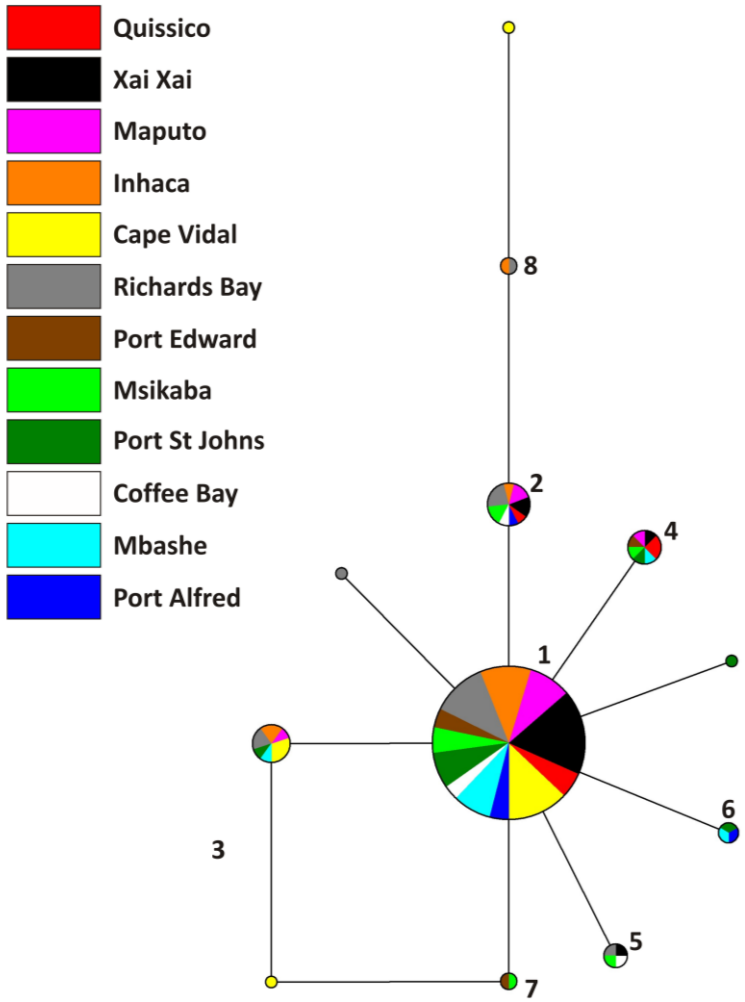


Figure 2.2: Minimum spanning haplotype network of the 269 sequences from 12 localities distributed amongst 12 unique cytochrome *b* haplotypes. Circle sizes are proportional to haplotype frequencies: the smallest circle represents one sample. Connecting branches between haplotypes are one mutational step while circle segment colours represent the geographical origin of the samples.

Genetic neutrality and demographic history

Tests for selective neutrality

Overall Tajima's *D* was not significant ($p = 0.056$) but was negative overall and for all localities (Table 2.3). However Tajima's *D* departed significantly from neutrality for eight of the 12 localities. Fu's *Fs* values were also negative and showed significant departure from neutrality for ten out of 12 localities as well as significant departure overall ($p = 0.027$).

Population size history by mismatch distribution

The expected mismatch distributions calculated under the demographic expansion model (population size increase) and the spatial expansion model (geographic range increase) were very similar (Figure 2.3). The mismatch distributions were both unimodal which suggests recent demographic and spatial expansion (Rogers and Harpending 1992; Ruegg and Smith 2002) of the *E. andersoni* population. The sum of squared deviations was non-significant ($p > 0.4$) for both the spatial and the demographic expansion models, indicating that the null hypotheses of spatial and demographic population expansion cannot be rejected. Harpending's (1994) raggedness index *r* was fairly low and non-significant for the demographic ($r = 0.239$, $p = 0.594$) and the spatial expansion models ($r = 0.239$, $p = 0.588$). Low Harpending's *r* values are generally associated with smooth, unimodal distributions (Excoffier and Lischer 2011) and the non-significance of this estimate suggests a good fit of the observed data to the models of population demographic and spatial expansion.

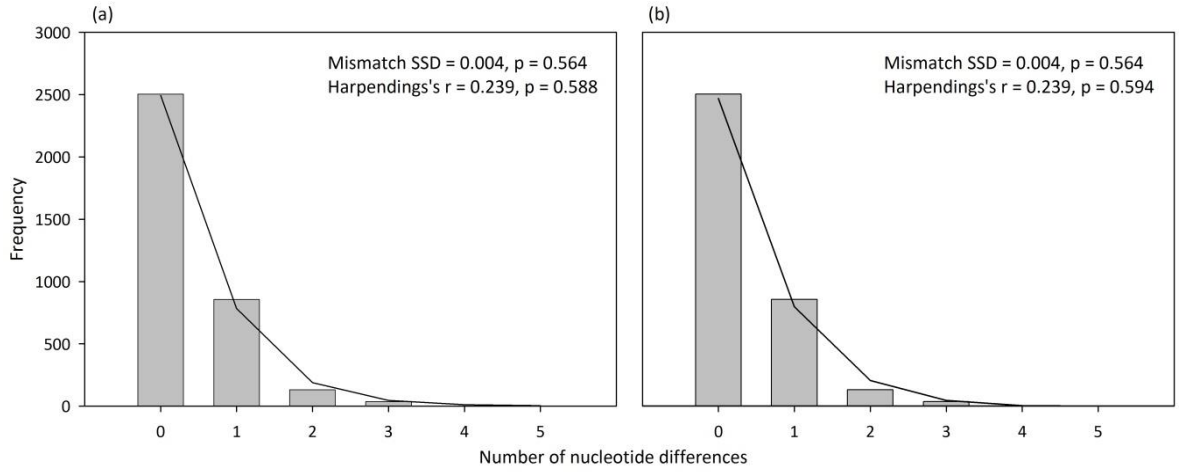


Figure 2.3: Observed (bars) and expected (line) mismatch distributions of the frequency of pairwise nucleotide differences between all pairs of haplotypes in the cytochrome *b* data, under the spatial expansion (a) and demographic expansion (b) models.

Population structure

Population differentiation

Exact tests of differentiation and pairwise population comparisons (using F_{ST}) only revealed significant comparisons for two comparisons associated with Cape Vidal. For exact tests of differentiation, the two significant comparisons were Cape Vidal versus Xai Xai and versus Msikaba ($\alpha = 0.05$) with only the first comparison being significant ($p = 0.018$) for the pairwise F_{ST} comparisons (Table 2.4). The fact that two rare haplotypes are represented in the Cape Vidal samples (Figure 2.2) could be contributing to the significant comparisons found between Cape Vidal and the two other localities.

Table 2.4: Pairwise F_{ST} values (above diagonal) and exact tests of genetic differentiation *p*-values (below diagonal) between localities for 771 bp cytochrome *b* data and the AMOVA grouping of localities that maximised F_{CT}.(Groups 1 to 6) are defined. Significant *p*-values (α = 0.05) are indicated in bold.

	AMOVA groups (6)	1	2	3	4	5	6	7	8	9	10	11	12
1. Quissico	1	*	0.013	-0.025	0.016	0.038	0.008	-0.042	-0.033	-0.008	-0.010	-0.007	-0.008
2. Xai Xai	2	0.296	*	-0.015	0.004	**0.038**	0.008	0.015	-0.001	0.006	-0.010	0.001	-0.004
3. Maputo	2	0.875	0.589	*	-0.027	0.002	-0.023	-0.009	-0.027	-0.012	-0.031	-0.015	-0.032
4. Inhaca	2	0.317	0.265	0.790	*	-0.014	-0.024	0.016	-0.003	-0.004	-0.019	-0.006	-0.021
5. Cape Vidal	3	0.084	**0.045**	0.340	0.844	*	0.005	-0.002	0.016	0.002	0.013	-0.001	0.011
6. Richards Bay	4	0.495	0.231	0.985	0.943	0.215	*	0.017	-0.020	0.012	-0.048	0.011	-0.029
7. Port Edward	5	0.866	0.274	0.625	0.356	0.317	0.459	*	-0.037	-0.025	0.002	-0.023	0.000
8. Msikaba	5	1.000	0.130	0.623	0.154	**0.032**	0.662	0.925	*	0.006	-0.071	0.008	-0.037
9. Port St Johns	5	0.776	0.162	0.700	0.637	0.474	0.411	0.864	0.427	*	0.005	-0.043	-0.025
10. Coffee Bay	5	0.533	0.332	0.692	0.372	0.236	0.842	0.701	1.000	0.605	*	0.016	-0.051
11. Mbashe	6	0.658	0.318	0.753	0.791	0.640	0.526	0.773	0.360	1.000	0.487	*	-0.023
12. Port Alfred	6	0.560	0.333	0.767	0.459	0.333	0.757	1.000	0.948	0.857	0.848	0.766	*

Analysis of molecular variance

The AMOVA was conducted between six geographic groups with only three of these containing multiple localities (group 2, 5 and 6) (see Table 2.4). Pairwise population comparisons were used to guide the grouping of localities and therefore Cape Vidal was grouped separately since this locality was involved in the two significant comparisons. The highest percentage of variation (100.19%) was within localities, with negligible percentages of variation among the six geographic groups and among localities within groups (Table 2.5). Observed genetic differentiation within localities (p = 0.632), among localities within groups (p = 0.912) and among geographic groups (p = 0.074) did not differ significantly from the null distribution of no genetic structure. Fixation indices were low also supporting the null hypothesis of no genetic structure, with F_{CT} being the only positive fixation index (F_{CT} = 0.013). The non-significance of F_{CT} suggests that levels of genetic variation within the cytochrome *b* region may be too low to distinguish differences between geographic groups.

Table 2.5: AMOVA results for 269 cytochrome *b* sequences of *Epinephelus andersoni*. Genetic variance among geographic groups (6); among localities within groups (12 sampling localities); and among individuals within localities (α = 0.05).

Source of variation	df	Sum of squares	% Variation	Fixation index	*p*
groups	5	1.074	1.260	F_{CT} 0.013	0.074
localities within groups	6	0.789	-1.460	F_{SC} -0.015	0.912
within populations	257	47.304	100.200	F_{ST} -0.002	0.632
Total	*268*	*49.167*	*100.000*		

Spatial analysis of molecular variance

The highest possible number of groups (*K*) for this study was 11, therefore 10 analyses were run with *K* ranging from two to 11 and the grouping resulting in the highest F_{CT} was then identified (Dupanloup *et al.* 2002). The SAMOVA analyses revealed significant variation among groups (α = 0.05) for all *K* values except for *K* = 2 (Cape Vidal consisted of one group with the remaining localities being grouped) (Table 2.6). Cape Vidal was consistently grouped separately from all other localities for all *K* values. The among group fixation indices (F_{CT}) were, however, relatively low (range: 0.025 - 0.045) for all values of *K*, with the lowest F_{CT} value for *K* = 2 and the highest F_{CT} value for *K* = 10. Similarly, the percent variation explained by the groups was consistently low for all *K* groupings (range: 2.50 - 4.49), and was lowest for *K* = 2 and highest for *K* = 10. For the *K* = 10 grouping (which had the highest among groups variability), eight localities were represented as individual groups, while the Msikaba and Coffee Bay localities were grouped, as were the Port St Johns and Mbashe localities. Port St Johns and Mbashe were consistently grouped together except for *K* = 11. Similarly Maputo and Port Alfred were consistently grouped together except for *K* = 10 and *K* = 11.

Table 2.6: Summary of SAMOVA results showing the level of cytochrome *b* genetic variation among geographic groups with varying numbers of groups (*K*). The percent variation explained by the groups, F_{CT} and corresponding *p*-values are indicated for all *K*. Significant *p*-values (α = 0.05) are indicated in bold.

Locality	*K*									
	2	3	4	5	6	7	8	9	10	11
Quissico	1	1	1	1	1	1	1	1	1	1
Xai Xai	1	2	2	2	2	2	2	2	2	2
Maputo	1	2	2	3	3	3	3	3	3	3
Inhaca	1	2	2	3	3	3	4	4	4	4
Cape Vidal	2	3	3	4	4	4	5	5	5	5
Richards Bay	1	2	2	3	5	5	6	6	6	6
Port Edward	1	1	1	1	1	6	7	7	7	7
Msikaba	1	2	2	1	1	7	8	8	8	8
Port St Johns	1	1	4	5	6	6	7	9	9	9
Coffee Bay	1	2	2	3	5	7	8	8	8	10
Mbashe	1	1	4	5	6	6	7	9	9	11
Port Alfred	1	2	2	3	3	3	3	3	10	10
% Variation	2.500	2.550	2.800	2.900	3.050	3.320	3.660	4.100	4.490	4.050
F_{CT}	0.025	0.026	0.028	0.029	0.031	0.033	0.037	0.041	0.045	0.041
p	0.092	**0.000**	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.003**	**0.020**

Spatial genetic patterns

Isolation by distance

The isolation by distance analysis (Mantel Test) showed no association between genetic differentiation (pairwise F_{ST} values) and geographic distance between localities since the linear regression had no significant correlation (r = -0.015, *p* = 0.691) (Figure 2.4). There was still no correlation (r = 0.003, *p* = 0.420) even when genetic and geographic distances were log-transformed.

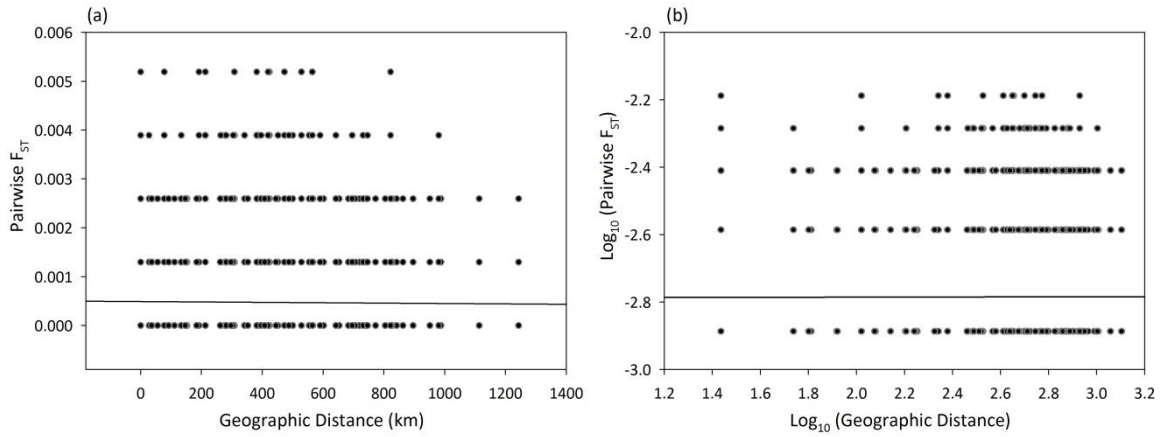


Figure 2.4: Scatterplots of pairwise F_{ST} comparisons versus geographic distance (r = -0.015, p = 0.691) (a) and log-transformed pairwise F_{ST} versus geographic distances (r = 0.003, p = 0.420) (b) between all pairs of samples for the mitochondrial cytochrome *b* region.

Spatial autocorrelation

Fourteen equal-sized 95 km distance classes were chosen for the final analysis since they gave rise to the highest number of significant correlations between genetic and geographic distance. However, the average genetic distance (*Ay*) calculated between all pairs of individuals using these distance classes still only showed a significant negative correlation (p = 0.036) within the second distance class (between 95 and 190 km, Figure 2.5). This would have included the following locality comparisons: Xai Xai versus Inhaca, Maputo and Quissico; Coffee Bay versus Port Edward and Msikaba; Mbashe versus Port Edward and Msikaba. Otherwise, there was no correlation between geographic and genetic distance at any other distance suggesting that geographic proximity does not imply genetic similarity for *E. andersoni*.

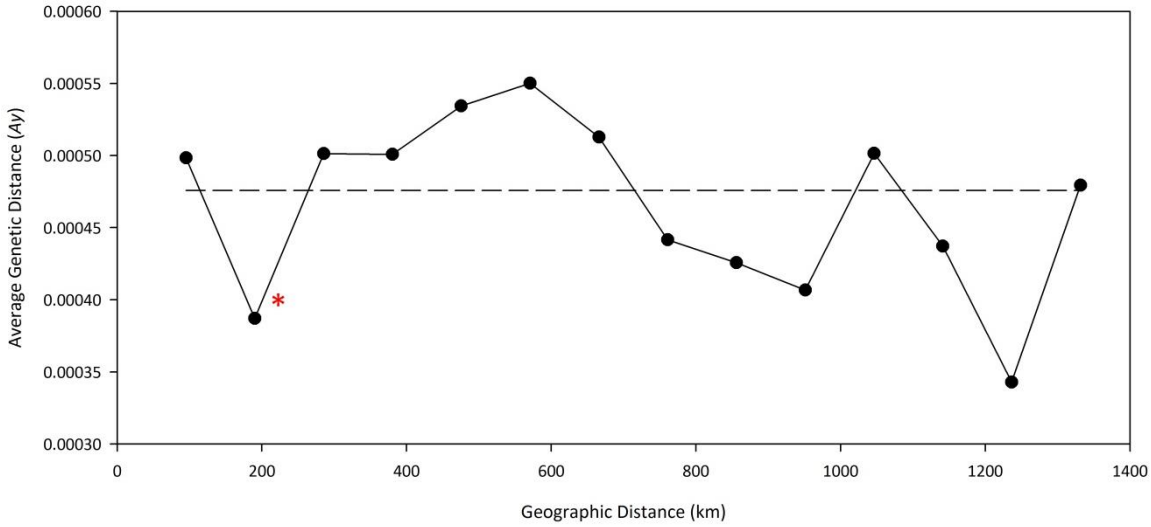


Figure 2.5: Spatial autocorrelation plot for the mitochondrial cytochrome *b* data using 14 distance classes of 95 km throughout the sampling range of *E. andersoni*. The dotted line indicates overall mean genetic distance and the solid line indicates genetic distance (*Ay*) within each distance class. Significant correlation is indicated by a star.

Location of barriers using Monmonier's algorithm

Monmonier's (1973) algorithm identified areas of low gene flow around Msikaba and Richards Bay. Nine iterations were conducted with all barriers separating out a single locality at a time as follows: I Msikaba; II Richards Bay; III Cape Vidal; IV Coffee Bay; V Quissico; VI Port St Johns; VII Port Alfred; VIII Port Edward; and IX Inhaca. Of these barriers, the ones that are likely to represent actual barriers to gene flow, based mainly on known oceanographic features, are only barriers III, V, VII and IX. This test may not have, therefore, been effective at detecting actual barriers to gene flow for the mitochondrial cytochrome *b* data, possibly due to the low levels of genetic variation present in these data.

Seascape shape

Genetic seascape interpolation indicated that neither the weighting parameter (a = 0.5; a = 1; a = 2) nor the grid sizes (50 × 50 and 100 × 100) qualitatively affected the resulting surface plot. Therefore only one surface plot with a = 1 and with X and Y axes as a 100 × 100 grid, is represented here (Figure 2.6). Although the surface plot shows the patterns of any existing genetic differentiation, it is important to note that the level of differentiation represented may still be low. For instance, the genetic differentiation measured as residual genetic distances and indicated as surface heights on the plot ranges from -0.00013 to 0.00010. This surface plot of the cytochrome *b* data has a general pattern of lower genetic differentiation (indicated by troughs) among samples in the more northern limit localities (Quissico, Xai Xai, Maputo and Inhaca). There was also fairly low genetic

differentiation among southern limit samples from Port Alfred, Mbashe and Coffee Bay. In the more central localities from the Transkei (including the Port St Johns, Msikaba and Port Edward localities) to KZN (including the Richards Bay and Cape Vidal localities), there was higher genetic differentiation which is indicated by peaks in the surface plot.

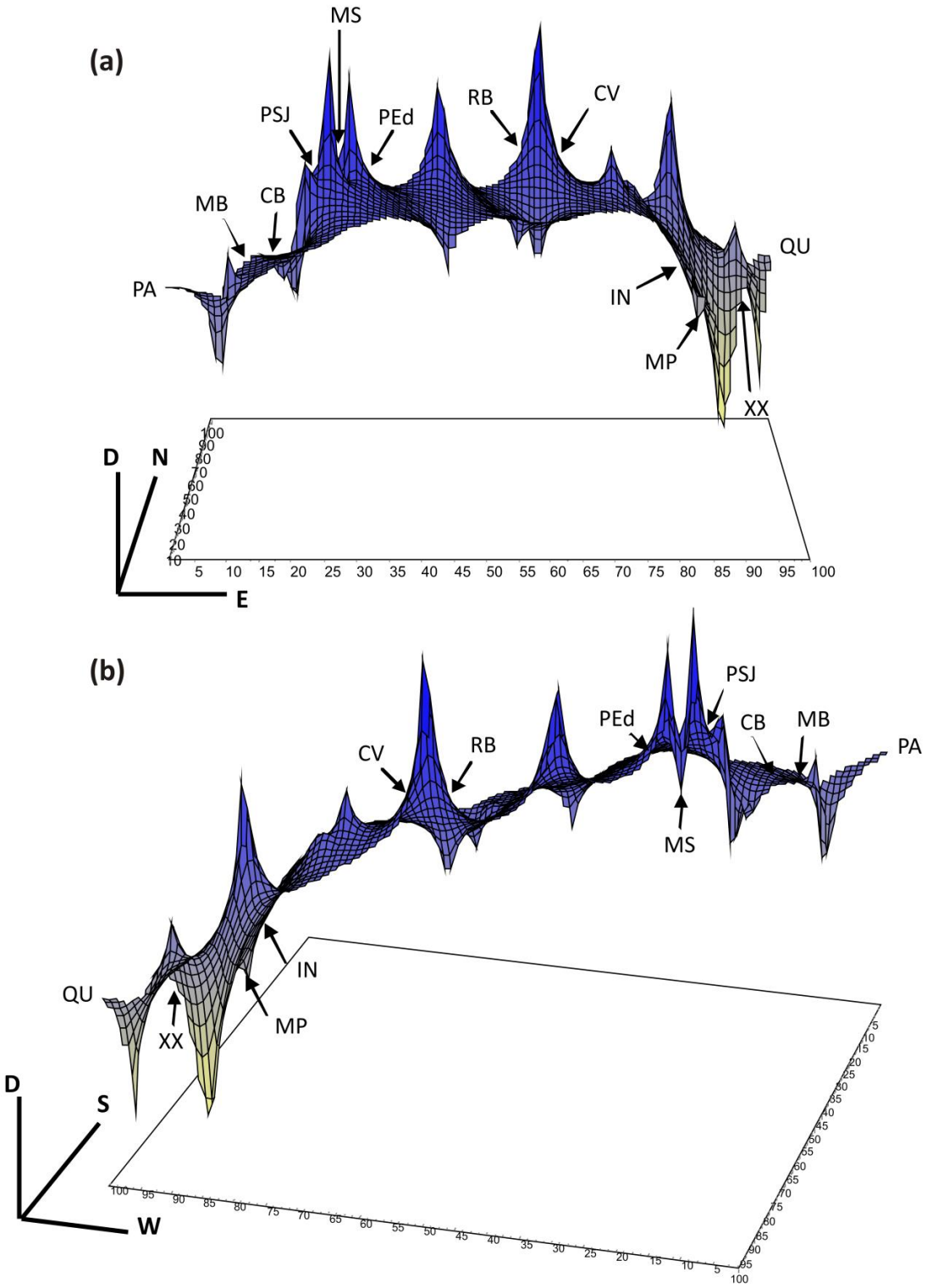


Figure 2.6: Cytochrome *b* genetic seascape interpolation with distance weighting parameter *a* = 1 and 100 × 100 gridded latitude and longitude axes. Surface plot heights (on the D axis) represent residual genetic distances (range: -0.00013 - 0.00010). Compass directions are indicated on the X and Y axes as S or N and E or W. Views from the southern (a) and northern (b) edges of the plot are shown. Localities are indicated on the plot: QU = Quissico; XX = Xai Xai; MP = Maputo; IN = Inhaca; CV = Cape Vidal; RB = Richards Bay; PEd = Port Edward; MS = Msikaba; PSJ = Port St Johns; CB = Coffee Bay; MB = Mbashe; PA = Port Alfred.

2.3.3 Nuclear DNA - RPS7-1

Genetic diversity

The primer pair S7RPEX1F and S7RPEX3R produced long and high quality 706 bp sequences from 240 samples. The remaining 64 samples had sequences that dropped off after approximately 450 bp due to slippage during PCR extension, presumably caused by a series of five or more Thymine (T) nucleotide bases and were excluded from the analyses. The remaining 240 longer sequences were phased using the PHASE algorithm (Stephens *et al.* 2001; Stephens and Donnelly 2003) resulting in 480 sequence alleles. A total of 29 segregating or polymorphic sites were present in the data which resulted in 39 haplotypes (Table 2.7). The numbers of segregating sites varied among the localities with the lowest number of segregating sites present in samples from Coffee Bay (S = 4) and the most present in samples from Port St Johns (S = 11). Inhaca Island also had a high number of segregating sites (S = 10) and had the highest number of haplotypes (H = 14) and rare haplotypes (Hp = 6). Each locality had at least one rare haplotype except for Maputo, Port Edward, Coffee Bay and Port Alfred. The mean number of nucleotide differences (k) varied between the localities (range: 0.567 - 1.784) with the lowest k in Xai Xai and the highest k in Msikaba. The overall k (1.935) and haplotype diversity (Hd = 0.801) were relatively high. However, haplotype diversities varied across the localities (range: 0.284 - 0.883) with the highest associated with Msikaba while the lowest was Coffee Bay. Nucleotide diversities were relatively low among the localities (range: 0.001 - 0.002) and overall (π = 0.001). No particular pattern of genetic diversity among the localities is evident from these indices. However, the haplotype and nucleotide diversities were relatively higher for the RPS7-1 than the cytochrome *b* region.

Table 2.7: RPS7-1 (706 bp) diversity indices and the test statistics and corresponding *p*-values for the tests for selective neutrality. Diversity indices represented include haplotype (*Hd*) and nucleotide (π) diversities as well as the numbers of sequences (*N*), polymorphic sites (*S*), haplotypes (*H*), private haplotypes (*Hp*) and mean number of nucleotide differences between sequences (*k*) for each of the 12 sampling localities. Significant values ($p < 0.05$) are indicated in bold.

								Tajima's		Fu's	
Locality	*N*	*S*	*H*	*Hp*	*k*	*Hd*	π	*D*	*p*	*Fs*	*p*
Quissico	24	8	6	1	1.649	0.819	0.001	-1.488	0.055	-0.512	0.388
Xai Xai	62	9	8	5	0.567	0.322	0.001	-1.897	**0.008**	-5.247	**0.002**
Maputo	36	6	9	0	1.286	0.763	0.001	-0.323	0.382	-3.644	**0.020**
Inhaca	66	10	14	6	0.881	0.492	0.001	-1.965	**0.004**	-11.836	**0.000**
Cape Vidal	64	8	13	4	1.687	0.875	0.001	-0.835	0.222	-5.087	**0.014**
Richards Bay	64	6	10	1	1.745	0.844	0.001	-0.179	0.468	-2.071	0.204
Port Edward	20	5	6	0	1.526	0.842	0.001	0.238	0.651	-1.024	0.246
Msikaba	30	5	9	2	1.784	0.883	0.001	1.507	0.948	-2.637	0.060
Port St Johns	40	11	8	5	1.400	0.473	0.002	-1.392	0.066	-2.047	0.119
Coffee Bay	20	4	4	0	0.579	0.284	0.000	-1.513	**0.036**	-1.237	0.061
Mbashe	36	7	12	3	1.286	0.787	0.001	-0.458	0.348	-7.736	**0.000**
Port Alfred	18	4	6	0	1.529	0.810	0.001	-0.191	0.374	-1.203	0.180
Overall	*480*	*29*	*39*		*1.935*	*0.801*	*0.001*	*-0.708*	*0.297*	*-3.690*	*0.108*

Allele network

The minimum spanning allele network for the nuclear RPS7-1 data (Figure 2.7) had seven possible equally parsimonious genealogies among the 39 allele sequences. There were a high number of shared alleles (12) but also a relatively high number of rare alleles (27). All the geographic localities are represented in the central common allele (allele 1), which contains 183 individual samples, with branches radiating from it giving rise to a star-like topology. Alleles 2 and 3 contain samples from the same 10 geographic localities, with Xai Xai and Port St Johns excluded from these. Alleles 4 and 5 also contain samples from the same six localities (Maputo, Cape Vidal, Richards Bay, Port Edward, Msikaba, Mbashe), while alleles 6 and 7 included samples from eight of the localities (Xai Xai, Port Edward, Port St Johns and Coffee Bay were excluded). These three pairs of haplotypes contain allele phases of the same individual sequences that are all separated by one mutational step. Port St Johns and Xai Xai were not represented in the lineage that included alleles 2 to 7.

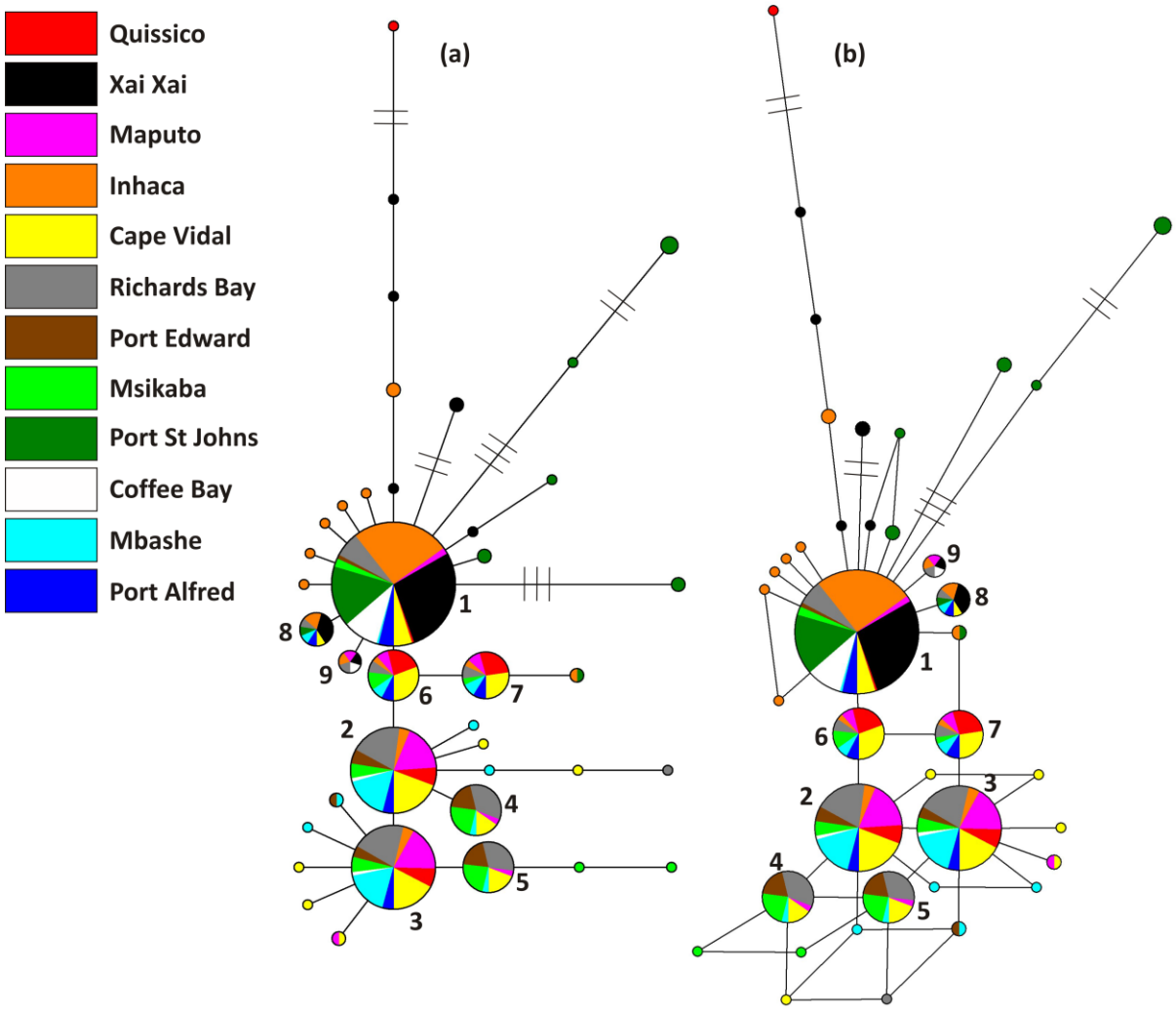


Figure 2.7: Minimum spanning haplotype network of 39 alleles represented from the phased 480 nuclear RPS7-1 sequences from the 12 localities. Nine of the shared alleles are numbered and circle sizes are proportional to the haplotype frequencies: the smallest circle represents one sample and the largest allele (allele 1) represents 183 samples. Colours of the circle segments represent the geographical origin of the samples. Connecting branches between the centres of haplotypes represent one mutational step except when transverse bars are present, each of which represent an additional step. The network representing the most likely genealogy (a), as well as the full allele network with all seven possible genealogies (b), are shown.

Genetic neutrality and demographic history

Tests for selective neutrality

Tajima's *D* was negative for all localities except for Port Edward and Msikaba and showed significant departure from neutrality for only three of the 12 localities (Xai Xai, Inhaca and Coffee Bay). However, overall Tajimas's *D* was negative and not significant ($p = 0.297$) for this data set (Table 2.7). Fu's *Fs* values were negative for all localities with the overall value being non-significant ($p = 0.108$).

However, Fu's *Fs* showed significant departure from neutrality for five (Xai Xai, Maputo, Inhaca, Cape Vidal and Mbashe) of the 12 localities. Although these results are not entirely congruent with results of the neutrality tests on the cytochrome *b* data, they were fairly similar, and these two sets of results agree that the data did not depart significantly from neutrality.

Population size history by mismatch distribution

The mismatch distributions for the RPS7-1 data were similar to the cytochrome *b* in that they were unimodal under both the demographic and spatial expansion models (Figure 2.8), supporting the hypothesis of a recent population expansion. The sum of squared deviations was non-significant for both the demographic (*SSD* = 0.007, *p* = 0.425) and the spatial (*SSD* = 0.058, *p* = 0.308) expansion models, therefore the null hypotheses of spatial and demographic population expansion cannot be rejected. Harpending's *r* values were low and non-significant for demographic (*r* = 0.140, *p* = 0.394) and spatial (*r* = 0.140, *p* = 0.413) expansion, which further supports the expansion hypothesis.

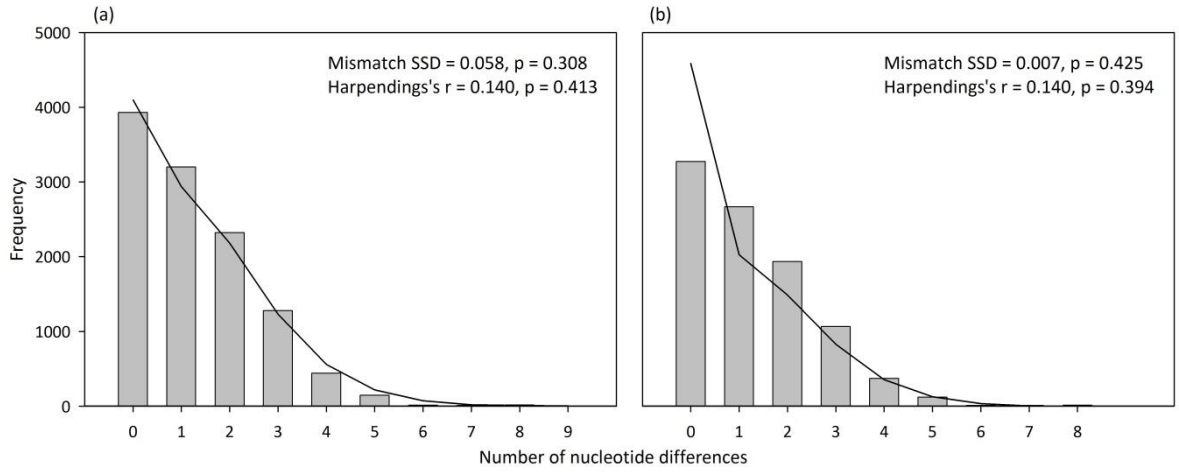


Figure 2.8: Observed (bars) and expected (line) mismatch distributions of the frequency of pairwise nucleotide differences between all pairs of haplotypes in the nuclear RPS7-1 data, under the spatial (a) and demographic expansion models (b).

Population structure

Population differentiation

Exact tests of differentiation and pairwise population comparisons using F_{ST}, revealed that most of the locality comparisons were significant (Table 2.8). The majority of the comparisons based on exact tests were significant (α = 0.05), with 23 being non-significant. Similarly, most of the pairwise F_{ST} comparisons were significant (α = 0.05) with 13 non-significant pairwise comparisons. For both tests, Quissico, Xai Xai, Coffee Bay and Port St Johns were different from most other localities. However, Quissico was not different from Cape Vidal and Port Alfred, while Coffee Bay was also not different from Xai Xai, Inhaca and Port St Johns. Localities in the centre of the *E. andersoni*

distribution (Richards Bay to Msikaba) were similar to each other for both exact tests and pairwise F_{ST}. However, Cape Vidal was not different from Richards Bay, Port Edward and Msikaba (the central localities) in the exact tests but was only similar to Richards Bay in the pairwise F_{ST} comparisons. With the pairwise F_{ST} comparisons, both Port Alfred and Mbashe were different from all localities except for northern localities, Quissico and Maputo, respectively. Although many of the localities in the northern distribution were not different from localities in the south, these localities (north and south) showed differences when compared with adjacent localities. Furthermore, localities in the central part of the *E. andersoni* range show similarities between each other.

Table 2.8: Pairwise F_{ST} values (above diagonal) and exact tests of genetic differentiation *p*-values (below diagonal) between localities for 706 bp RPS7-1 data and the AMOVA grouping of localities that maximised F_{CT}.(Groups 1 to 5) are defined. Significant *p*-values (α = 0.05) are indicated in bold.

	AMOVA groups (5)	1	2	3	4	5	6	7	8	9	10	11	12
1. Quissico	1	*	**0.584**	**0.054**	**0.445**	0.019	**0.077**	**0.202**	**0.131**	**0.455**	**0.438**	**0.082**	0.064
2. Xai Xai	1	**0.000**	*	**0.654**	**0.031**	**0.526**	**0.532**	**0.719**	**0.630**	**0.060**	0.022	**0.682**	**0.375**
3. Maputo	1	0.056	**0.000**	*	**0.532**	0.004	0.022	**0.105**	**0.081**	**0.544**	**0.537**	-0.019	**0.174**
4. Inhaca	1	**0.000**	0.064	**0.000**	*	**0.416**	**0.427**	**0.609**	**0.519**	**0.066**	-0.026	**0.569**	**0.197**
5. Cape Vidal	2	0.368	**0.000**	0.576	**0.000**	*	0.002	**0.083**	**0.041**	**0.453**	**0.396**	0.025	**0.084**
6. Richards Bay	2	**0.000**	**0.000**	0.111	**0.000**	0.285	*	0.025	0.003	**0.460**	**0.403**	**0.041**	**0.121**
7. Port Edward	2	**0.000**	**0.000**	**0.011**	**0.000**	0.134	0.613	*	-0.019	**0.589**	**0.609**	**0.102**	**0.297**
8. Msikaba	2	**0.003**	**0.000**	**0.006**	**0.000**	0.362	0.475	0.876	*	**0.517**	**0.495**	**0.093**	**0.200**
9. Port St Johns	3	**0.000**	**0.010**	**0.000**	0.065	**0.000**	**0.000**	**0.000**	**0.000**	*	0.040	**0.576**	**0.255**
10. Coffee Bay	3	**0.000**	0.404	**0.000**	0.997	**0.000**	**0.000**	**0.000**	**0.000**	0.404	*	**0.574**	**0.205**
11. Mbashe	4	0.092	**0.000**	0.992	**0.000**	0.202	**0.019**	**0.017**	**0.004**	**0.000**	**0.000**	*	**0.223**
12. Port Alfred	5	0.081	**0.000**	0.069	0.061	0.716	0.099	**0.004**	**0.047**	**0.000**	**0.018**	**0.040**	*

Analysis of molecular variance

Five geographic groups were defined for the RPS7-1 data based on pairwise population comparisons and the group structure that gave the highest significant F_{CT} (Table 2.8). Three of these groups contained multiple localities (groups 1, 2 and 3, Table 2.8). Although the pairwise F_{ST} comparisons revealed that all the localities within the northern *E. andersoni* distribution showed differences when compared with each other, there were some similarities with the exact tests of differentiation and therefore the northern localities were grouped since this also made the most geographic sense in terms of proximity.

All hierarchical levels (among geographic groups; among localities within groups; and within localities) were found to be significant (α = 0.05) when five geographic groups were defined (Table 2.9). The within localities genetic variation contributed the most (61.70%) to the total genetic variation, although the variation among geographic groups also contributed substantially (20.39%) to the total variation. Fixation indices were relatively high, the highest being F_{ST} (0.383), followed by F_{SC} (0.225) and then F_{CT} (0.204).

Table 2.9: AMOVA results for 480 nuclear RPS7-1 sequences of for *Epinephelus andersoni*. Genetic variance among geographic groups (5); among localities within groups (12 sampling localities); and among individuals within localities were all significant (α = 0.05).

Source of variation	df	Sum of squares	% Variation	Fixation index	*p*
groups	4	100.881	20.390	F_{CT} 0.204	0.046
localities within groups	7	58.468	17.910	F_{SC} 0.225	0.000
within localities	468	304.145	61.700	F_{ST} 0.383	0.000
Total	*479*	*463.494*	*100.000*		

Spatial analysis of molecular variance

A number of locality groupings were tested with SAMOVA analyses to further determine if spatial groupings of localities other than the ones tested for the AMOVA, revealed significant structuring. The SAMOVA analyses revealed significant variation among groups (α = 0.05) for all *K* values (Table 2.10). The among group fixation indices (F_{CT}) were relatively high (range: 0.36 - 0.463) for all values of *K*, with the lowest F_{CT} value for *K* = 11 and highest for *K* = 2. The percentage variation explained by the groups was also consistently high (range: 36.04 - 46.26), and was lowest for *K* = 11 and highest for *K* = 2. The two groups when *K* = 2 had non-adjacent localities being grouped together: Xai Xai and Inhaca were grouped with Port St Johns and Coffee Bay and the rest of the localities were in the other group. Port Alfred was grouped separately for all *K* values except when *K* = 2. Xai Xai and

Inhaca, as well as Cape Vidal and Richards Bay, were consistently grouped together except for when $K = 11$.

Table 2.10: Summary of SAMOVA results showing the level of RPS7-1 genetic variation among geographic groups with varying numbers of groups (*K*). The percent variation explained by the groups, F_{CT} and corresponding *p*-values are shown for all *K* values. Significant *p*-values ($\alpha = 0.05$) are indicated in bold.

Locality	*K*									
	2	3	4	5	6	7	8	9	10	11
Quissico	1	1	1	1	1	1	1	1	1	1
Xai Xai	2	2	2	2	2	2	2	2	2	2
Maputo	1	1	3	1	1	3	3	3	3	3
Inhaca	2	2	2	2	2	2	2	2	2	4
Cape Vidal	1	1	3	1	1	3	4	3	4	5
Richards Bay	1	1	3	1	1	3	4	3	4	6
Port Edward	1	1	3	1	3	4	5	4	5	7
Msikaba	1	1	3	1	1	5	6	5	6	7
Port St Johns	2	2	2	3	4	2	2	6	7	8
Coffee Bay	2	2	2	4	5	2	2	7	8	9
Mbashe	1	1	3	1	1	6	7	8	9	10
Port Alfred	1	3	4	5	6	7	8	9	10	11
% Variation	46.26	45.29	43.60	41.78	40.62	39.33	38.14	37.34	36.44	36.04
F_{CT}	0.463	0.453	0.436	0.418	0.406	0.393	0.381	0.373	0.364	0.360
p	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.001**	**0.038**

Spatial genetic patterns

Isolation by distance

The isolation by distance analysis showed some correlation between genetic differentiation and geographic distance in the RPS7-1 data. Linear regression showed significant correlation between genetic and geographic distances ($r = 0.032$, $p = 0.029$) and when log-transformed data were used ($r = 0.092$, $p = 0.001$) (Figure 2.9). These results suggest that some of the population structure revealed in the RPS7-1 data by the pairwise population comparisons, the AMOVA and the SAMOVA can be attributed to the geographic distance between localities.

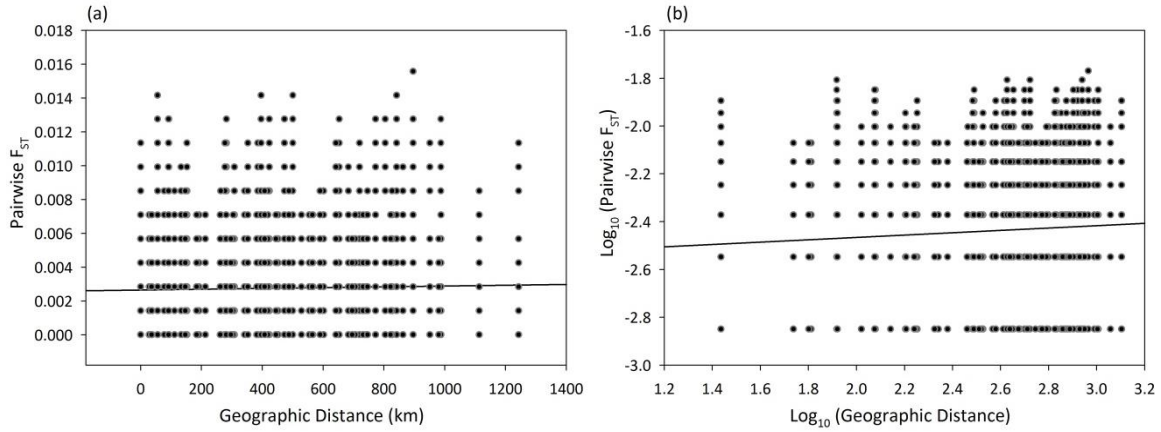


Figure 2.9: Scatterplots of pairwise F_{ST} comparisons versus geographic distance (r = 0.032, p = 0.029) (a) and log-transformed pairwise F_{ST} versus geographic distances (r = 0.092, p = 0.001) (b) between all pairs of samples for the nuclear RPS7-1 region.

Spatial autocorrelation

Fourteen equal-sized distance classes (95 km) were selected for the final analysis since they gave rise to the highest number of significant correlations between genetic and geographic distance. The average genetic distances (*Ay*) calculated between all pairs of RPS7-1 data samples in each distance class had significant positive and negative correlations (Figure 2.10) at different geographic distances (Table 2.11). Significant negative correlation was found within the first distance class for locality comparisons around 95 km apart (p = 0.000) as well as for localities between 1046 and 1141 km apart (distance class 12, p = 0.000) and between 1141 and 1236 km apart (distance class 13, p = 0.001). The locality comparisons within 1046 to 1141 km and 1141 to 1236 km classes included Mbashe versus Quissico, Port Alfred versus Xai Xai, and Port Alfred versus Maputo. Locality comparisons within the 95 km distance class included: Inhaca versus Maputo; Richards Bay versus Cape Vidal; Msikaba versus Port Edward; Port St Johns versus Port Edward, Msikaba and Coffee Bay; and Mbashe versus Coffee Bay. It is evident that negative correlation occurs when proximate localities are compared and when distant localities are compared. This indicates that adjacent localities were not genetically similar, while some distant localities were genetically similar.

Significant positive correlation between genetic and geographic distance occurred among localities that were 380 to 476 km apart (distance class 5, p = 0.000); 476 to 571 km apart (distance class 6, p = 0.000); 571 to 666 km apart (distance class 7, p = 0.046); and 951 to 1046 km apart (distance class 11, p = 0.001). This indicates that localities that are moderate distances apart show correlation between genetic and geographic distance. There was a significant positive correlation within the 11th distance class that included only three locality comparisons (Quissico versus Port St Johns and Coffee Bay, and Inhaca versus Port Alfred) that were identified to be genetically different in the pairwise F_{ST}

tests. The 12th and 13th distance classes showed significant negative correlation. The positive and negative spatial autocorrelation shows no consistent pattern of correlation between genetic and geographic distance since other localities that are distant from each other do not show positive autocorrelation (e.g. Quissico versus Port Alfred) (Table 2.11). It is likely that significant spatial autocorrelation was caused by specific locality comparisons that gave the appearance of a correlation between genetic and geographic correlation within certain distance classes.

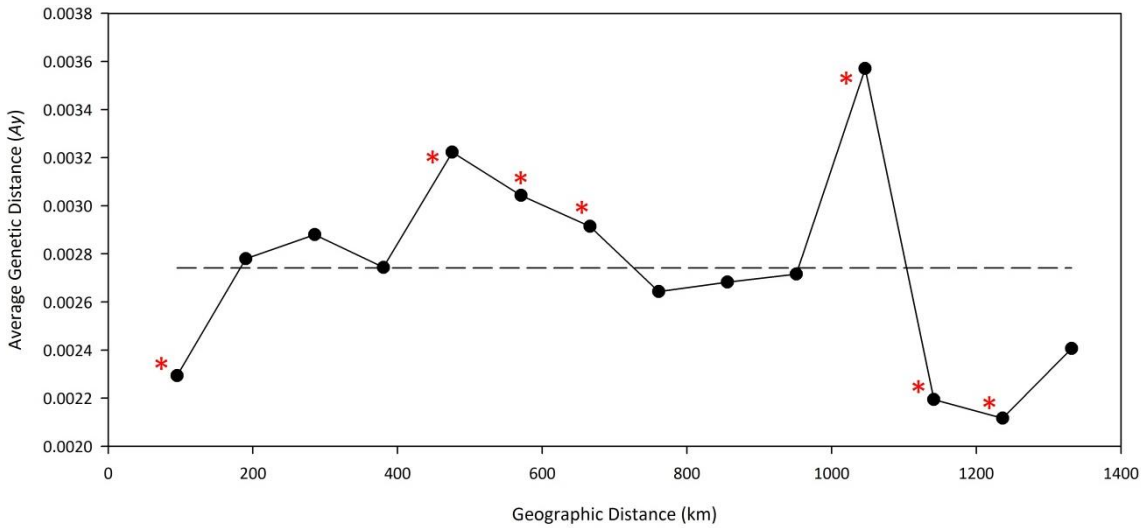


Figure 2.10: Spatial autocorrelation analysis for the RPS7-1 data using 14 distance classes of 95 km throughout the sampling range of *Epinephelus andersoni*. The dotted line indicates the overall mean genetic distance and the solid line indicates genetic distance (*Ay*) within each distance class. Significant correlation is indicated by a star.

Table 2.11: Matrix of pairwise geographic distances (km) between localities with significantly negative spatial autocorrelation between RPS7-1 genetic and geographic distances (α = 0.05) indicated by an asterisk (*). Significantly positive autocorrelations (α = 0.05) are indicated by two asterisks (**).

Locality	1	2	3	4	5	6	7	8	9	10	11	12
1. Quissico	0											
2. Xai Xai	151.7	0										
3. Maputo	284.4	132.7	0									
4. Inhaca	311.7	160.0	27.3*	0								
5. Cape Vidal	503.6**	351.9	219.2	191.9	0							
6. Richards Bay	581.2**	429.5**	296.8	269.5	77.6*	0						
7. Port Edward	885.4	733.7	601.0**	573.7**	381.8**	304.2	0					
8. Msikaba	922.8	771.1	638.4**	611.1**	419.2**	341.6	37.4*	0				
9. Port St Johns	978.2**	826.5	693.8	666.5	474.6**	397.1**	92.9*	55.5*	0			
10. Coffee Bay	1034.2**	882.5	749.8	722.5	530.6**	453.0**	148.8	111.4	55.9*	0		
11. Mbashe	1070.1*	918.4	785.7	758.4	566.5**	489.0**	184.8	147.4	91.9*	36.0*	0	
12. Port Alfred	1331.5	1179.8*	1047.1*	1019.8**	827.9	750.4	446.2**	408.8**	353.3	297.4	261.4	0

Location of barriers using Monmonier's algorithm

Monmonier's (1973) algorithm identified areas of low gene flow within the *E. andersoni* distribution. As with the cytochrome *b* results, nine iterations were conducted (Table 2.12). The first barrier that was detected separated Port St Johns, Coffee Bay, Mbashe and Port Alfred from the rest of the samples, suggesting low levels of gene flow between these southern and more northern localities. The second barrier separated Inhaca Island and Xai Xai from the rest of the localities, even though these localities are not adjacent. The rest of the barrier results were similar to the cytochrome *b* results and separated out a single locality at a time, identifying low levels of gene flow throughout the distribution. The higher levels of genetic variation in the RPS7-1 data when compared to the cytochrome *b* data could mean that the barriers detected by Monmonier's algorithm for the RPS7-1 data are more accurate than those detected within the cytochrome *b* data.

Table 2.12: The groups of localities resulting from each of the barriers generated from the nine iterations of Monmonier's algorithm using RPS7-1 data are shown.

Locality	Iteration Number								
	1	2	3	4	5	6	7	8	9
Quissico	1	1	1	1	1	1	1	1	1
Xai Xai	1	3	3	3	3	3	3	3	3
Maputo	1	1	1	1	6	6	6	6	6
Inhaca	1	3	3	3	3	3	3	3	3
Cape Vidal	1	1	1	1	1	1	1	9	9
Richards Bay	1	1	4	4	4	4	4	4	4
Port Edward	1	1	1	1	1	1	8	8	8
Msikaba	1	1	1	1	1	7	7	7	7
Port St Johns	2	2	2	5	5	5	5	5	5
Coffee Bay	2	2	2	2	2	2	2	2	10
Mbashe	2	2	2	2	2	2	2	2	2
Port Alfred	2	2	2	2	2	2	2	2	2

Seascape shape

Similar to the cytochrome *b* interpolation, neither the weighting parameter nor the grid sizes qualitatively affected the resulting surface plot, therefore only the surface plot with $a = 1$ and with X and Y axes as a 100 × 100 grid, is represented for the RPS7-1 data (Figure 2.11). The residual genetic distances that are represented as the surface plot heights (range: -0.00085 - 0.00157), are larger than the cytochrome *b* surface plot heights (range: -0.00013 - 0.00010), indicating a higher level of genetic differentiation in the RPS7-1 data. The surface plot shows generally lower genetic differentiation among the northern and the southern localities by the presence of a distinct trough among the Mozambique localities and troughs between Port Alfred and Mbashe, and Coffee Bay and

Port St Johns. However, there is a small peak between Mbashe and Coffee Bay and a large peak between Port St Johns and Msikaba within the Transkei localities. There is a peak between the Transkei localities and the KZN localities, indicating that there is genetic differentiation between these localities. There is also a small peak between Inhaca and Maputo within the Mozambique localities. The patterns seen are similar to those found in the cytochrome *b* genetic seascape interpolation.

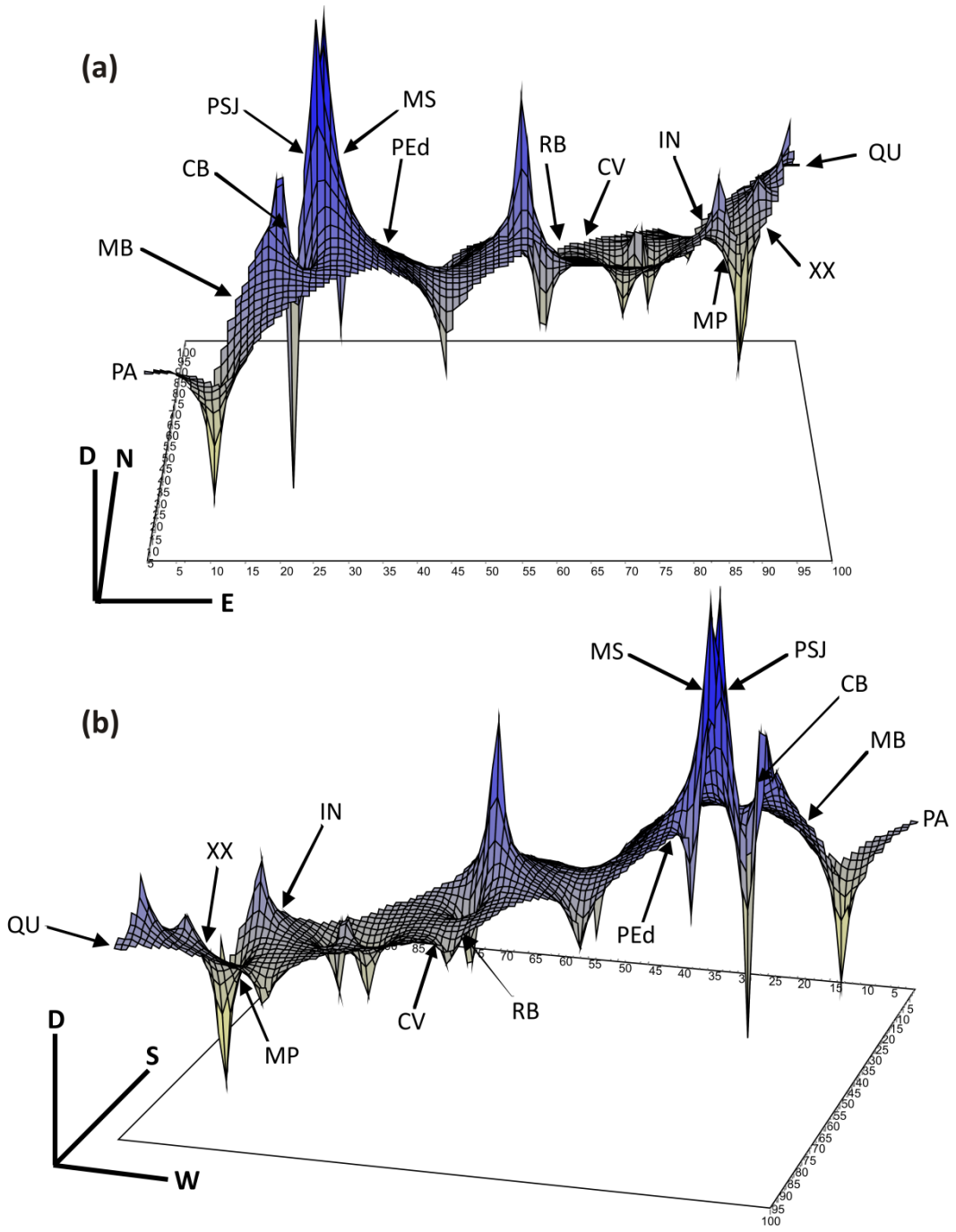


Figure 2.11: Nuclear RPS7-1 genetic seascape interpolation with distance weighting parameter *a* = 1 and 100 × 100 gridded latitude and longitude axes. Surface plot heights (on the D axis) represent residual genetic distances (range: -0.00085 - 0.00157). Compass directions are indicated on the X and Y axes as S or N and E or W. Views from the southern (a) and northern (b) edges of the plot are shown. Localities are indicated on the plot: QU = Quissico; XX = Xai Xai; MP = Maputo; IN = Inhaca; CV = Cape Vidal; RB = Richards Bay; PEd = Port Edward; MS = Msikaba; PSJ = Port St Johns; CB = Coffee Bay; MB = Mbashe; PA = Port Alfred.

2.4 Discussion

The *E. andersoni* population on the south-eastern African coast is characterised by highly complex and convoluted patterns of genetic diversity. Genetic diversity does not show strict structuring in terms of the known biology of the species or the oceanographic characteristics of the study area. This phenomenon has been encountered in genetic studies of marine species in the past and has been termed "chaotic genetic patchiness" (Johnson and Black 1982, 1984; Larson and Julian 1999; Selkoe *et al.* 2006, 2010; Hogan *et al.* 2010). Selkoe *et al*. (2010) found that if these genetic patterns are viewed in relation to a number of environmental variables, significant correlations can be revealed which explain the genetic structure observed in a number of species with varying life histories within the same distribution. This could become an important part of population genetic studies in the future, enabling genetic findings to be more comprehensively contextualised within the environment surrounding populations.

The complex patterns of genetic diversity present in the *E. andersoni* stock may also be a reflection of the variable movement habits of the species. Although *E. andersoni* are said to be predominantly non-migratory, pioneering behaviour has been reported in which the species undergoes ranging movements, supposedly in search of vacant habitat to occupy (Maggs 2011). No known studies have been done on the genetics of *Epinephelus* species on the south-eastern African coast, and therefore the patterns of genetic diversity found in *E. andersoni* were compared with findings for other marine species within a similar distributional range that have been studied. The levels of genetic variation and diversity for *E. andersoni* were, however, compared with those of other *Epinephelus* species studied in other regions (Table 2.13).

Genetic diversity and population demography

The *E. andersoni* mitochondrial cytochrome *b* region generally had low levels of diversity when compared to that of the control region of other *Epinephelus* species and other marine species in general. The mitochondrial control region is commonly used as a genetic marker since it is non-coding and therefore not limited by functional constraints, generally resulting in a much higher mutation rate than other gene regions (Brown *et al.* 1979; Matthee *et al.* 2007). This could explain the relatively high haplotype and nucleotide diversities in other studies on Epinephelus species in which the control region was used (Table 2.13) (Matthee *et al.* 2007).

Table 2.13: Haplotype diversity (*H*) and nucleotide diversity (π) indices for population studies on selected grouper species, based on mitochondrial DNA sequencing. The gene region studied is shown, as well as whether genetic structure or isolation by distance (IBD) was found and the proposed reason for genetic structure. All species presented have long-lived planktonic larvae and are either generally reef-dependent or completely reliant upon rocky reefs, similar to *Epinephelus andersoni*.

Common name	**Species name**	**Gene region**	***H***	**π**	**Genetic Structure**	**Reason for structure**	**IBD**	**Geographic range of study**	**Reference**
catface rockcod	*Epinephelus andersoni*	cytochrome *b*	0.174 - 0.507	0.00023 - 0.00074	No		No	Quissico to Port Elizabeth, endemic	Current study
red spotted grouper	*E. akaara*	control region	0.720 - 1.000	0.006 - 0.017	Yes	Post-glacial sea level rise	-	north west Pacific	Chen *et al*. 2008
goliath grouper	*E. itajara*	control region	0.8	0.004	No, only F_{ST} significant	Territoriality, habitat preferences, fidelity to reproductive shoaling sites	-	northern Brazil	Silva-oliveira *et al*. 2008
rock hind	*E. adscensionis*	cytochrome *b*	0.586	0.008	Yes	Not defined. Differentiated populations are only 250km apart	-	Caribbean, Gulf of Mexico, west Atlantic	Carlin *et al*. 2003

Low nucleotide diversity, or low levels of sequence divergence between haplotypes is a recurring pattern in marine fishes (Table 2.13) (Grant and Bowen 1998; Muss *et al.* 2001; Rocha *et al.* 2002). The extremely low nucleotide diversity and low to moderate haplotype diversity of the *E. andersoni* cytochrome *b* gene when compared with other *Epinephelus* species (Table 2.13) could indicate that a recent population bottleneck existed, or a founder event took place, causing the coalescence of mtDNA lineages and shallow evolutionary histories (Grant and Bowen 1998). Population bottlenecks in marine fishes have been attributed to a number of factors, the first of these being historical shifts in climate and oceanographic conditions, especially those associated with the sea level rise that occurred after the last glacial maximum (approximately 20,000 years ago) around the Pliocene and early Pleistocene (Grant and Bowen 1998; Gasse 2000; Hutchings *et al.* 2002; Jacobs *et al.* 2004; Dawson and Hamner 2005; Chen *et al.* 2008). Secondly, variance in reproductive success can also affect the diversity of many marine fishes (Grant and Bowen 1998). Like *E. andersoni*, many marine species are highly fecund but rely on a narrow range of favourable oceanographic conditions for the successful dispersal and recruitment of their planktonic larvae (Cowen *et al.* 2007). This could cause an annual cohort to be parented by only a fraction of the spawning adult population (Larson and Julian 1999; Selkoe *et al.* 2006, 2010; Hogan *et al.* 2010), also known as Hedgecock's (1994) "sweepstakes-chance matching hypothesis." The consequence of this would be a loss of genetic variability among recruits and therefore a population bottleneck in years when there is poor recruitment (Hedgecock 1994; Larson and Julian 1999).

The haplotype and nucleotide diversities for the first intron of the RPS7-1 gene of *E. andersoni* were much higher than those for the cytochrome *b* region. However, these low mtDNA estimates are comparable with the diversities found in studies done on other *Epinephelus* species using various mtDNA markers (Table 2.13). This is contrary to what would be expected since mtDNA is usually thought to have a higher mutation rate than nuclear DNA and should therefore have more diversity (Brown *et al.* 1979; Hare and Palumbi 2003; Caccone *et al.* 2004; Matthee *et al.* 2007; Zink and Barrowclough 2008). These relatively high nuclear diversity indices were similar to those of organisms such as corals and plants that have been shown to have slowly evolving mtDNA when compared to nuclear markers (Hellberg 2006). They were also similar to the bluntnose klipfish *Clinus cottoides*, for which the second intron of the RPS7-1 gene was found to have haplotype and nucleotide diversities which were higher than the diversities that were found in the ND2 and control regions from the mtDNA (von der Heyden *et al*. 2008). It is possible that the first intron of the RPS7-1 gene of *E. andersoni* mutates faster than the cytochrome *b* region, possibly due to the fact that the RPS7-1 region is non-coding with no limits to mutation whereas cytochrome *b* is a coding region which may limit its mutation rate (Matthee *et al.* 2007; Zink and Barrowclough 2008). This may mean

that the RPS7-1 region shows the more recent signal of diversification following the bottleneck event that was evident in the cytochrome *b* data. Another possible reason for the lower cytochrome *b* diversity could be the fact that males have been found to be more abundant than females within the *E. andersoni* population (Fennessy and Sadovy 2002). This could cause a lower effective population size based on mtDNA when compared to the RPS7-1 since the cytochrome *b* region is maternally inherited, resulting in decreased cytochrome *b* diversity (Allendorf 2013).

The RPS7-1 region has low nucleotide diversity similar to the cytochrome *b* region, but its haplotype diversity ranges from low to high among localities. This is also a common pattern among marine fishes and indicates that a population bottleneck has occurred, followed by rapid population expansion (Grant and Bowen 1998). If the RPS7-1 region has a faster mutation rate than the cytochrome *b* region as argued above, the low nucleotide diversity of the RPS7-1 data could be an indication of the same bottleneck suggested by the cytochrome *b* data and the variable levels of haplotype diversity could be an indication of subsequent mutations and diversification since the bottleneck, only evident in the faster mutating RPS7-1 data.

The unimodal mismatch distributions, non-significant raggedness index and the neutrality tests all indicated a recent population expansion typically following a population bottleneck (Tajima 1989a; Rogers and Harpending 1992; Rand 1996; Fu 1997; Ruegg and Smith 2002; Excoffier and Lischer 2011). Although the neutrality tests were not consistently significant at the same localities, the Tajima's tests and Fu's *Fs* tests consistently showed evidence of a population bottleneck or a selective sweep. The typical star-like topology of the haplotype networks also indicates that a population bottleneck has taken place in this species. It is likely that the common central haplotype represents an ancestral haplotype with more recently evolved haplotypes branching from it after population expansion (Hudson 1990; Slatkin and Hudson 1991; Grant and Bowen 1998; Buehler and Baker 2005; von der Heyden *et al.* 2008; Opgenoorth *et al.* 2010). The branching haplotypes are only up to five mutational steps from the central haplotypes in the nuclear data, signifying small amounts of recent diversification (Grant and Bowen 1998). There is no significant grouping of haplotypes which represent similar localities and many of the haplotypes are common, indicating a lack of regional genetic structure.

Cytochrome b genetic structure and barriers to gene flow

A population that has recently experienced a bottleneck would generally be expected to be fairly homogeneous due to a loss of rare alleles and of much of the genetic diversity that would have structured the population before the bottleneck (Grant and Bowen 1998). This could be the case

with the cytochrome *b* data for *E. andersoni* since there is no (or very little) differentiation between localities. This was shown by the very few significant comparisons in the pairwise population comparisons and the non-significant AMOVA analysis. There was no isolation by distance and no positive spatial autocorrelation, indicating that geographic distance had no influence on genetic distance. Despite these apparently low levels of differentiation, a few of the exact tests and F_{ST} comparisons were significant; all of the SAMOVA groupings had significant among group genetic variation except when $K = 2$; and the seascape shape interpolation identified areas that were more genetically differentiated than others (although these genetic distances were still low). The genetic differentiation detected by these analyses could be an indication of increasing genetic diversity caused by population expansion after the bottleneck event.

Pairwise population comparisons that were significant, pinpointed Cape Vidal as different from Xai Xai in both the exact tests and F_{ST} comparisons, and from Msikaba in the exact tests. This could be explained by Cape Vidal being the northern-most "population" that was sampled that contributes to genetic diversity within the Agulhas Current portion of the *E. andersoni* distribution. Most of the coastal localities south of the point where the Agulhas Current begins are relatively isolated from the Mozambique Channel waters. According to Lutjeharms (2006), the Mozambique Channel waters contribute little to the Agulhas Current proper and when they do it is only intermittently. This, combined with the transient Delagoa Bight cyclonic eddy off Maputo, and the anticyclonic Mozambique Channel eddies and associated upwelling of cold water (Lutjeharms 2006, *sensu* Lamont *et al*. 2012), may limit the dispersal and recruitment of fish larvae from the Mozambique spawning population into more southern rocky reef habitats. There is also localised upwelling near Richards Bay and just north of Richards Bay where the continental shelf begins to widen at the northern margin of the Natal Bight (Lutjeharms 2006). The combination of the possibly limited connectivity between the Mozambique and the Agulhas Currents; the evidence suggesting that *E. andersoni* are non-migratory (Heemstra and Randall 1993; Maggs 2011); and the Richards Bay upwelling cell, could effectively isolate Cape Vidal fish from all other populations. The seascape shape interpolation also shows genetic differentiation in the form of peaks between Cape Vidal and its adjacent localities. The SAMOVA results also provide some evidence for the isolation of the Cape Vidal fish from the rest of the population since Cape Vidal is grouped separately from other localities for all *K* and with significant among group variation for *K* values from 3 to 11.

Contradictory to the hypothesis that the Mozambique Channel populations are isolated from the more southern distribution, the northern localities are frequently grouped with southern localities in the SAMOVA analyses. This suggests that irregular dispersal events from the Mozambique Channel

populations can sometimes result in successful recruitment of larvae into habitat further south during an intermittent connection between the Mozambique Channel eddies and the Agulhas Current. These dispersal events may bypass the Cape Vidal and Richards Bay localities due to a combination of a long larval phase of *E. andersoni* (30 days or more) (Connell 2012), and the widening of the continental shelf just south of Cape Vidal (the Natal Bight). The Natal Bight pushes the mainstream of the Agulhas Current offshore and gives rise to variable flow regimes on the continental shelf in the area, referred to as the Natal Pulse, which meanders southwards along the coast and occurs about five times a year (Lutjeharms 2006; Roberts *et al.* 2010). These varying oceanographic currents could result in variable dispersal patterns of *E. andersoni* larvae from spawning sites and could explain the complex and apparently geographically meaningless groupings produced by the SAMOVA analyses. Although one may assume that there is no significant genetic structure in the cytochrome *b* data from the AMOVA and pairwise population comparisons, other tests reveal very low structuring and levels of differentiation. The seascape shape interpolation visualises the genetic differentiation within the data and it is apparent that although there are areas of low genetic differentiation in the southern and northern localities, the localities in the central distribution are separated by peaks indicating higher levels of genetic differentiation. Although the observed patterns of genetic diversity appear chaotic, there are underlying causes behind the chaos. It is likely that patterns of genetic diversity will become more evident once the population has recovered from the bottleneck event that was detected from the cytochrome *b* data.

RPS7-1 genetic structure and barriers to gene flow

A population that has experienced a recent population expansion typically shows evidence of new mutations becoming fixed in the population and an increase in genetic diversity and differentiation since the beginning of the expansion (Grant and Bowen 1998). Genetic differentiation was evident in the RPS7-1 from the pairwise population comparisons, AMOVA analysis, SAMOVA analyses, isolation by distance, spatial autocorrelation and the seascape shape interpolation analysis. Genetic variation was found to be significant at all hierarchical levels for the AMOVA and among group variation can be explained by oceanographic currents and features. As discussed above, the northern Mozambique Channel localities could be largely isolated from the southern localities in the Agulhas Current waters (Lutjeharms 2006; *sensu* Lamont *et al*. 2012). This would explain the significant F_{CT} when Mozambique Channel localities were grouped, separately from more southern Agulhas Current localities in the AMOVA. The pairwise population comparisons among Mozambique Channel localities indicated that some significant differentiation existed among these localities. This was further supported by peaks between these localities in the RPS7-1 seascape interpolation and could be explained by the Delagoa Bight upwelling and the variable flow regimes of the Mozambique

Channel eddies (Lutjeharms 2006) that may impede gene flow among adjacent localities. However, the SAMOVA locality groupings contradicted the separation of northern and southern localities since Mozambique Channel localities were consistently grouped with more distant Agulhas Current localities in southern Transkei and toward the southern distribution limit. This could be an indication of irregular historical dispersal events from the Mozambique Channel that result in the recruitment of larvae in southern localities, as proposed earlier using the cytochrome *b* results.

The AMOVA groups numbered two (Cape Vidal, Richards Bay, Port Edward and Msikaba) and three (Port St Johns and Coffee Bay) (see Table 2.8), which included the KZN and northern Transkei localities were similarly grouped in some of the SAMOVA results (Table 2.10). This consistent grouping of localities between the two tests can be explained using existing oceanographic features and currents. The grouping of Cape Vidal, Richards Bay, Port Edward and Msikaba could possibly be explained by oceanic currents such as the Natal Pulse and the associated northward flowing counter-current on the continental shelf which may provide some level of connectivity between the KZN localities in the Natal Bight and the northern Transkei localities (Roberts *et al.* 2010). Port St Johns and Coffee Bay are also frequently grouped together in the SAMOVA results, separately from the above-mentioned more northern localities. This could be explained by the slightly wider continental shelf in the Port St Johns area that alters the flow of the Agulhas Current and may limit dispersal between northern localities and Port St Johns and Coffee Bay (Lutjeharms 2006). The similarity between Port St Johns and Coffee Bay and the separation of these two localities from others, is also evident on the seascape interpolation: viz. the trough between Port St Johns and Coffee Bay, and the high peaks between these two localities and their adjacent localities respectively.

Port Alfred and Mbashe are frequently grouped separately in the SAMOVA results. The differentiation among these southern localities could possibly be explained by the presence of the biogeographic break near the Mbashe River that separates the warm-temperate and subtropical zones (Sink *et al.* 2004). This break also marks the distribution limit for other marine organisms such as abalone and east coast rock lobster, and marks a clear break in estuarine fish communities (*sensu* Sink *et al*. 2004). The Port Alfred upwelling cell (Lutjeharms *et al.* 2000; Hutchings *et al.* 2002; Lutjeharms 2006) could also possibly create a barrier to gene flow, allowing genetic differentiation to be maintained. The fact that the majority of the Port Alfred samples (excluding one) originated from juvenile fish caught in the Kowie estuary, could also explain the differentiation between Port Alfred and other localities. On the seascape interpolation, however, there is a trough between Port Alfred and Mbashe that indicates that they are genetically similar. Despite this, there is a peak between Mbashe and Coffee Bay which could indicate reduced gene flow between the two localities

that can be attributed to the Mbashe biogeographic break. Despite the seascape interpolation results, Monmonier's algorithm identified barriers to gene flow that reinforced the groupings of the SAMOVA analyses. Similar to the cytochrome *b* results, the RPS7-1 patterns of diversity indicate that dispersal is affected by complex relationships between gene flow and oceanographic features.

The RPS7-1 Mantel test revealed a positive correlation between genetic distance and geographic distance although this correlation was only significant when genetic and geographic distances were log-transformed. This suggests that gene flow may be affected by geographic distance. Although the spatial autocorrelation analysis supports the result of the Mantel test, it shows both significant negative and positive correlations between genetic and geographic distance at varying geographic distances. This complex pattern could possibly be explained by the specific habitat preferences of *E. andersoni*. Since rocky reef habitat within the preferred depth range of *E. andersoni* is not continuous throughout the distributional range of the species and appropriate habitat patches are not likely to be equidistant from one another, it is likely that there is a simple linear relationship between genetic and geographic distances. Further complications arise when appropriate habitat patches are distributed among areas of differing oceanographic currents and environmental variables, giving rise to complex patterns of genetic diversity that are difficult to explain (Wiens 1976; Pelc *et al.* 2009; Hogan *et al.* 2010). The seascape shape interpolation visualises these complex patterns of genetic differentiation, with the lowest levels of genetic differentiation (deepest troughs) near the range limits of *E. andersoni* and the highest differentiation (highest peaks) among the central localities.

2.5 Conclusions

The results suggest that the *E. andersoni* population is characterised by the phenomenon known as chaotic genetic patchiness, coined by Johnson and Black (1982), therefore rejecting the null hypothesis of no genetic structure. The null hypothesis can be rejected since, although the genetic patterns are complex and seemingly chaotic, there is some underlying genetic structure. These genetic patterns may have been caused by a combination of factors, therefore making them complex to explain. Chaotic genetic patchiness could have been caused by any or all of the following factors: pre-settlement selection acting on larvae; post-settlement selection; variable sources of spawning adult populations (Hogan *et al.* 2010); sweepstakes recruitment (Hedgecock 1994); patchy habitat distributed among different biogeographic regions and among areas with differing levels of exploitation; the pioneer behaviour of *E. andersoni* and associated, apparently random ranging movements shown by tagging data (Maggs 2011), among others. In order to more fully understand

the patterns of genetic diversity in *E. andersoni*, further study needs to be done on the dispersal patterns of its planktonic eggs and larvae, these being the predominant avenue of dispersal since adults are apparently non-migratory (Fennessy and Sadovy 2002). A genetics study that incorporates a number of different spatial and temporal scales may help to reveal patterns of genetic diversity: for example a study that incorporates egg, larval, juvenile and adult samples from the entire distribution of the species could possibly help to unveil the patterns of larval dispersal. Using different and faster-mutating mitochondrial gene regions such as the D loop may also show more recent differentiation. It may also be enlightening to correlate genetic patterns with a number of the above-mentioned environmental variables that occur during the spawning season when larvae are dispersed, such as oceanic current direction or upwelling, habitat type, or additional variables such as sea surface temperature, to enable a more confident explanation of patterns of genetic differentiation within the context of the surrounding environment (Selkoe *et al.* 2010).

CHAPTER 3: Modelling the distribution of *Epinephelus andersoni* with future sea surface temperature changes

3.1 Introduction

Although global climate change is a topic of much current political controversy, it is not a new topic in terms of biological research (Parmesan 2006). The importance of climatic thresholds for restricting the geographic distributions of many species (Grinnell 1917) was recognised before climate change and species distribution models (SDMs) became popular research topics. Current research now predicts that future projected climatic changes will cause shifts in the distributions of a number of species globally (Thomas *et al.* 2006), including those of marine organisms (e.g. Fields *et al*. 1993; Clark 2006; Harley *et al*. 2006; Brander 2007; Hoegh-Guldberg *et al*. 2007; Lasram *et al*. 2010). Climate-induced changes in the marine environment are anticipated to have substantial effects on global fisheries, a number of them negative, especially when combined with the high levels of exploitation experienced by virtually all wild marine stocks (Perry *et al.* 2005; Harley *et al.* 2006; Brander 2007).

In a global review on the numbers of modelling studies that have been conducted to date, Robinson *et al*. (2011) found that relatively few studies have used SDMs to determine the effect of climate change on marine species relative to terrestrial studies. Of the marine SDM studies that have been conducted, the majority focus on the effect of climate change on commercially exploited fishes, reflecting the commercial value attached to marine fish stocks (Robinson *et al.* 2011). These SDM studies have proven effective at mapping the distributions of important fisheries species world-wide and have provided useful information for resource managers (Kupschus 2003; Eastwood *et al.* 2003; Hedger *et al.* 2004; Maxwell *et al.* 2009). Specifically in southern Africa, SDMs have been used to predict the effects of climate change on the distribution of terrestrial species (e.g. Erasmus *et al*. 2002; Pearson *et al*. 2006; Keith *et al*. 2009; Coetzee *et al*. 2009) but no known SDM studies have been published on coastal and marine species (but see Clark *et al*. 2000; Lutjeharms *et al*. 2001; Bakun and Weeks 2004; Clark 2006; James *et al*. 2008 for suggested climate change impacts on southern African marine and coastal environments). Since SDMs model species' distributions, it is useful to define such a distribution as: the complex expression of their ecology and evolutionary history (Brown 1995), which are determined by a multitude of factors, working with different intensities and at different spatial scales (Pearson and Dawson 2003). Factors affecting distributions include: abiotic conditions such as climatic and environmental conditions; biotic factors such as inter and intra-specific interactions; dispersal; and evolutionary capacity (Soberón and Peterson 2005).

Although numerous environmental variables are needed to accurately predict terrestrial species occurrences as suggested by the definition of a species' distribution, in the marine environment it has been shown that distributions can be determined by fewer variables, one of the most influential being water temperature (Sabatés *et al.* 2006; Hiddink and Hofstede 2008; Dulvy *et al.* 2008; Cheung *et al.* 2009; Lasram *et al.* 2010; Sunday *et al.* 2012). This can be supported by the fact that many marine fishes live closer to their thermal tolerance limits than terrestrial species (Harley *et al.* 2006; Sunday *et al.* 2012). Bathymetry (especially bathymetric complexity) has also been shown to be influential in determining marine fish distributions (Pittman *et al.* 2007; Beger and Possingham 2008; Pittman and Brown 2011).

There are two main modelling approaches which differ in their complexity (i.e. in the extent to which they include the variables that affect species' distributions), namely correlative and mechanistic models (Pearson and Dawson 2003). Correlative models are relatively basic and correlate observed presence (and sometimes absence) data with background environmental variables, under the assumption that the best indicator of a species' environmental requirements is its current distribution (Pearson and Dawson 2003). Mechanistic models are physiologically-based models that make fewer assumptions than correlative models (Pearson and Dawson 2003) and can include physiological variables such as constraints to growth, regeneration (Sykes *et al.* 1996) and reproduction (Kearney and Porter 2004). Although mechanistic models are generally more robust under climate change scenarios than correlative models (Pearson and Dawson 2003), data limitations make correlative models a more popular approach (Pearson and Dawson 2003; Kearney and Porter 2004).

Many studies have shown that the type of correlative model used to predict species' distributions can have profound effects on model results, and models' performances have also been found to vary under different circumstances (Thuiller 2003, 2004; Araújo *et al.* 2005; Hijmans and Graham 2006; Pearson *et al.* 2006; Pearson 2007). There are two key areas where models differ: firstly their data input requirements vary (e.g. presence-only versus presence-absence data); and secondly there are variable model extrapolation assumptions when potential species' distributions are being projected into future "unknown" environmental conditions that were not present for the current projection (Thuiller 2003; Thuiller *et al.* 2004; Pearson *et al.* 2006). Generally, presence-absence data are preferred since they provide more information on the current distribution and the prevalence of the species than presence-only data (Phillips *et al.* 2009). However, obtaining detailed occurrence data for marine organisms is not always practical or possible. This is especially true when studying marine species that are mobile and difficult to detect (Kaschner *et al.* 2006) such as fishes (Maxwell *et al.*

2009), necessitating the use of pseudo-absence data since all models require some form of absence data (Phillips *et al*. 2009).

Another aspect that affects the accuracy of model results is the geographical extent to which the environmental data are used to project the future distributions of a species; if the extent is excessively restricted it could result in skewed model results (Thuiller *et al.* 2004). Furthermore, model outputs are likely to represent only one of a number of possible future scenarios (Pearson and Dawson 2003). An increasingly popular approach is to combine a number of individual models into an ensemble model in an attempt to reduce the uncertainty implicit in correlative SDMs (e.g. Lawler *et al*. 2006; Araújo and New 2007; Diniz-filho *et al*. 2009; Thuiller *et al*. 2009a; Lasram *et al*. 2010). Ensemble modelling approaches have been used to determine the potential effects of climate change on various species, providing useful information for conservation and the management of natural resources (Araújo *et al.* 2005; Diniz-filho *et al.* 2009; Marmion *et al.* 2009; Capinha and Anastácio 2011). Although results need to be interpreted with caution due to the uncertainty associated with correlative models (Araújo *et al.* 2005; Heikkinen *et al.* 2006; Pearson *et al.* 2006; Diniz-filho *et al.* 2009; Maxwell *et al.* 2009), the use of models as a first approximation of future species' distribution changes can play a role in helping resource managers prepare for future climate- induced changes (Pearson and Dawson 2003; Lawler *et al.* 2006; Elith and Leathwick 2009).

The catface rockcod *Epinephelus andersoni* is a range-restricted, commercially important species that is endemic to rocky reefs in the inshore zone along the southern African coast (Fennessy and Sadovy 2002). Since *E. andersoni* is a range-restricted endemic (Heemstra and Randall 1993; Heemstra and Heemstra 2004; Craig *et al.* 2011) and it likely to be over-exploited (Fennessy and Mann 2013) it is probable that climatic changes will impact its distribution. It is clear that research is needed if this species is to be managed effectively, especially in light of future climatic changes. The aim of this chapter was to determine possible range shifts or changes in range size of *E. andersoni* with future changes in sea surface temperature (SST). The findings of this study may contribute to future management and conservation of *E. andersoni*. It is also hoped that this study will lead to further research on climate-related impacts on other range-restricted endemics and important fisheries species in southern Africa.

3.2 Materials & Methods

3.2.1 *Epinephelus andersoni* occurrence data

Epinephelus andersoni occurrence records were collected from a number of sources including the National Marine Linefish System (NMLS) database and the ORI/WWF South African Marine Linefish Tagging database. These databases are both compiled by the Oceanographic Research Institute (ORI). The ORI/WWF Tagging Project (van der Elst and Bullen 1993) was initiated by ORI in 1984 and involves the enlisting of recreational fishers that voluntarily tag and then release their catch. All details related to their catch such as fish species, length, capture locality and date are recorded together with subsequent recaptures of tagged fish. The NMLS database includes data collected during hook and line fisheries monitoring such as observer-based shore angling inspections and boat angling inspections; it also includes commercial catch records. Mozambican occurrence data consisted of commercial fishing records and were provided by the Instituto de Investigação Pesqueira (IIP). Occurrence records from recreational skiboat fishermen in Mozambique (samples used in the genetic analysis) were also included. When only locality names or codes were given in the database as was the case with the NMLS and ORI/WWF Tagging data, these data were converted to GPS coordinates using Google Earth®. All point localities were recorded as being at least five kilometres offshore. This was done to ensure that all presence points would lie within grid cells containing background environmental data. Data points were then collated into a single database and visualized in ArcMap version 10 (ESRI 2011). Questionable presence points, such as any located inland, were removed. The presence points were further edited to ensure that each grid cell of the environmental layers contained no more than a single presence point. This was done to avoid spatial autocorrelation during the modelling process (Dormann *et al.* 2007; Valavanis *et al.* 2008).

Furthermore, all presence points south of Jeffreys Bay (Figure 3.1) were removed from the model since many of these occurrence records consisted of fishes smaller than the recorded length at 50% maturity for females (Fennessy and Sadovy 2002). Due to an edge of range effect, possibly related to colder temperatures (Fennessy and Sadovy 2002), it was assumed that *E. andersoni* was not common south of Jeffreys Bay. Fish able to survive south of this point were assumed to reside largely in estuaries where conditions were favourable for survival. Lipcius *et al*. (1997) suggest that areas where juveniles are abundant but adults are scarce can be classified as sink habitats. This is because postlarval and juvenile supply may be decoupled from adult abundance by the dispersal of planktonic larvae in ocean currents, and by the effects of habitat quality on the survival of postlarvae and juveniles. These factors could suggest that individuals south of Jeffreys Bay may be from sink populations that are maintained by larval input from further north but are not self-sustaining

populations (Pulliam 2000), which justifies excluding these occurrences from the model. This resulted in 210 occurrence records being used in the models (see Figure 3.1).

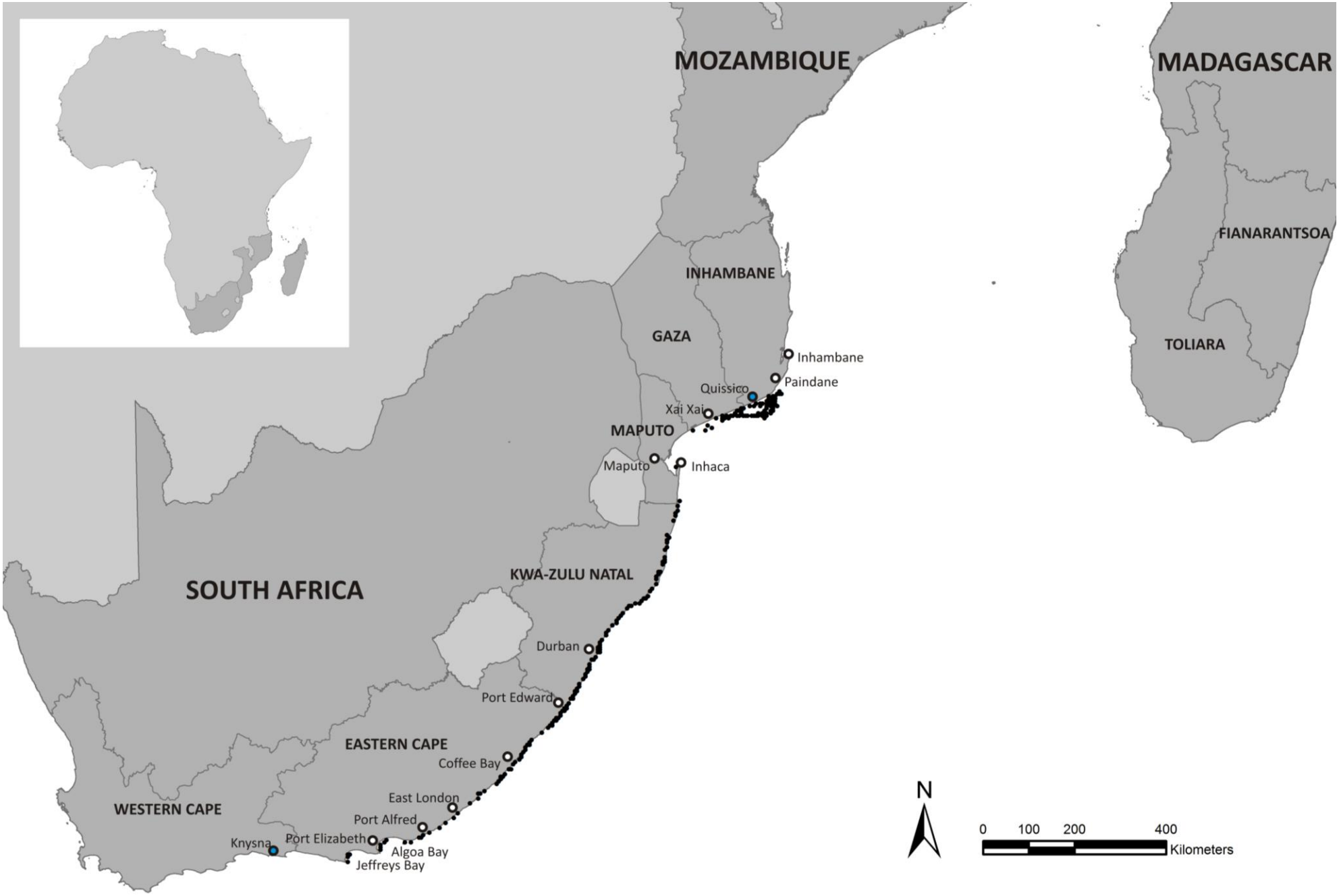


Figure 3.1: Map of the study area showing the 210 *Epinephelus andersoni* presence points used for modelling. Blue dots show previously recorded distributional limits.

3.2.2 Environmental data

Historic

Seasonal sea surface temperature and bathymetry were the environmental variables used to calibrate the SDM. Sea surface temperature (SST) data are freely available from the National Oceanic and Atmospheric Administration's (NOAA) office of Oceanic and Atmospheric Research (OAR) and Earth System Research Laboratory (ESRL) in the Physical Sciences Division (PSD), Boulder, Colorado, USA, on their website (National Oceanic and Atmospheric Administration 2005). The data used were 1° × 1° resolution optimally interpolated (OI) SST (Reynolds' SST, Reynolds *et al*. 2002). Reynolds' 30 year monthly means between 1971 and 2000 were extracted. Despite being of a relatively lower resolution than other freely available SST products, Reynolds' SST were preferred since 30 years of this form of climatology data was available and there were no data gaps from cloud cover. Reynolds SST combines remotely sensed satellite data, in situ data and SST simulated from ice cover. Satellite data are also corrected for biases using methods developed by Reynolds (1988) and Reynolds and Marsico (1993) which makes this data fairly reliable when compared with other SST products. Rouault *et al*. (2010) caution that Reynolds SST may not capture all the SST features of the Agulhas Current, especially in coastal areas, because of the resolution and the interpolation scheme. However, Reynolds SST can be used relatively accurately when averaged over a large domain (Rouault and Lutjeharms 2003; Rouault *et al.* 2010).

Seasonal maximum and minimum SST layers were generated in ArcMap from the averaged monthly climatologies. The four seasons used were: summer (January, February, and March); autumn (April, May, and June); winter (July, August, and September); and spring (October, November, and December). Bathymetry data were downloaded from the African Marine Atlas website (African Marine Atlas 2006). The bathymetry data used were a blend of the Smith and Sandwell (Smith and Sandwell 1997) and General Bathymetric Charts of the Oceans (GEBCO Relief, British Oceanographic Data Centre 2003) datasets. Smith and Sandwell (1997) is a two-minute resolution dataset that blends ship bathymetry data with bathymetry derived from remote-sensing satellite gravity data (Marks and Smith 2006). The GEBCO Relief Dataset is a one-minute resolution dataset created from hand-drawn bathymetric contours at 500 m intervals at 1:10 million scale. These hand-drawn maps originated from digital and analog ship surroundings resulting in data that may be sparse or irregular since it relies solely on ship track coverage (Marks and Smith 2006).

Future

IDRISI Earth Trends Modeller version 17 was then used to calculate linear regressions of SST, using monthly Reynolds' SST data from January 1982 to December 2010, in which the slope of the linear regression indicates the change in SST over time. This method was chosen since global climate models (GCMs) have problems with biases in cooler upwelling areas (Stock *et al.* 2010) and so the best way to forecast SST is to project past trends into the future. SST trends were first calculated for each month. The monthly SST trends were then averaged to calculate seasonal SST trends (°C per decade). These layers were used to create seasonal maximum and minimum SST layers for the years 2020 and 2030.

Raster processing

All raster layers were interpolated inland using focal statistics in ArcMap to fill in data gaps along the coast. All layers were then resampled to a smaller grid size of 0.05 degrees using a distance weighted average between points. Finally all layers were clipped between the shore and the 1000 m depth contour to exclude large areas where *E. andersoni* is not likely to occur.

Restricting the range of environmental conditions over which niche-based models are developed when projecting the potential future distributions of species with future changes in climate, has been shown to negatively affect the performance of these models (Thuiller *et al.* 2004). Therefore, the environmental layers were not clamped to represent only the possible distributional range of *E. andersoni* and instead large stretches of the eastern and western coasts of southern Africa were included in the modelling area, as well as areas with depths between 100 m and 1000 m. Most models projected populations of *E. andersoni* near the tropics on the western coast, however, which affected the calculations of range size changes. The continental shelf waters of the western coast north of the Angolan border were therefore excluded since *E. andersoni*, which is a subtropical to warm-temperate Indian Ocean species, is extremely unlikely to traverse the cold upwelled waters of the Benguela Current to reach the warmer waters near the tropics, in the form of larvae or as adults. The Benguela Current has been documented as being an effective biogeographic barrier to the movement of a number of fish species (Henriques 2011).

3.2.3 Ensemble modelling

Creating pseudo absences

To ensure that the input data were in a format that could be used within the BIOMOD packages, pseudo-absences (PAs) or background data had to be selected since no true absence data were available (Thuiller *et al.* 2009b; R Development Core Team 2012). Although different models ideally require differing numbers of PAs, in order to combine the models into a single ensemble model, all models must use the same PAs. Ten times the number of oceanographically and environmentally stratified PAs were simulated as there were presences (i.e. 2100 PAs were simulated) as recommended by Barbet-Massin *et al*. (2012). Barbet-Massin *et al*. (2012) demonstrated that randomly selected PAs produced the most predictive models (i.e. the most accurate predictions) but that the Surface Range Envelope (SRE) approach could be beneficial when sampling of occurrence data was biased. Therefore PAs were generated both randomly within the modelling area where no presence points occurred, as well as using the SRE approach. The SRE approach was used since occurrence data collection methods differed between sampling regions and may have resulted in biased occurrence data. Different PA generation techniques were used to assess whether the method of PA generation had an effect on model results. For the SRE approach, PAs were only extracted from the environmental area outside a Surface Range Envelope that contained 99.99% of the presence points, under the assumption that the entire environmental area within the SRE is favourable for the species and PAs should not be extracted from this area (Thuiller *et al.* 2009a). This resulted in two modelling runs with the same number of PAs (2100 PAs) but different methods of PA generation: 1) SRE, 2) randomly generated. The main results of the first model run are presented in the results section of this chapter. Supplementary results for the first model run and the results of the second model run are presented in Appendix I.

Input variables and variable importances

Autocorrelation between the environmental variables was assessed in ArcMap using a scatterplot matrix, since the results of some niche models are affected by correlating variables (Dormann *et al.* 2007; Valavanis *et al.* 2008). Summer maximum and minimum, and winter maximum and minimum SST were found to be highly correlated. Fish species distributions are likely to be constrained by the highest and lowest temperatures that the species will experience in its distributional range, which are generally close to the thermal tolerance limits for fishes (Sunday *et al.* 2012), and therefore summer maximum and winter minimum SST layers were retained for inclusion in the models as opposed to summer minimum and winter maximum. This resulted in a total of seven environmental variables that were used for modelling: bathymetry; summer maximum (summer max); winter

minimum (winter min); autumn maximum (autumn max) and minimum (autumn min); and spring maximum (spring max) and minimum (spring min) SST.

The relative variable importances with regard to the degree to which each "explained" the distribution of *E. andersoni* were determined using the method developed by Thuiller *et al*. (2009b) for the BIOMOD packages. This method uses a randomisation procedure to estimate the importance of each variable, independent of model technique. Firstly, a standard model prediction is created using all the environmental variables. Secondly, BIOMOD produces a set of new predictions, each with one of the independent (environmental) variables being randomized. Pearson correlations are then done between the standard predictions and predictions where the variable under investigation has been randomly permutated. High correlation between the predictions indicates that the variable permutated is considered not important for the model projections. The mean correlation coefficient over a number of runs is recorded and BIOMOD then gives a ranking of the variables for each of the models selected by subtracting the correlation score obtained from one (Thuiller *et al.* 2009b, 2010). Values closer to zero are therefore of low importance and values closer to one are of higher importance (Thuiller *et al.* 2010; Capinha and Anastácio 2011).

Model calibration and evaluation

The correlative models were run in the BIOMOD (Thuiller 2003; Thuiller *et al.* 2009b) packages for ensemble forecasting in R version 2.15.1 (R Development Core Team 2012). An ensemble modelling approach was used. Seven models available in the BIOMOD packages were used: (1) Generalized linear model (GLM); (2) Generalized Additive Model (GAM); (3) Classification Tree Analysis (CTA); (4) Random Forest (RF); (5) Boosted Regression Trees (BRT); (6) Multiple Adaptive Regression Splines (MARS); and (7) Maximum Entropy (MAXENT) (Table 3.1). These models represent a broad range of the available modelling methods and have been commonly used, or have been shown to perform well in some cases when modelling the distributions of species.

Table 3.1: Strengths and weaknesses of all models used in the ensemble and general comments on the class of each model and how it functions.

Model	Class of model	Strengths	Weaknesses	Comments
MAXENT (Maximum Entropy)	➢ Machine learning - refers to algorithms that enable computers to "learn" from experience and improve their performance over time (Elith *et al.* 2006; Termansen *et al.* 2006)	➢ Generally performs well ➢ Able to fit complex responses (Elith *et al.* 2006)	➢ It can predict high suitability of environmental conditions that are outside the range present in the study area (i.e. extrapolation). To alleviate this problem, values are clamped to the upper or lower values found in the study area (Pearson 2007)	➢ Finds the probability distribution of maximum entropy (the distribution that is most spread out or closest to uniform) subject to constraints imposed by the information available regarding the observed distribution of the species and environmental conditions across the study area. ➢ Uses background environmental data (i.e. pseudo absences) (Pearson 2007)
GLM (Generalized Linear Models)	➢ Standard regression method ➢ Extensions of linear regression (Valavanis *et al.* 2008) ➢ Fits parametric terms	➢ Relative contributions of variables can be determined (sensu Pearson 2007) ➢ Most commonly used niche model, showing good ability to predict current species distributions (Thuiller 2003)	➢ Unable to deal with complex response curves i.e. non-linear relationships between species and environmental predictors (Thuiller 2003; Leathwick *et al.* 2006)	➢ Using classification and regression trees (CARTs) complementary to GLMs and GAMs enables ecologically meaningful interactions to be identified (Valavanis *et al.* 2008)
GAM (Generalized Additive Models)	➢ Standard regression method ➢ Extension of GLM: linear and other parametric terms replaced by smoothing functions (Valavanis *et al.* 2008)	➢ Well developed and commonly used to model fish distributions ➢ Ecologically interpretable non-parametric response curves ➢ Statistically well defined allowing good inference but also flexible enough to fit the data closely (Valavanis *et al.* 2008)	➢ Computational complexity makes the generation of predictions for independent datasets such as in a GIS system cumbersome (Leathwick *et al.* 2006)	➢ Using classification and regression trees (CARTs) complementary to GLMs and GAMs enables ecologically meaningful interactions to be identified (Valavanis *et al.* 2008)

MARS (Multivariate Adaptive Regression Splines)	➢ Standard regression method ➢ Fits non-linear responses using piecewise linear fits rather than smooth functions (Elith *et al.* 2006; Valavanis *et al.* 2008)	➢ Faster to implement than GAMs and simpler to use in GIS applications (Elith *et al.* 2006) ➢ Combines strengths of regression trees and spline fitting by replacing step functions with piecewise linear basis functions. ➢ Can model complex relationships (Leathwick *et al.* 2006)	➢ Current implementations are fitted assuming normally distributed errors, so they need to be coupled with a GAM or GLM model, respectively, to properly analyse presence/absence or count data (Leathwick *et al.* 2006)	
RF (Random Forest)	➢ Model-averaging ➢ Algorithm builds multiple trees using randomly selected subsets of the observations and random subsets of the predictor variables. The predictions from the trees are then averaged (Lawler *et al.* 2006)	➢ Perform well due to the power derived from averaging hundreds of different models ➢ Tree-based models advantageous since they can model complex interactions without having to specify them a priori; and they allow the relationships between the response and the predictors to vary over the domain of the study (Lawler *et al.* 2006)	➢ Accuracy depends on the strength of the individual trees in the model and the correlation between them (Breiman 2001) ➢ Often little understanding of the mechanism of the RF "black box" (Breiman 2001)	

GBM/BRT (Generalized Boosting Model/ Boosted Regression Trees)	➢ Combines two algorithms: the boosting algorithm iteratively calls the regression-tree algorithm to construct a combination or ‘‘ensemble’’ of trees (Elith *et al.* 2006)	➢ Accommodates both different types of predictor variables and missing values. Immune to the effects of extreme outliers and the inclusion of irrelevant predictors. Can fit interactions between predictors (Valavanis *et al.* 2008) ➢ Good at selecting relevant variables and can model interactions. Boosting is used to overcome the inaccuracies inherent in a single tree model (Elith *et al.* 2006)	➢ Over-fitting can be a problem if cross validation is not used and if predictive accuracy is not adequately tested (Elith *et al.* 2006) ➢ Often little understanding of the mechanism of the GBM “black box”(Friedman 2001)	➢ Regression trees are fitted sequentially on weighted versions of the data set, where the weights continuously adjust to take account of observations that are poorly fitted by the preceding models. Boosting can be seen as a method for developing a regression model in a forward stage-wise fashion, at each step adding small modifications in parts of the model space to fit the data better (Elith *et al.* 2006)
CTA (Classification Tree Analysis)	➢ Machine learning ➢ Non-parametric modelling approach ➢ Involves the recursive binary partitioning of data (Lawler *et al.* 2006)	➢ Accurate and useful to describe hierarchical interactions between species (*sensu* Thuiller 2003)	➢ Tendency to over-fit during the calibration process (Thuiller 2003) ➢ Misclassification can be problematic with classification and regression trees unless they are used in conjunction with boosting algorithms (Leathwick *et al.* 2006)	➢ Function by way of recursive binary partitioning of data into increasingly homogenous groups with respect to the dependent variable. The two most homogenous groups of data with respect to the response variable are chosen (using the explanatory variables) and the resulting model is a tree-like structure consisting of a series of nodes (Valavanis *et al.* 2008)

Models were calibrated using a random sample of the initial data (80%) and each model was evaluated using the remaining 20% of the data. Models were assessed with three measures of model accuracy: Cohen's Kappa (Cohen 1960), the Area Under the Curve (AUC) of Receiver-Operating Characteristic (ROC) (Swets 1988), and the True Skills Statistic (TSS) (Allouche *et al.* 2006). A confusion matrix is first calculated that gives the number of true positives (***a***), false positives (***b***), false negatives (***c***) and true negatives (***d***) that are predicted by each model (Table 3.2).

Table 3.2: Measures of model's predictive accuracy calculated from a two × two confusion matrix, modified from McPherson *et al*. (2004) and Allouche *et al*. (2006).

		Observed		
		Present	**Absent**	**Total**
Predicted	**Present**	true positives (***a***)	false positives (commission) (***b***)	No. predicted presences
	Absent	false negatives (omission) (***c***)	true negatives (***d***)	No. predicted absences
	Total	No. of observed presences	No. of observed absences	***N*** = total observations

All three evaluation methods are affected by the values of sensitivity and specificity which are derived from the confusion matrix. Sensitivity refers to the proportion of observed presences that are correctly predicted as presences by the model (true positives, ***a***) and therefore low sensitivity indicates numerous omission errors (incorrect predictions of absences, ***c***) (McPherson *et al.* 2004). The formula for sensitivity is:

$$\text{sensitivity} = a/a + c \qquad \text{(Equation 1)}$$

Specificity measures the proportion of observed absences that are correctly predicted as absence (true negatives, ***d***) and therefore low specificity indicates high commission error (incorrect predictions of presence, ***b***) (McPherson *et al.* 2004). The formula for specificity is:

$$\text{specificity} = d/b + d \qquad \text{(Equation 2)}$$

KAPPA is calculated as:

$$\text{KAPPA} = \left(\frac{a}{a+c}\right) - \frac{(a+b)(a+c) + (c+d)(d+b)}{N^2} \Big/ 1 - \frac{(a+b)(a+c) + (c+d)(d+b)}{N^2}$$

(Equation 3)

and TSS is calculated as:

$$\mathrm{TSS} = ad - bc/(a+c)(b+d) = (\text{sensitivity} + \text{specificity}) - 1$$

(Equation 4)

ROC curves are constructed by using the entire range of possible thresholds to classify the scores into confusion matrices (such as the one shown in Table 3.2), obtaining the sensitivity and specificity for each matrix, and then plotting sensitivity against the corresponding proportion of false positives or commission error (i.e. 1 - specificity) (Allouche *et al.* 2006).

Kappa and AUC methods have been heavily criticised (McPherson *et al.* 2004; Termansen *et al.* 2006; Allouche *et al.* 2006; Peterson *et al.* 2007; Austin 2007; Lobo *et al.* 2008). Kappa values are affected by prevalence (which refers to the proportion of sampled sites in which a species is present), and this introduces statistical bias to estimates of predictive accuracy (Allouche *et al.* 2006). Although ROC measurements of model accuracy are independent of prevalence, ROC curves are constructed using all possible thresholds (Allouche *et al.* 2006). Therefore, when probability distributions are binary transformed into presence-absence predictions, the threshold used for binary transformation should be used to evaluate models rather than the threshold-independent ROC curves (Allouche *et al.* 2006). In accordance with recent studies, model evaluations were based on TSS (Freeman and Moisen 2008; Boitani *et al.* 2008; La Morgia *et al.* 2008) with Kappa and ROC evaluations provided for the interest of comparison. Prevalence was kept at 0.5 as is the procedure when using pseudo-absence data in BIOMOD, to ensure that presence data have the same weight as PA data even when a large number of PAs are extracted (Georges and Thuiller 2012).

The interpretation of TSS scores was based on the Landis and Koch (1977) accuracy classification scheme scores as used by Lasram *et al*. (2010): TSS > 0.8 is excellent; 0.6 < TSS < 0.8 is good; 0.4 < TSS < 0.6 is fair; 0.2 < TSS < 0.4 is poor; and TSS < 0.2 is no predictive ability. For ROC, a value of 0.5 indicates no apparent accuracy whereas 1 indicates perfect accuracy (Hanley and McNeil 1982). The detailed classification scheme is as follows: ROC > 0.9 is excellent; 0.8 < ROC < 0.9 is good; 0.7 < ROC < 0.8 is fair; 0.6 < ROC < 0.7 is poor; and 0.5 < ROC < 0.6 is no predictive ability (Swets 1988; Broennimann *et al.* 2006).

The accuracy of the ensemble model is reliant on the accuracy of the individual models that are included in the ensemble (Araújo and New 2007; Elith *et al.* 2010) and therefore only models that had a TSS score of above 0.85 were included in the final ensemble projections. A weighted average

approach was used to create the ensemble model projections (Georges and Thuiller 2012). This method weights individual models that are included in the ensemble according to their performance as measured by TSS, so that more accurate models are more influential in the final ensemble model (Georges and Thuiller 2012). After being weighted, the probabilities of occurrence of each individual model are then averaged to create a weighted mean of probabilities (Georges and Thuiller 2012). Weighted methods have been found to perform well in a number of different studies (Araújo *et al.* 2005; Araújo and New 2007; Thuiller *et al.* 2009b; Capinha and Anastácio 2011).

Model Projections

Both the binary transformed (presence-absence distributions) and the non-binary transformed (probability of occurrence distributions) projections of species distributions were plotted. The threshold used to transform probability values into binary presence-absence was the threshold that maximized model accuracy measured by TSS. For the probability distributions, the probability of occurrence classes that were used, were taken from Kumar and Stohlgren (2009) and were defined as: very low 0.1 - 0; low 0.1 - 0.4; medium 0.4 - 0.6; and high 0.6 - 1. Using the projected SST trends for the years 2020 and 2030, the potential thermal habitats for *E. andersoni* were projected into the future in both binary transformed and non-binary transformed formats.

Range Changes

Range changes were calculated using the range size function in BIOMOD. The proportion and relative number of pixels (or amount of habitat) that were lost, gained or that remained stable between the current and the 2020 and 2030 projections, were calculated for each of the individual models projections and for the ensemble model projections. Two scenarios are considered by BIOMOD when calculating range changes, one in which full dispersal is assumed to be possible and one in which there is no dispersal. Both these scenarios were presented in the results. Range loss and gain were converted into percentages and the overall range change was also calculated.

3.3 Results

3.3.1 Future SST trends

The maps of the seasonal decadal trends in SST between 1982 and 2010 are presented in Figure 3.2. The trends reveal a pattern of coastal warming throughout the year on the coast of southern Africa by as much as 0.5°C per decade in summer (Figure 3.2 A). However, pockets of cooling were also

observed along the south-east coast in the autumn, winter and spring months (Figure 3.2 B, C and D respectively). Within the current distributional range of *E. andersoni*, cooling occurred along the south coast of South Africa from Port Alfred to Struis Bay (up to -0.4°C between Knysna and Plettenberg Bay in winter) and along the southern Mozambican coast from the South Africa-Mozambique border to Xai Xai (up to -0.3°C in Maputo Bay in spring). Outside of the current distributional range of *E. andersoni*, cooling occurred on the east coast between Beira and the Zambezi Delta in winter and spring and on the west coast of South Africa between St Helena Bay and Stilbaai in summer, autumn and winter.

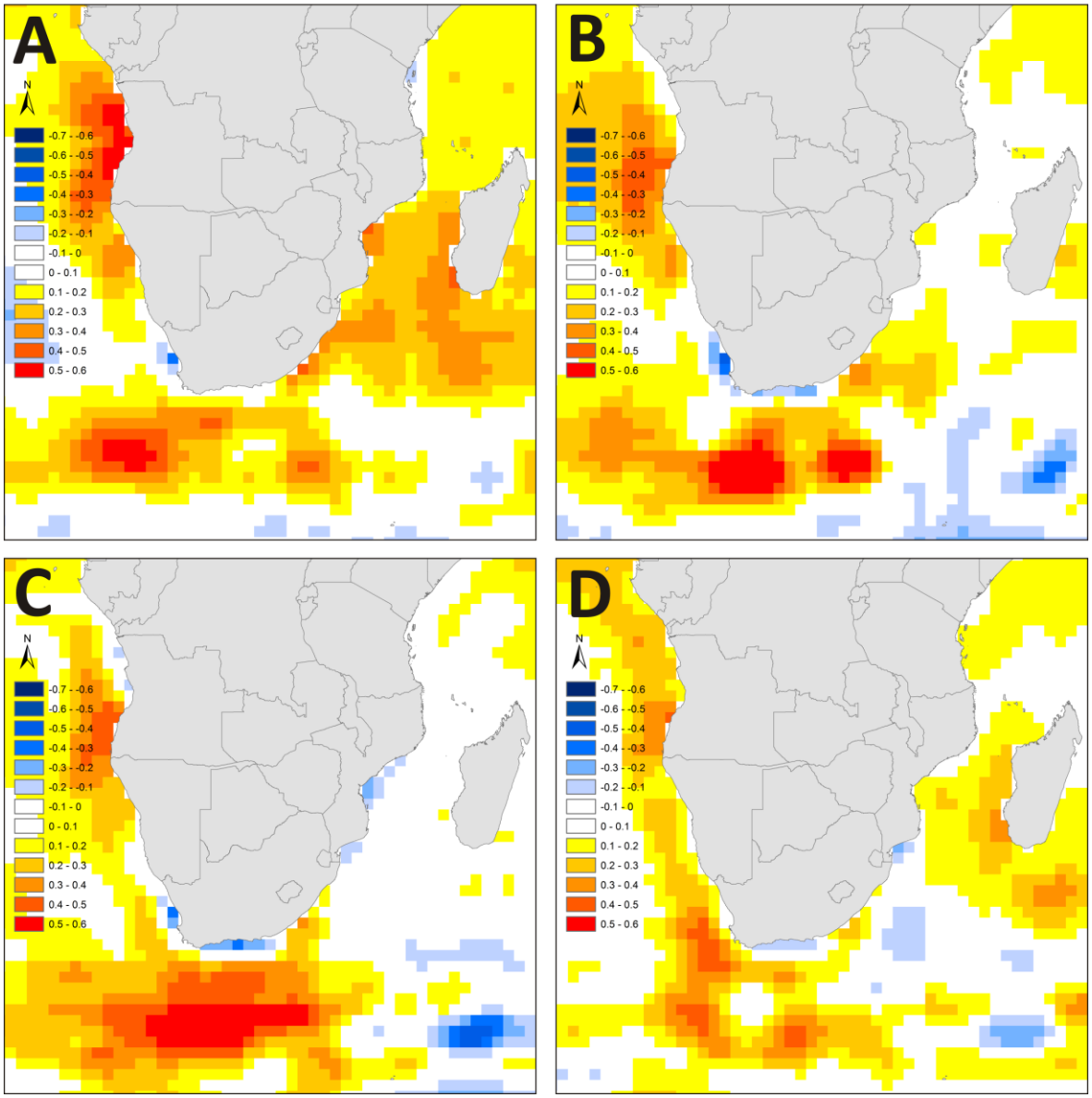


Figure 3.2: Decadal trends in NOAA OI SST between 1982-2010 projected for southern Africa for the four seasons: A) summer (January, February, March); B) autumn (April, May, June); C) winter (July, August, September); D) spring (October, November, December). The accompanying key represents decadal SST change (i.e. cooling is negative and warming is positive).

3.3.2 Variable importance

The standard errors and standard deviations were relatively high across different modelling algorithms for all environmental variables except for bathymetry (Table 3.3, Figure 3.3). The large standard deviations and standard errors of the SST variable importances indicate that the degree to which these variables influence model predictions is highly variable. Only winter minimum, autumn minimum and bathymetry have mean variable importances higher than the overall mean importance (0.306), as shown in Table 3.3 and Figure 3.3. Bathymetry was the most important environmental variable and had the lowest errors, reinforcing its predictive importance for all modelling techniques. Although SST variables had high variation in their importance values, colder temperatures were of greater importance when making model predictions (i.e. winter and autumn minimum SSTs were important); summer and spring maximums were, however also relatively important.

Table 3.3: Variable importances for each of the seven environmental variables (columns) when the *Epinephelus andersoni* distribution was modelled using seven different modelling algorithms (rows). Mean importance, standard deviation and standard error for each variable across modelling techniques are shown.

	Summer max	Winter min	Bathymetry	Autumn max	Autumn min	Spring max	Spring min
MAXENT	0.058	0.358	0.554	0.279	0.191	0	0
CTA	0.011	0.68	0.539	0.129	0.122	0	0
RF	0.011	0.015	0.496	0.021	0.104	0.025	0.019
MARS	0.91	0.486	0.51	0.029	0.651	0	0
GAM	0.94	0.925	0.514	0.01	0.663	0.745	1
GLM	0	0	0.479	0.69	1	0.779	0
GBM	0.006	0.129	0.518	0.172	0.198	0.024	0.008
Mean	0.277	0.370	0.516	0.190	0.418	0.225	0.147
SD	0.443	0.351	0.025	0.241	0.351	0.367	0.376
SE	0.168	0.133	0.009	0.091	0.133	0.139	0.142

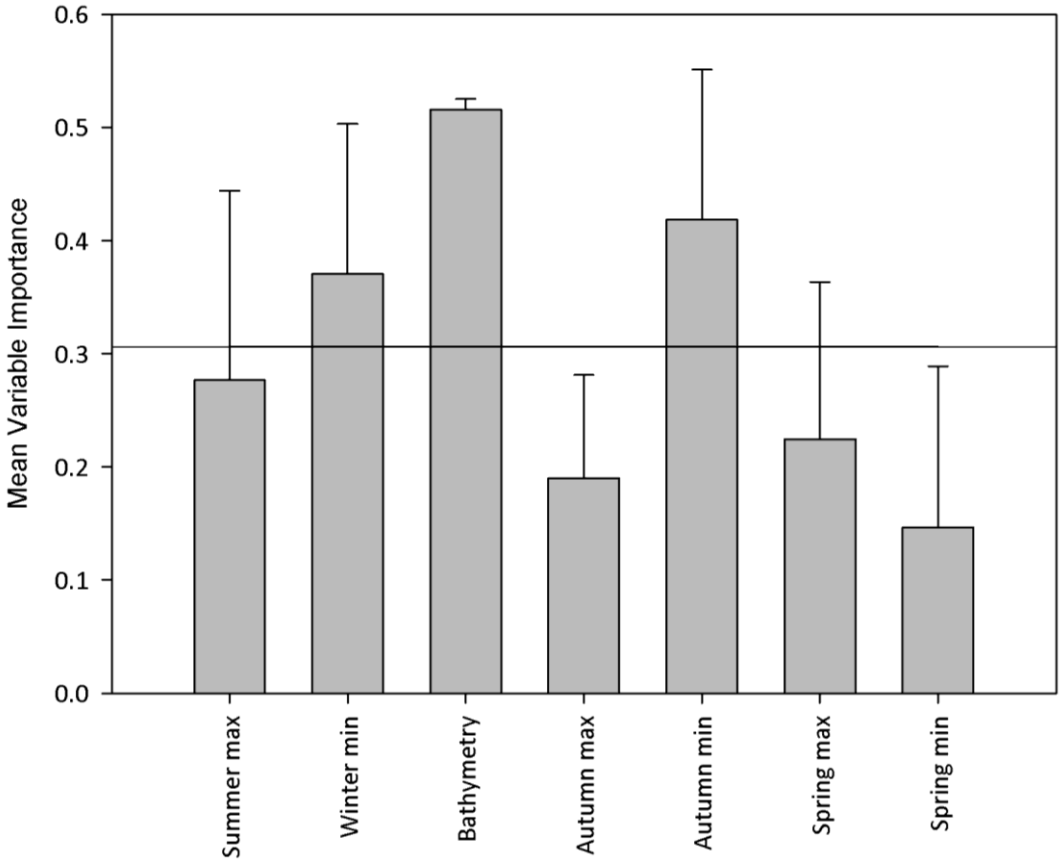


Figure 3.3: Mean variable importances for each of the seven environmental variables with standard error whiskers. The solid horizontal line indicates the overall mean variable importance (0.306).

3.3.3 Model performance

For the individual model projections, all models could be classified as excellent at predicting the distribution of *E. andersoni* (i.e. all TSS and KAPPA > 0.8, and all ROC > 0.9) (Table 3.4). Model accuracy for TSS ranged from 0.976 for MARS to 0.996 for both RF and CTA; KAPPA ranged from 0.944 for GLM to 0.995 for RF; and ROC ranged from 0.988 for GLM to 1.000 for RF, MAXENT and GBM. According to all methods of model evaluation, RF was one of the best-performing models and GLM and MARS were among the worst-performing models. All models had a TSS score higher than the set TSS threshold of 0.850 and consequently all models were included in the ensemble. For the weighted mean ensemble model projections, all model evaluation methods showed excellent predictive accuracy (TSS: 0.991, ROC: 1.000, KAPPA: 0.989).

Table 3.4: Three measures of model evaluation, True Skills Statistic (TSS), area under the Receiver Operating Curve (ROC), and Cohen's KAPPA are shown for each of the seven modelling algorithms used to model the distribution of *Epinephelus andersoni* into the future.

	MAXENT	CTA	RF	MARS	GAM	GLM	GBM
TSS	0.982	0.996	0.996	0.976	0.983	0.977	0.989
ROC	1.000	0.999	1.000	0.998	0.998	0.988	1.000
KAPPA	0.987	0.984	0.995	0.946	0.954	0.944	0.989

3.3.4 Ensemble model projections

Current

The current *E. andersoni* distribution projections were relatively similar for each of the seven individual modelling techniques and fitted the presence points fairly accurately (Figure 5.1, Appendix I). Similarly, the current probability weighted mean ensemble model projection fitted the presence points accurately as shown in Figure 3.4 A and Figure 3.5 A. Although southern Madagascar is considered to be part of the *E. andersoni* distribution, Heemstra and Randall (1993) state that reports of the species occurring there may be dubious. Since all individual models, as well as the probability weighted mean ensemble model, project a southern Madagascan distribution despite not including any occurrence records from this region, it is possible that this part of the distribution may exist.

2020

No highly noticeable change in the *E. andersoni* distribution takes place between the current ensemble projection and the 2020 ensemble projection as shown by Figure 3.4 B and Figure 3.5 B. The ensemble model results mirror the results for the individual models projected for 2020 (Figure 5.2, Appendix I). The small change in range by 2020 is supported by the range changes calculated by BIOMOD for the individual models (Table 5.1 and Table 5.5 in Appendix I). The range changes calculated for the ensemble model (Table 3.5) show that of the 2645 pixels that were occupied in the current projection, 2402 pixels remain occupied in 2020. Despite this, some range loss was still projected with 9.149% of the *E. andersoni* distribution predicted to be lost by 2020. This range loss was caused by contraction of the southern limit of the *E. andersoni* distribution (as projected by the binary transformed ensemble models) from Jeffreys Bay to just south of Algoa Bay (Figure 3.5 A - B). In the binary transformed projections, there was also a small contraction of the north-eastern Madagascan range in 2020.

2030

The individual model projections for 2030 generally indicate a more noticeable reduction in the *E. andersoni* range than was predicted for 2020 (Figure 5.3, Appendix I). There is also a noticeable change in the 2030 binary transformed ensemble projected distributions (Figure 3.5 C) and the area of high probability of occurrence in the non-binary transformed ensemble projections (Figure 3.4 C). In the binary transformed maps there is no change to the Madagascan distribution but there is further contraction at the southern limit of the *E. andersoni* distribution where the distribution offshore of Algoa Bay is lost. Furthermore the northern limit also contracts from just south of Inhambane to Paindane in Mozambique. These contractions result in overall 16.371% of the *E. andersoni* distribution being lost (Table 3.5). There was virtually no range gain since the current projection (2 pixels) but there was noticeable range loss (435 pixels, Table 3.5). From the non-binary transformed maps, it is evident that the area of high probability of occurrence shrinks at the southern and the northern range limits. In the south, the area of high probability moves north from the southern edge of Algoa Bay in 2020 (Figure 3.4 B) to the northern edge of Algoa Bay (just south of Port Alfred) in 2030 (Figure 3.4 C).

.

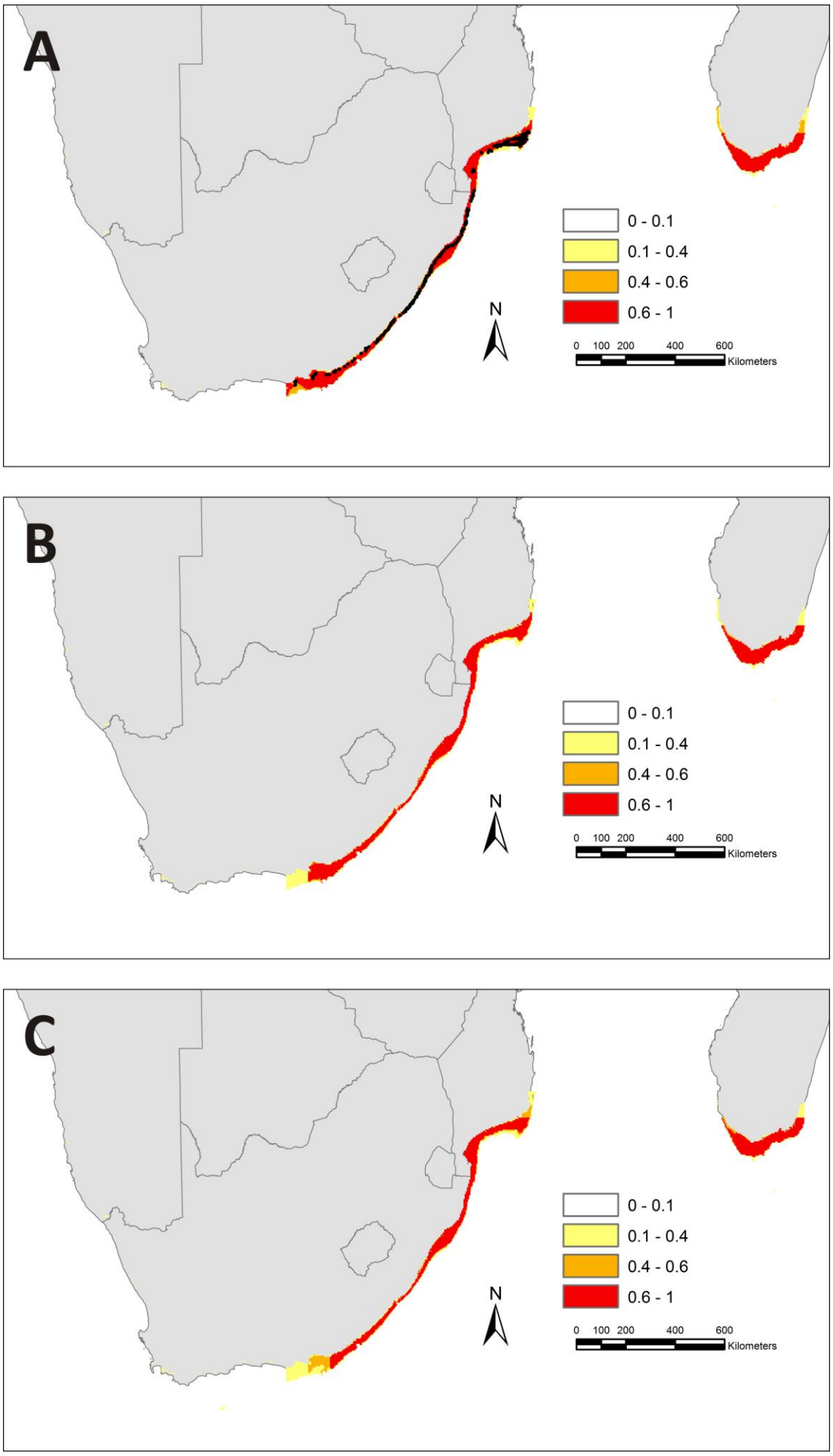


Figure 3.4: Probability mean weighted ensemble probability distributions of *E. andersoni* projected for the present (A), 2020 (B), and 2030 (C). Occurrences are plotted against the current projection.

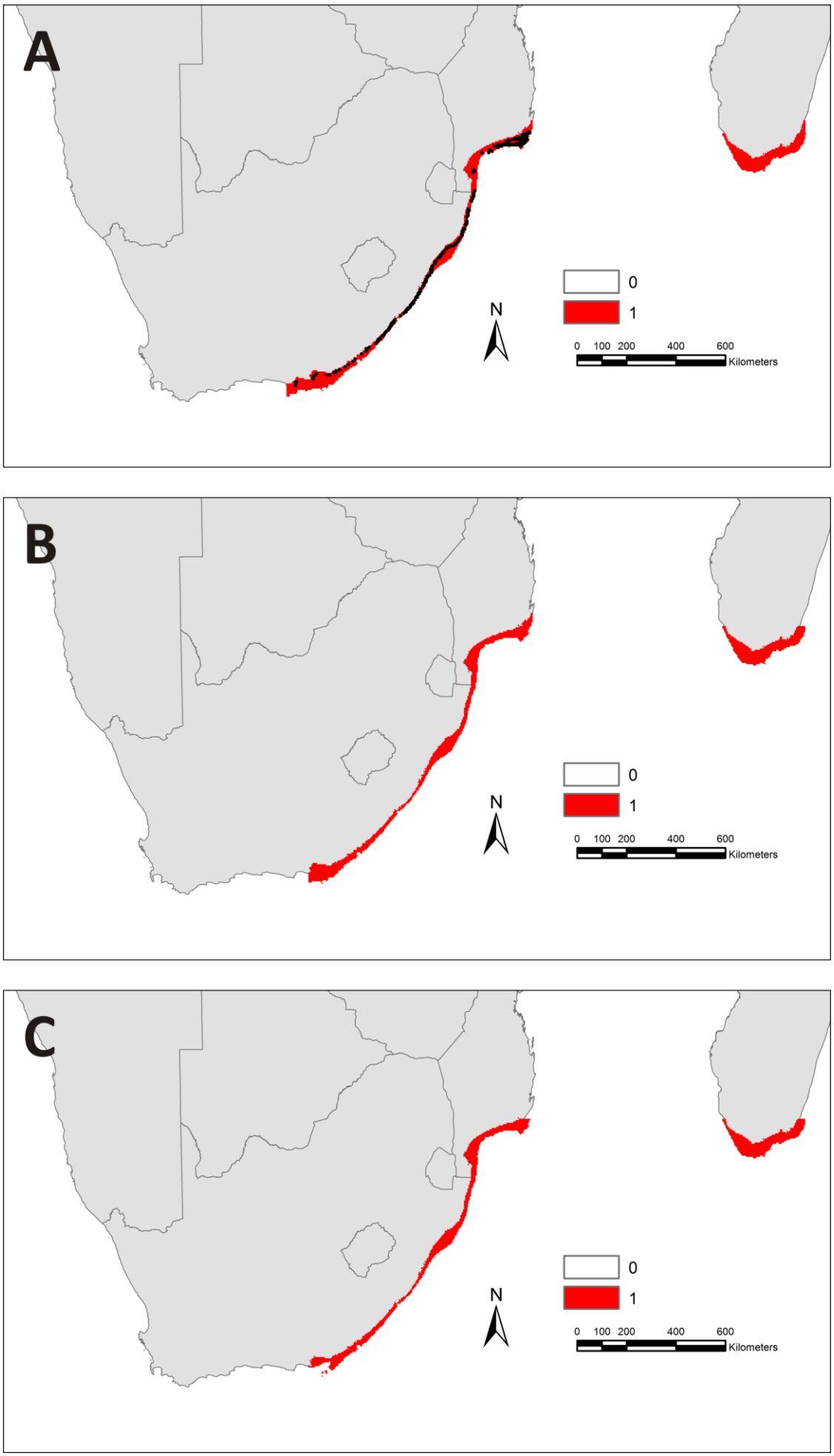


Figure 3.5: Binary probability mean weighted ensemble distributions of *E. andersoni* projected for the present (A), 2020 (B), and 2030 (C). Occurrences are plotted against the current projection.

Table 3.5: Summary of the range changes between the Probability Weighted Mean ensemble current and the 2020 and 2030 projections. Figures represent number of pixels except where percentages are indicated. “Stable0” indicates the number of pixels that are unoccupied in the current projection and that remain unoccupied in the future projections. “Stable1” indicates the number of pixels that are occupied in the current projection and are predicted to remain occupied in the future.

	Current - 2020	Current - 2030
Loss	243	435
Stable0	25450	25449
Stable1	2402	2210
Gain	1	2
%Loss	9.187	16.446
%Gain	0.038	0.076
Range Change	-9.149	-16.371
Current Range	2645	2645
Future Range Size		
No Dispersal	2402	2210
Full Dispersal	2403	2212

3.4 Discussion

This study presents the first known attempt to predict future climate change induced distributional changes of southern African marine fish fauna using correlative SDMs. Documented intensification of upwelling zones on the south-east African coast causing localised cooling and intensification of the Agulhas Current causing general warming (Rouault *et al.* 2010), were supported by the seasonal decadal SST trends (Figure 3.2). This confirms the likelihood of future climatic changes in the Indian Ocean. Fisheries are highly important globally, especially in southern Africa where they support many livelihoods and are important food sources (van der Elst *et al*. 2005) therefore making their sustainable use a necessity in light of the projected effects of climate change. Furthermore, the south-east coast of Africa is an area of high endemism which is connected to the unique environment of the southern tip of continental Africa and consequently this region is pinpointed as a conservation priority (van der Elst *et al*. 2005).

Since *E. andersoni* is a range-restricted endemic in an area of high endemism it fits into the climatic rarity hypothesis, which states that areas (or “islands”) of high endemicity are likely to contract with future climatic changes (Ohlemüller *et al.* 2008). Model results indicate that the northern limits of the range will contract further southward with no expansion at the southern limits of the *E. andersoni* distribution. This is despite predictions of pole-ward distribution shifts with future projected warming for tropical, warm-temperate and temperate species (Parmesan and Yohe 2003;

Perry *et al.* 2005; Sabatés *et al.* 2006; Hiddink and Hofstede 2008; Cheung *et al.* 2009). The lack of southward range expansion could be explained by the regional cooling projected by the decadal SST trends, south and offshore of Algoa Bay (Figure 3.2 B, C and D). Southward contraction of the northern distribution limit to Paindane could be explained by future projected warming during summer in this region (Figure 3.2 A). If this contraction from both range limits continues, and if further southward movement is limited by cooling, this could result in the species becoming "sandwiched" within an ever decreasing margin of favourable temperature conditions. This, coupled with the relatively specific habitat requirements of *E. andersoni*, being predominantly exclusive to rocky reefs (Fennessy 2004; Craig *et al.* 2011) and being non-migratory (Fennessy 2004), may reduce the habitable area available to the species even more than the conservative contraction predicted by the ensemble model indicates.

In many instances SDM studies are forced to generate pseudo-absences (PAs) data when true absence data is not available for the species being modelled (Stokland *et al.* 2011). Pseudo-absences have been found to affect the accuracy of SDMs and different methods of PA generation are recommended in different circumstances (Barbet-Massin *et al.* 2012). The two methods of PA generation that were tested for the *E. andersoni* SDMs were found to perform differently. PAs generated using the Surface Range Envelope (SRE) approach resulted in models that performed better according to the measures of model evaluation (Table 3.4) than the models using the randomly generated PAs which did not perform as well (Table 5.3). The difference in performance between the methods of PA generation could be explained if there was bias in the occurrence records, in which case SRE generated PAs have been found to perform better (Barbet-Massin *et al.* 2012). A noticeable difference between the two model runs is that models using SRE simulated PAs consistently projected a Madagascan *E. andersoni* distribution whereas models using random PAs did not. This may have been due to the fact that no occurrence records from the Madagascan coast were available and therefore random PAs were randomly allocated to the Madagascan coastline, preventing the projection of any *E. andersoni* distribution. For the SRE approach, environmental areas of the Madagascan coastline that coincided with the environmental envelope that contained 99.99% of the existing occurrence records may have been considered as suitable habitat, which would explain the projected *E. andersoni* distribution along the southern Madagascan coastline for the SRE model run (Figure 5.1, Figure 5.2 and Figure 5.3).

Pseudo-absences are one of many influences on model results which are also reliant on the assumption that changes in abiotic variables (i.e. SST in this case) are the primary factors responsible for species distribution shifts (Pearson and Dawson 2003; Lawler *et al.* 2006). Although it is widely

accepted that ocean temperature plays an important role in determining marine fish distributions (Pörtner *et al.* 2001; Parmesan and Yohe 2003; Perry *et al.* 2005; Parmesan 2006; Sabatés *et al.* 2006; Pörtner and Knust 2007; Hiddink and Hofstede 2008; Dulvy *et al.* 2008; Rijnsdorp *et al.* 2009; Cheung *et al.* 2009; Lasram *et al.* 2010), the exact physiological impacts of temperature changes on fishes in the ocean are unclear (Rijnsdorp *et al.* 2009; Pankhurst and Munday 2011; Sunday *et al.* 2012). Reactions to changes in temperature may be complicated by other climate-induced changes such as increases in CO_2, hypoxia, salinity changes, eutrophication and ocean acidification (Harley *et al.* 2006; Pörtner and Farrell 2008). Some have speculated that changes in fish distributions may be due to adult thermal tolerance windows (Pörtner *et al.* 2001). It is, however, likely that an adult fish's responses to changes in temperature will not be as straightforward as assumed by correlative models since most fish, including *E. andersoni*, are thermal conformers (Clark *et al.* 2003). This means that temperature changes will have indirect effects on the higher functions of fishes such as *E. andersoni*, including effects on secondary production and reproductive output (Pörtner *et al.* 2001; Pörtner and Knust 2007; Pörtner and Farrell 2008). Laboratory data for marine fish and invertebrates show that the maximum temperature at which oxygen delivery becomes limiting under experimental conditions, closely matches the upper temperature limit in the natural environment beyond which growth performance is restricted and abundance decreases (Pörtner and Knust 2007). Pörtner *et al*. (2001) found that temperatures below optimal levels in the North Atlantic shifted the thermal tolerance windows in common eelpout (*Zoarces viviparus*) and Atlantic cod (*Gadus morhua*); however this came at a cost as it significantly decreased growth performance, fecundity and recruitment.

Conventional correlative models also do not consider any existing ontogenetic shifts, which are especially common in many marine organisms (Robinson *et al.* 2011) such as marine fishes with pelagic larvae (Potts *et al*. in press; Colton Jr. 1959; Berlinsky *et al*. 2004; Harley *et al*. 2006; Pörtner and Farrell 2008; Rijnsdorp *et al*. 2009; Pankhurst and Munday 2011). It has been proposed that despite adult fish being able to tolerate a relatively wide range of temperatures, larvae generally have a much narrower range of tolerable temperatures (Pankhurst and Munday 2011). Larval mortality as a result of increased oceanic temperatures has been documented in the field for whiting and flounder larvae (Colton Jr. 1959) and serranid larvae have been reported to be particularly fragile and susceptible to environmental fluctuations (Tucker Jr 1994, 1999; Berlinsky *et al.* 2004). Since unusually high SSTs have been shown to increase larval mortality (Berlinsky *et al*. 2004; Pörtner and Farrell 2008; Rijnsdorp *et al*. 2009; Pankhurst and Munday 2011), future projected SST increase is likely to affect *E. andersoni* larval survival. Furthermore, contrary to popular belief, mounting evidence suggests that reef fish larval dispersal may be limited by various factors, resulting in

complex connectivity patterns (Cowen *et al.* 2006). The more constricted larval thermal limits are not considered by SDMs despite this being the life-history stage that is likely to be critically affected by climatic changes (Rijnsdorp *et al.* 2009). If future temperatures do not remain within the *E. andersoni* larval thermal limits, larval survival and therefore dispersal and recruitment could be reduced, which may ultimately compromise the survival of localised populations or of the species as a whole (Sheaves 2006; Pankhurst and Munday 2011). In such cases, the assumption of full dispersal that is used by the correlative models may not be realistic, resulting in over-optimistic model projections of future distributional ranges.

Authors also postulate that temperature acts as a proximate cue for adult fish to spawn in order to maximise the chances of larval survival (Potts *et al*. in press; Sheaves 2006; Rijnsdorp *et al*. 2009; Pankhurst and Munday 2011). It is not always clear whether this is to maximise the dispersal potential of larvae in currents usually persisting at the time of spawning (Rijnsdorp *et al.* 2009) or whether it is to ensure that larvae are not exposed to temperatures exceeding their thermal tolerance limits (Sheaves 2006; Pankhurst and Munday 2011). Potts *et al*. (in press) demonstrated that reproductive activity of blacktail seabream *Diplodus sargus capensis*, in the warming hotspot of the northern Benguela system (Atlantic Ocean), only occurred at temperatures below 20°C and that this species has begun shifting its spawning season and spawning localities in response to warming. It is predicted that if current warming trends in the Benguela continue, the reproductive output of blacktail will decrease by 20% per decade and completely cease within 60 years (Potts *et al*. in press).

Many groupers have been shown to synchronise spawning events with specific temperatures or seasons. For example Nemeth *et al*. (2007) documented the red hind *Epinephelus guttatus* spawning within a narrow temperature range between 26°C and 27.5°C, in the U.S. Virgin Islands, and when current speeds are low. Similarly, in the Bahamas and Belize, *Epinephelus striatus* spawns when temperatures range between 25°C and 26°C. *Epinephelus andersoni* have also been shown to spawn within a defined season (i.e. during warmer austral spring and summer months, November-January) and the species does not spawn as readily in southern regions where it is cooler (Fennessy 2000; Fennessy and Sadovy 2002). It is likely that *E. andersoni* seeks preferred spawning temperatures that occur during this time. If this is the case, the timing of reproductive activity may change in response to changing temperatures (Sheaves 2006; Pörtner and Farrell 2008; Rijnsdorp *et al.* 2009) possibly resulting in a mismatch between the time of spawning and the time that maximizes larval dispersal potential and/or survival (Pankhurst and Munday 2011). This could be especially concerning for *E. andersoni*, considering that it is a resident fish (Fennessy and Sadovy 2002; Maggs 2011) and

therefore possibly less capable of shifting its range to more favourable spawning grounds than migratory fishes, which could possibly result in compromised spawning (e.g. Pankhurst and Munday 2011). The model results appear to indicate sandwiching of the *E. andersoni* population within a narrowing range of favourable conditions in the future, which could limit any potential expansion of appropriate spawning areas, and possibly compounding the narrowing of spawning windows in time or space. This reinforces the possibility that the model's assumption of full dispersal may be unrealistic and therefore that model results are optimistic.

As well as the complex physiological aspects of the species being modelled, there are numerous factors other than the abiotic variables generally used to train correlative models, that could potentially influence the distribution of a species (Soberón and Peterson 2005; Pearson 2007). These include: biotic interactions (e.g. intraspecific interactions such as competition and interspecific interactions such as predation) (Pearson and Dawson 2003; Harley *et al.* 2006); oceanographic features and circulation patterns (although these are linked to climate and temperature) (Harley *et al.* 2006); dispersal (Pearson and Dawson 2003) which is linked to ocean circulation patterns for marine fishes (Cowen *et al.* 2006); and genetic variability, phenotypic plasticity and evolutionary changes (Pearson and Dawson 2003; Elith and Leathwick 2009). Another factor not considered by SDMs is human resource management and exploitation (Pearson and Dawson 2003). Fishing pressure has substantial impacts on grouper populations (Sadovy de Mitcheson *et al.* 2013) and the added impact of climatic changes on exploited marine fish stocks is becoming a major concern globally (Perry *et al.* 2005; Brander 2007; Rijnsdorp *et al.* 2009). The synergistic effects of climatic changes and fishing pressure were fatal for the cod *Gadus morhua* fishery off western Greenland which collapsed in the 1990s due to increased fishing pressure at a time of alternating periods of warming and cooling which affected recruitment (Hamilton *et al.* 2000). *Epinephelus andersoni* have been subjected to extensive fishing pressure which has reduced their spawner biomass-per-recruit levels to below 50% of pristine levels (Fennessy and Mann 2013). Judging by the sensitivity of *E. andersoni* to exploitation, the added impacts of climatic changes could prove even more detrimental to the species than model predictions suggest.

Correlative SDMs are also based on a number of assumptions that are often violated, which may result in inaccurate model projections. One example is the assumption that the species being modelled is currently in equilibrium with the environments used to train or fit the models (Pearson *et al.* 2006; Elith *et al.* 2010) i.e. that the population is not expanding or contracting (Lasram *et al.* 2010). This assumption is often violated, especially in non-equilibrium settings such as climate change when species distributions are projected into novel environments that are different to those

that were used to train the model (Elith and Leathwick 2009). Another assumption of SDMs is that all relevant environmental gradients have been adequately sampled (Elith and Leathwick 2009) which is often not a realistic assumption, especially when using coarse environmental data which is often the case (Pearson and Dawson 2003; Araújo and Guisan 2006; Elith and Leathwick 2009). Furthermore, the occurrence data used to calibrate the distribution models for *E. andersoni* may also be biased due to differences in sampling methods. The majority of the Mozambique occurrence data was obtained from commercial fishing boats which fish mainly offshore, targeting *Chrysoblephus puniceus* on deep reefs down to about 100 m depth (Fennessy *et al*. 2011), resulting in some occurrence records that are below 75 m depth (Figure 3.6). On the other hand, the majority of the occurrence records in South Africa were obtained from ORI taggers who fished predominantly from the shore, resulting in no occurrence records below the 75 m depth contour south of the Mozambique border.

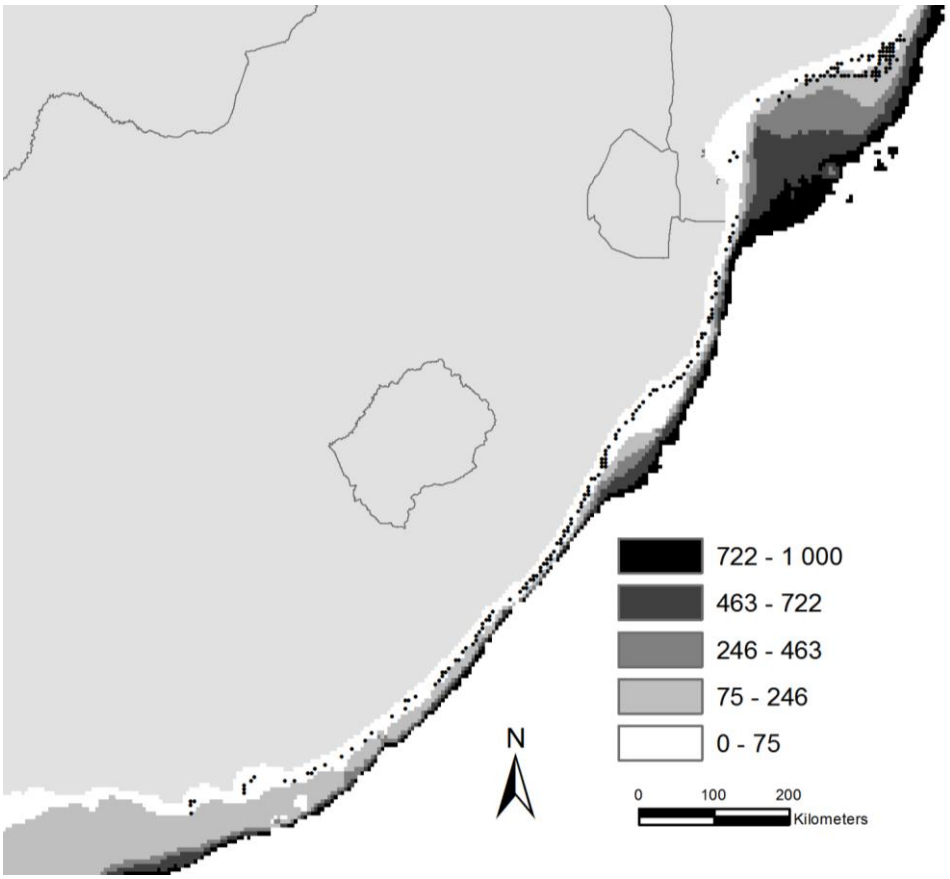


Figure 3.6: Bathymetry (scale in metres below sea level) of the study area with overlaid occurrences.

3.5 Conclusions

Epinephelus andersoni is a non-migratory range-restricted endemic which is exclusive to rocky reef habitats and is classified as near threatened due to extensive fishing pressure and its narrow range

(Fennessy 2004). Contrastingly, there is evidence to suggest that *E. andersoni* may exhibit pioneer behaviour enabling it to take advantage of any available habitats (Maggs 2011), and it is likely that it has a long larval phase possibly enabling widespread dispersal (Heemstra and Randall 1993; Connell 2012). Under circumstances of general ocean temperature increase however, larval mortality is likely to increase (Colton Jr. 1959) and optimal spawning windows are likely to shift (Potts *et al*. in press; Sheaves 2006; Rijnsdorp *et al*. 2009; Pankhurst and Munday 2011) even when temperatures are tolerable to adult fish (Rijnsdorp *et al.* 2009). This may result in compromised larval dispersal which would reduce the ability of *E. andersoni* to shift its range to escape unfavourable temperatures since adult fish are presumably non-migratory. This, coupled with the indication from the model results that the favourable range of *E. andersoni* has already expanded as much as possible, could mean that the species has already used up all its options and cannot expand its range if projected SST trends continue.

The ensemble model results indicate limited range contractions, although these estimates may be conservative. Many aspects of the ecosystem and the species being modelled are not taken into account by correlative SDMs and therefore model results have to be interpreted with due uncertainty. If coastal pockets along the south coast continue to cool as suggested by current trends (Rouault *et al.* 2010), preventing a southward shift of the *E. andersoni* distribution, and if the northern limit of the distribution in Mozambique continues to contract further south, the species may become sandwiched within a dwindling refuge of favourable conditions. The physiology of *E. andersoni* at all life-history stages and its interactions with the surrounding environment need to be resolved and considered in order to determine if model projections are realistic. Determining the source-sink dynamics of the population is vital if more accurate modelling of future distributions of *E. andersoni* is to be achieved using mechanistic models. Although model results of simple correlative models may not be completely reliable, currently available data for *E. andersoni* did not allow for the use of mechanistic models, making the current results the most accurate that could be achieved. Despite model results only representing one of a number of possible future scenarios, the results make sense when reference is made to what is known of the physiology of *E. andersoni*. Future range contraction of the *E. andersoni* distribution is therefore highly plausible if projected SST trends continue.

CHAPTER 4: General Discussion

Rijnsdorp *et al*. (2009) hypothesised that narrow-range endemics with specific habitat requirements and that are under exploitation would be most affected by climate change. *Epinephelus andersoni* has these characteristics (Fennessy 2004), with a restricted extent and area of occurrence (Fennessy 2004), and therefore this species can be classified as a conservation concern in the face of future projected climate change. Conservation organisations globally have recognised the threat of future climate change and variability and are increasingly developing conservation strategies that take climate change into account, and that aim to conserve biodiversity within a changing climate (Hannah *et al.* 2002; Willis and Birks 2006). This study is therefore timely, as it begins to address the issue of long-term conservation for a species that is likely to be negatively impacted by climate change.

4.1 Will genetic diversity be lost with future climate change?

A multidisciplinary approach using genetic data and species distribution models (SDMs) is becoming increasingly useful in assessing the impacts of future climate change on the genetic diversity of populations. Such integrated studies have provided useful information for the conservation and management of species. For example Weaver *et al*. (2006) used SDMs in conjunction with phylogenetic data to speculate on the biogeographical forces that influenced the observed patterns of genetic diversity and the observed distribution of the Black Hills mountain snails *Oreohelix cooperi* of Wyoming and provided useful information for their conservation and management. Similarly, Alsos *et al*. (2009), using genetic data and future climate change models, predicted a 5% loss of the genetic diversity of the dwarf willow *Salix herbacea*, due to range contractions. The glacial history of this species (using fossil data and model predictions of the Last Glacial Maximum) also provided information on how *S. herbacea* had dealt with past climatic changes. This information was useful in determining the possible reactions of the species to future projected climate changes (Alsos *et al.* 2009).

The modelling results of this study similarly indicate a range contraction in the distribution of *E. andersoni* if current changes in sea surface temperature continue into the future. These contractions are predicted to mainly occur at the range limits, notably the southern and northern range limits. Although there was a lot of mixing throughout the species distribution (approximately 39% of the sampled diversity was mixed among localities), both range limit areas also contained range-restricted genetic diversity, indicating that some degree of isolation exists at both range limits.

Occurrence records used for the SDM included localities that were much further south of Port Alfred and north of Xai Xai, which were the southern and northern limits of the genetic data (see Figure 4.1), i.e. it is possible that unique populations, that were not represented in the genetic samples, exist even closer to the range limits. This may be unlikely, however, considering the general declines in abundance of species towards their range limits (Bahn *et al*. 2006). The predicted range contractions did not indicate a complete loss of *E. andersoni* at any of the localities where genetic samples were collected, precluding the making of precise inferences about the degree of genetic diversity that might be lost. However, it is fair to assume that if these range limits continue to contract at the projected rate, there is the potential for unique alleles within the populations near these range limits to be lost.

Although the cytochrome *b* data showed little restricted genetic diversity toward and at the northern and southern range limits of *E. andersoni*; the RPS7-1 region indicated that private alleles were present close to the northern range limit at Quissico (1 private allele), Xai Xai (5 private alleles) and Inhaca (6 private alleles) (Table 4.1). The *E. andersoni* distribution projected for 2030 indicates that the northern range limit will contract from Inhambane to close to Quissico (Figure 4.1). This equates to a stretch of distribution approximately 80 km long being lost. If further contraction occurs at the same rate, it is likely that the northern range limit will contract to Xai Xai by 2040 or 2050. If this occurs, the loss of the private alleles present in samples from Quissico would equate to 2.6% of the genetic diversity of *E. andersoni* being lost. If further contraction occurs at the projected rate, in the long-term, the northern range limit could shrink to south of Xai Xai and result in a further significant loss of up to 15.4% of the genetic diversity.

The vulnerability of northern range limit populations is exacerbated by the fact that the fishery in Mozambique is essentially unrestricted (van der Elst *et al*. 2003). Compounding the gravity of these findings, is the fact that studies have observed that rear range limits (the low-latitude species limit), being the northern range limit in this study, are of critical importance as they act as a store of genetic diversity and are foci of speciation (Hampe and Petit 2005). This is particularly true where these populations have originated from historical remnants of the species existing in refugia habitats that existed during the Last Glacial Maximum. These populations could consequently have been the source which gave rise to much of the current distribution of these species (Hampe and Petit 2005). This may be true for *E. andersoni* considering the oceanographic conditions within its range that would promote northern spawning populations to supply populations further south, since larvae are presumably carried southward in association with the southward-flowing eastern boundary Agulhas Current and Mozambique Channel eddies (Hutchings *et al.* 2002; Lutjeharms 2006).

Table 4.1: The private haplotypes and alleles (*Hp*), and haplotype diversities (*Hd*) for the cytochrome *b* and RPS7-1 regions respectively, for each locality. The total number of haplotypes (*H*) for each gene region (bottom of *Hp* columns) and overall *Hd* is also shown.

	RPS7-1		Cytochrome *b*	
Locality	*Hp*	*Hd*	*Hp*	*Hd*
Quissico	1	0.819	0	0.362
Xai Xai	5	0.322	0	0.174
Maputo	0	0.763	0	0.308
Inhaca	6	0.492	0	0.267
Cape Vidal	4	0.875	2	0.271
Richards Bay	1	0.844	1	0.414
Port Edward	0	0.842	0	0.345
Msikaba	2	0.883	0	0.507
Port St Johns	5	0.473	1	0.352
Coffee Bay	0	0.284	0	0.417
Mbashe	3	0.787	0	0.271
Port Alfred	0	0.810	0	0.345
Overall	*12*	*0.801*	*39*	*0.309*

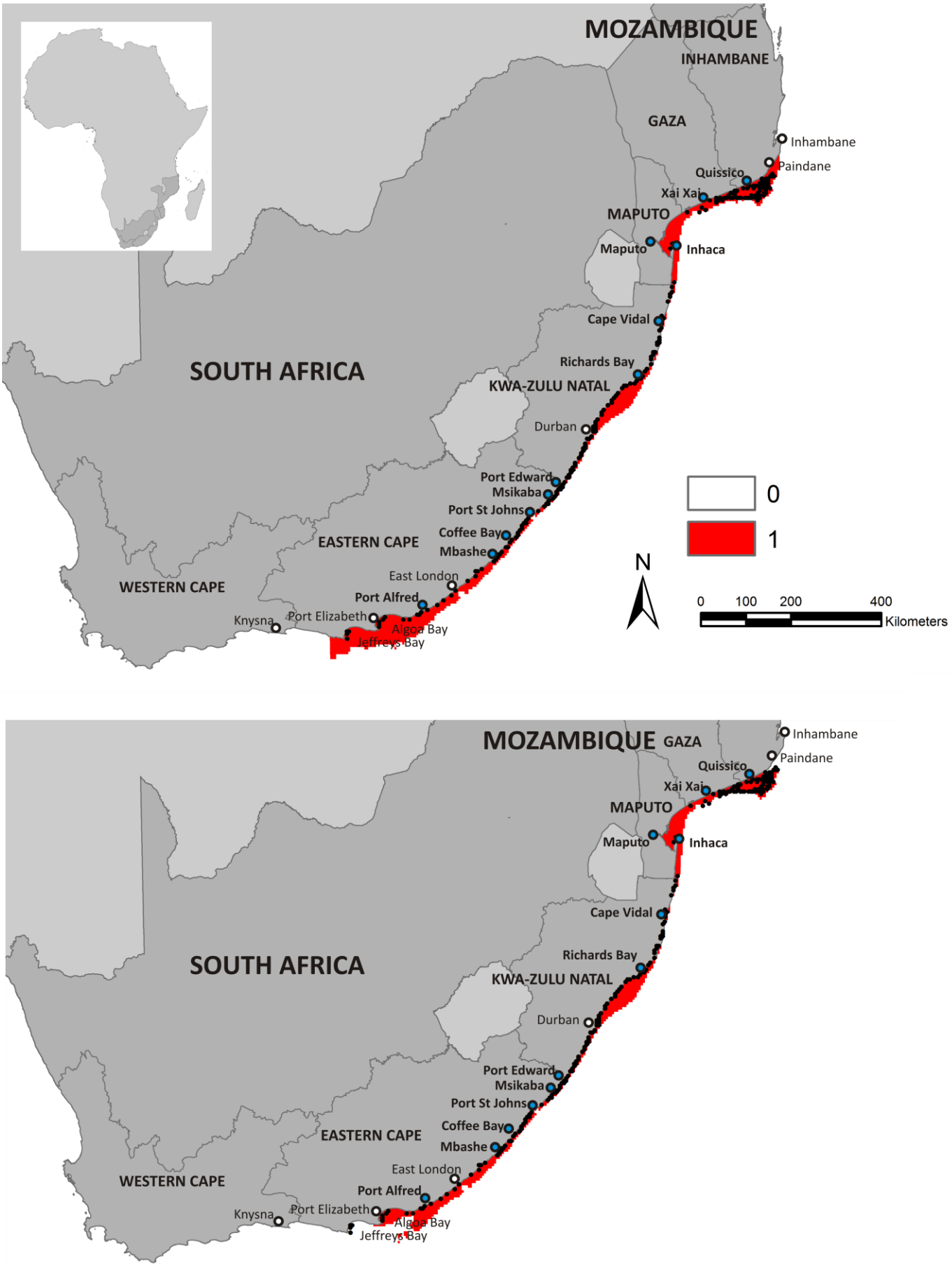


Figure 4.1: Binary current (top) and 2030 (bottom) projected distributions for *Epinephelus andersoni*. Occurrence records are shown (black dots) and localities that were sampled for genetic data are indicated (blue-filled dots, bolded names).

Marine fish populations were originally thought to be characterised by relatively high connectivity (Hauser and Carvalho 2008), as fishes are generally thought to be able to migrate to escape unfavourable conditions (Ebersole *et al.* 2001; Rijnsdorp *et al.* 2009). However, there is increasing evidence that isolation does occur in marine populations (Cowen *et al.* 2007; Hauser and Carvalho 2008). The fact that private or isolated alleles were found in certain localities within the *E. andersoni* population, suggests that this species may be isolated in these localities due to various environmental factors. These factors appear to be similar to the life-history factors that have been found to cause genetic structuring in other groupers (e.g. De Innocentiis *et al*. 2001; Carlin *et al*. 2003; Rivera *et al*. 2004; Antoro *et al*. 2006; Chen *et al*. 2008; Silva-oliveira *et al*. 2008; Buchholz-sørensen and Vella 2010; Wang *et al*. 2011).

Firstly, *E. andersoni* has been shown to be a relatively sedentary species and the majority of the population does not appear to migrate (Dunlop and Mann 2012a), which may play a role in the isolation among populations. Secondly, larvae are more fragile and susceptible to unfavourable environmental conditions and they are also less capable of avoiding them (Potts *et al*. in press; Pörtner and Farrell 2008; Rijnsdorp *et al*. 2009; Pankhurst and Munday 2011). Therefore, larval mortality is likely to increase with projected climate change (Berlinsky *et al.* 2004; Pörtner and Farrell 2008; Rijnsdorp *et al.* 2009; Pankhurst and Munday 2011), which would result in reduced larval dispersal (Hoegh-Guldberg and Bruno 2010) and, in turn, further decrease connectivity between genetically distinct populations (Cowen *et al.* 2007) (such as those on the northern and southern range limits). The projected intensification of upwelling cells and the Agulhas Current (Rouault *et al.* 2010) could potentially augment this fragmentation, which will eventually reduce the resilience of affected populations (Jones *et al.* 2007) making them less able to deal with further climate-induced changes (Brander 2007). Despite some differentiation being present within the *E. andersoni* stock, there is no clear separation of Mozambique from South African localities. The sustainability of the shared *E. andersoni* stock, therefore relies to some extent, on the co-operation between South African and Mozambican linefish managers (Lamberth *et al*. 2009)

4.2 Management and implications for fisheries

Past management measures for *E. andersoni* in South Africa had no biological basis (Fennessy and Sadovy 2002): 300 mm TL size limit since 1974 (Fennessy 2000); 400 mm TL size limit since 1991 (Fennessy and Sadovy 2002); and five fish per person per day bag limit only for recreational fishers since 1984 (Fennessy 2000). Considering that the species is currently likely to be exploited beyond optimal levels (Fennessy and Mann 2013), these past catch restrictions for the species may not have

been sufficient to conserve the stock. This seems particularly likely when one considers that, with a size limit of 400 mm TL (which is smaller than the length at 50% maturity for females), the age at first capture was 2.5 years (Fennessy 1998), which is younger than both the male (2.7 years) and female (3 years) ages at first maturity (Fennessy and Sadovy 2002). Catch restrictions for *E. andersoni* have been amended to a bag limit of five for both recreational and commercial sectors and a size limit of 500 mm (larger than male and female size at 50% maturity) has been implemented for all fishery sectors (Fennessy 2004; Dunlop 2011). Commercial fishing effort has also been generally reduced and restricted by limiting the number of permits (Dunlop 2011). A current stock assessment should be conducted for *E. andersoni* to determine whether the amended regulations are allowing the stock to regenerate.

Although a number of fisheries regulations are in place in South Africa, and fishers are relatively knowledgeable of regulations, compliance is poor since there is insufficient enforcement and monitoring (Fennessy *et al.* 2003; Dunlop 2011). Under-reporting also occurs in some areas in South Africa, mainly due to fishers attempting to avoid limited catches and higher taxes (Fennessy *et al.* 2003). Furthermore, the management of South African fisheries is difficult because it is not managed on a quota basis and instead the commercial fisheries are controlled by limiting the size of the fishing force (Griffiths *et al.* 1999). A suite of species-specific bag limits, size limits and closed seasons are also used for both the recreational and commercial sectors which are relatively difficult to enforce (Griffiths *et al.* 1999). The deterioration of the linefishery due to poor management and over-exploitation has resulted in socio-economic hardships for those who depend on this fishery for food security and employment opportunities (McGrath *et al.* 1997). Due to the importance of linefisheries on the east coast of South Africa, particularly in terms of the number of livelihoods they support, sustainable management of these resources is imperative (Britz *et al*. 2001; Lamberth *et al*. 2009). Serranids form a particularly economically important part of the KZN linefishery, with a value of R5 million and they contribute 9% to the total value of the KZN recreational and commercial boat-based linefishery (Lamberth *et al*. 2009).

The linefishery in Mozambique is essentially unregulated (Fennessy *et al*. 2011) and there is no species-specific management since the linefishery targets multiple species (van der Elst *et al*. 2003). The commercial fishery is only controlled through permit issue which is not effective at limiting effort (van der Elst *et al*. 2003). Artisanal catches are particularly unmanaged and consist mostly of fish that would be considered of sub-legal size for *E. andersoni* within the South African fishery - 91% of all *E. andersoni* sampled at artisanal landing sites for this study were smaller than 500 mm TL. The recreational skiboat fishery is also likely to have a substantial impact on linefish resources but is not

subject to many restrictions; recreational fishers exploit many of the same species as the commercial fishery and have similar seagoing capabilities (Fennessy *et al*. 2011). A recent assessment (Fennessy *et al*. 2011) indicated that the linefishery in southern Mozambique, which includes *E. andersoni* in its catches, is over-exploited. These catch trends indicate that most life stages of *E. andersoni* currently experience exploitation in Mozambique.

Numerous sources have recommended marine protected areas (MPAs) as a method of conserving the over-exploited linefish resources of Mozambique (e.g. Pereira 2000, 2001) and South Africa (e.g. von der Heyden 2009; Teske *et al*. 2010; Dunlop and Mann 2012b). Specifically for *E. andersoni*, Fennessy and Mann (2013) recommended that an MPA should be declared north of Maputo in Mozambique which would protect the species at the northern limit of its range. It is thought that MPAs are easier to enforce and monitor, and are successful in rebuilding damaged stocks, provided that sufficient large-scale marine spatial planning is done and all stakeholders are included in the implementation process (Fennessy *et al.* 2003; Dunlop 2011). Population genetics is also an important tool for marine conservation planning and needs to be considered when planning MPAs (von der Heyden 2009). Using population genetics, the existing networks of MPAs can be assessed, levels of connectivity between MPAs can be determined and genetically important populations and regions can be identified (von der Heyden 2009).

Currently, *E. andersoni* receives protection in the St Lucia and Maputaland MPAs in a 155 km long and 5.6 km wide MPA in northern KZN (Figure 4.2). *Epinephelus andersoni* are caught commonly in the St Lucia MPA (Pradervand *et al*. 2007), however the only *Epinephelus* species that has been recorded in the Maputaland MPA in high numbers is *E. rivulatus* (Fennessy 2000). The Pondoland MPA in Transkei provides protection in an 80 km stretch of coastline which includes a 40 × 10 km no-take area. Maggs (2011) found *E. andersoni* to be more abundant in numbers and biomass in the adjacent exploited areas, although the mean size of fish in the exploited area decreased between 2006 and 2010. The higher abundance in exploited areas leads to the assumption that *E. andersoni* is resilient to fishing pressure and is able to colonise habitat that has been vacated by other predatory fishes that are targeted by fishers (Maggs 2011). However, the overall reduction in mean size suggests that *E. andersoni* may not be resilient to the effects of exploitation in the long-term if fish cannot reach sexual maturity. *Epinephelus andersoni* was also reported to undergo ranging movements that made them available to exploitation outside of protected areas, but it was assumed that the no-take area would provide temporary protection at some point during these movements (Maggs 2011).

The suitability of other South African and Mozambican MPAs for protecting *E. andersoni* has not been assessed (see Figure 4.2 for South African MPAs and Figure 4.3 for Mozambican MPAs). However, the cytochrome *b* data suggested that Cape Vidal was relatively isolated from all other localities. This could indicate that there would be low connectivity between populations within the Maputaland and St Lucia MPAs which encompass Cape Vidal, and other MPAs, limiting the usefulness of these two reserves for conserving *E. andersoni*. However, the spatial autocorrelation results of the RPS7-1 data indicate some level of connectivity between the northern localities in Mozambique (Quissico, Xai Xai and Maputo) and southern localities in the Eastern Cape (Mbashe and Port Alfred). This provides support for the proposed declaration of an MPA north of Maputo as it may further benefit populations in Mbashe and Port Alfred. There is a focus on implementing MPAs and conservation areas in Mozambique, and several conservation areas have recently been declared in Mozambique: the Maputo Special Reserve (established in 2013), Primeiras and Segundas MPA (established in 2013), and Quirimbas National Park (established in 2002). This indicates that an MPA north of Maputo could be effectively implemented and supported.

Although Mozambican MPAs are large, there are few that are effective for conserving *E. andersoni* since the majority of the MPAs are north of the northern range limit of the species (Figure 4.3). The successful planning of MPAs within the *E. andersoni* range will need to consider the habitat preferences (predominantly rocky reefs) of *E. andersoni* and the source-sink dynamics of the population (Crowder *et al.* 2000), as well as the patterns of larval dispersal and recruitment. Larval duration and thermal tolerance limits will also inform marine conservation planning, especially in terms of current climate change. Ensuring that source rather than sink populations receive protection is particularly important since the placement of MPAs may simply displace fishing effort rather than reduce it (if restrictions on fishing effort are not increased) (Crowder *et al*. 2000; Hilborn *et al*. 2004). If MPAs protect sink populations this may then intensify pressure on surrounding source populations which are the lifeblood of the stock (Crowder *et al*. 2000; Hilborn *et al*. 2004). Determining whether spawning is triggered by certain temperature regimes could also benefit future management, allowing more careful protection in the future if predicted trends in temperature continue and cause spawning seasons to change in time and space.

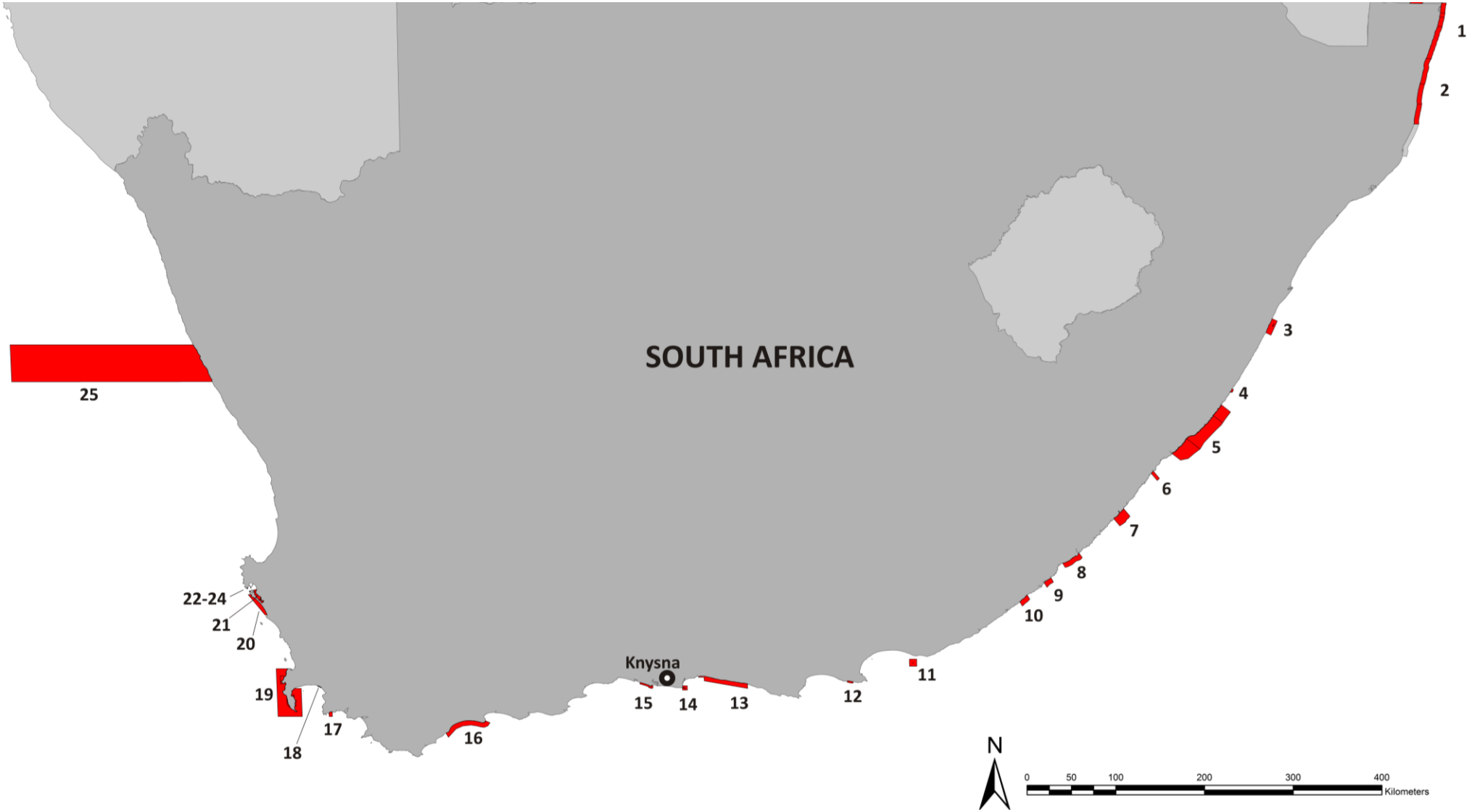


Figure 4.2: South African MPAs are shown in red. Knysna, the accepted southern *E. andersoni* distribution limit is shown. MPAs shown include: 1 Maputaland sanctuary/restricted areas; 2 St Lucia sanctuary/restricted areas; 3 Aliwal Shoal controlled area; 4 Trafalgar MPA; 5 Pondoland controlled/restricted areas; 6 Hluleka MPA; 7 Dwesa-Cwebe MPA; 8 Nyara River to Great Kei River; 9 Nahoon Point to Gonubie Point; 10 Christmas Rock to Gxulu River; 11 Bird Island MPA; 12 Sardinia Bay MPA; 13 Tsitsikamma MPA; 14 Robberg MPA; 15 Goukamma MPA; 16 De Hoop MPA; 17 Betty's Bay MPA; 18 Helderberg MPA; 19 Table Mountain National Park MPA; 20 Sixteen Mile Beach MPA; 21 Langebaan Lagoon MPA; 22 Jutten Island MPA; 23 Malgas Island MPA; 24 Marcus Island MPA; 25 Namaqualand MPA.

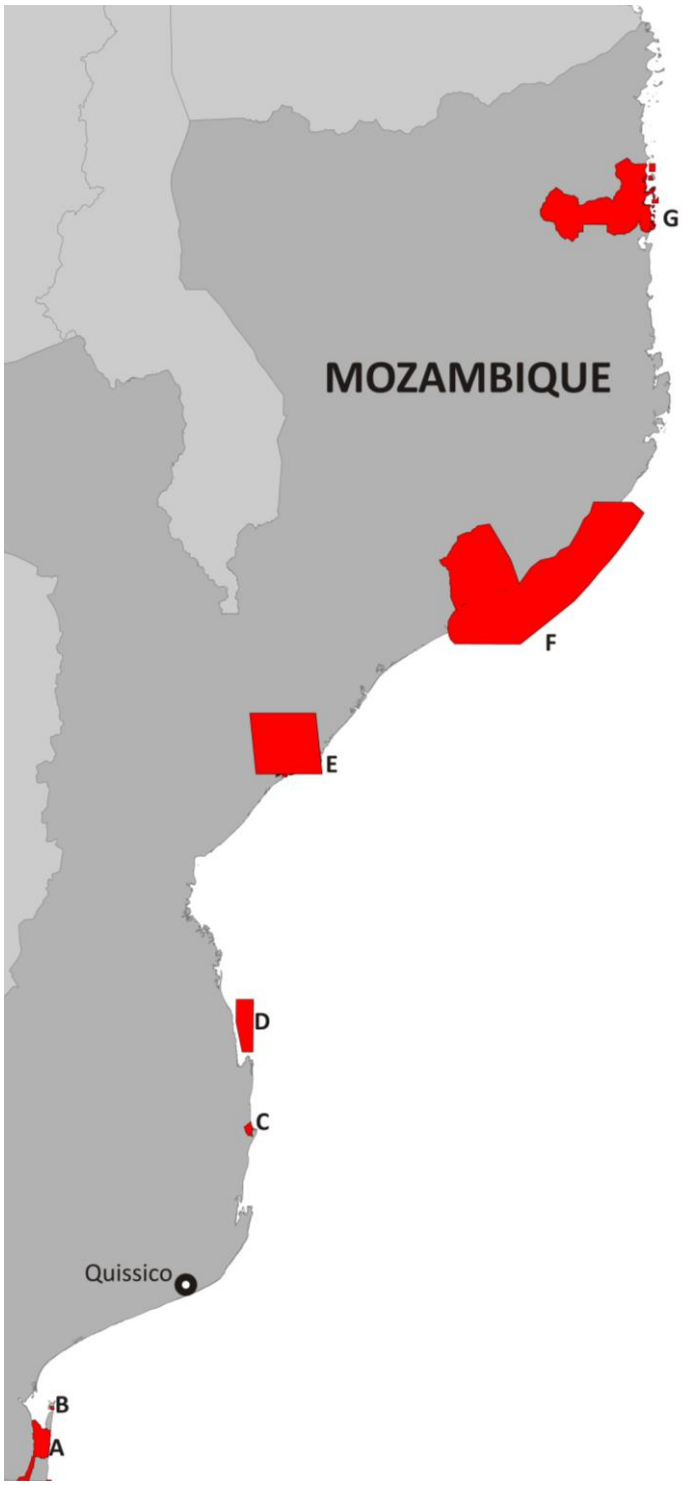


Figure 4.3: MPAs are shown in red. Quissico, the accepted northern *E. andersoni* distribution limit is shown. Mozambique MPAs shown include: A) Maputo Special; B) Ilhas da Inhaca e dos Portuguesas; C) Pomene; D) Bazaruto; E) Marromeu Complex; F) Primeiras and Segundas; G) Quirimbas.

Marine protected areas (MPAs) can provide sufficient protection for largely sedentary species that are less likely to move beyond the borders of suitable reserves and thus get caught be fishers (Hilborn *et al.* 2004). This indicates that MPAs can be an effective method of conserving *E. andersoni* since it is a relatively sedentary species. MPAs can also be more cost effective than stock assessments and catch regulations (Hilborn *et al.* 2004), and are more likely to benefit multispecies fisheries (Hilborn *et al.* 2004). Despite the benefits, MPAs are likely to be more effective in conjunction with the implementation of catch regulations, especially when fishing pressure is merely displaced by MPAs and not reduced (Hilborn *et al.* 2006). The Mozambique fishery would therefore benefit more noticeably from the implementation of a combination of MPAs and management aimed at reducing fishing effort.

A decline in the abundance of this species and others likely to be affected by climate change could compromise the fishery and the livelihoods of local fishers (van der Elst *et al.* 2005). Good science is needed if Western Indian Ocean fisheries are to be managed sustainably; resourceful enforcement mechanisms also need to be applied, especially to the recreational and artisanal fisheries in the region which are rarely subject to management (van der Elst *et al*. 2005). Fisheries management is also inconsistent in the Western Indian Ocean, with fishery sectors even being defined differently in neighbouring countries (van der Elst *et al.* 2005). Since there is no clear stock separation between Mozambique and South African populations of *E. andersoni*, catch regulations should ideally be consistent throughout the entire range of this species (i.e. in Mozambique and South Africa) in order for these regulations to be fully effective (Gulland 1980).

4.3 Further study

Although the information from the current study provides important information on the species, there is a need for more comprehensive biological data to enable the development of conservation measures that will withstand future climatic changes. Such comprehensive information may also assist in the conservation of other similarly-distributed species. To address these knowledge gaps, further research is required in the following areas:

- Increased understanding of the biology, ecology and life-history of *E. andersoni* in terms of its spawning habits, larval dispersal (as well as length of the larval phase), larval and adult thermal tolerance limits, and migration habits. More clarity is also needed on the reasons for the pioneer behaviour of *E. andersoni* that was suggested by Maggs (2011), since this would

aid in determining their resilience to future climate change and would inform the implementation of MPAs.

- Spawning behaviour should also be examined, as it may be of particular importance for the species' conservation: there is anecdotal evidence that *E. andersoni* aggregates to spawn (Maggs 2011). If this is the case, aggregations need to be identified and protected and a closed season enforced during the spawning season in KZN where spawning occurs regularly. Further study could also clarify which spawning populations contribute to genetic diversity, so as to ensure that these receive protection in MPAs, especially where these spawning sites coincide with sites of unique genetic diversity.
- Use phylogeography to uncover the origin of the current observed patterns of genetic diversity of *E. andersoni* (e.g. possibly confinement in refugia during the Last Glacial Maximum) to aid in determining the impact of future projected climate change. This would assist managers in determining where best to focus conservation efforts, by identifying possible regugia.
- Increasing genetic sample sizes in KZN to enable a more comprehensive genetic study of the *E. andersoni* stock would also provide more information on areas of important genetic diversity and connectivity patterns within the species' distribution. There were no genetic samples collected between Port Edward and Richards Bay and sampling sites between Richards Bay and Inhaca were widely spaced and so adding sample locations between Port Edward and Inhaca may increase the resolution of the genetic data.
- Use paleoecological records (e.g. fossils) to determine past distributions of *E. andersoni* or other marine fishes that have such data available, and then use SDMs to project past distributions over the same time period, so that the models used in this study can be validated. This would assist in determining whether the model projections of the future distributions of marine fishes such as *E. andersoni* are accurate, which in turn would benefit long-term management plans for marine species. Paleoecology can contribute to conservation by addressing questions such as what constitutes natural variability and can provide an historical perspective to the distributions of species at the present time and in the future (Willis and Birks 2006).
- Determine the detailed colonisation and demographic history of *E. andersoni* which could reveal more about the species' possible reactions to future climatic changes by correlating the historical population fluctuations with past environmental changes (e.g. Alsos *et al.* 2009).
- Given the findings of this study, an MPA encompassing the Xai Xai *E. andersoni* population of would be useful for the conservation of the species. Xai Xai is relatively distant from the

contracting northern range limit of the species and the Xai Xai *E. andersoni* population contains a reasonable amount of genetic diversity, indicated by a high number of private haplotypes. The genetic data also indicates that there is connectivity between Xai Xai *E. andersoni* population and southern populations in the Eastern Cape.

4.4 Conclusions

The findings of this study in terms of the population genetics of the stock indicate low mitochondrial diversity and minimal geographic structuring of the population, which could be attributed to a past bottleneck event in the species' demographic history (possibly caused by the Last Glacial Maximum). Although a large amount of genetic mixing was evident from both the nuclear and mitochondrial DNA datasets, the faster-evolving nuclear data suggested some strong genetic structure within the population which could be due to population expansion following the historical founder event. The SDMs suggest that the range of *E. andersoni* will contract and that populations near the northern and southern range limits in particular will be threatened if the observed climate trends over the past 30 years continue. Some unique genetic diversity of the species is restricted to the northern populations specifically, which means that this diversity could potentially be threatened by the predicted range contractions at the northern range limit. Fisheries management of *E. andersoni* and other range-restricted endemics will need to become increasingly conservative to allow for the effects of climate change on fish stocks, as research suggests that the exploitation of fish stocks will compromise their genetic resilience and their ability to deal with environmental changes (e.g. Brander 2007; Rijnsdorp *et al.* 2009).

The *E. andersoni* stock should be managed as a single stock since there is mixing among populations, but conservation measures should ideally be increased in areas where the population contains high numbers of private haplotypes and high genetic diversity (i.e. Quissico, Xai Xai and Inhaca in Mozambique; Cape Vidal, Richards Bay, Msikaba, Port St Johns, and Mbashe in South Africa). Important populations to protect are the Cape Vidal, Richards Bay, and Port St Johns populations since there was a high number of private haplotypes present in both the RPS7-1 and cytochrome *b* data. The northern populations (Quissico, Xai Xai and Inhaca) are also conservation priorities since RPS7-1 data identified relatively high numbers of private haplotypes among these localities, and low-latitude species' limits are likely to act as critical foci of speciation (Hampe and Petit 2005). Furthermore, spawning sites that are unlikely to be under threat in future climate change scenarios should be a conservation priority since these populations are more likely to provide long-term protection for the species.

Although it is impractical for fisheries management to focus on a single species, MPAs that protect the priority Mozambican and KZN *E. andersoni* populations, could be justified by the fact that species diversity is high in this tropical region and therefore MPAs sited in these areas could provide protection for a number of other species (Turpie *et al*. 2000). Furthermore, priority *E. andersoni* populations located within the Transkei will invariably provide protection for a number of other endemic species since this region is a centre of marine fish endemism (Turpie *et al*. 2000). Multispecies assessments would need to be conducted before an MPA is sited, to ensure that its' location reaps the maximum benefit to a range of species that is as large as possible (Turpie *et al*. 2000).

References

African Marine Atlas (2006) omap.africanmarineatlas.org/BASE/pages/relief_grids.htm (accessed 3 Sept 2012).

Allendorf, F., Luikart, G. and Aitken, S.N. (2013) *Conservation and the genetics of populations*, 2nd edn. Wiley-Blackwell, West Sussex pp 609.

Allouche, O., Tsoar, A. and Kadmon, R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* **43**, 1223–1232.

Alsos, I.G., Alm, T., Normand, S. and Brochmann, C. (2009) Past and future range shifts and loss of diversity in dwarf willow (*Salix herbacea* L.) inferred from genetics, fossils and modelling. *Global Ecology and Biogeography* **18**, 223–239.

Antoro, S., Na-Nakorn, U. and Koedprang, W. (2006) Study of genetic diversity of orange-spotted grouper, *Epinephelus coioides*, from Thailand and Indonesia using microsatellite markers. *Marine Biotechnology* **8**, 17–26.

Araújo, M.B. and Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography* **33**, 1677–1688.

Araújo, M.B. and New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution* **22**, 42–47.

Araújo, M.B., Whittaker, R.J., Ladle, R.J. and Erhard, M. (2005) Reducing uncertainty in projections of extinction risk from climate change. *Global Ecology and Biogeography* **14**, 529–538.

Austin, M. (2007) Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling* **200**, 1–19.

Ayre, D.J. and Hughes, T.P. (2004) Climate change, genotypic diversity and gene flow in reef-building corals. *Ecology Letters* **7**, 273–278.

Bahn, V., O'Connor, R.J. and Krohn, W.B. (2006) Effect of dispersal at range edges on the structure of species ranges. *Oikos* **115**, 89–96.

Bakun, A. and Weeks, S.J. (2004) Greenhouse gas buildup, sardines, submarine eruptions and the possibility of abrupt degradation of intense marine upwelling ecosystems. *Ecology Letters* **7**, 1015–1023.

Bandeira, S.O., Silva, R.P. De, Paula, J., Macia, A., Hernroth, L., Guissamulo, A.T. and Gove, D.Z. (2002) Marine biological research in Mozambique: past, present and future. *Ambio* **31**, 606–609.

Bandelt, H.J., Forster, P. and Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37–48.

Banks, S.C., Piggott, L.M.P., Williamson, J.E., Bové, U., Holbrook, N.J. and Beheregaray, L.B. (2007) Oceanic variability and coastal topography shape genetic structure in a long-dispersing sea urchin. *Ecology* **88**, 3055–64.

Barbet-Massin, M., Jiguet, F., Albert, C.H. and Thuiller, W. (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* **3**, 327–338.

Beaumont, A. and Hoare, K. (2003) Genetic structure in natural populations. In: *Biotechnology and genetics in fisheries and aquaculture*. (eds A. Beaumont and K. Hoare). Blackwell Science Ltd., Cornwall, pp 47–72.

Beger, M. and Possingham, H. (2008) Environmental factors that influence the distribution of coral reef fishes: modeling occurrence data for broad-scale conservation and management. *Marine Ecology Progress Series* **361**, 1–13.

Bennett, R.H. (2012) *Movement patterns, stock delineation and conservation of an overexploited fishery species,* Lithognathus lithognathus *(Pisces: Sparidae)*. PhD thesis, Rhodes University, 374 pp.

Berlinsky, D.L., Taylor, J.C., Howell, R.A. and Bradley, T.M. (2004) The effects of temperature and salinity on early life stages of black sea bass *Centropristis striata*. *Journal of the World Aquaculture Society* **35**, 335–344.

Bohonak, A.J. (1999) Dispersal, gene flow, and population structure. *The Quarterly Review of Biology* **74**, 21–45.

Boitani, L., Sinibaldi, I., Corsi, F., De Biase, A., Carranza, I. d'I., Ravagli, M., Reggiani, G., Rondinini, C., Trapanese, P. (2008) Distribution of medium- to large-sized African mammals based on habitat suitability models. *Biodiversity and Conservation* **17**, 605–621.

Brander, K.M. (2007) Global fish production and climate change. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 19709–19714.

Brassel, K.E. and Reif, D. (1979) A procedure to generate Thiessen Polygons. *Geographical Analysis* **11**, 31–36.

Breiman, L. (2001) Random forests. *Machine Learning* **45**, 5–32.

British Oceanographic Data Centre (2003) Centenary Edition of the GEBCO Digital Atlas [CDROM] *Published on behalf of the Intergovernmental Oceanographic Commission and the International Hydrographic Organization as part of the General Bathymetric Chart of the Oceans*. pp 171.

Britz, P.J., Sauer, W.H.H., Mather, D., Oellerman, L., Cowley, P.D., Ter Morshuizen, L. and Bacela, N. (2001) Baseline study of the utilisation of living marine resources in the Eastern Cape province. *Prepared for the Department of Econmonic Affairs, Environment and Tourism* pp 103.

Broennimann, O., Thuiller, W., Hughes, G., Midgley, G.F., Alkemade, J.M.R. and Guisan, A. (2006) Do geographic distribution, niche property and life form explain plants' vulnerability to global change? *Global Change Biology* **12**, 1079–1093.

Brouns, G., De Wulf, A. and Constales, D. (2003) Delaunay triangulation algorithms useful for multibeam echosounding. *Journal of Surveying Engineering* **129**, 79–84.

Brouwer, S.L. and Buxton, C.D. (2002) Catch and effort of the shore and skiboat linefisheries along the South African Eastern Cape coast. *South African Journal of Marine Science* **24**, 341–354.

Brown, J.H. (1995) *Macroecology.* University of Chicago Press, Chicago pp 268.

Brown, W.M., George, M. and Wilson, A.C. (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 1967–1971.

Buchholz-sørensen, M. and Vella, A. (2010) Preliminary population genetic structure of the endangered dusky grouper, *Epinephelus marginatus*, in Malta as revealed by six microsatellite markers. *Rapport Commission International pour l'exploration scientifique de la Mer Mediterranee* **39**, 463.

Buehler, D.M. and Baker, A.J. (2005) Population divergence times and historical demography in red knots and dunlins. *The Condor* **107**, 497–513.

Caccone, A., Gentile, G., Burns, C.E., Sezzi, E., Bergman, W., Ruelle, M., Saltonstall, K., and Powell, J. R. (2004) Extreme difference in rate of mitochondrial and nuclear DNA evolution in a large ectotherm, Galápagos tortoises. *Molecular phylogenetics and evolution* **31**, 794–8.

Cantatore, P., Roberti, M., Pesole, G., Ludovico, A., Milella, F., Gadaleta, M.N. and Saccone, C. (1994) Evolutionary analysis of cytochrome b sequences in some Perciformes : evidence for a slower rate of evolution than in mammals. *Journal of Molecular Evolution* **39**, 589–597.

Capinha, C. and Anastácio, P. (2011) Assessing the environmental requirements of invaders using ensembles of distribution models. *Diversity and Distributions* **17**, 13–24.

Carlin, J.L., Robertson, D.R. and Bowen, B.W. (2003) Ancient divergences and recent connections in two tropical Atlantic reef fishes *Epinephelus adscensionis* and *Rypticus saponaceous* (Percoidei: Serranidae). *Marine Biology* **143**, 1057–1069.

Chen, S., Liu, T., Li, Z. and Gao, T. (2008) Genetic population structuring and demographic history of red spotted grouper (*Epinephelus akaara*) in South and East China Sea. *African Journal of Biotechnology* **7**, 3554–3562.

Cheung, W.W.L., Lam, V.W.Y., Sarmiento, J.L., Kearney, K., Watson, R. and Pauly, D. (2009) Projecting global marine biodiversity impacts under climate change scenarios. *Fish and Fisheries* **10**, 235–251.

Chow, S. and Hazama, K. (1998) Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology* **7**, 1247–1263.

Clark, B.M. (2006) Climate change: a looming challenge for fisheries management in southern Africa. *Marine Policy* **30**, 84–95.

Clark, B.M., Steffani, N.C., Young, S., Richardson, A.J. and Lombard, A.T. (2000) The effects of climate change on marine biodiversity in South Africa. *Report prepared for the Foundation for Research*

Development, South African country study on climate change, vulnerability and adaptation asessment, marine biodiversity section pp 130.

Clark, R.A., Fox, C.J., Viner, D. and Livermore, M. (2003) North Sea cod and climate change – modelling the effects of temperature on population dynamics. *Global Change Biology* **9**, 1669–1680.

Coetzee, B.W.T., Robertson, M.P., Erasmus, B.F.N., van Rensburg, B.J. and Thuiller, W. (2009) Ensemble models predict important bird areas in southern Africa will become less effective for conserving endemic birds under climate change. *Global Ecology and Biogeography* **18**, 701–710.

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **XX**, 37–46.

Colton Jr., J.B. (1959) A field observation of mortality of marine fish larvae due to warming. *Limnology and Oceanography* **4**, 219–222.

Connell, A. (2012) Marine fish eggs and larvae from the east coast of South Africa. Available at: http://www.fisheggsandlarvae.com/LIIIB4 Serranidae.htm (Accessed December 18, 2012).

Cowen, R.K., Gawarkiewicz, G., Pineda, J., Thorrold, S.R. and Werner, F.E. (2007) Population connectivity in marine systems: an overview. *Oceanography* **20**, 14–21.

Cowen, R.K., Paris, C.B. and Srinivasan, A. (2006) Scaling of connectivity in marine populations. *Science* **311**, 522–527.

Craig, M.T., Graham, R.T., Torres, R.A., Hyde, J. R., Freitas, M. O., Ferreira, B. P., Gerhardinger, L. C., Bertoncini, A. A., and Robertson, D. R. (2008) How many species of Goliath grouper are there? Cryptic genetic divergence in a threatened marine fish and the resurrection of a geopolitical species. *Endangered Species Research* **7**, 167–174.

Craig, M.T., Sadovy de Mitcheson, Y.J. and Heemstra, P.C. (2011) *Groupers of the world: a field and market guide.* National Inquiry Service Centre (Pty) Ltd, Grahamstown.

Crowder, L.B., Lyman, S.J., Figueira, W.F. and Priddy, J. (2000) Source-sink population dynamics and the problem of siting marine reserves. *Bulletin of Marine Science* **66**, 799–820.

DAFF (2012) Strategic plan for the 2012/13-2016/17. *Department of Agriculture, Forestry and Fisheries* pp 108.

Dawson, M.N. and Hamner, W.M. (2005) Rapid evolutionary radiation of marine zooplankton in peripheral environments. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 9235–9240.

Dengo, A. and David, B. (1993) The linefishery in southern Mozambique. In: *Fish, fishers and fisheries* (Proceedings of the second South African marine linefish symposium, Durban, 23-24 October 1992). (eds L.E. Beckley and R.P. van der Elst). pp 104–106.

Diniz-filho, J.A.F., Bini, L.M., Rangel, T.F., Loyola, R.D., Hof, C., Nogués-Bravo, D. and Araújo, M.B. (2009) Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography* **32**, 897–906.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kühn, I., Ohlemüller, R., Peres-Neto, P. R., Reineking, B., Schröder, B., Schurr, F. M., and Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* **30**, 609–628.

Dulvy, N.K., Rogers, S.I., Jennings, S., Stelzenmüller, V., Dye, S.R. and Skjoldal, H.R. (2008) Climate change and deepening of the North Sea fish assemblage: a biotic indicator of warming seas. *Journal of Applied Ecology* **45**, 1029–1039.

Dunlop, S.W. (2011) *An assessment of the shore-based and offshore boat-based linefisheries of KwaZulu-Natal, South Africa*. MSc thesis, Univeristy of KwaZulu-Natal, 220 pp.

Dunlop, S.W. and Mann, B.Q. (2012a) ORI Tagging Project: summary of tag and recapture data for catface rockcod (*Epinephelus andersoni*) caught along the Southern African coastline: 1984-2011. *Oceanographic Research Institute* (ORI Data Report 2012/05), Durban: pp 4.

Dunlop, S.W. and Mann, B.Q. (2012b) An assessment of participation, catch and effort in the KwaZulu-Natal shore-based marine linefishery, with comments on management effectiveness. *African Journal of Marine Science* **34**, 479–496.

Dunlop, S.W. and Mann, B.Q. (2013) An assessment of participation, catch and effort in the offshore boat-based linefishery in KwaZulu-Natal, South Africa. *African Journal of Marine Science* **35**, 79–97.

Dupanloup, I., Schneider, S. and Excoffier, L. (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* **11**, 2571–2581.

Eastwood, P.D., Meaden, G.J., Carpentier, A. and Rogers, S.I. (2003) Estimating limits to the spatial extent and suitability of sole (*Solea solea*) nursery grounds in the Dover Strait. *Journal of Sea Research* **50**, 151–165.

Ebersole, J., Liss, W. and Frissell, C. (2001) Relationship between stream temperature, thermal refugia and rainbow trout *Oncorhynchus mykiss* abundance in arid-land streams in the northwestern United States. *Ecology of Freshwater Fish* **10**, 1–10.

Ehlers, A., Worm, B. and Reusch, T.B.H. (2008) Importance of genetic diversity in eelgrass *Zostera marina* for its resilience to global warming. *Marine Ecology Progress Series* **355**, 1–7.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. McC., Peterson, A. T., Phillips, S. J., Richardson, K., Scachetti-pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S., and Zimmermann, N. E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129–151.

Elith, J., Kearney, M. and Phillips, S. (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution* **1**, 330–342.

Elith, J. and Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* **40**, 677–697.

Erasmus, B.F.N., van Jaarsveld, A.S., Chown, S.L., Kshatriya, M. and Wessels, K.J. (2002) Vulnerability of South African animal taxa to climate change. *Global Change Biology* **8**, 679–693.

ESRI (2011) ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

Evans, B., Sweijd, N., Bowie, R., Cook, P. and Elliott, N. (2004) Population genetic structure of the perlemoen *Haliotis midae* in South Africa: evidence of range expansion and founder events. *Marine Ecology Progress Series* **270**, 163–172.

Ewens, W.J. (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**, 87–112.

Excoffier, L. and Lischer, H. (2011) Arlequin version 3.5. An integrated software package for population genetics data analysis. *User Manual*. pp 174.

Excoffier, L. and Lischer, H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564–567.

Excoffier, L., Smouse, P.E. and Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.

Felsenstein, J. (2007) *Theoretical evolutionary genetics*. Department of Genome Sciences and Department of Biology, University of Washington, Washington pp 408.

Fennessy, S.T. (1998) Biology and stock assessment of Serranidae. *Unpublished Report of the Oceanographic Research Institute* **153**, 35–41.

Fennessy, S.T. (2000) *Comparative life histories and stock assessments of rockcods (family Serranidae) from the east coast of South Africa*. PhD thesis, University of Natal, 174 pp.

Fennessy, S.T. (2004) *Epinephelus andersoni*. In: IUCN 2012. IUCN Red List of Threatened Species version 2012.2 (accessed 18 Mar 2013).

Fennessy, S.T. and Mann, B.Q. (2013) Catface rockcod (*Epinephelus andersoni*). In: *Southern African marine linefish species profiles.* (ed B.Q. Mann). Special Publication Oceanographic Research Institute **9**: 191–192.

Fennessy, S.T., McDonald, A.M., Mann, B.Q. and Everett, B.I. (2003) An assessment of the recreational and commercial skiboat fishery in the Transkei. *African Journal of Marine Science* **25**, 61–78.

Fennessy, S.T., Mutombene, R., Simango, A., Cuco, C. and van der Elst, R.P. (2011) Mozambique linefish assessment 2011. *Relatório Interno de Investigação Pesqueira* No 14 pp 18.

Fennessy, S.T., Radebe, P.V. and Mann, B.Q. (2000) Catface rockcod (*Epinephelus andersoni*). In: *Southern African marine linefish status reports*. Mann, B.Q. (Ed.). Special Publication Oceanographic Research Institute **7**: 113-114.

Fennessy, S.T. and Sadovy, Y. (2002) Reproductive biology of a diandric protogynous hermaphrodite, the serranid *Epinephelus andersoni*. *Marine And Freshwater Research* **53**, 147–158.

Fields, P. A., Graham, J.B., Rosenblatt, R.H. and Somero, G.N. (1993) Effects of expected global climate change on marine faunas. *Trends in Ecology & Evolution* **8**, 361–367.

Ford, M.J. (2002) Applications of selective neutrality tests to molecular ecology. *Molecular Ecology* **11**, 1245–1262.

Fourmanoir, P. (1957) Poissons téléostéens des eaux Malgaches du Canal de Mozambique. *Mémoires de l'Institut Scientifique de Madagascar. Série F: Océanographie* **1**, 1 –316. [In French.]

Freeman, E.A. and Moisen, G.G. (2008) A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling* **217**, 48–58.

Friedman, J.H. (2001) Greedy function approximation: a gradient boosting machine. *The Annals of Statistics* **29**, 1189–1232.

Fu, Y.-X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925.

Funk, W.C., Forsman, E.D., Mullins, T.D. and Haig, S.M. (2008) Landscape features shape genetic structure in threatened northern spotted owls. *Geological Survey Open-file Report* 2008-1240 pp 1–11.

Gaffney, P.M. (2000) Molecular tools for understanding population structure in Antarctic species. *Antarctic Science* **12**, 288–296.

Gasse, F. (2000) Hydrological changes in the African tropics since the Last Glacial Maximum. *Quaternary Science Reviews* **19**, 189–211.

Georges, D. and Thuiller, W. (2012) An example of species distribution modelling with biomod2. *R version 2.15.1, biomod2 version 1.0 user manual* pp 23.

Grant, W.S. and Bowen, B.W. (1998) Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *The American Genetic Association* **89**, 415–426.

Griffiths, M., Attwood, C. and Thompson, R. (1999) New protocol for the South African linefishery. In: *South African Network for Coastal and Oceanic Research Occasional Report No. 5* (Proceedings of the third Southern African marine linefish symposium, Arniston, 28 April- 1 May 1999). pp 145–156.

Grinnell, J. (1917) Field tests of theories concerning distributional control. *The American Naturalist* **51**, 115–128.

Gulland, J.A. (1980) Some problems of the management of shared stocks. *FAO Fisheries Technical Papers*.

Hall, T.A. (1999) BioEdit: a user friendly biological sequence alignment editor analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95–98.

Hamilton, L., Lyster, P. and Otterstad, O. (2000) Social change, ecology and climate in 20th-century Greenland. *Climatic Change* **47**, 193–211.

Hampe, A. and Petit, R.J. (2005) Conserving biodiversity under climate change: the rear edge matters. *Ecology letters* **8**, 461–467.

Hanley, J.A. and McNeil, B.J. (1982) The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. *Radiology* **143**, 29–36.

Hannah, L., Midgley, G.F. and Millar, D. (2002) Climate change-integrated conservation strategies. *Global Ecology and Biogeography* **11**, 485–495.

Hare, M.P. and Palumbi, S.R. (2003) High intron sequence conservation across three mammalian orders suggests functional constraints. *Molecular Biology and Evolution* **20**, 969–978.

Harley, C.D.G., Randall Hughes, A., Hultgren, K.M., Miner, B. G., Sorte, C. J. B., Thornber, C. S., Rodriguez, L. F., Tomanek, L., and Williams, S. L. (2006) The impacts of climate change in coastal marine systems. *Ecology letters* **9**, 228–241.

Harpending, H.C. (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* **66**, 591–600.

Hateley, J.G. (2005) Preliminary results of a protein electrophoretic analysis of genetic variation, population structure and gene flow in the Nassau grouper, *Epinephelus striatus*. In: *Proceedings of the Forty Seventh Annual Gulf and Caribbean Fisheries Institute*. Fort Pierce, Florida, pp 888–905.

Hauser, L. and Carvalho, G.R. (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries* **9**, 333–362.

Hedgecock, D. (1994) Does variance in reproductive success limit effective population sizes of marine organisms? In: *Genetics and evolution of aquatic organisms*. (ed A.R. Beaumont), 1st edn. Chapman & Hall, London, pp 122–134.

Hedger, R., McKenzie, E., Heath, M., Wright, P., Scott, B., Gallego, A. and Andrews, J. (2004) Analysis of the spatial distributions of mature cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*) abundance in the North Sea (1980–1999) using generalised additive models. *Fisheries Research* **70**, 17–25.

Heemstra, P.C. and Heemstra, E. (2004) *Coastal fishes of southern Africa.* National Inquiry Service Centre and the South African Institute for Aquatic Biodiversity, Grahamstown.

Heemstra, P.C. and Randall, J.E. (1993) FAO species catalogue. Vol. 16: groupers of the world (family Serranidae, Subfamily Epinephelinae). An annotated and illustrated catalogue of the grouper, rockcod, hind, coral grouper and lyretail species known to date. pp 382.

Heikkinen, R.K., Luoto, M., Araújo, M.B., Virkkala, R., Thuiller, W. and Martin, T. (2006) Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography* **6**, 1–27.

Hellberg, M.E. (2006) No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. *BMC Evolutionary Biology* **6**, 24.

Henriques, R.P.N.L. (2011) *Influence of the Benguela Current in genetic sub-structuring of commercially exploited fish species*. PhD thesis, Royal Holloway University of London, pp 288.

Hiddink, J.G. and Hofstede, R. (2008) Climate induced increases in species richness of marine fishes. *Global Change Biology* **14**, 453–460.

Hijmans, R.J. and Graham, C.H. (2006) The ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biology* **12**, 2272–2281.

Hilborn, R., Micheli, F. and De Leo, G.A. (2006) Integrating marine protected areas with catch regulation. *Canadian Journal of Fisheries and Aquatic Sciences* **63**, 642–649.

Hilborn, R., Stokes, K., Maguire, J.-J., Smith, T., Botsford, L. W., Mangel, M., Orensanz, J., Parma, A., Rice, J., Bell, J., Cochrane, K. L., Garcia, S., Hall, S. J., Kirkwood, G. P., Sainsbury, K., Stefansson, G., and Walters, C. (2004) When can marine reserves improve fisheries management? *Ocean & Coastal Management* **47**, 197–205.

Hoegh-Guldberg, O. and Bruno, J.F. (2010) The impact of climate change on the world's marine ecosystems. *Science* **328**, 1523–1528.

Hoegh-Guldberg, O., Mumby, P.J., Hooten, A.J., Steneck, R. S., Greenfield, P., Gomez, E., Harvell, C. D., Sale, P. F., Edwards, A. J., Caldeira, K., Knowlton, N., Eakin, C. M., Iglesias-Prieto, R., Muthiga, N., Bradbury, R. H., Dubi, A., and Hatziolos, M. E. (2007) Coral reefs under rapid climate change and ocean acidification. *Science* **318**, 1737–1742.

Hogan, J., Thiessen, R. and Heath, D. (2010) Variability in connectivity indicated by chaotic genetic patchiness within and among populations of a marine fish. *Marine Ecology Progress Series* **417**, 263–275.

Hudson, R.R. (1990) Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* **7**, 1–44.

Hutchings, L., Beckley, L.E., Griffiths, M.H., Roberts, M.J., Sundby, S. and van der Lingen, C. (2002) Spawning on the edge: spawning grounds and nursery areas around the southern African coastline. *Marine and Freshwater Research* **53**, 307–318.

De Innocentiis, S., Sola, L., Cataudella, S. and Bentzen, P. (2001) Allozyme and microsatellite loci provide discordant estimates of population differentiation in the endangered dusky grouper (*Epinephelus marginatus*) within the Mediterranean Sea. *Molecular Ecology* **10**, 2163–2175.

Ivetić, V., Isajev, V. and Mladenović Drinić, S. (2008) "Landscape shape interpolation" for defining spatial pattern of beech genetic diversity in Serbia. *Genetika* **40**, 283–292.

Ivetić, V., Isajev, V., Stavretović, N. and Mladenović-Drinić, S. (2010) Implementation of Monmonier's algorithm of maximum differences for the regionalization of forest tree populations as a basis for the selection of seed sources. *Archives of Biological Sciences* **62**, 425–430.

Jacobs, D.K., Haney, T. A. and Louie, K.D. (2004) Genes, diversity, and geologic process on the Pacific coast. *Annual Review of Earth and Planetary Sciences* **32**, 601–652.

James, N.C., Whitfield, A.K. and Cowley, P.D. (2008) Preliminary indications of climate-induced change in a warm-temperate South African estuarine fish community. *Journal of Fish Biology* **72**, 1855–1863.

Johnson, M.S. and Black, R. (1982) Chaotic genetic patchiness in an intertidal limpet, *Siphonaria* sp. *Marine Biology* **70**, 157–164.

Johnson, M.S. and Black, R. (1984) Pattern beneath the chaos: the effect of recruitment on genetic patchiness in an intertidal limpet. *Evolution* **38**, 1371–1383.

Jones, G.P., Srinivasan, M. and Almany, G.R. (2007) Population connectivity and conservation of marine biodiversity. *Oceanography* **20**, 100–111.

Kaschner, K., Watson, R., Trites, A.W. and Pauly, D. (2006) Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. *Marine Ecology Progress Series* **316**, 285–310.

Kearney, M. and Porter, W.P. (2004) Mapping the fundamental niche: physiology, climate, and the distribution of a nocturnal lizard. *Ecology* **85**, 3119–3131.

Keith, D.A., Akçakaya, H.R., Thuiller, W., Midgley, G. F., Pearson, R. G., Phillips, S. J., Regan, H. M., Araújo, M. B., and Rebelo, T. G. (2009) Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. *Biology Letters* **4**, 560–563.

Kenchington, E.L., Harding, G.C., Jones, M.W. and Prodöhl, P. A. (2009) Pleistocene glaciation events shape genetic structure across the range of the American lobster, *Homarus americanus*. *Molecular Ecology* **18**, 1654–67.

Kimura, M. (1968a) Evolutionary rate at the molecular level. *Nature* **217**, 624–626.

Kimura, M. (1968b) Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genetical Research* **11**, 247–269.

Kimura, M. (1983) The neutral mutation-random drift hypothesis as an evolutionary paradigm. In: *The neutral theory of molecular evolution*, 1st edn. Cambridge University Press, New York, pp 366.

Kimura, M. (1969) The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* **61**, 893–903.

Kimura, M. and Crow, J.F. (1964) The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

Kimura, M. and Ohta, T. (1974) On some principles governing molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 2848–2852.

Kingman, J.F.C. (1982a) On the genealogy of large populations. *Journal of Applied Probability* **19**, 27–43.

Kingman, J.F.C. (1982b) The coalescent. *Stochastic Processes and their Applications* **13**, 235–248.

Kocher, T.D. and White, T.J. (1989) Evolutionary analysis via PCR. In: *PCR Technology: Principles and Applications for DNA Amplification* (ed H.A. Erlich). Stockton Press, New York, pp 137–147.

Kumar, S. and Stohlgren, T.J. (2009) Maxent modeling for predicting suitable habitat for threatened and endangered tree *Canacomyrica monticola* in New Caledonia. *Journal of Ecology and Natural Environment* **1**, 094–098.

Kupschus, S. (2003) Development and evaluation of statistical habitat suitability models: an example based on juvenile spotted seatrout *Cynoscion nebulosus*. *Marine Ecology Progress Series* **265**, 197–212.

Lamberth, S.J., Drapeau, L. and Branch, G.M. (2009) The effects of altered freshwater inflows on catch rates of non-estuarine-dependent fish in a multispecies nearshore linefishery. *Estuarine, Coastal and Shelf Science* **84**, 527–538.

Lamont, T., Roberts, M.J., Barlow, R.G., Morris, T. and van den Berg, M.A. (2012) Circulation patterns in the Delagoa Bight, Mozambique, and the influence of deep ocean eddies. *African Journal of Marine Science* **32**, 553–562.

La Morgia, V., Bona, F. and Badino, G. (2008) Bayesian modelling procedures for the evaluation of changes in wildlife habitat suitability: a case study of roe deer in the Italian Alps. *Journal of Applied Ecology* **45**, 863–872.

Landis, J.R. and Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174.

Larson, R.J. and Julian, R.M. (1999) Spatial and temporal genetic patchiness in marine populations and their implications for fisheries management. *California Cooperative Oceanic Fisheries Investigations Report* **40**, 94–99.

Lasram, F.B.R., Guilhaumon, F., Albouy, C., Somot, S., Thuiller, W. and Mouillot, D. (2010) The Mediterranean Sea as a “cul-de-sac” for endemic fishes facing climate change. *Global Change Biology* **16**, 3233–3245.

Lawler, J.J., White, D., Neilson, R.P. and Blaustein, A.R. (2006) Predicting climate-induced range shifts: model differences and model reliability. *Global Change Biology* **12**, 1568–1584.

Leathwick, J.R., Elith, J. and Hastie, T. (2006) Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological Modelling* **199**, 188–196.

Librado, P. and Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452.

Lipcius, R.N., Stockhausen, W.T., Eggleston, D.B., Marshall Jr, L.S. and Hickey, B. (1997) Hydrodynamic decoupling of recruitment, habitat quality and adult abundance in the Caribbean spiny lobster: source–sink dynamics? *Marine & Freshwater Research* **48**, 807–815.

Lloyd, P., Plagányi, É.E., Weeks, S.J., Magno-Canto, M. and Plagányi, G. (2012) Ocean warming alters species abundance patterns and increases species diversity in an African sub-tropical reef-fish community. *Fisheries Oceanography* **21**, 78–94.

Lobo, J.M., Jiménez-Valverde, A. and Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* **17**, 145–151.

Lutjeharms, J.R.E. (2006) The coastal oceans of south-eastern Africa. In: *The sea*, Vol. 14B. (eds A.R. Robinson and K.H. Brink). Harvard University Press, Cambridge, pp 783–834.

Lutjeharms, J.R.E., Cooper, J. and Roberts, M. (2000) Upwelling at the inshore edge of the Agulhas Current. *Continental Shelf Research* **20**, 737–761.

Lutjeharms, J.R.E., Monteiro, P.M.S., Tyson, P.D. and Obura, D. (2001) The oceans around southern Africa and regional effects of global change. *South African Journal of Science* **97**, 119–120.

Maggs, J.Q. (2011) *Fish surveys in exploited and protected areas of the Pondoland marine protected area with consideration of the impact of the MPA on coastal fisheries*. MSc thesis, University of KwaZulu-Natal, pp 140.

Malaney, J.L., Frey, J.K. and Cook, J.A. (2012) The biogeographic legacy of an imperilled taxon provides a foundation for assessing lineage diversification, demography and conservation genetics. *Diversity and Distributions* **18**, 689–703.

Manel, S., Schwartz, M.K., Luikart, G. and Taberlet, P. (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189–197.

Mann, B.Q., Celliers, L., Fennessy, S.T., Bailey, S. and Wood, A.D. (2006) Towards the declaration of a large marine protected area: a subtidal ichthyofaunal survey of the Pondoland coast in the Eastern Cape, South Africa. *African Journal of Marine Science* **28**, 535–551.

Mann, B.Q., Scott, G.M., Mann-Lang, J.B., Brouwer, J.B., Lamberth, S.J., Sauer, W.H.H. and Erasmus, C. (1997) An evaluation of participation in and management of the South African spearfishery. *South African Journal of Marine Science* **18**, 179–193.

Manni, F., Guérard, E. and Heyer, E. (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Human Biology* **76**, 173–190.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209–220.

Maree, R.C., Whitfield, A.K. and Booth, A.J. (2000) Effect of water temp on the biogeography of South African estuarine fishes associated with the subtropical/warm temperate subtraction zone. *South African Journal of Science* **96**, 184–188.

Marks, K.M. and Smith, W.H.F. (2006) An evaluation of publicly available global bathymetry grids. *Marine Geophysical Researches* **27**, 19–34.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. and Thuiller, W. (2009) Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions* **15**, 59–69.

Maruyama, T. (1983) Stochastic theory of population genetics. *Bulletin of Mathematical Biology* **45**, 521–554.

Matthee, C.A., Eick, G., Willows-Munro, S., Montgelard, C., Pardini, A.T. and Robinson, T.J. (2007) Indel evolution of mammalian introns and the utility of non-coding nuclear markers in eutherian phylogenetics. *Molecular Phylogenetics and Evolution* **42**, 827–837.

Maxwell, D.L., Stelzenmüller, V., Eastwood, P.D. and Rogers, S.I. (2009) Modelling the spatial distribution of plaice (*Pleuronectes platessa*), sole (*Solea solea*) and thornback ray (*Raja clavata*) in UK waters for marine management and planning. *Journal of Sea Research* **61**, 258–267.

McGrath, M.D., Horner, C.C.M., Brouwer, S.L., Lamberth, S.J., Mann, B.Q., Sauer, W.H.H. and Erasmus, C. (1997) An economic valuation of the South African linefishery. *South African Journal of Marine Science* **18**, 203–211.

McPherson, J.M., Jetz, W. and Rogers, D.J. (2004) The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology* **41**, 811–823.

Miller, M.P. (2005) Alleles in space (AIS): computer software for the joint analysis of interindividual spatial and genetic information. *The Journal of Heredity* **96**, 722–724.

Miller, M.P., Bellinger, M.R., Forsman, E.D. and Haig, S.M. (2006) Effects of historical climate change, habitat connectivity, and vicariance on genetic structure and diversity across the range of the red tree vole (*Phenacomys longicaudus*) in the Pacific Northwestern United States. *Molecular Ecology* **15**, 145–159.

Monmonier, M.S. (1973) Maximum-difference barriers: an alternative numerical regionalization method. *Geographical Analysis* **3**, 245–261.

Moran, P.A.P. (2012) Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23.

Mullis, K.B. (1990) The unusual origin of the polymerase chain reaction. *Scientific American* **262**, 56-61.

Muss, A., Robertson, D.R., Stepien, C.A., Wirtz, P. and Bowen, B.W. (2001) Phylogeography of *Ophioblennius*: the role of ocean currents and geography in reef fish evolution. *Evolution* **55**, 561–572.

National Oceanic and Atmospheric Administration (2005) http://www.esrl.noaa.gov/psd/ (accessed 1 Sept 2012).

Neethling, M., Matthee, C.A., Bowie, R.C.K. and von der Heyden, S. (2008) Evidence for panmixia despite barriers to gene flow in the southern African endemic, *Caffrogobius caffer* (Teleostei: Gobiidae). *BMC Evolutionary Biology* **8**, 325–334.

Nei, M. and Tajima, F. (1981) DNA polymorphism detectable by restrictions endonucleases. *Genetics* **97**, 145–163.

Nemeth, R.S., Blondeau, J., Herzlieb, S. and Kadison, E. (2007) Spatial and temporal patterns of movement and migration at spawning aggregations of red hind, *Epinephelus guttatus*, in the U.S. Virgin Islands. *Environmental Biology of Fishes* **78**, 365–381.

Ohlemüller, R., Anderson, B.J., Araújo, M.B., Butchart, S.H.M., Kudrna, O., Ridgely, R.S. and Thomas, C.D. (2008) The coincidence of climatic and species rarity: high risk to small-range species from climate change. *Biology Letters* **4**, 568–572.

Ohta, T. (1992) The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics* **23**, 263–286.

Ohta, T. and Kimura, M. (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research* **22**, 201–204.

Opdam, P. and Wascher, D. (2004) Climate change meets habitat fragmentation: linking landscape and biogeographical scale levels in research and conservation. *Biological Conservation* **117**, 285–297.

Opgenoorth, L., Vendramin, G.G., Mao, K., Miehe, G., Miehe, S., Liepelt, S., Liu, J., and Ziegenhagen, B. (2010) Tree endurance on the Tibetan Plateau marks the world's highest known tree line of the Last Glacial Maximum. *The New Phytologist* **185**, 332–342.

Palsbøll, P.J., Bérubé, M. and Allendorf, F.W. (2007) Identification of management units using population genetic data. *Trends in Ecology & Evolution* **22**, 11–16.

Palumbi, S.R. and Baker, C.S. (1994) Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Molecular Biology and Evolution* **11**, 426–435.

Pankhurst, N.W. and Munday, P.L. (2011) Effects of climate change on fish reproduction and early life history stages. *Marine and Freshwater Research* **62**, 1015–1026.

Parmesan, C. (2006) Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics* **37**, 637–671.

Parmesan, C. and Yohe, G. (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42.

Pearson, R.G. (2007) Species' distribution modeling for conservation educators and practitioners (synthesis). *American Museum of Natural History. Available at http://ncep.amnh.org.*, pp 54.

Pearson, R.G. and Dawson, T.P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography* **12**, 361–371.

Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T. P., and Lees, D. C. (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography* **33**, 1704–1711.

Pelc, R. A., Warner, R.R. and Gaines, S.D. (2009) Geographical patterns of genetic structure in marine species with contrasting life histories. *Journal of Biogeography* **36**, 1881–1890.

Pereira, M.A.M. (2000) A review on the ecology, exploitation and conservation of reef fish resources in Mozambique. In: *2nd National Conference on Coastal Zones Research, Maputo* (27-29 September 2000) pp 9.

Pereira, M.A.M. (2001) Reef-associated fishes of Mozambique: what is known about them. Maputo pp 1–8.

Perry, A.L., Low, P.J., Ellis, J.R. and Reynolds, J.D. (2005) Climate change and distribution shifts in marine fishes. *Science* **308**, 1912–1914.

Peterson, A.T., Papeş, M. and Eaton, M. (2007) Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* **30**, 550–560.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. and Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* **19**, 181–197.

Pittman, S.J. and Brown, K. A. (2011) Multi-scale approach for predicting fish species distributions across coral reef seascapes. *PloS one* **6**, e20583.

Pittman, S.J., Christensen, J.D., Caldow, C., Menza, C. and Monaco, M.E. (2007) Predictive mapping of fish species richness across shallow-water seascapes in the Caribbean. *Ecological Modelling* **204**, 9–21.

Pörtner, H.O., Berdal, B., Blust, R., Brix, O., Colosimo, A., De Wachter, B., Giuliani, A., Johansen, T., Fischer, T., Knust, R., Lannig, G., Naevdal, G., Nedenes, A., Nyhammer, G., Sartoris, F. J., Serendero, I., Sirabella, P., Thorkildsen, S., and Zakhartsev, M. (2001) Climate induced temperature effects on growth performance, fecundity and recruitment in marine fish: developing a hypothesis for cause and effect relationships in Atlantic cod (*Gadus morhua*) and common eelpout (*Zoarces viviparus*). *Continental Shelf Research* **21**, 1975–1997.

Pörtner, H.O. and Farrell, A.P. (2008) Physiology and climate change. *Science* **322**, 690–692.

Pörtner, H.O. and Knust, R. (2007) Climate change affects marine fishes through the oxygen limitation of thermal tolerance. *Science* **315**, 95–97.

Posada, D. and Crandall, K. (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology & Evolution* **16**, 37–45.

Potts, W.M., Booth, A.J., Richardson, T.J. and Sauer, W.H. Ocean warming affects the distribution and abundance of resident fishes by changing their reproductive scope. *Reviews in Fish Biology and Fisheries* (in press).

Pradervand, P. (2004) Long-term trends in the shore fishery of the Transkei coast, South Africa. *African Zoology* **39**, 247–261.

Pradervand, P., Mann, B.Q. and Bellis, M.F. (2007) Long-term trends in the competitive shore fishery along the KwaZulu-Natal coast, South Africa. *African Zoology* **42**, 216–236.

Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters* **3**, 349–361.

R Development Core Team (2012) R: a language and environment for statistical computing.

Rand, D.M. (1996) Neutrality tests of molecular markers and the connection between DNA polymorphism, demography and conservation biology. *Conservation Biology* **10**, 665–671.

Raymond, M. and Rousset, F. (1995) An exact test for population differentiation. *Evolution* **49**, 1280–1283.

Reusch, T.B.H., Ehlers, A., Hämmerli, A. and Worm, B. (2005) Ecosystem recovery after climatic extremes enhanced by genotypic diversity. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 2826–31.

Reynolds, R.W. (1988) A real-time global sea surface temperature analysis. *Journal of Climate* **1**, 75–86.

Reynolds, R.W. and Marsico, D.C. (1993) An improved real-time global sea surface temperature analysis. *Journal of Climate* **6**, 114–119.

Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C. and Wang, W. (2002) An improved in situ and satellite SST analysis for climate. *Journal of Climate* **15**, 1609–1625.

Rijnsdorp, A.D., Peck, M.A., Engelhard, G.H., Möllmann, C. and Pinnegar, J.K. (2009) Resolving the effect of climate change on fish populations. *ICES Journal of Marine Science* **66**, 1570–1583.

Rivera, M.A.J., Kelley, C.D. and Roderick, G.K. (2004) Subtle population genetic structure in the Hawaiian grouper, *Epinephelus quernus* (Serranidae) as revealed by mitochondrial DNA analyses. *Biological Journal of the Linnean Society* **81**, 449–468.

Roberts, M.J., van der Lingen, C.D., Whittle, C. and van den Berg, M. (2010) Shelf currents, lee-trapped and transient eddies on the inshore boundary of the Agulhas Current, South Africa: their relevance to the KwaZulu-Natal sardine run. *African Journal of Marine Science* **32**, 423–447.

Robinson, L.M., Elith, J., Hobday, A.J., Pearson, R.G., Kendall, B.E., Possingham, H.P. and Richardson, A.J. (2011) Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Global Ecology and Biogeography* **20**, 789–802.

Rocha, L. A., Bass, A.L., Robertson, D.R. and Bowen, B.W. (2002) Adult habitat preferences, larval dispersal, and the comparative phylogeography of three Atlantic surgeonfishes (Teleostei: Acanthuridae). *Molecular Ecology* **11**, 243–52.

Roessig, J.M., Woodley, C.M., Cech, J.J. and Hansen, L.J. (2004) Effects of global climate change on marine and estuarine fishes and fisheries. *Reviews in Fish Biology and Fisheries* **14**, 251–275.

Rogers, A.R. and Harpending, H. (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**, 552–569.

Rouault, M. and Lutjeharms, J.R.E. (2003) Estimation of sea-surface temperature around southern Africa from satellite-derived microwave observations. *South African Journal of Science* **99**, 489–494.

Rouault, M., Penven, P. and Pohl, B. (2009) Warming in the Agulhas Current system since the 1980's. *Geophysical Research Letters* **36**, 1–5.

Rouault, M., Pohl, B. and Penven, P. (2010) Coastal oceanic climate change and variability from 1982 to 2009 around South Africa. *African Journal of Marine Science* **32**, 237–246.

Ruegg, K.C. and Smith, T.B. (2002) Not as the crow flies: a historical explanation for circuitous migration in Swainson's thrush (*Catharus ustulatus*). *Proceedings of the Royal Society B* **269**, 1375–1381.

Sabatés, A., Martín, P., Lloret, J. and Raya, V. (2006) Sea warming and fish distribution: the case of the small pelagic fish, *Sardinella aurita*, in the western Mediterranean. *Global Change Biology* **12**, 2209–2219.

Sadovy de Mitcheson, Y., Craig, M.T., Bertoncini, A.A., Carpenter, K. E., Cheung, W. W. L., Choat, J. H., Cornish, A. S., Fennessy, S. T., Ferreira, B. P., Heemstra, P. C., Liu, M., Myers, R. F., Pollard, D. A., Rhodes, K. L., Rocha, L. A., Russell, B. C., Samoilys, M. A., and Sanciangco, J. (2013) Fishing groupers towards extinction: a global assessment of threats and extinction risks in a billion dollar fishery. *Fish and Fisheries* **14**, 119–136.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., Erlich, H. A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487–491.

Schneider, S. and Excoffier, L. (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites : application to human mitochondrial DNA. *Genetics* **152**, 1079–1089.

Schwartz, M.K. and McKelvey, K.S. (2008) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics* **10**, 441–452.

Selkoe, K. A., Gaines, S.D., Caselle, J.E. and Warner, R.R. (2006) Current shifts and kin aggregation explain genetic patchiness in fish recruits. *Ecology* **87**, 3082–94.

Selkoe, K. A., Watson, J.R., White, C., Horin, T. B., Iacchei, M., Mitarai, S., Siegel, D. A., Gaines, S. D., and Toonen, R. J. (2010) Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Molecular ecology* **19**, 3708–26.

Sheaves, M. (2006) Is the timing of spawning in sparid fishes a response to sea temperature regimes? *Coral Reefs* **25**, 655–669.

Shimono, A., Ueno, S., Gu, S., Zhao, X., Tsumura, Y. and Tang, Y. (2010) Range shifts of *Potentilla fruticosa* on the Qinghai-Tibetan Plateau during glacial and interglacial periods revealed by chloroplast DNA sequence variation. *Heredity* **104**, 534–542.

Silva-oliveira, G.C., Rêgo, P.S., Schneider, H., Sampaio, I. and Vallinoto, M. (2008) Genetic characterisation of populations of the critically endangered Goliath grouper (*Epinephelus itajara*, Serranidae) from the northern Brazilian coast through analyses of mtDNA. *Genetics and Molecular Biology* **31**, 988–994.

Simonsen, K.L., Churchill, G.A. and Aquadro, C.F. (1995) Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**, 413–429.

Sink, K., Harris, J. and Lombard, A. (2004) South African national spatial biodiversity assessment 2004: technical report Vol. 4 marine component. *Appendix 1. South African marine bioregions*. pp 97–108.

Slatkin, M. (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**, 264–279.

Slatkin, M. and Hudson, R.R. (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**, 555–562.

Smith, W.H.F. and Sandwell, D.T. (1997) Global sea floor topography from satellite altimetry and ship depth soundings. *Science* **277**, 1956–1962.

Soberón, J. and Peterson, A.T. (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* **2**, 1–10.

Song, C.B., Near, T.J. and Page, L.M. (1998) Phylogenetic relations among Percid fishes as inferred from mitochondrial cytochrome b DNA sequence data. *Molecular Phylogenetics and Evolution* **10**, 343–353.

Stephens, M. and Donnelly, P. (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* **73**, 1162–1169.

Stephens, M., Smith, N.J. and Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**, 978–989.

Stock, C.A., Alexander, M.A., Bond, N.A., Brander, K. M., Cheung, W. W. L., Curchitser, E. N., Delworth, T. L.Dunne, J. P., Griffies, S. M., Haltuch, M. A., Hare, J. A., Hollowed, A. B., Lehodey, P., Levin, S. A., Link, J. S., Rose, K. A., Rykaczewski, R. R., Sarmiento, J. L., Stouffer, R. J., Schwing, F. B., Vecchi, G. A., and Werner, F. E. (2010) On the use of IPCC-class models to assess the impact of climate on living marine resources. *Progress in Oceanography* **88**, 1–27.

Stokland, J.N., Halvorsen, R. and Støa, B. (2011) Species distribution modelling—Effect of design and sample size of pseudo-absence observations. *Ecological Modelling* **222**, 1800–1809.

Sunday, J.M., Bates, A.E. and Dulvy, N.K. (2012) Thermal tolerance and the global redistribution of animals. *Nature Climate Change* **2**, 1–5.

Swets, J.A. (1988) Measuring the accuracy of diagnostic systems. *Science* **240**, 1285–1293.

Sykes, M.T., Prentice, I.C. and Cramer, W. (1996) A bioclimatic model for the potential distributions of north European tree species under present and future climates. *Journal of Biogeography* **23**, 203–233.

Taberlet, P., Meyer, A. and Bouvet, J. (1992) Unusually large mitochondrial variation in populations of the blue tit, *Parus caeruleus*. *Molecular Ecology* **1**, 27–36.

Tajima, F. (1996) Infinite-allele model and infinite-site model in population genetics. *Journal of Genetics* **75**, 27–31.

Tajima, F. (1989a) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

Tajima, F. (1989b) The effect of change in population size on DNA polymorphism. *Genetics* **123**, 597–601.

Termansen, M., McClean, C.J. and Preston, C.D. (2006) The use of genetic algorithms and Bayesian classification to model species distributions. *Ecological Modelling* **192**, 410–424.

Teske, P., Forget, F., Cowley, P., von der Heyden, S. and Beheregaray, L. (2010) Connectivity between marine reserves and exploited areas in the philopatric reef fish *Chrysoblephus laticeps* (Teleostei: Sparidae). *Marine Biology* **157**, 2029–2042.

Teske, P., Froneman, P., Barker, N. and McQuaid, C. (2007) Phylogeographic structure of the caridean shrimp *Palaemon peringueyi* in South Africa: further evidence for intraspecific genetic units associated with marine biogeographic provinces. *African Journal of Marine Science* **29**, 253–258.

Thomas, C.D., Franco, A.M.A. and Hill, J.K. (2006) Range retractions and extinction in the face of climate warming. *Trends in Ecology & Evolution* **21**, 415–416.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

Thuiller, W. (2003) BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology* **9**, 1353–1362.

Thuiller, W. (2004) Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology* **10**, 2020–2027.

Thuiller, W., Brotons, L., Araújo, M.B. and Lavorel, S. (2004) Effects of restricting environmental range of data to project current and future species distributions. *Ecography* **27**, 165–172.

Thuiller, W., Lafourcade, B. and Araújo, M. (2009a) ModOperating manual for BIOMOD. pp 90.

Thuiller, W., Lafourcade, B. and Araújo, M. (2010) Presentation manual for BIOMOD. pp 35.

Thuiller, W., Lafourcade, B., Engler, R. and Araújo, M.B. (2009b) BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography* **32**, 369–373.

Tollefsrud, M.M., Sønstebø, J.H., Brochmann, C., Johnsen, Ø., Skrøppa, T. and Vendramin, G.G. (2009) Combined analysis of nuclear and mitochondrial markers provide new insight into the genetic structure of North European *Picea abies*. *Heredity* **102**, 549–62.

Tucker Jr, J.W. (1999) Species profile: grouper aquaculture. *Southern Regional Aquaculture Center Publication* No. 721 pp 11.

Tucker Jr, J.W. (1994) Spawning by captive serranid fishes: a review. *Journal of the World Aquaculture Society* **25**, 345–359.

Turner, M.G., Gardner, R.H. and O'Neill, R. V (2001) *Landscape ecology in theory and practice: pattern and process.* (Vol. 1). Springer Science & Business Media, Inc., New York.

Turpie, J.K., Beckley, L.E. and Katua, S.M. (2000) Biogeography and the selection of priority areas for conservation of South African coastal fishes. *Biological Conservation* **92**, 59–72.

Valavanis, V.D., Pierce, G.J., Zuur, A.F., Palialexis, A., Saveliev, A., Katara, I. and Wang, J. (2008) Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. *Hydrobiologia* **612**, 5–20.

van der Elst, R., Everett, B., Jiddawi, N., Mwatha, G., Afonso, P.S. and Boulle, D. (2005) Fish, fishers and fisheries of the Western Indian Ocean: their diversity and status. *Philosophical Transactions of the Royal Society* **363**, 263–284.

van der Elst, R.P. and Bullen, E. (1993) ORI/Sedgwicks marine linefish tagging project yields results. *Oceanographic Research Institute Special Publication* **2**, 64–67.

van der Elst, R.P., Fennessy, S.T., Torres, R.G.A. and Cuco, C.A. (2003) Linefish resources annual report for the year 2000. Part 1: fisheries assessment. *Boletim de Divulgação, Instituto de Investigação Pesqueira, Mozambique 38*, pp 29.

von der Heyden, S., Prochazka, K. and Bowie, R.C.K. (2008) Significant population structure and asymmetric gene flow patterns amidst expanding populations of *Clinus cottoides* (Perciformes, Clinidae): application of molecular data to marine conservation planning in South Africa. *Molecular Ecology* **17**, 4812–4826.

Wakeley, J. (2005) The limits of theoretical population genetics. *Genetics* **169**, 1–7.

Waldman, J.R. (1999) The importance of comparative studies in stock analysis. *Fisheries Research* **43**, 237–246.

Wang, L., Meng, Z., Liu, X., Zhang, Y. and Lin, H. (2011) Genetic diversity and differentiation of the orange-spotted grouper (*Epinephelus coioides*) between and within cultured stocks and wild populations inferred from microsatellite DNA analysis. *International Journal of Molecular Sciences* **12**, 4378–4394.

Waples, R.S. (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity* **89**, 438–450.

Ward, R.D. (2000) Genetics in fisheries management. *Hydrobiologia* **420**, 191–201.

Watson, D.F. (1992) *Contouring: a guide to the analysis and display of spatial data (with programs on diskette).* Pergamon Press, New York. pp 327.

Weaver, K.F., Anderson, T. and Guralnick, R. (2006) Combining phylogenetic and ecological niche modeling approaches to determine distribution and historical biogeography of Black Hills mountain snails. (Oreohelicidae). *Diversity and Distributions* **12**, 756–766.

Wiens, J.A. (1976) Population responses to patchy environments. *Annual Review of Ecology and Systematics* **7**, 81–120.

Willis, K.J. and Birks, H.J.B. (2006) What is natural? The need for a long-term perspective in biodiversity conservation. *Science* **314**, 1261–1265.

Wright, S. (1978) *Evolution and the genetics of populations. Variability within and among natural populations* (Vol. 4). University of Chicago Press, Chicago.

Wright, S. (1931) Evolution in Mendelian populations. *Genetics* **16**, 97–159.

Wright, S. (1951) The genetical structure of populations. *Annals of Eugenics* **15**, 323–354.

Zardi, G.I., McQuaid, C.D., Teske, P.R. and Barker, N.P. (2007) Unexpected genetic structure of mussel populations in South Africa: indigenous *Perna perna* and invasive *Mytilus galloprovincialis*. *Marine Ecology Progress Series* **337**, 135–144.

Zatcoff, M.S., Ball, A.O. and Sedberry, G.R. (2004) Population genetic analysis of red grouper, *Epinephelus morio*, and scamp, *Mycteroperca phenax*, from the southeastern U.S. Atlantic and Gulf of Mexico. *Marine Biology* **144**, 769–777.

Zink, R.M. and Barrowclough, G.F. (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular ecology* **17**, 2107–21.

Appendix I

Supplemental results for CHAPTER 3 are presented.

5.1 Model run 1: Surface Range Envelope (SRE) projected pseudo-absences

Individual model results

Current

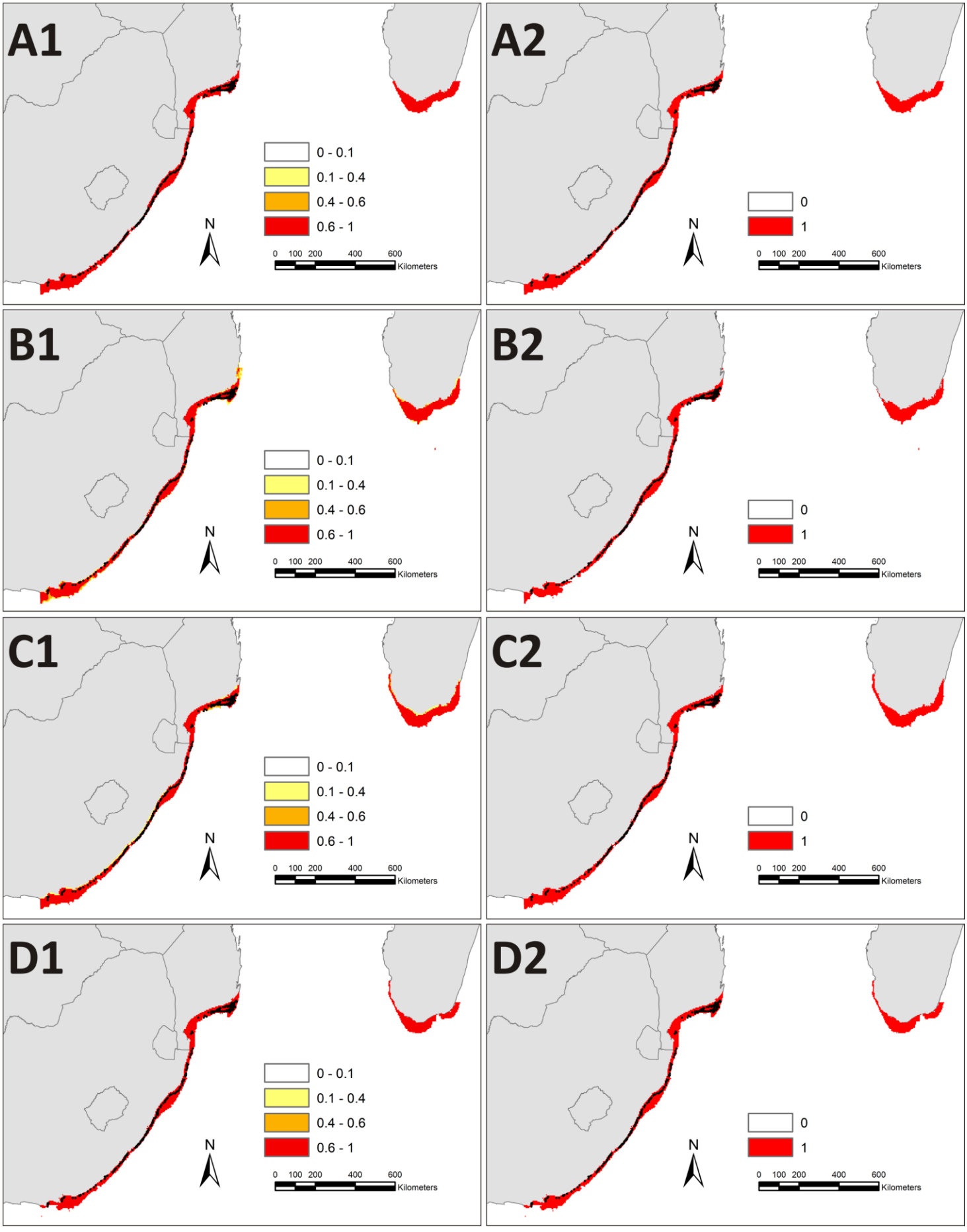
Before the ensemble models could be projected, each of the individual models included in the ensemble needed to be projected. Each model predicted a slightly different change in range but the overall consensuses were averaged in an ensemble model which is presented in Figure 3.4 and Figure 3.5 of CHAPTER 3. Presented here, are the current range maps projected by each of the individual models (Figure 5.1). This model run used 2100 pseudo-absences (PAs) that were projected using the Surface Range Envelope (SRE) approach, as explained in the methods of CHAPTER 3. Each of the individual models for the current distribution fit the presence points relatively closely indicating that the models performed well and the projections did not differ significantly between models.

2020

The individual model projections for 2020 (Figure 5.2) show slightly more variation in their projected distributions and shrinkage of the distribution between the current and the 2020 projections is evident for some of the models. Range loss was predicted by all models in 2020 except the GAM model (Figure 5.2) which projected a 7.574% increase in the *E. andersoni* range (Table 5.1).

2030

Further shrinkage of the distribution between the 2020 and the 2030 projections is evident for most of the models (Figure 5.3), also shown by the range change calculations in Table 5.1. Again, all models predict range loss in 2030 relative to the current distribution but GAM predicts a small range increase, although the increase is less than projected for 2020, implying a range decrease between 2020 and 2030 for GAM.


A1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
A2
0
1
N
0 100 200 400 600
Kilometers
B1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
B2
0
1
N
0 100 200 400 600
Kilometers
C1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
C2
0
1
N
0 100 200 400 600
Kilometers
D1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
D2
0
1
N
0 100 200 400 600
Kilometers

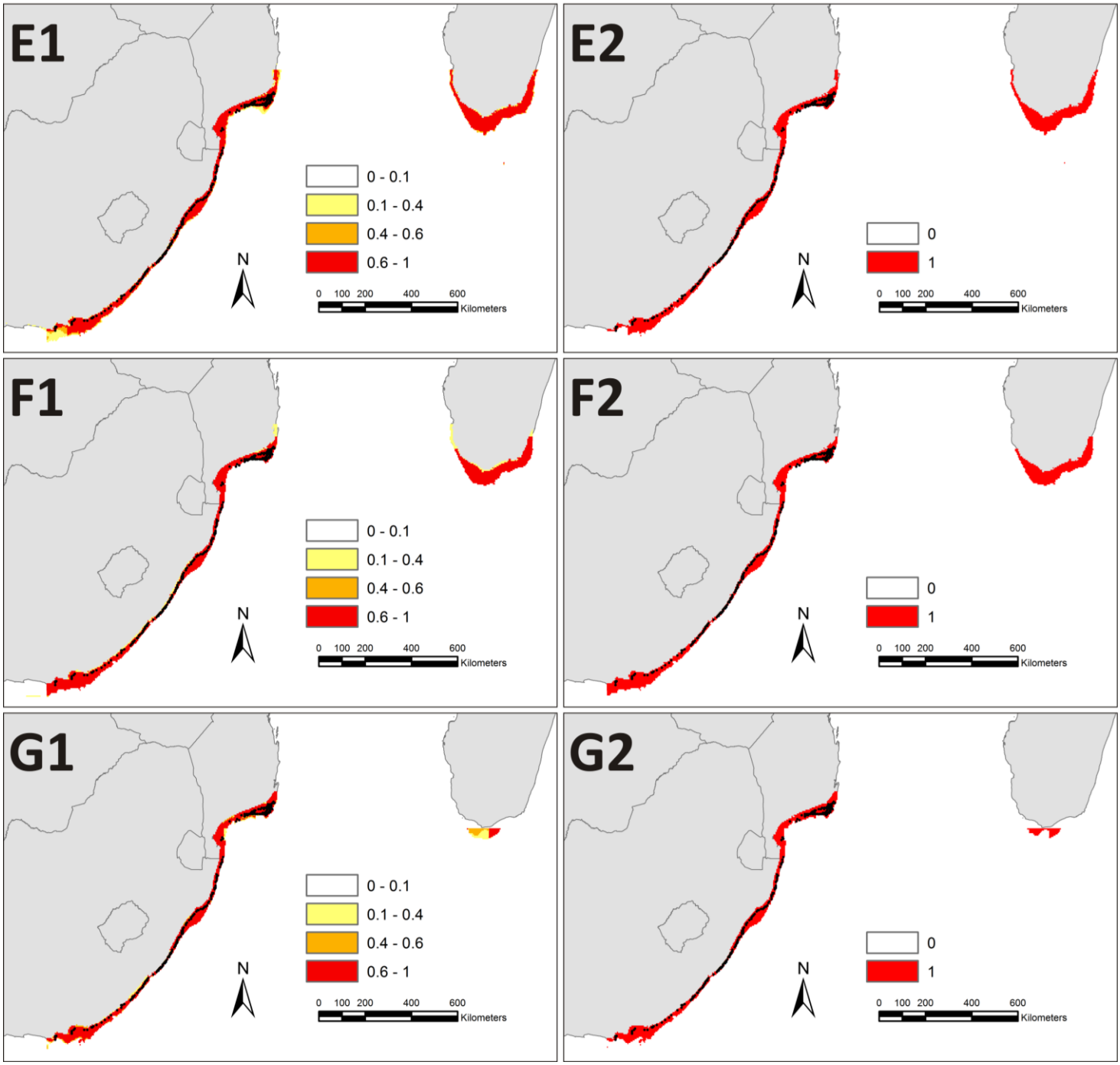
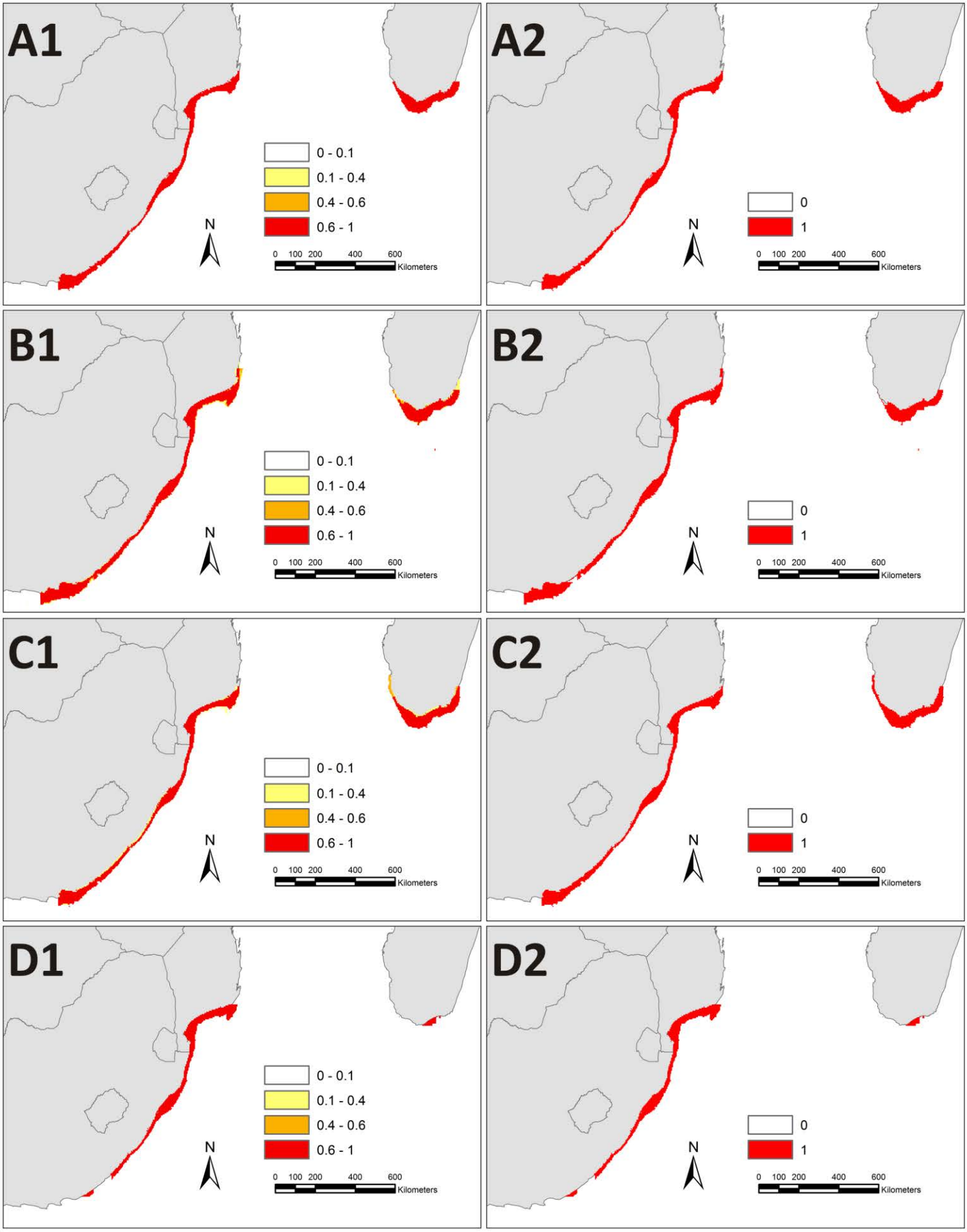


Figure 5.1: Individual model projections for the current *E. andersoni* distribution using 2100 pseudo-absences projected using the SRE approach. The probability of occurrence distributions are shown on the left (all figures labelled 1) and the binary transformed distributions are shown on the right (all figures labelled 2). The results of the following models are shown: A) CTA, B) GAM, C) GBM, D) GLM, E) MARS, F) MAXENT, G) RF.


A1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
A2
0
1
N
0 100 200 400 600
Kilometers
B1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
B2
0
1
N
0 100 200 400 600
Kilometers
C1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
C2
0
1
N
0 100 200 400 600
Kilometers
D1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
D2
0
1
N
0 100 200 400 600
Kilometers

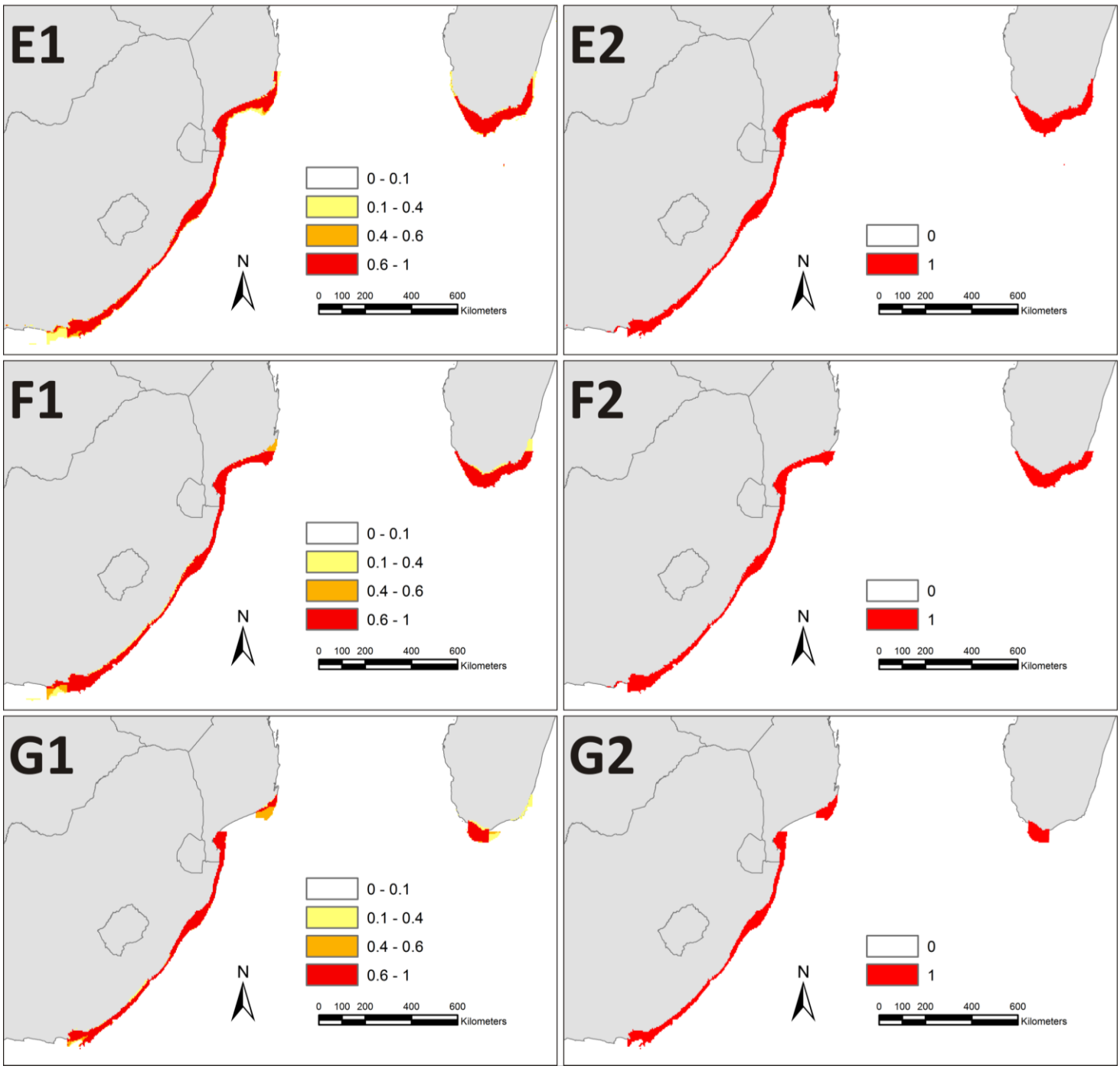
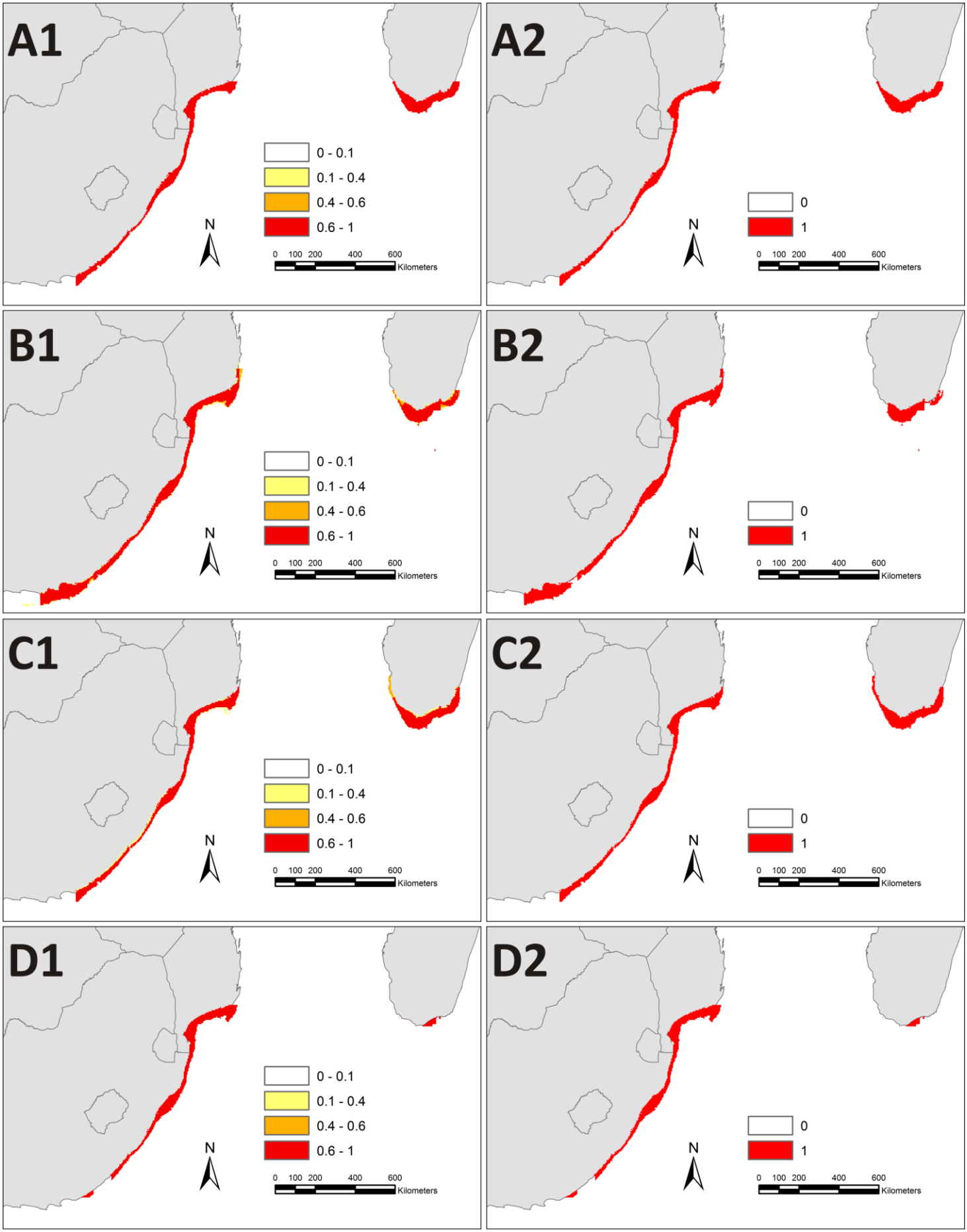


Figure 5.2: Individual model projections for the 2020 *E. andersoni* distribution using 2100 pseudo-absences projected using the SRE approach. The probability of occurrence distributions are shown on the left (all figures labelled 1) and the binary transformed distributions are shown on the right (all figures labelled 2). The results of the following models are shown: A) CTA, B) GAM, C) GBM, D) GLM, E) MARS, F) MAXENT, G) RF.


A1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
A2
0
1
N
0 100 200 400 600
Kilometers
B1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
B2
0
1
N
0 100 200 400 600
Kilometers
C1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
C2
0
1
N
0 100 200 400 600
Kilometers
D1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
D2
0
1
N
0 100 200 400 600
Kilometers

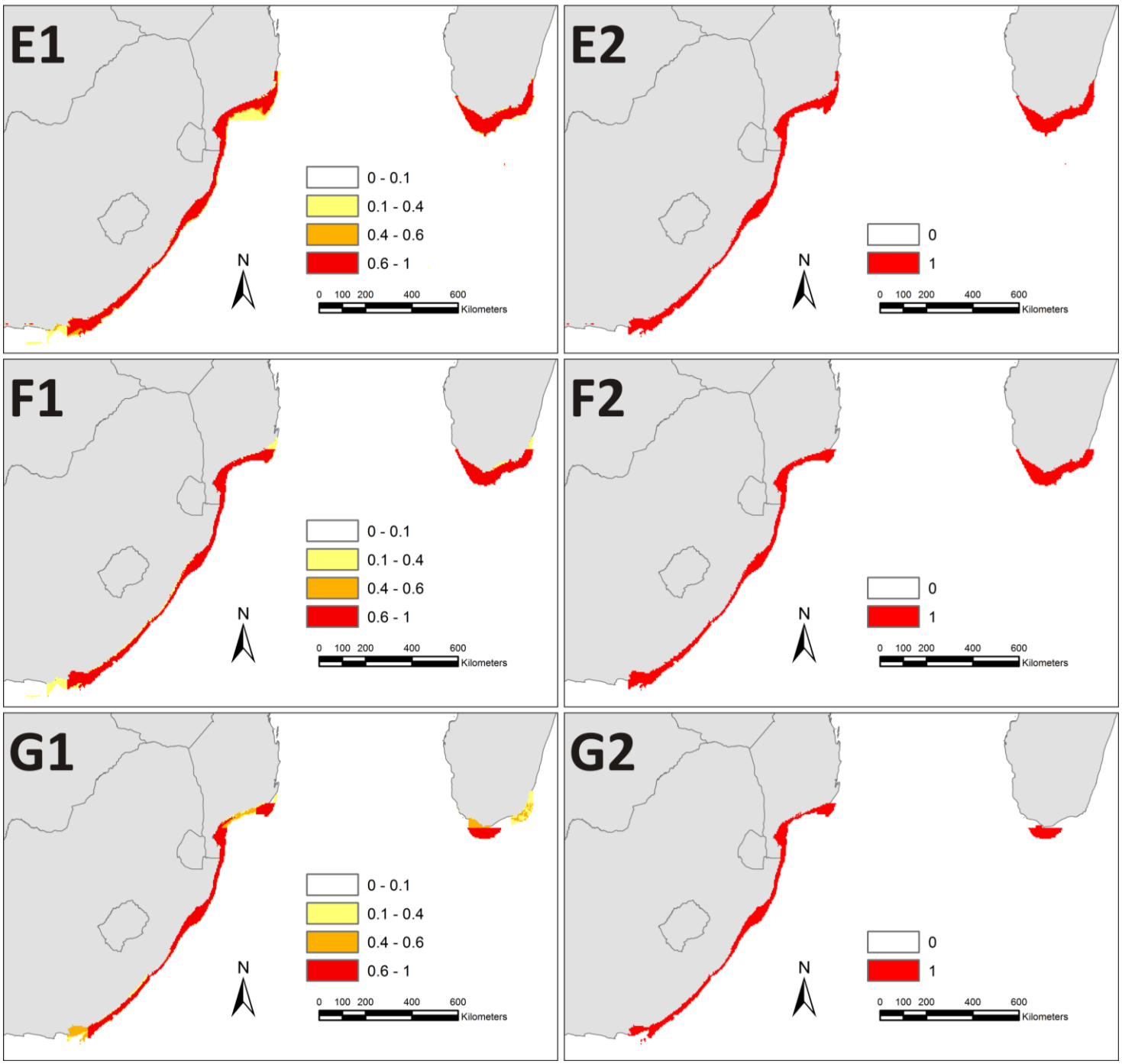


Figure 5.3: Individual model projections for the 2030 *E. andersoni* distribution using 2100 pseudo-absences projected using the SRE approach. The probability of occurrence distributions are shown on the left (all figures labelled 1) and the binary transformed distributions are shown on the right (all figures labelled 2). The results of the following models are shown: A) CTA, B) GAM, C) GBM, D) GLM, E) MARS, F) MAXENT, G) RF.

Table 5.1: Range changes between the current and the 2020, and the current and the 2030 binary transformed distributions for each model. Figures represent number of pixels except where percentages are indicated. Stable0 indicates the number of pixels that are unoccupied in the current projection and that remain unoccupied in the future projections. Stable1 indicates the number of pixels that are occupied in the current projection and are predicted to remain occupied in the future.

										Future Range Size	
		Loss	Stable0	Stable1	Gain	%Loss	%Gain	Range Change	Current Range	No Dispersal	Full Dispersal
Current - 2020	MAXENT	288	25478	2330	0	11.001	0	-11.001	2618	2330	2330
	CTA	270	25437	2385	4	10.169	0.151	-10.019	2655	2385	2389
	RF	456	26007	1431	202	24.165	10.705	-13.461	1887	1431	1633
	MARS	150	25272	2629	45	5.398	1.619	-3.778	2779	2629	2674
	GAM	74	25295	2461	266	2.919	10.493	7.574	2535	2461	2727
	GLM	1054	25613	1391	38	43.108	1.554	-41.554	2445	1391	1429
	GBM	248	25269	2575	4	8.785	0.142	-8.643	2823	2575	2579
Current - 2030	MAXENT	342	25478	2276	0	13.063	0	-13.063	2618	2276	2276
	CTA	581	25439	2074	2	21.883	0.075	-21.808	2655	2074	2076
	RF	391	26077	1496	132	20.721	6.995	-13.725	1887	1496	1628
	MARS	204	25224	2575	93	7.341	3.347	-3.994	2779	2575	2668
	GAM	165	25234	2370	327	6.509	12.899	6.391	2535	2370	2697
	GLM	1055	25592	1390	59	43.149	2.413	-40.736	2445	1390	1449
	GBM	499	25271	2324	2	17.676	0.071	-17.605	2823	2324	2326

5.2 Model run 2: Randomly projected pseudo-absences

The purpose of the second model run was to determine if using a different method of PA generation affected the resulting projection of the *E. andersoni* distribution. Barbet-Massin *et al*. (2012) demonstrated that randomly generated PAs produced the most predictive models and therefore this method was tested to determine if models performed better using this approach.

Variable importance

The standard errors and standard deviations were fairly high across different modelling algorithms for all environmental variables except for summer max (Table 5.2, Figure 5.4), indicating that the predictive importance of these variables varied across model algorithms. Bathymetry, autumn min, spring max and spring min had mean variable importances higher than the overall mean importance (0.281). Similar to the CHAPTER 3 results for the SRE projected PAs (Table 3.3 and Figure 3.3) bathymetry was the most important environmental variable and also had relatively low errors. Autumn max was on average the most important SST variable for model projections and summer max was noticeably of low importance. All other SST variables were close to the overall mean importance value.

Table 5.2: Variable importances for each of the seven environmental variables (columns) when the *Epinephelus andersoni* distribution was modelled using seven different modelling algorithms (rows). Mean importance, standard deviation and standard error for each variable across modelling techniques are shown.

	Summer max	Winter min	Bathymetry	Autumn max	Autumn min	Spring max	Spring min
MAXENT	0.039	0.243	0.483	0.363	0.311	0.065	0.149
CTA	0.063	0.708	0.568	0.313	0.000	0.000	0.000
RF	0.039	0.098	0.557	0.054	0.067	0.109	0.080
MARS	0.000	0.000	0.000	0.429	0.004	0.424	0.000
GAM	0.027	0.073	0.394	1.000	0.635	0.674	1.000
GLM	0.000	0.757	0.385	0.495	0.790	0.734	0.771
GBM	0.033	0.055	0.540	0.152	0.073	0.022	0.013
Mean	0.029	0.276	0.418	0.401	0.269	0.290	0.288
SD	0.023	0.321	0.199	0.306	0.324	0.317	0.417
SE	0.009	0.121	0.075	0.115	0.122	0.120	0.158

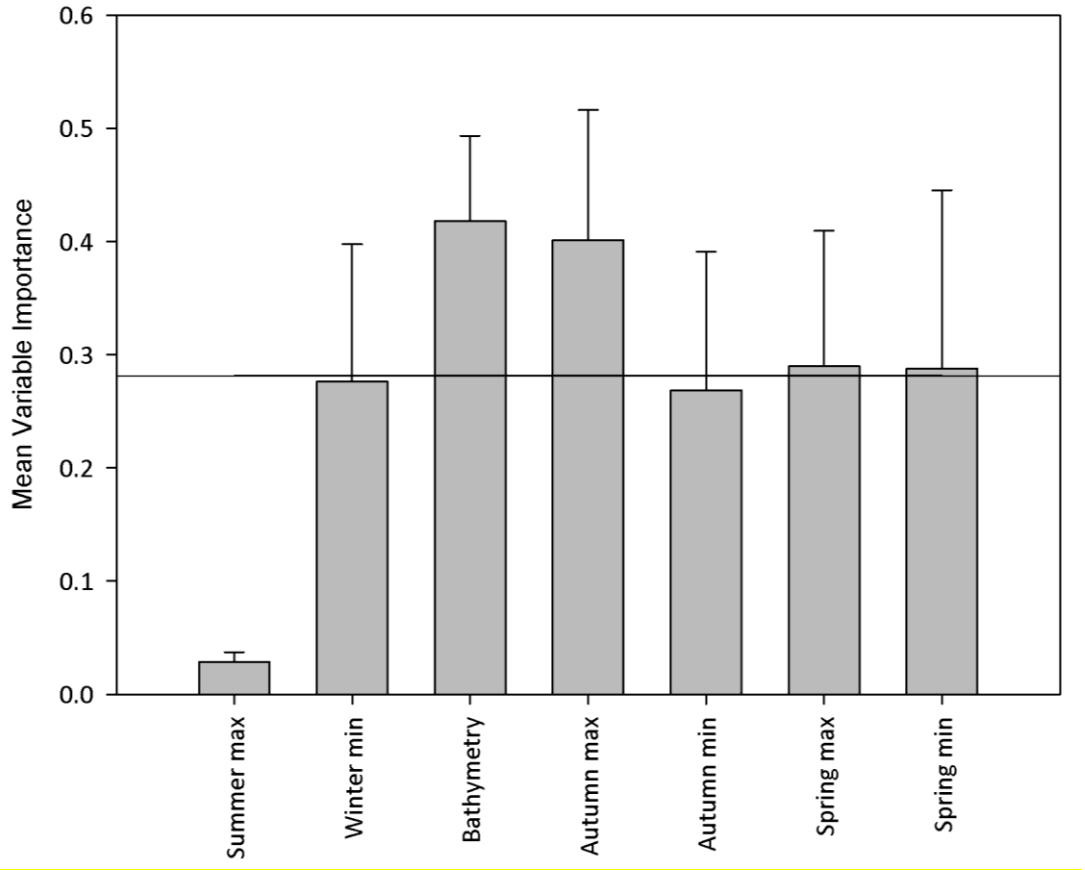


Figure 5.4: Mean variable importances for each of the seven environmental variables with standard error whiskers. The solid horizontal line indicates the overall mean variable importance (0.281).

Model performance

For the individual model projections, all models except for MARS were classified as excellent (i.e. TSS > 0.8) at predicting the distribution of *E. andersoni* according to TSS. The KAPPA statistic only classified MAXENT, CTA, RF and GBM as performing excellently (KAPPA > 0.8) where MARS performed poorly and the performance of GAM and GLM models was classified as "good" (Table 5.3). All models performed excellently according ROC, except for MARS whose performance was classified as "good". According to both ROC and KAPPA, RF was one of the best performing models and GLM and MARS were among the worst performing models. According to TSS, CTA had the highest model performance. All models except MARS had a TSS score higher than the set TSS threshold of 0.850 and consequently all models except MARS were included in the ensemble. For the weighted mean ensemble model projections, all model evaluation methods showed excellent predictive accuracy (TSS: 0.959, ROC: 0.993, KAPPA: 0.870). Generally, model performance was much lower for the second model run using random PAs.

Table 5.3: Three measures of model evaluation, True Skills Statistic (TSS), area under the Receiver Operating Curve (ROC), and Cohen's KAPPA are shown for each of the seven modelling algorithms used to model the distribution of *Epinephelus andersoni* into the future.

	MAXENT	CTA	RF	MARS	GAM	GLM	GBM
TSS	0.925	0.969	0.959	0.624	0.894	0.894	0.935
ROC	0.989	0.989	0.995	0.811	0.975	0.947	0.990
KAPPA	0.828	0.851	0.879	0.333	0.724	0.647	0.838

Ensemble model projections

Current

The current probability weighted mean and the binary transformed ensemble projection fitted the presence points fairly closely as shown in Figure 5.5 A and Figure 5.6 A respectively. Again, there is no area of high probability of occurrence on the Madagascan coast and no Madagascan distribution is projected by the binary transformed distributions. This is unlike the model results projected using the PAs simulated with the SRE approach.

2020

The southern limit of the area of high probability of occurrence decreased from Jeffreys Bay to Port Elizabeth and the area of high probability of occurrence in Mozambique thinned, from the current probability weighted mean distribution (Figure 5.5 A) to the 2020 projected distribution (Figure 5.5 B). Despite this, the area of high probability of occurrence on the Madagascan coast increased. These range changes were mirrored by the binary transformed distributions (Figure 5.6 A and B). Overall, 242 pixels of the *E. andersoni* range was lost by 2020 which equates to a 9.149% range loss (Table 5.4).

2030

Further contraction of the area of high probability of occurrence from Port Elizabeth to just south of Port Alfred and further thinning of the Mozambique distribution was noticeable in 2030 (Figure 5.5 C). Moreover, the northern limit of the area of high probability of occurrence of *E. andersoni* was projected to contract from Inhambane to just south of Paindane. There was not much change in the Madagascan area of high probability of occurrence although the binary transformed maps showed an increase in the Madagascan distribution (Figure 5.6 C). Overall there was further range loss in 2030 with 16.371% of the range being lost between the current and the 2030 projections (Table 5.4).

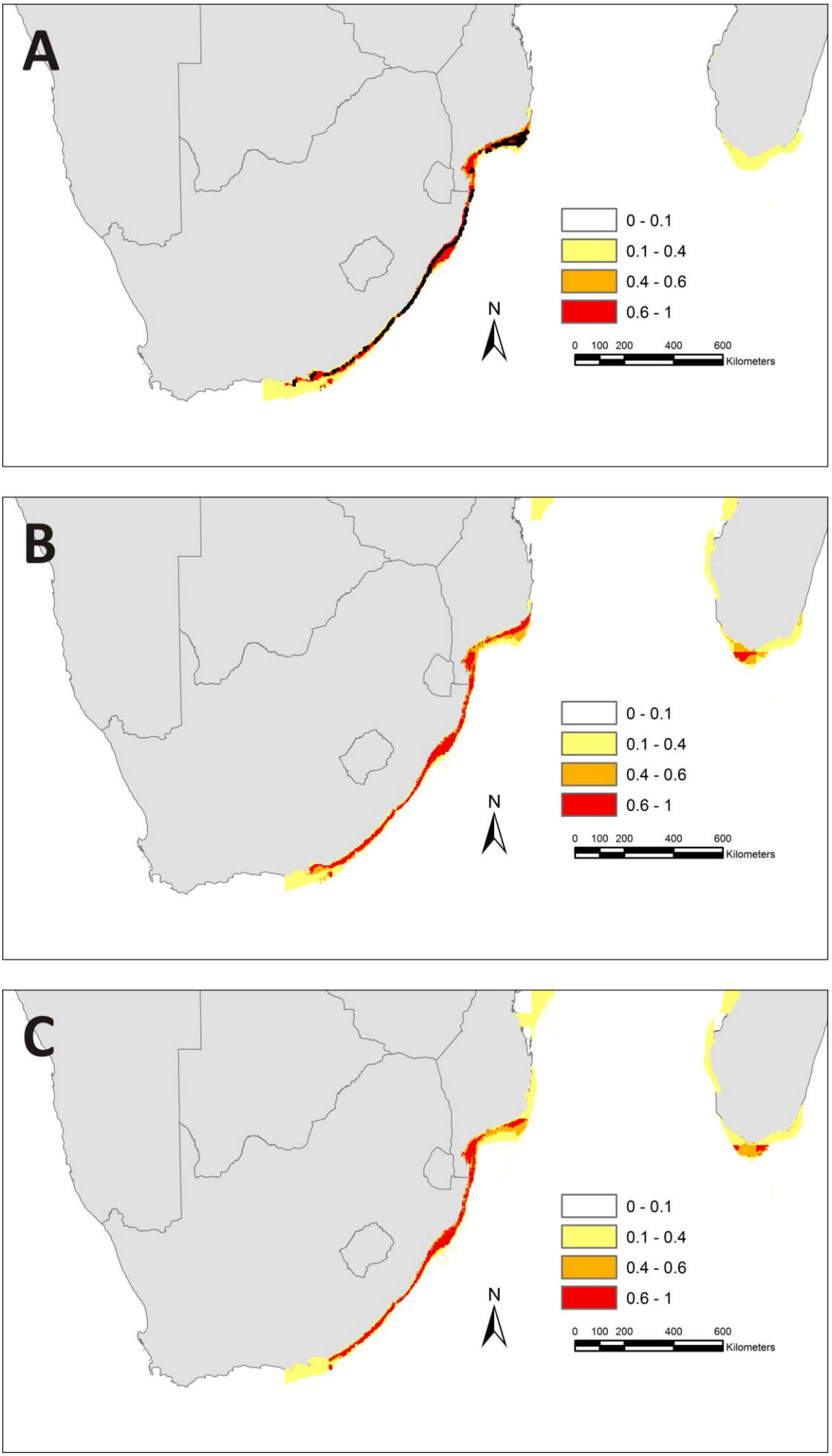


Figure 5.5: Probability mean weighted ensemble probability distributions of *E. andersoni* projected for the present (A), 2020 (B), and 2030 (C). Occurrences are plotted against the current projection.

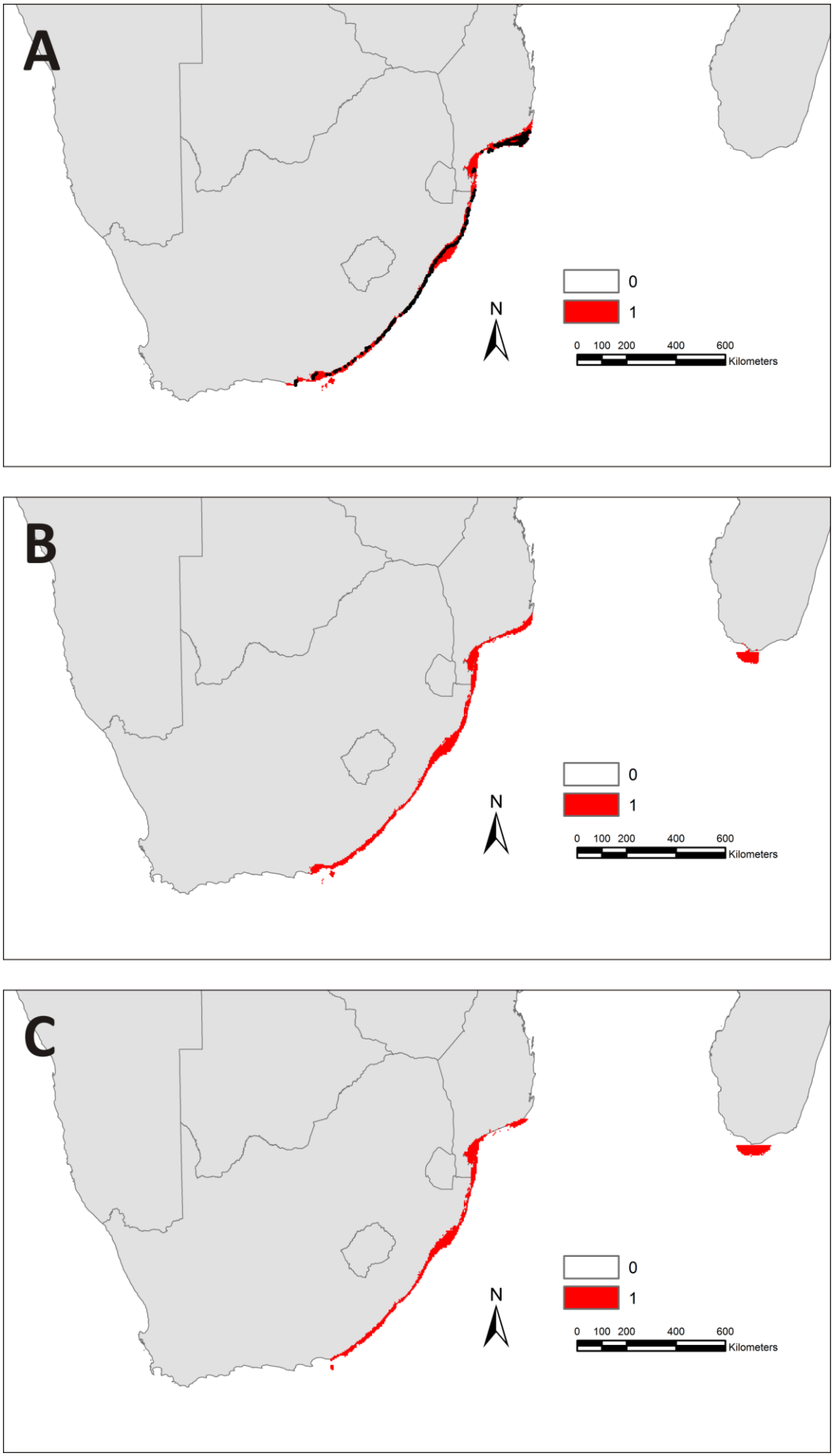


Figure 5.6: Binary probability mean weighted ensemble distributions of *E. andersoni* projected for the present (A), 2020 (B), and 2030 (C). Occurrences are plotted against the current projection.

Table 5.4: Summary of the range changes between the Probability Weighted Mean ensemble current and the 2020 and 2030 projections. Figures represent number of pixels except where percentages are indicated. Stable0 indicates the number pixels that are unoccupied in the current projection and that remain unoccupied in the future projections. Stable1 indicates the number of pixels that are occupied in the current projection and are predicted to remain occupied in the future.

	Current - 2020	Current - 2030
Loss	243	435
Stable0	25450	25449
Stable1	2402	2210
Gain	1	2
%Loss	9.187	16.446
%Gain	0.038	0.076
Range Change	-9.149	-16.371
Current Range	2645	2645
Future Range Size		
No Dispersal	2402	2210
Full Dispersal	2403	2212

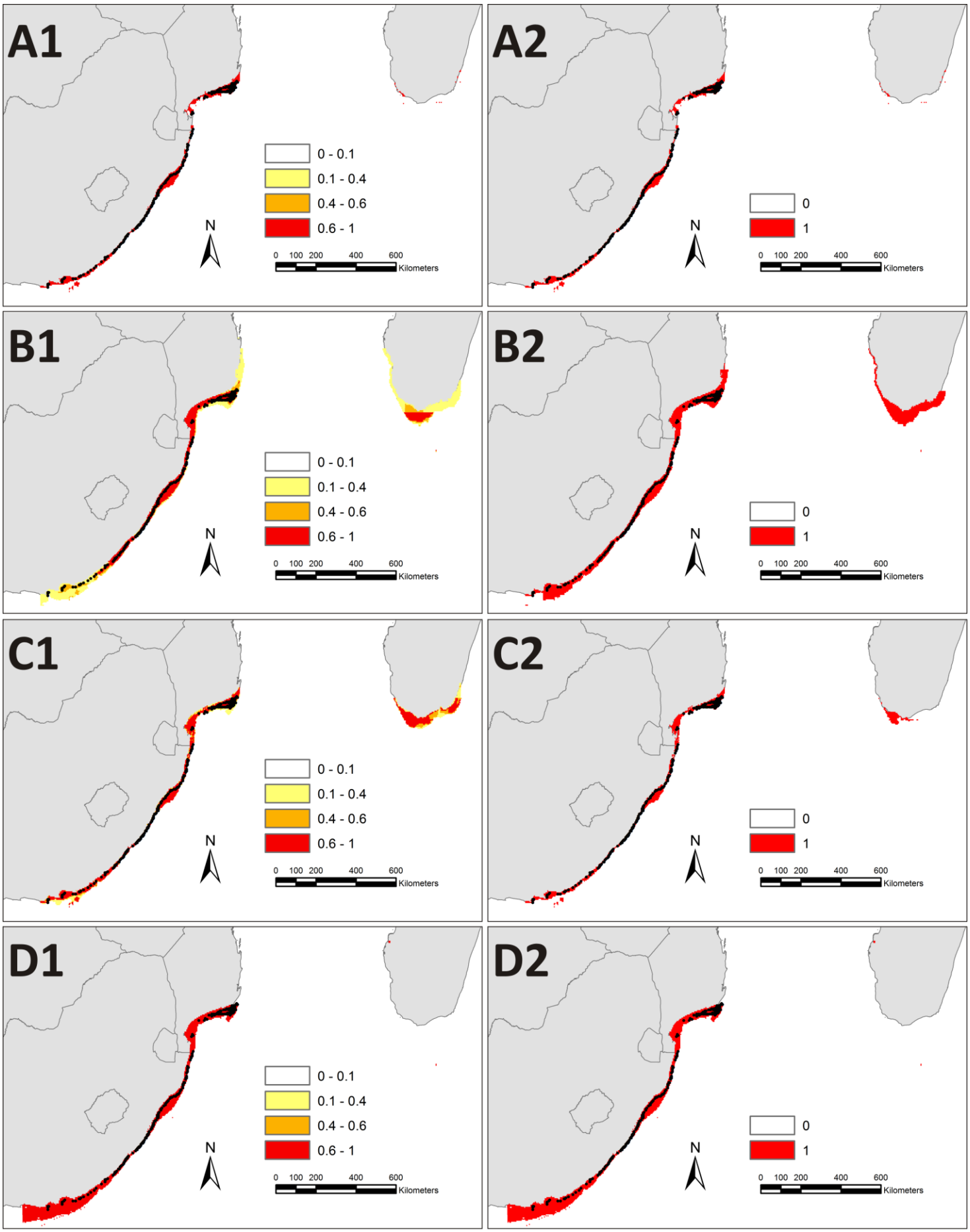
Individual model results

Current

The results of the individual models reflect their accuracy. For example, it is noticeable from the output of the MARS model that this model did not perform well since the projected distribution includes almost the entire area that has background environmental data (Figure 5.7 E2). This result is reinforced by the low performance of this model algorithm as measured by the three model evaluation methods (Table 5.3).

Future

The individual model projections for 2020 (Figure 5.8) and 2030 (Figure 5.9), mostly predict range contractions except for the GAM and GLM models which projected total range expansions of 3.443% and 71.352% respectively (Table 5.5). Despite not all models showing a Madagascan distribution of *E. andersoni* for the current projection, the 2020 and 2030 projections all predicted at least some distribution of *E. andersoni* on the Madagascan coastline.


A1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
A2
0
1
N
0 100 200 400 600
Kilometers
B1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
B2
0
1
N
0 100 200 400 600
Kilometers
C1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
C2
0
1
N
0 100 200 400 600
Kilometers
D1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
D2
0
1
N
0 100 200 400 600
Kilometers

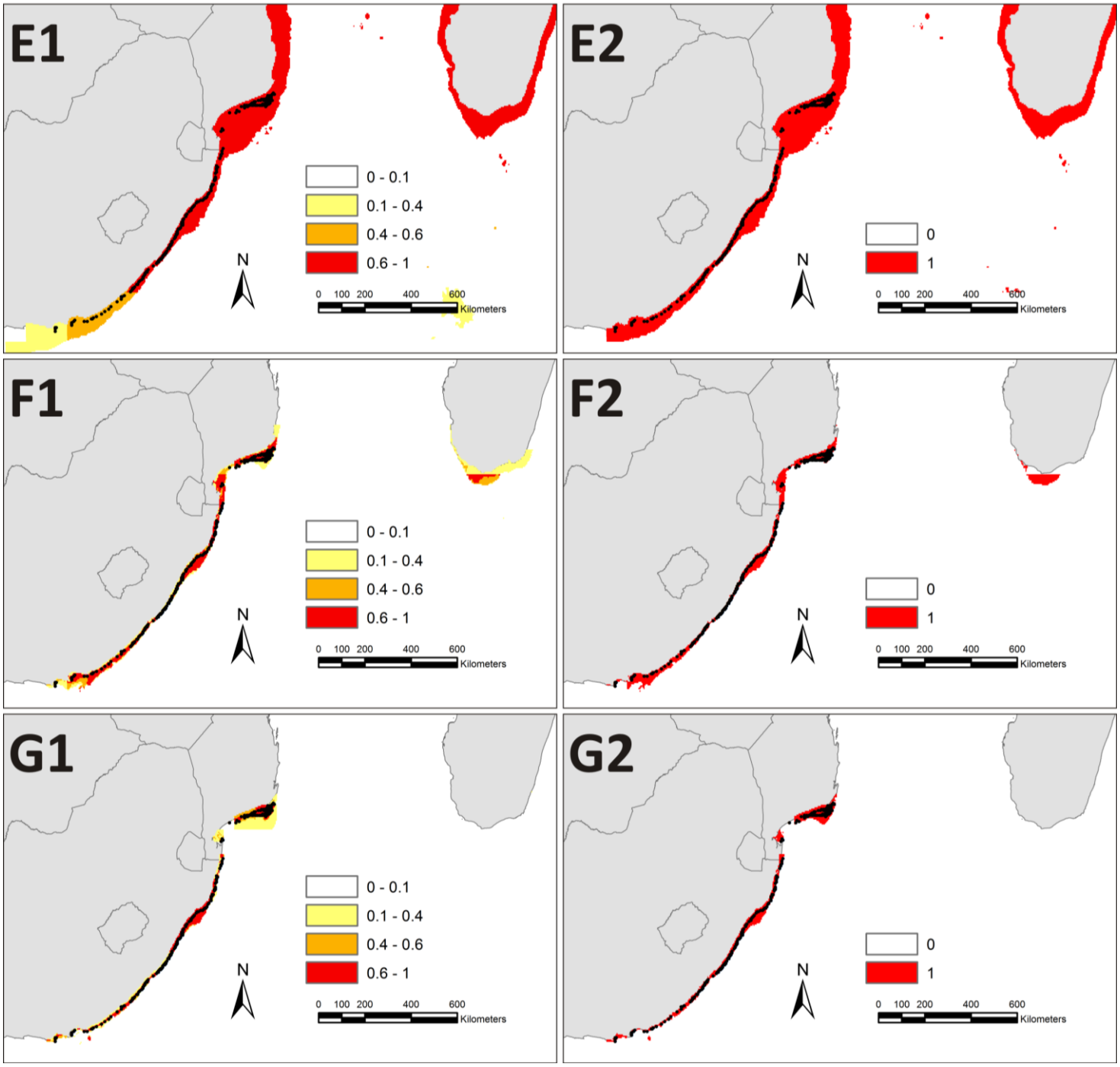
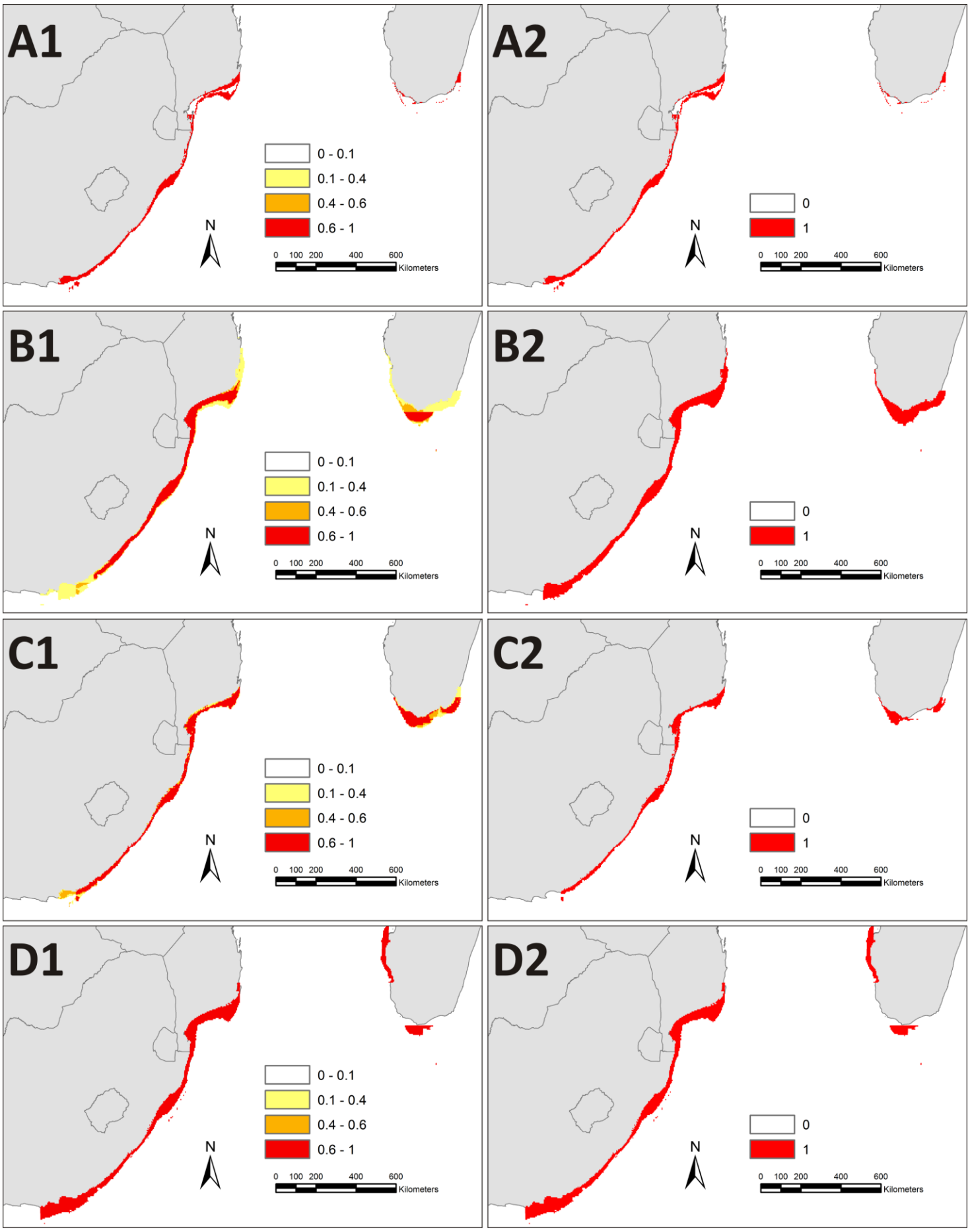


Figure 5.7: Individual model projections for the current *E. andersoni* distribution using 2100 pseudo-absences projected randomly. The probability of occurrence distributions are shown on the left (all figures labelled 1) and the binary transformed distributions are shown on the right (all figures labelled 2). The results of the following models are shown: A) CTA, B) GAM, C) GBM, D) GLM, E) MARS, F) MAXENT, G) RF.


A1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
A2
0
1
N
0 100 200 400 600
Kilometers
B1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
B2
0
1
N
0 100 200 400 600
Kilometers
C1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
C2
0
1
N
0 100 200 400 600
Kilometers
D1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
D2
0
1
N
0 100 200 400 600
Kilometers

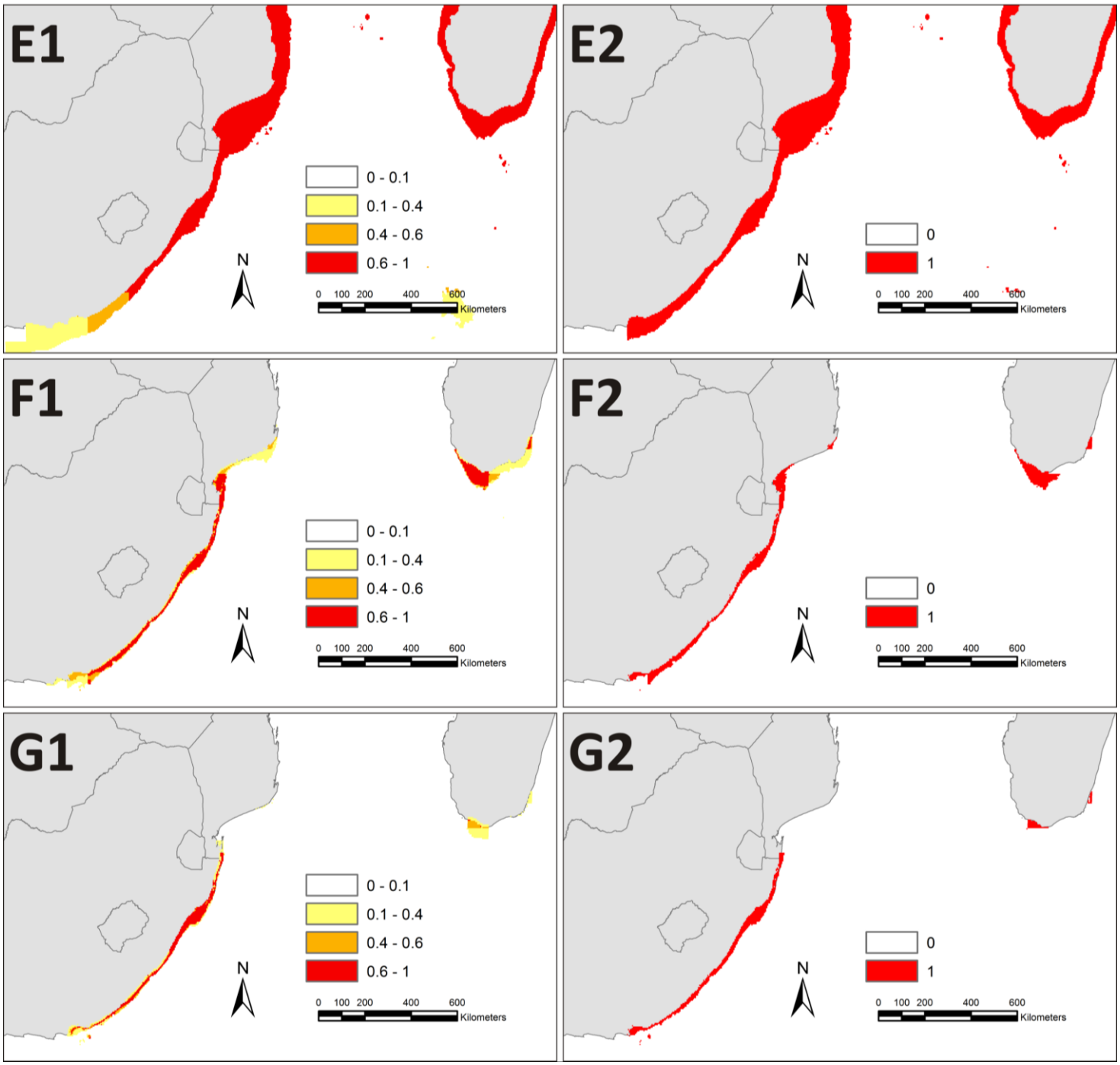
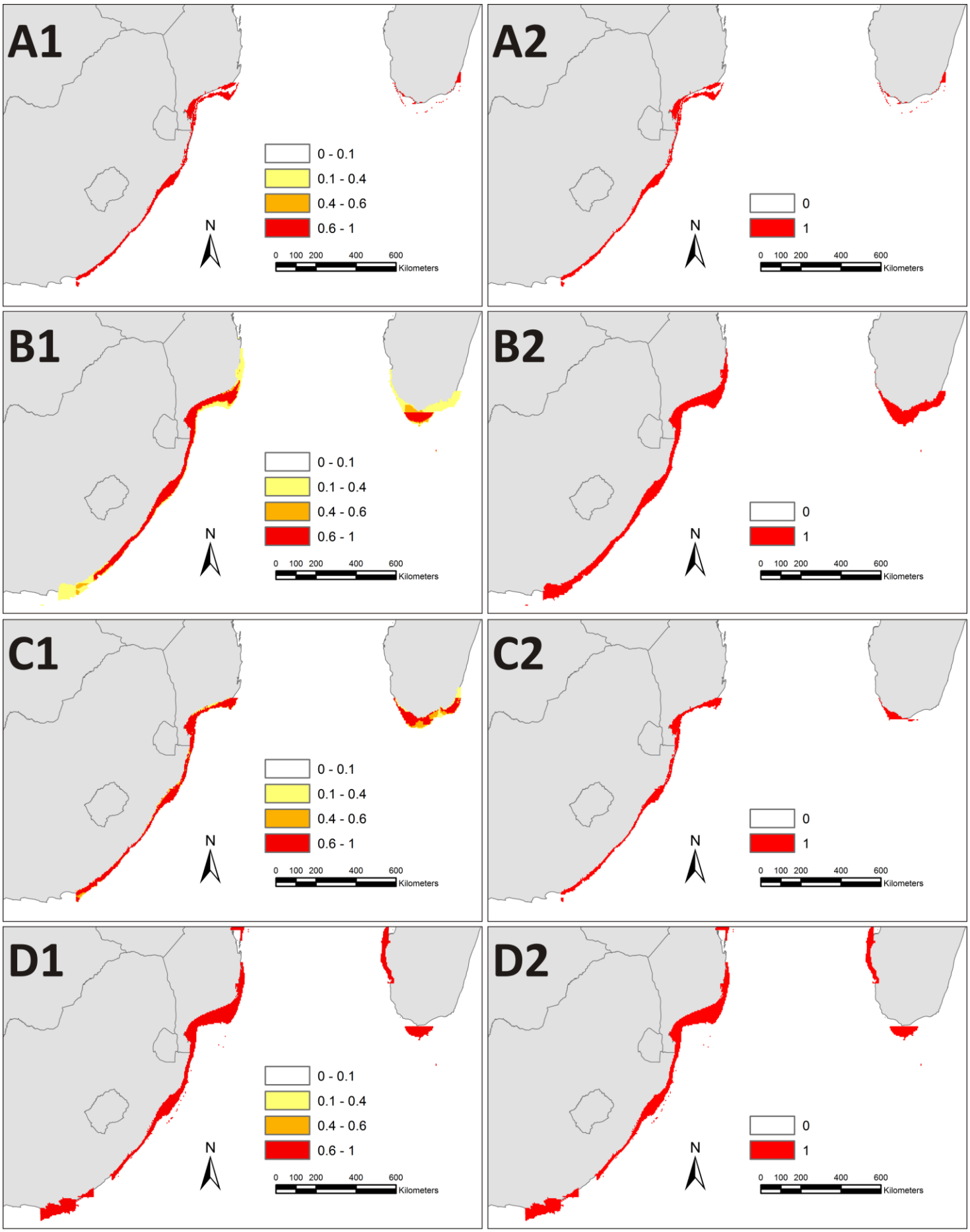


Figure 5.8: Individual model projections for the 2020 *E. andersoni* distribution using 2100 pseudo-absences projected randomly. The probability of occurrence distributions are shown on the left (all figures labelled 1) and the binary transformed distributions are shown on the right (all figures labelled 2). The results of the following models are shown: A) CTA, B) GAM, C) GBM, D) GLM, E) MARS, F) MAXENT, G) RF.


A1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
A2
0
1
N
0 100 200 400 600
Kilometers
B1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
B2
0
1
N
0 100 200 400 600
Kilometers
C1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
C2
0
1
N
0 100 200 400 600
Kilometers
D1
0 - 0.1
0.1 - 0.4
0.4 - 0.6
0.6 - 1
N
0 100 200 400 600
Kilometers
D2
0
1
N
0 100 200 400 600
Kilometers

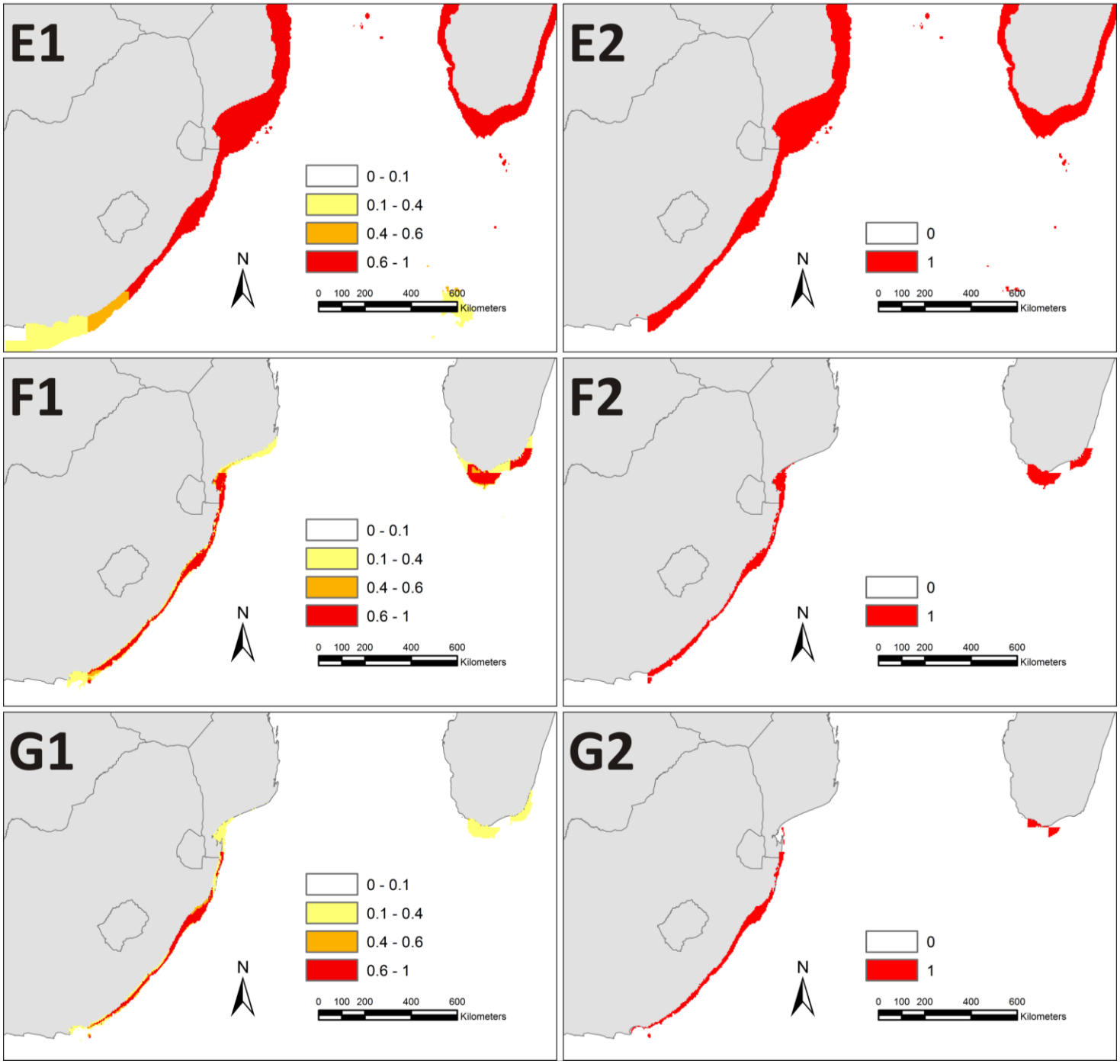


Figure 5.9: Individual model projections for the 2030 *E. andersoni* distribution using 2100 pseudo-absences projected randomly. The probability of occurrence distributions are shown on the left (all figures labelled 1) and the binary transformed distributions are shown on the right (all figures labelled 2). The results of the following models are shown: A) CTA, B) GAM, C) GBM, D) GLM, E) MARS, F) MAXENT, G) RF.

Table 5.5: Range changes between the current and the 2020, and the current and the 2030 binary transformed distributions for each model. Figures represent number of pixels except where percentages are indicated. Stable0 indicates the number of pixels that are unoccupied in the current projection and that remain unoccupied in the future projections. Stable1 indicates the number of pixels that are occupied in the current projection and are predicted to remain occupied in the future.

										Future Range Size	
		Loss	Stable0	Stable1	Gain	%Loss	%Gain	Range Change	Current Range	No Dispersal	Full Dispersal
Current - 2020	MAXENT	408	26177	1296	215	23.944	12.617	-11.326	1704	1296	1511
	CTA	245	26663	973	215	20.115	17.652	-2.463	1218	973	1188
	RF	500	26657	776	163	39.185	12.774	-26.411	1276	776	939
	MARS	262	17077	10757	0	2.378	0	-2.378	11019	10757	10757
	GAM	103	24759	3063	171	3.253	5.401	2.148	3166	3063	3234
	GLM	448	23951	2177	1520	17.067	57.905	40.838	2625	2177	3697
	GBM	204	26562	1164	166	14.912	12.135	-2.778	1368	1164	1330
Current - 2030	MAXENT	566	26081	1138	311	33.216	18.251	-14.965	1704	1138	1449
	CTA	397	26591	821	287	32.594	23.563	-9.031	1218	821	1108
	RF	526	26674	750	146	41.223	11.442	-29.781	1276	750	896
	MARS	579	17059	10440	18	5.255	0.163	-5.091	11019	10440	10458
	GAM	133	24688	3033	242	4.201	7.644	3.443	3166	3033	3275
	GLM	700	22898	1925	2573	26.667	98.019	71.352	2625	1925	4498
	GBM	301	26646	1067	82	22.003	5.994	-16.009	1368	1067	1149