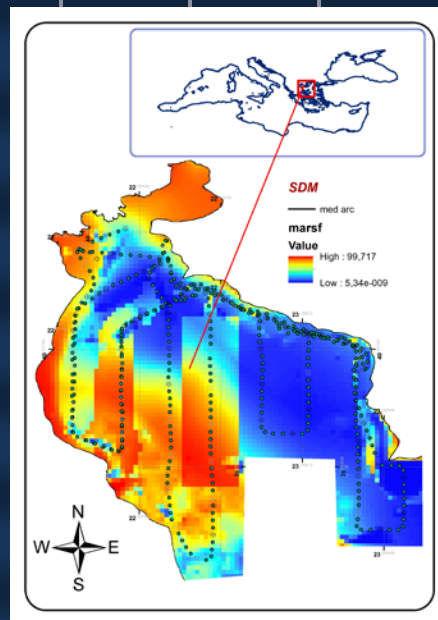




UNIVERSITY OF CRETE
DEPARTMENT OF BIOLOGY



Μοντέλα κατανομής ιχθυομάζας πελαγικών ειδών:
Αξιολόγηση των διαδομένων τεχνικών, εισαγωγή νέων, έλεγχος της αποτελεσματικότητας τους σε αλιευτικά δεδομένα, διερεύνηση σχέσεων ειδών-περιβάλλοντος



Διδακτορική Διατριβή

ΠΑΛΙΑΛΕΞΗΣ ΑΝΔΡΕΑΣ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΚΑΙ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

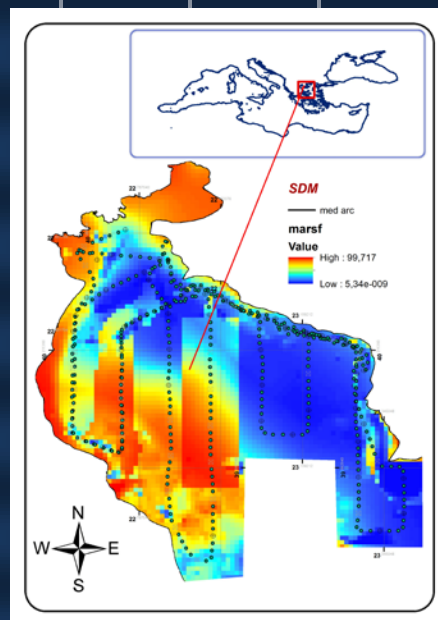
Ηράκλειο 2011



UNIVERSITY OF CRETE
DEPARTMENT OF BIOLOGY



Small pelagic species distribution models: Evaluating of well established modelling techniques, introducing new innovative approaches, assessing their efficiency to fisheries data, and identifying species-environment relationships



PhD Thesis

PALIALEXIS ANDREAS

UNIVERSITY OF CRETE
FACULTY OF SCIENCES AND ENGINEERING
DEPARTMENT OF BIOLOGY

Heraklion Greece 2011

"Μοντέλα κατανομής ιχθυομάζας πελαγικών ειδών: Αξιολόγηση των διαδεδομένων τεχνικών, εισαγωγή νέων, έλεγχος της αποτελεσματικότητας τους σε αλιευτικά δεδομένα, διερεύνηση σχέσεων ειδών-περιβάλλοντος"

Παλιαλέξης Ανδρέας

Διδακτορική Διατριβή

2011

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΚΑΙ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΕΛΛΗΝΙΚΟ ΚΕΝΤΡΟ ΘΑΛΑΣΣΙΩΝ ΕΡΕΥΝΩΝ, ΙΝΣΤΙΤΟΥΤΟ ΘΑΛΑΣΣΙΩΝ ΒΙΟΛΟΓΙΚΩΝ ΠΟΡΩΝ

Επιβλέπων διδακτορικής διατριβής

Καθηγητής Καρακάσης Ιωάννης, Τμήμα Βιολογίας Πανεπιστημίου Κρήτης

Τριμελής Συμβουλευτική επιτροπή

Καρακάσης Ιωάννης, Καθηγητής, Τμήμα Βιολογίας Πανεπιστημίου Κρήτης

Γεωργακαράκος Ευστράτιος, Επίκουρος Καθηγητής, Τμήμα Επιστημών της Θάλασσας Πανεπιστημίου Αιγαίου

Λύκα Κωνσταντίνα, Επίκουρη Καθηγήτρια, Τμήμα Βιολογίας Πανεπιστημίου Κρήτης

Επταμελής Εξεταστική Επιτροπή

Καρακάσης Ιωάννης, Καθηγητής, Τμήμα Βιολογίας Πανεπιστημίου Κρήτης

Γεωργακαράκος Ευστράτιος, Επίκουρος Καθηγητής, Τμήμα Επιστημών της Θάλασσας Πανεπιστημίου Αιγαίου

Λύκα Κωνσταντίνα, Επίκουρη Καθηγήτρια, Τμήμα Βιολογίας Πανεπιστημίου Κρήτης

Κεντούρη Μαρουδιώ, Καθηγήτρια, Τμήμα Βιολογίας Πανεπιστημίου Κρήτης

Παυλίδης Μιχάλης, Επίκουρος Καθηγητής, Τμήμα Βιολογίας Πανεπιστημίου Κρήτης

Γιαννουλάκη Μαριάννα, Ερευνήτρια Β', Ελληνικό Κέντρο Θαλάσσιων Ερευνών

Βασιλοπούλου Βασιλική, Ερευνήτρια Α', Ελληνικό Κέντρο Θαλάσσιων Ερευνών

Ημερομηνία Δημόσιας Παρουσίασης Διατριβής: 31-10-2011

Στην οικογένεια μου,
ιδιαίτερα στην Άντα και στο γιό μου.



Πρόλογος

Τα Μοντέλα Κατανομής Ειδών (ΜΚΕ) αποτελούν ένα ραγδαία αναπτυσσόμενο τομέα στο χώρο της οικολογίας και βιογεωγραφίας, όπου η συνεισφορά τους αναγνωρίζεται διαρκώς και αντίστοιχα αυξάνει ο αριθμός των σχετικών δημοσιεύσεων. Η παρούσα διατριβή επιχειρεί να συνεισφέρει:

Α) στον τομέα των Μοντέλων Κατανομής Ειδών, την ανάπτυξη νέων εφαρμογών και την αξιολόγηση τους, όπως και σε ζητήματα που απασχολούν την ορθή χρήση και διαχείριση τους

Β) στον τομέα της θαλάσσιας βιολογίας με την ανάδειξη της ικανότητας τους να αποτυπώνουν πολύπλοκες σχέσεις ειδών και περιβάλλοντος

Γ) στην αξιολόγηση καθιερωμένων και νέων προσεγγίσεων μοντελοποίησης μέσα από μια ευρεία σύγκριση τεχνικών για τον προσδιορισμό της καταλληλότητας τους ανάλογα με τη διαθεσιμότητα διαφορετικών τύπων δεδομένων και ανάλογα με το στόχο της διεξαγόμενης μελέτης.

Η διατριβή διεξήχθη στο εργαστήριο Θαλάσσιων Γεωγραφικών Πληροφοριακών Συστημάτων του Ινστιτούτου Θαλάσσιων Βιολογικών Πόρων του Ελληνικού Κέντρου Θαλάσσιων Ερευνών σε συνεργασία με το τμήμα Βιολογίας του Πανεπιστημίου Κρήτης. Η εμβάθυνση στα μοντέλα κατανομής ειδών και στη χρήση τους για τον προσδιορισμό σημαντικών ενδιαιτημάτων ψαριών (essential fish habitats) πραγματοποιήθηκε με το πρόγραμμα του FP6: 2004-SSP-4 DG Fisheries and Maritime Affairs: EnviEFH: Environmental Approach to Essential Fish Habitat Designation μεταξύ 2005-2007, του οποίου συντονιστής ήταν ο Β. Βαλαβάνης. Στη συνέχεια υπήρξε καταλυτική η συμμετοχή στο σεμινάριο Habitat Modelling of Marine Population, IPIMAR, Lisbon, Portugal και η προσωπική επαφή με ερευνητές που ειδικεύονται στο χώρο των ΜΚΕ όπως οι J. Leathwick, A. Lehmann, B. Planque, αλλά και οι επαφές με ερευνητές που αφιέρωσαν πολύτιμο χρόνο για την ορθή εφαρμογή των τεχνικών που ειδικεύονταν όπως οι I. Tetko, Σ. Γεωργακαράκος και Μ. Γιαννουλάκη. Η πρόταση της διατριβής και ο ορισμός της τριμελούς επιτροπής πραγματοποιήθηκε το Φεβρουάριο του 2008 στα πλαίσια του μεταπτυχιακού προγράμματος "Περιβαλλοντική Βιολογία. Διαχείριση χερσαίων και υδάτινων βιολογικών πόρων" του τμήματος Βιολογίας του Πανεπιστημίου Κρήτης.

Παλιαλέξης Ανδρέας

Νοέμβριος 2011

Ευχαριστίες

Ιδιαίτερα θερμές ευχαριστίες οφείλω στην τριμελή επιτροπή για την προσφορά της στην ολοκλήρωση της διατριβής. Συγκεκριμένα στον επιβλέπων Καθηγητή Καρακάση Ιωάννη, για τις πολύτιμες συμβουλές του και άμεση ανταπόκριση σε κάθε ζήτημα που προέκυπτε. Οφείλω πολλά στον επίκουρο καθηγητή Γεωργακαράκο Ευστράτιο καταρχήν για τη προσφορά των δεδομένων πάνω στα οποία βασίστηκε η διατριβή, αλλά και την αμέριστη υποστήριξη και συνεργασία καθ' όλη τη διάρκεια και τις υποδείξεις και υλικό που μου παρείχε στα θέματα όπου ειδικεύεται. Η παιδαγωγική ικανότητα του, η αμεσότητα του και η εύκολη μετάδοση των ειδικών γνώσεων του σε καίρια ζητήματα της διατριβής θεωρώ ότι βελτίωσαν σημαντικά την ποιότητά της και την ταχύτητα περαίωσης της. Ιδιαίτερες ευχαριστίες οφείλω επίσης στην, Επίκουρη Καθηγήτρια, Κωνσταντίνα Λύκα για τις εύστοχες υποδείξεις της και τις νέες ιδέες της που με βοήθησαν να δω πιο σφαιρικά την επιστημονική προσέγγιση της διατριβής, καθώς επίσης και για την άριστη συνεργασία της και άμεσες αποκρίσεις της σε κάθε ζήτημα που προέκυπτε. Μετά την ολοκλήρωση της διατριβής και της δημόσιας παρουσίασης της θα μπορούσα να αναγνωρίσω την άριστη και ουσιαστική συνεισφορά της τριμελούς επιτροπής και το γεγονός ότι στο διάστημα περαίωσης της το επίπεδο συνεργασίας και επικοινωνίας ήταν πολύ υψηλό.

Ευχαριστώ ιδιαίτερα τα υπόλοιπα μέλη της επταμελούς εξεταστικής επιτροπής για την προσφορά τους στη βελτίωση της διατριβής και ιδιαίτερα του κειμένου, αλλά και για την ταχύτατη απόκριση τους ώστε να πραγματοποιηθεί η δημόσια παρουσίαση. Ειδικότερα, ευχαριστώ θερμά την Ερευνήτρια Β' Μαριάννα Γιαννουλάκη για το χρόνο που αφιέρωσε και τη διάθεση της να με κατευθύνει ως την τελική μορφή της διατριβής. Θεωρώ ότι η συμβολή της σε όλη τη διάρκεια της διατριβής βελτίωσε ποιοτικά το επίπεδο της έρευνας και του κειμένου. Ευχαριστώ ιδιαίτερα την Καθηγήτρια Μαρουδιώ Κεντούρη για την πολυετή επιτυχημένη συνεργασία που είχαμε από προπτυχιακό επίπεδο έως την ολοκλήρωση της διατριβής και για τα κίνητρα που μου παρείχε ως ακαδημαϊκός για να επιδιώκω την συνεχή βελτίωση του έργου μου. Πολύτιμη τέλος ήταν η προσφορά της Ερευνήτριας Α' Βασιλικής Βασιλοπούλου στη βελτίωση του κειμένου της διατριβής αλλά και στην άμεση απόκριση της σε όλα τα ζητήματα που προέκυψαν ως τη δημόσια παρουσίαση της διατριβής.

Ιδιαίτερα υπόχρεος είμαι απέναντι στον Βασίλη Βαλαβάνη, που ηγείται του εργαστηρίου Θαλάσσιων Γεωγραφικών Πληροφοριακών Συστημάτων του Ινστιτούτου Θαλάσσιων Βιολογικών Πόρων του Ελληνικού Κέντρου θαλάσσιων Ερευνών. Με φιλοξένησε στο εργαστήριο του για σχεδόν δέκα χρόνια, όπου δημιούργησε ιδανικές συνθήκες συνεργασίας, μου έδωσε πολλές ευκαιρίες για να εξελίξω τις γνώσεις μου και τις τεχνικές μου και μου μετέδιδε τις δικές του γνώσεις ώστε να μπορούμε να παρακολουθούμε και να συμμετέχουμε σε έρευνες αιχμής. Τα παραπάνω μετουσιώθηκαν σε αρκετές δημοσιεύσεις, συμμετοχές σε ερευνητικά προγράμματα και εργασίες σε συνέδρια. Η παρουσία του ήταν καταλυτική για την εξέλιξη μου αυτά τα χρόνια, αλλά και για την ολοκλήρωση της διατριβής.

Τέλος οφείλω να ευχαριστήσω την οικογένεια μου και το ευρύτερο περιβάλλον μου που μου στάθηκαν το διάστημα αυτό, δημιουργώντας τις καλύτερες συνθήκες για την ευκολότερη διεξαγωγή της δουλειάς μου.

Παλιαλέξης Ανδρέας

Νοέμβριος 2011

Συντομογραφίες

AIC	κριτήριο πληροφορίας του Akaike, Akaike's Information Criteria
ANN	τεχνητά νευρωνικά δίκτυα, artificial neural networks
ANNE	σύνολα τεχνητών νευρωνικών δικτύων, artificial neural networks ensemble
ASE	μέσο τυπικό σφάλμα, Average Standard Error
ASNN	συνδεδεμένα νευρωνικά δίκτυα, Associative Neural Networks
AUC	περιοχή κάτω από την καμπύλη, Area Under Curve
Ave	μέση τιμή, average
AverMod	Μοντέλο Ιδανικευμένης Πρόβλεψης
BIC	Μπαεσιανό κριτήριο πληροφορίας Bayesian Information Criteria
BIOCLIM	βιοκλιματικός φάκελος, bioclimatic envelope model
BRT	ενισχυμένα δέντρα παλινδρόμησης, Boosted Regression Trees
CART	δέντρα ταξινόμησης και παλινδρόμησης classification and regression trees
CHL	επιφανειακή συγκέντρωση χλωροφύλλης-α
COR	συσχέτιση, correlation
DCoast	απόσταση από την ακτή
DDND	χρονική κατηγορική μεταβλητή που εκφράζει σούρουπο, μέρα, δύση, βράδυ
DefKrig	το πρώτο μοντέλο που αναπτύχθηκε και επιλέχθηκε από τη μέθοδο Kriging
DEP	βάθος
DEPsl	κλίση του βυθού
DFA	Discriminant Function Analysis
DOMAIN	μέθοδος τύπου φακέλου για τη μοντελοποίηση κατανομής ειδών
DS1	σύνολο δεδομένων εκπαίδευσης
DS2	σύνολο δεδομένων κάθε σημείου στον κάρναβο της περιοχής μελέτης
DS3	σύνολο δεδομένων επαλήθευσης
ENFA	ανάλυση παράγοντα οικολογικού θώκου, ecological niche factor analysis
EnvDist	μέθοδος των περιβαλλοντικών αποστάσεων, Environmental Distance
EnvDistChe	μέθοδος των περιβαλλοντικών αποστάσεων με την εφαρμογή της απόστασης Chebyshev
EnvScore	Περιβαλλοντική βαθμολογία, Envelope Score
ESE	Πρόωρο σταμάτημα συνόλου, Early Stopping over Ensemble
FN	αρνητικό σφάλμα πρόβλεψης, false negative
FP	σφάλμα πρόβλεψης, false positive
GAM	γενικευμένα προσθετικά μοντέλα, generalized additive models

GAMM	Γενικευμένα Προσθετικά Μεικτά Μοντέλα, <i>generalized additive mixed models</i>
GARP	Algorithm for Rule-set Prediction
GARP	γενετικός αλγόριθμος <i>genetic algorithm of rule-set prediction</i>
GAS	λογισμικό γεωστατιστικής ανάλυσης, <i>Geostatistical Analyst Software</i>
GIS	Γεωγραφικών Πληροφοριακών Συστημάτων, <i>Geographical Information Systems</i>
GLM	γενικευμένα γραμμικά μοντέλα, <i>generalized linear models</i>
IDW	στάθμισης με αντίστροφη απόσταση, <i>Inverse Distance Weighted</i>
Krig	το δεύτερο μοντέλο που αναπτύχθηκε και επιλέχθηκε από τη μέθοδο <i>Kriging</i>
MAE	μέσο απόλυτο σφάλμα, <i>Mean Absolute Error</i>
MARS	<i>Multivariate Analysis and Regression Splines</i>
MAXENT	μοντέλο μέγιστης εντροπίας, <i>maximum entropy</i>
ME	μέσο σφάλμα, <i>Mean Error</i>
MSqE	τετράγωνο του μέσου σφάλματος, <i>Mean Square Error</i>
MStE	κανονικοποιημένο μέσο σφάλμα <i>Mean Standardized Error</i>
PAR	διαθέσιμη φωτοσυνθετική ακτινοβολία
Q-Q plots	γραφήματα <i>Quantile-Quantile</i>
RMSE	ρίζα του μέσου τετραγώνου του σφάλματος πρόβλεψης, <i>root-mean-square prediction error</i>
RMSqSt	μέση τετραγωνική ρίζα του τετραγώνου των κανονικοποιημένων σφαλμάτων, <i>root-mean-square standardized prediction error</i>
ROC	<i>Receiver Operating Characteristics</i>
ROC-AUC	<i>Receiver Operating Characteristics και Area Under Curve</i>
s_A	συντελεστής ακουστικής ανάκλασης ανά μονάδα επιφάνειας, <i>area backscattering coefficient</i>
SLA	υψομετρία ή αλτιμετρία <i>sea level anomaly</i>
SST	επιφανειακή θερμοκρασία της θάλασσας, <i>sea surface temperature</i>
SSTsl	κλίση της επιφανειακής θερμοκρασίας της θάλασσας, <i>sea surface temperature slope</i>
StDev	τυπική απόκλιση, <i>standard deviation</i>
SVM	μηχανές υποστήριξης διανυσμάτων, <i>Support Vector Machines</i>
SVM-Nu	μηχανές υποστήριξης διανυσμάτων, <i>Support Vector Machines</i>
VCCLAB	<i>Virtual Computational Chemistry Laboratory</i>
MKE	μοντέλα κατανομής ειδών, <i>SDM Species Distribution Models</i>

I. Περιεχόμενα

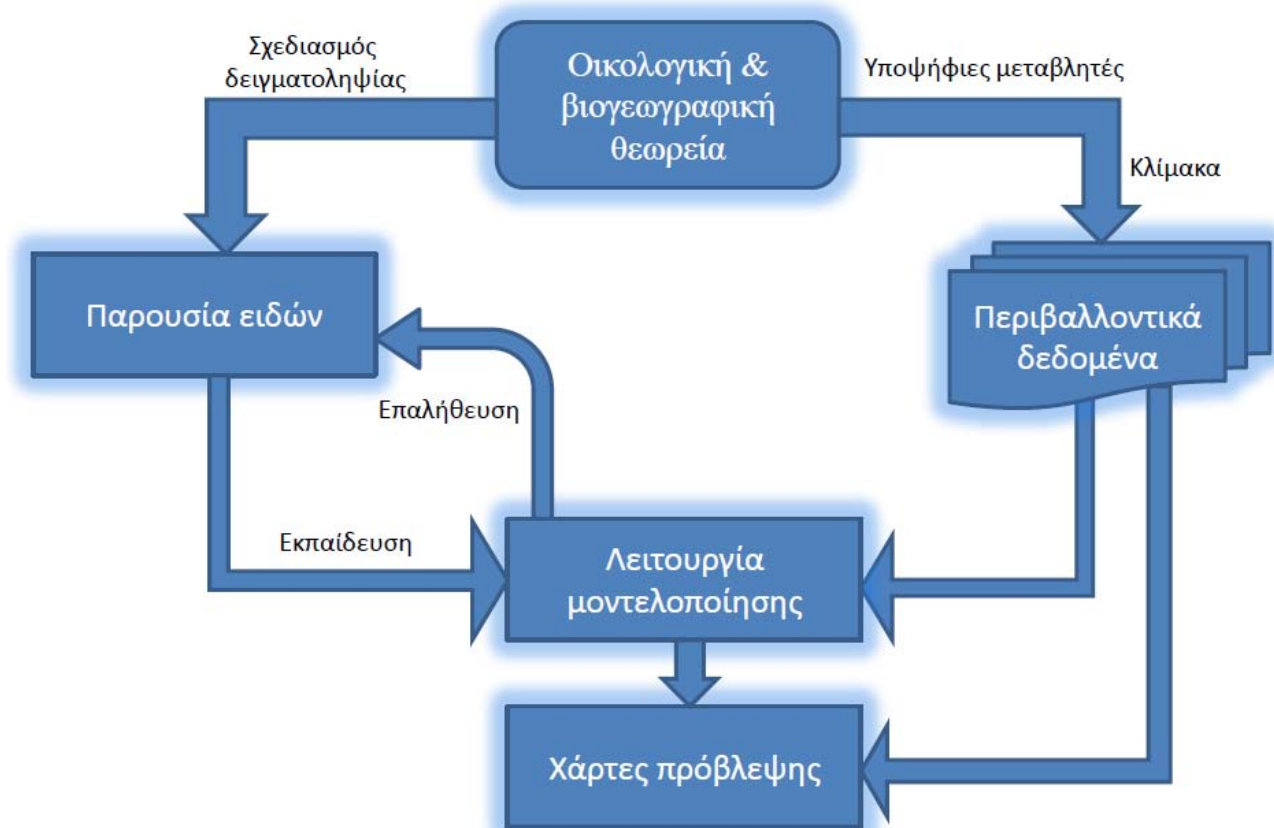
	Περιεχόμενα	i
	Συνοπτομογραφίες	iii
1.	Εισαγωγή	1
1.1	Σημασία και χρήση των μοντέλων κατανομής ατόμων ειδών	1
1.2	Η οικολογική θεωρία πίσω από τα ΜΚΕ	3
1.3	Ανάπτυξη των μοντέλων κατανομής ειδών	6
1.4	Εφαρμογή μοντέλων κατανομής ειδών σε υδροακουστικά δεδομένα μικρών πελαγικών ειδών ψαριών	8
1.5	Σκοπός της διατριβής	9
2.	Υλικά και μέθοδοι	12
2.1	Περιοχή μελέτης	12
2.2	Δεδομένα	13
2.2.1	Δεδομένα αφθονίας πελαγικών ειδών	13
2.2.1.1	Δημιουργία χρονοσειρών σε κάθε σύνολο δεδομένων και κριτήρια σύνθεσης του συνόλου επαλήθευσης	14
2.2.2	Περιβαλλοντικά και χωροχρονικά δεδομένα	14
2.3	Μοντέλα κατανομής ειδών	17
2.3.1	Μοντέλα αφθονίας	18
2.3.1.1	Μοντέλα παλινδρόμησης	18
2.3.1.2	Συνδεδετικά Νευρωνικά Δίκτυα (Associative Neural Networks)	19
2.3.1.3	Kriging	19
2.3.1.4	Μοντέλο Ιδανικευμένης Πρόβλεψης	20
2.3.2	Μοντέλα παρουσίας-απουσίας και μόνο-παρουσίας	21
2.4	Μέθοδοι επαλήθευσης των μοντέλων κατανομής ειδών	25
2.4.1	Δεδομένα αφθονίας	25
2.4.2	Δεδομένα παρουσίας-απουσίας	26
3.	Αποτελέσματα	29
3.1	Μοντέλα Αφθονίας	29
3.1.1	Προβλέψεις κατανομής ειδών	29
3.1.1.1	Μοντέλα παλινδρόμησης	29
3.1.1.2	Συνδυασμένα νευρωνικά δίκτυα	30
3.1.1.3	Kriging	31
3.1.1.4	Μοντέλο Ιδανικευμένης Πρόβλεψης	32
3.1.2.	Σύγκριση των μοντέλων αφθονίας	33
3.1.2.1	Προσαρμογή των μοντέλων στα δεδομένα εκπαίδευσης	33
3.1.2.2	Σύγκριση των προβλέψεων του συνόλου επαλήθευσης	35
3.1.2.3	Σύγκριση των καννάβων πρόβλεψης κατανομής των ειδών	37
3.2	Μοντέλα παρουσίας-απουσίας	41
3.2.1	Χωρικά πρότυπα και σύγκριση μεταξύ συνόλου δεδομένων εκπαίδευσης και επαλήθευσης	41
3.2.2	Απόδοση Εφαρμογής	42
3.2.3	Ικανότητα πρόβλεψης	44
3.2.4	Χάρτες Πιθανοτήτων	45
4.	Συζήτηση	51
4.1	Απόδοση μοντέλων αφθονίας	52
4.1.1	Σύγκριση και αξιολόγηση των τεχνικών μοντελοποίησης δεδομένων αφθονίας	55
4.2	Απόδοση μοντέλων που εκπαιδεύονται με δεδομένα παρουσίας-απουσίας ή μόνο-παρουσίας	57
4.2.1	Σύγκριση των ΜΚΕ για τα δεδομένα παρουσίας-απουσίας και μόνο-παρουσίας	58
4.2.2	Αξιολόγηση των ΜΚΕ για τα δεδομένα παρουσίας-απουσίας και μόνο-παρουσίας	60
4.3	Διαδικασία επαλήθευσης ΜΚΕ	64
4.4	Χωρικά πρότυπα στα δεδομένα	65

4.5	Περιβαλλοντικοί παράγοντες που επηρεάζουν την κατανομή των μικρών πελαγικών ψαριών	66
5.	Συμπεράσματα	69
5.1	Μοντέλα αφθονίας	69
5.2	Μοντέλα παρουσίας-απουσίας	69
6.	Βιβλιογραφία	71
	Περίληψη/Summary	84
7.	Παράρτημα I	86
8.	Παράρτημα II	88
9.	Παράρτημα III	128
10.	Παράρτημα IV	139

1. Εισαγωγή

1.1 Σημασία και χρήση των μοντέλων κατανομής ατόμων ειδών

Η γνώση της οικολογικής και γεωγραφικής κατανομής των ειδών είναι απαραίτητη για το σχεδιασμό της διατήρησης των πληθυσμών (Ferrer *et al.*, 2002) καθώς και για την αξιολόγηση της επίδρασης των εξελικτικών παραγόντων στα χωρικά πρότυπα της βιοποικιλότητας (Graham *et al.*, 2006). Η ανάγκη για επιστημονικά τεκμηριωμένες μελέτες, σχετικά με το χαρακτηρισμό των θαλάσσιων οικοσυστημάτων, τη διακύμανση της βιοποικιλότητας και αφθονίας των ειδών και τις επιπτώσεις της αλλαγής του κλίματος στα θαλάσσια οικοσυστήματα, έχει οδηγήσει στην αύξηση των μελετών που αφορούν στην πρόβλεψη της κατανομής των ειδών. Η πρόβλεψη της κατανομής των ειδών μπορεί να πραγματοποιηθεί με τη χρήση πολλών και διαφορετικών προσεγγίσεων για τη μοντελοποίηση των ενδιαιτημάτων και την αναγνώριση της ετερογένειας τους (Guisan *et al.*, 2002; Elith *et al.*, 2006; Valavanis *et al.*, 2008). Πολλές τεχνικές έχουν αναπτυχθεί για τον εντοπισμό της κατανομής των ειδών, χρησιμοποιώντας δειγματοληπτικά δεδομένα. Αυτές αναφέρονται ως **μοντέλα κατανομής ειδών (ΜΚΕ)** και πρόκειται για στατιστικά μοντέλα που συσχετίζουν δειγματοληπτικά δεδομένα κατανομής ειδών με τα αντίστοιχα περιβαλλοντικά και γεωγραφικά χαρακτηριστικά της περιοχής δειγματοληψίας (Elith & Leathwick, 2009). Η Εικόνα 1.1 παρουσιάζει ένα διάγραμμα που περιλαμβάνει όλα τα στάδια που συνθέτουν την εφαρμογή ενός ΜΚΕ, από την οικολογική θεωρία ως τη δημιουργία των χαρτών πρόβλεψης της κατανομής των ατόμων ειδών. Τα εργαλεία αυτού του τύπου θα πρέπει να είναι εύκολα προσβάσιμα στους διαχειριστές και τους ενδιαφερόμενους φορείς που εμπλέκονται στη διαδικασία λήψης αποφάσεων και το σχεδιασμό και οργάνωση διαχειριστικών σχεδίων.



Εικόνα 1.1 Διάγραμμα που παρουσιάζει τα συστατικά της μοντελοποίησης της κατανομής ατόμων ειδών (τροποποιημένο από Franklin, 2009).

Η ανάπτυξη ισχυρών αναλυτικών εργαλείων και η εξέλιξη των Γεωγραφικών Πληροφοριακών Συστημάτων (GIS) και της Τηλεπισκόπησης, σε συνδυασμό με τη διαθεσιμότητα εκτεταμένων χρονοσειρών, αποτελούν ένα πολλά υποσχόμενο υπόβαθρο για την εξέλιξη των στατιστικών τεχνικών μοντελοποίησης και την εξαγωγή ρεαλιστικών αποτελεσμάτων. Η εύκολη πρόσβαση σε δορυφορικά δεδομένα τα οποία καλύπτουν εκτεταμένες γεωγραφικές περιοχές αποτελεί το βασικότερο λόγο για την ολοένα και ευρύτερη χρήση των ΜΚΕ.

Οι θεωρητικοί λόγοι που έχουν κάνει απαραίτητη τη χρήση και εξέλιξη των ΜΚΕ αφορούν κυρίως:

(α) Την κατανόηση των σχέσεων ενός είδους με βιοτικούς και αβιοτικούς παράγοντες χρησιμοποιώντας δειγματοληπτικά δεδομένα παρουσίας του είδους. Οι σχέσεις αυτές χρησιμοποιούνται στα πλαίσια οικολογικών θεωρήσεων και στην εξέταση οικολογικών ή βιογεωγραφικών υποθέσεων που σχετίζονται με το εύρος και την κατανομή του είδους. Ο προσδιορισμός του προτύπου κατανομής και αφθονίας ενός είδους και η κλίμακα επίδρασης από αβιοτικούς παράγοντες είναι πρωτίτως αντικείμενο της Οικολογίας (Elton, 1927; Scott *et al.*, 2002) και σε αυτές τις περιπτώσεις δεν απαιτείται πάντα η δημιουργία χαρτών πρόβλεψης.

(β) Τη δυνατότητα των μοντέλων κατανομής ειδών να παρεμβάλουν ή να εξάγουν τιμές αφθονίας ή παρουσίας των ειδών που προέρχονται από σημειακές παρατηρήσεις, στο χώρο ή στο χρόνο. Οι περιοχές παρεμβολής ή εξαγωγής αντιστοιχούν σε περιοχές, όπου δεν υπάρχουν δειγματοληπτικά δεδομένα. Οι χάρτες καταλληλότητας των ενδιαιτημάτων ή οι χάρτες κατανομής των ειδών είναι χρήσιμοι στους ελέγχους υπόθεσης για τα χαρακτηριστικά του εύρους κατανομής, τη διαμερισματοποίηση του οικολογικού θώκου ή το συντηρητισμό του.

Ταυτόχρονα οι ουσιαστικοί λόγοι που έχουν κάνει απαραίτητη την χρήση και εξέλιξη των ΜΚΕ είναι οι εξής:

- Σχεδιασμός βιολογικών αποθεμάτων και πλάνων διαχείρισης. Εδώ εντάσσονται και παγκόσμιες εφαρμογές που απαιτούν εκτιμήσεις βιοποικιλότητας. Τα ΜΚΕ μπορούν να χρησιμοποιηθούν στην ανάπτυξη λεπτομερών χαρτών πρόβλεψης για την καταλληλότητα ή ποιότητα των ενδιαιτημάτων (Wu & Smeins, 2000; Gibson *et al.*, 2004). Αυτοί οι χάρτες μπορούν να συγκριθούν με την αναμενόμενη χωρική κατανομή παραγόντων επίδρασης στα οικοσυστήματα, όπως η χρήση γης (π.χ. Wisz *et al.*, 2008).
- Οικολογική αποκατάσταση και οικολογική μοντελοποίηση. Τα ΜΚΕ έχουν χρησιμοποιηθεί για τον προσδιορισμό κατάλληλων περιοχών και ενδιαιτημάτων, ώστε να επιτευχθεί η επιτυχημένη επανεισαγωγή ειδών (Pearce & Lindenmayer, 1998; Schadt *et al.*, 2002; Carroll *et al.*, 2003; Hirzel *et al.*, 2004; Martínez-Meyer *et al.*, 2006).
- Κίνδυνοι και επιδράσεις των ειδών εισβολέων. Τα είδη εισβολείς έχουν μεγάλη οικονομική και οικολογική επίδραση στα οικοσυστήματα που εισέρχονται. Τα ΜΚΕ δύνανται να προβλέψουν τις πιθανές περιοχές, όπου τα είδη αυτά ενδέχεται να εγκαταστήσουν πληθυσμούς (Peterson & Vieglais, 2001; Peterson, 2003; Andersen *et al.*, 2004, Siapatis *et al.*, 2008).
- Επίδραση της κλιματικής αλλαγής στη βιοποικιλότητα και τα οικοσυστήματα (Heikkinen *et al.*, 2006; Pearson *et al.*, 2006; Thuiller *et al.*, 2006; Lawler *et al.*, 2009).

Ως εξαρτημένες μεταβλητές στα ΜΚΕ χρησιμοποιούνται συνήθως δεδομένα παρουσίας-απουσίας, δεδομένα μόνο-παρουσίας και δεδομένα αφθονίας, που προέρχονται από διάφορες στρατηγικές δειγματοληψίας. Οι Zaniewski *et al.* (2002) υποστήριξαν ότι η μοντελοποίηση που βασίζεται σε δεδομένα παρουσίας-απουσίας είναι πιο πιθανό να αντανakλά την παρούσα φυσική κατανομή των ειδών, δηλαδή τον πραγματικό οικολογικό θώκο, ενώ οι μέθοδοι που βασίζονται σε δεδομένα μόνο-παρουσίας προβλέπουν πιθανές κατανομές, πιο κοντά στον πιθανό ή θεωρητικό οικολογικό θώκο. Γενικότερα, τα ΜΚΕ έχουν κατά καιρούς χαρακτηριστεί ως "μοντέλα θώκου ειδών", "μοντέλα οικολογικού θώκου" ή ακόμα και "μοντέλα θεωρίας του θώκου". Αυτά τα μοντέλα έχουν περιγραφεί ως εκτιμητές του (α) θεωρητικού (πιθανού) οικολογικού

θώκου, (β) του πραγματικού θώκου, (γ) του πολυδιάστατου θώκου (Rotenberry *et al.*, 2006) ή και (δ) του κλιματικού θώκου, όταν το μοντέλο βασίζεται αποκλειστικά σε περιβαλλοντικές μεταβλητές. Διάφοροι ερευνητές διακρίνουν τα μοντέλα οικολογικού θώκου, που αντικατοπτρίζουν την πιθανή κατανομή των ειδών, από τα μοντέλα κατανομής ειδών, που αντικατοπτρίζουν την πραγματική κατανομή των ατόμων τους (Peterson *et al.*, 2008).

Για τη μοντελοποίηση ενδιαιτημάτων ή για την πρόβλεψη της κατανομής ειδών χρησιμοποιούνται συνήθως τεχνικές που ομαδοποιούνται ως (α) Λειτουργίες Επιλογής Πόρων (Resource Selection Function) (Valavanis *et al.*, 2008), (β) ποσοτικές σχέσεις δομής-δραστηκότητας (Quantitative Structure-Activity Relationships) (Tetko *et al.*, 1995), (γ) γεωστατιστική (Petitgas, 2001). Εκτεταμένες αναφορές σε μοντέλα κατανομής ειδών είναι διαθέσιμες σε πολυάριθμες σχετικές ανασκοπήσεις (Guisan & Zimmermann, 2000; Redfern *et al.*, 2006; Richards *et al.*, 2007; Schröder, 2008; Valavanis *et al.*, 2008; Elith & Leathwick, 2009). Αυτές οι προσεγγίσεις θα μπορούσαν να χρησιμοποιηθούν εναλλακτικά, για τον προσδιορισμό των σχέσεων ειδών-περιβάλλοντος αλλά και των ενδο- και δια-ειδικών σχέσεων αξιοποιώντας δεδομένα του κύκλου ζωής των ειδών. Νέες προσεγγίσεις, που κατατάσσονται πλέον στα ΜΚΕ, έχουν εφαρμοστεί πρόσφατα σε θαλάσσια είδη, ενώ συνεχίζεται η εισαγωγή καινοτόμων τεχνικών περισσότερο αποδοτικών (Leathwick *et al.*, 2005; Leathwick *et al.*, 2006a; Phillips *et al.*, 2006; Palialexis *et al.*, 2011). Τα ΜΚΕ έχουν επίσης χρησιμοποιηθεί για τη μελέτη των σχέσεων μεταξύ περιβαλλοντικών μεταβλητών και της παρουσίας ειδών (Amara *et al.*, 2004; Giannoulaki *et al.*, 2008; Lefkaditou *et al.*, 2008; Martin *et al.*, 2008), για τον προσδιορισμό βασικών ενδιαιτημάτων των ειδών (Planque *et al.*, 2007), καθώς και για την πρόβλεψη της κατανομής ειδών ως απόκριση στις κλιματικές αλλαγές (Siaratis *et al.*, 2008).

1.2 Η οικολογική θεωρία πίσω από τα ΜΚΕ

Η βιογεωγραφική και οικολογική θεωρία οριοθετούν ζήτημα τα οποία πραγματεύονται και επιλύουν τα ΜΚΕ. Βάση των δύο θεωριών προσδιορίζονται τα χαρακτηριστικά των περιβαλλοντικών δεδομένων και των δεδομένων του είδους ή των ειδών που μελετούνται. Επίσης, προσδιορίζονται οι σχέσεις μεταξύ των προαναφερθέντων δεδομένων και των απαιτούμενων για την εκπαίδευση ενός ΜΚΕ. Η εκπαίδευση και επαλήθευση του ΜΚΕ οδηγεί στη δημιουργία του βέλτιστου μοντέλου, το οποίο τελικά χρησιμοποιείται για τη δημιουργία χαρτών πρόβλεψης της παρουσίας των ατόμων του είδους ή χαρτών κατάλληλων ενδιαιτημάτων για τα είδη. Η παραπάνω διαδικασία αποτελεί μια γενική επισκόπηση της μεθοδολογίας ανάπτυξης ΜΚΕ (Εικ. 1.1) η οποία καταλήγει στην εφαρμογή του ΜΚΕ σε χάρτες περιβαλλοντικών και άλλων παραμέτρων για την πρόβλεψη της πιθανής γεωγραφικής κατανομής. Οι χάρτες που αποτυπώνουν την πρόβλεψη έχουν κατά καιρούς περιγραφεί ως: επιφάνειες οικολογικής απόκρισης (Lenihan, 1993), βιογεωγραφικά μοντέλα κατανομής ειδών (Guisan *et al.*, 2006), χωρική πρόβλεψη της κατανομής ειδών (Austin, 2002), χάρτες πρόβλεψης (Franklin, 1995), πρόβλεψη ύπαρξης-παρουσίας του είδους (Rushton *et al.*, 2004), χάρτες πρόβλεψης κατανομής (Rodríguez *et al.*, 2007).

Καίρια σημασία για τη λειτουργικότητα και εφαρμογή των ΜΚΕ έχει η κατανόηση και επίγνωση των δυνητικών αποτελεσμάτων τους. Συνήθως, οι χάρτες πρόβλεψης της κατανομής ειδών αποτυπώνουν τη διακύμανση μιας συνεχούς μεταβλητής, σχετικής με την πιθανότητα παρουσίας του είδους ή την καταλληλότητα του ενδιαιτήματος, ακόμα και στην περίπτωση που η κύρια μεταβλητή βάση της οποίας εκπαιδεύεται το μοντέλο είναι διακριτή (π.χ. παρουσία-απουσία). Εάν δημιουργείται ένα μοντέλο γεωγραφικής κατανομής με μία συνεχή οικολογική μεταβλητή, όπως βιομάζα ή αφθονία ειδών (π.χ. Meentemeyer *et al.*, 2001; Cumming *et al.*, 2000; Thogmartin *et al.*, 2004; Bellis *et al.*, 2008), τότε αυτή η "εξαρτημένη μεταβλητή" είναι ο παράγοντας που αναγεννάται στην πρόβλεψη. Πέρα από τις οικολογικές μεταβλητές που είναι εύκολα κατανοητές, πολλές φορές, για λόγους που αναλύονται παρακάτω, χρησιμοποιούνται στα ΜΚΕ διωνυμικές μεταβλητές. Η ανάπτυξη των ΜΚΕ που βασίζονται σε διωνυμικές

κατανομές παρέχει μεγαλύτερη ποικιλία εργαλείων για τα διάφορα στάδια ανάπτυξης του ΜΚΕ, από ότι τα ΜΚΕ που χρησιμοποιούν συνεχείς μεταβλητές. Οι χάρτες πρόβλεψης που απεικονίζουν την πιθανότητα μιας διωνυμικής κατάστασης (π.χ. παρουσία-απουσία) έχουν διατυπωθεί ως:

- Η πιθανότητα παρουσίας ατόμων ειδών ή η πιθανή κατανομή τους (Scott *et al.*, 2002)
- Τα θεωρητικά δεσμευμένα ενδαιτήματα (Guisan & Zimmermann, 2000; Edwards *et al.*, 2006)
- Η καταλληλότητα ή ποιότητα των ενδαιτημάτων (Hirzel & Guisan, 2002; Hirzel *et al.*, 2002; Gibson *et al.*, 2004)
- Η υποθετική πιθανότητα της χρήσης ενδαιτήματος ή η κατάταξη της καταλληλότητας του ενδαιτήματος (Keating & Cherry, 2004)
- Τα γεωγραφικά όρια κατανομής του είδους (Dunk *et al.*, 2004)
- Μια λειτουργία επιλογής πόρων ή τιμή ανάλογη ως προς την πιθανότητα της χρήσης μιας μονάδας πόρου (Boyce *et al.*, 2002; Manly *et al.*, 2002)
- Ο πραγματικός οικολογικός θώκος (Peterson, 2003)
- Ο θεωρητικός οικολογικός θώκος (Soberón & Peterson, 2004)

Σύμφωνα με τους Guisan & Zimmermann, (2000) η ποσοτικοποίηση της σχέσης ειδών-περιβάλλοντος αποτελεί τον πυρήνα των μοντέλων γεωγραφικής πρόβλεψης στην οικολογία. Η μοντελοποίηση της κατανομής των ατόμων των ειδών έχει τις ρίζες της στην ανάλυση οικολογικής διαβάθμισης (Whittaker, 1960; Whittaker *et al.*, 1973), τη βιογεωγραφία (Box, 1981), την τηλεπισκόπηση και τα πληροφοριακά συστήματα.

Πολλές οικολογικές θεωρίες σχετίζονται με τις αιτίες που ευθύνονται για τα πρότυπα κατανομής των ειδών, τη βιοποικιλότητα και τη δομή των βιοκοινοτήτων και είναι σε άμεση σχέση με τα ΜΚΕ. Ωστόσο, ιδιαίτερη έμφαση έχει δοθεί σε πρόσφατες εργασίες για τη σχέση των ΜΚΕ με τη θεωρία του οικολογικού θώκου των ειδών (Austin & Smith, 1989; Austin, 2002; Guisan & Thuiller, 2005; Araújo & Guisan, 2006; Kearney, 2006; Soberón, 2007; Hirzel & Le Lay, 2008; Jiménez-Valverde *et al.*, 2008). Η θεωρία του θώκου στην περιβαλλοντική και γεωγραφική διάσταση έχει σχεδόν ένα αιώνα ζωής, με ανάλογη εξέλιξη στο χρόνο και πολλές ερμηνείες. Σύμφωνα με τον Hutchinson (1987) ο θώκος αποτελεί την υπερδιάσταση που ορίζεται από τις περιβαλλοντικές διαστάσεις μέσα σε αυτή, όπου το είδος μπορεί να επιβιώσει και να αναπαραχθεί. Ο Hutchinson διακρίνει τον θεωρητικό (φυσιολογικό ή πιθανό) θώκο, που ορίζεται ως η απόκριση του είδους στο περιβάλλον (βιολογικοί πόροι), όπου απουσιάζουν βιοτικές αλληλεπιδράσεις (ανταγωνισμός, θήρευση), από τον ρεαλιστικό (οικολογικό, πραγματικό) θώκο, όπου το είδος μπορεί να επιβιώσει και να αναπαραχθεί, συμπεριλαμβάνοντας τις βιοτικές αλληλεπιδράσεις (Εικ. 1.2) (Austin & Smith, 1989; Austin, 2002). Ο θώκος ορίζεται ως χωρική διάσταση που αναφέρεται στο είδος και όχι στο περιβάλλον (δες Pulliam, 2000). Τα ΜΚΕ βασίζονται στη θεωρία του θώκου όσον αφορά τις απαιτήσεις του είδους, κυρίως δε (αλλά όχι αποκλειστικά) στην προσομοίωση των αβιοτικών παραγόντων, που ελέγχουν την κατανομή των ατόμων του είδους. Η υπερδιάσταση του Hutchinson προτείνει πολλαπλούς αιτιατούς παράγοντες για την κατανομή των ειδών, που αντανakλάται και επιχειρησιακά στην εφαρμογή των ΜΚΕ, όπου συνήθως χρησιμοποιούνται πολλές επεξηγηματικές μεταβλητές. Αντίθετα η θεωρητική αντιμετώπιση του θώκου συνήθως δίνει έμφαση σε λίγους βιολογικούς πόρους, στις απαιτήσεις του είδους ή σε περιοριστικούς παράγοντες για την κατανομή του.

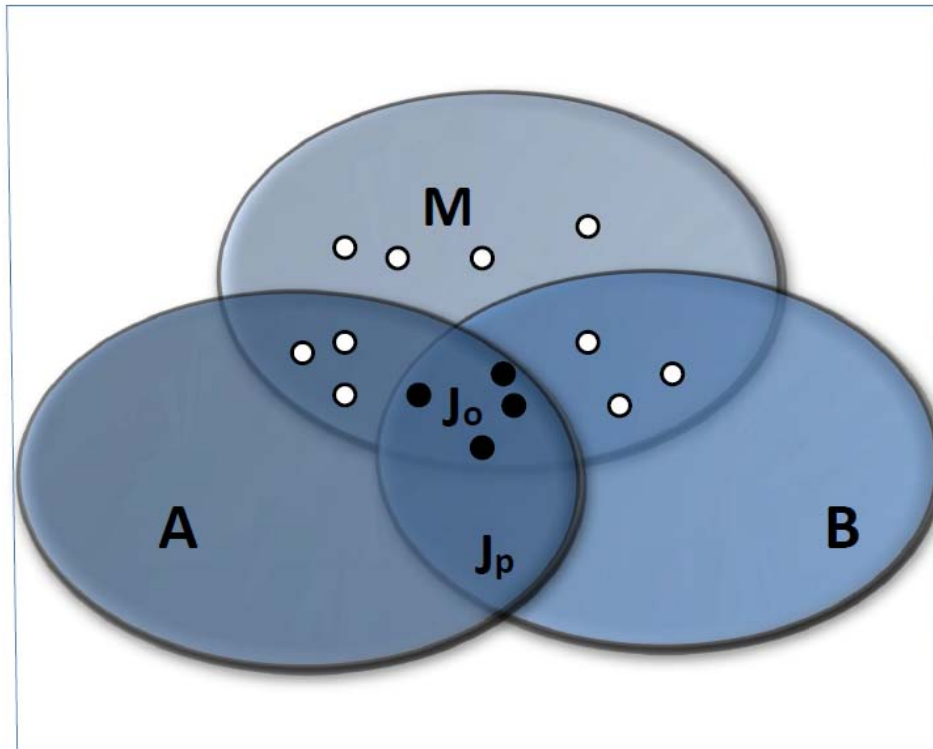


Εικόνα 1.2 Η θεωρία του οικολογικού θώκου από τον Hutchinson. Ο βιότοπος αντιπροσωπεύει το εύρος των περιβαλλοντικών συνθηκών που επικρατούν σε μια περιοχή. Ο θεωρητικός θώκος των ειδών αντιπροσωπεύει το εύρος των φυσιολογικών ορίων των ειδών στην απουσία αλληλεπιδράσεων με άλλα είδη (ανταγωνισμός, θήρευση). Ο πραγματικός θώκος αντιστοιχεί στην περιβαλλοντική διάσταση, όπου συνυπολογίζεται τόσο η καταλληλότητα των περιβαλλοντικών παραγόντων, όσο και οι βιοτικές αλληλεπιδράσεις που οριοθετούν την κατανομή των ατόμων του είδους (τροποποιημένο από Franklin, 2009).

Ο Pulliam (2000) τονίζει τη σημασία της μετρούμενης προσαρμοστικότητας, ειδικά το ρυθμό αύξησης του πληθυσμού, ώστε να προσδιορίσει τον οικολογικό θώκο του είδους (όπου ο ρυθμός αύξησης είναι θετικός). Προτείνει ότι η θεωρία του θώκου του Hutchinson, η θεωρία των μεταπληθυσμών (Hanski, 1999) και η θεωρία source-sink (Pulliam, 1988), ταυτόχρονα, μπορούν να εξηγήσουν τη σχέση μεταξύ της κατανομής των ειδών και των κατάλληλων ενδιαιτημάτων. Σύμφωνα με τον Pulliam (2000) το πηγαίο ενδιαιτήμα, όπου τοπικά η αναπαραγωγή προηγείται της θνησιμότητας, διαχωρίζεται από το δευτερεύον ενδιαιτήμα, όπου βρίσκονται άτομα του είδους αλλά δεν συνεισφέρουν στο ρυθμό αύξησης του πληθυσμού. Το πλαίσιο των μεταπληθυσμών (Hanski, 1999) χρησιμοποιείται για να περιγράψει και να μοντελοποιήσει τους πληθυσμούς ως ομάδα διακριτών υποπληθυσμών. Ένα είδος μπορεί να απουσιάζει από ένα σημείο κατάλληλου ενδιαιτήματος ως αποτέλεσμα της δυναμικής των μεταπληθυσμών ή περιορισμών της διασποράς του. Λαμβάνοντας υπόψη τις παραπάνω θεωρίες εξάγεται ότι ένα είδος μπορεί να βρεθεί σε ακατάλληλο ενδιαιτήμα (δευτερεύον) ή μπορεί να απουσιάζει από κατάλληλο ενδιαιτήμα. Αυτή η ιδιαιτερότητα της θεωρίας του θώκου χρήζει ιδιαίτερης προσοχής στην εφαρμογή των ΜΚΕ, επειδή τα περισσότερα στηρίζονται σε δεδομένα παρουσίας του είδους. Η εξίσωση των κατελημμένων ενδιαιτημάτων με τα κατάλληλα ενδιαιτήματα μπορεί να οδηγήσει σε υπεραπλούστευση της καταλληλότητας.

Ο ρεαλιστικός θώκος απεικονίζεται συχνά ως υποσύνολο του θεωρητικού θώκου (Εικ. 1.2) αν και δεν είναι πάντα έτσι, όπως στην περίπτωση του δευτερεύοντος ενδιαιτήματος ή στην περίπτωση θετικών βιοτικών αλληλεπιδράσεων (π.χ. συμβίωση), όπου η παρουσία ενός άλλου είδους είναι προαπαιτούμενη. Η Εικ. 1.3 αποτυπώνει σε γεωγραφική διάσταση το θεώρημα του Pulliam, παρά σε περιβαλλοντικό χώρο. Ο χώρος Α είναι η γεωγραφική περιοχή, όπου ο εγγενής ρυθμός ανάπτυξης είναι θετικός (αντιστοιχεί στο θεωρητικό θώκο). Στο χώρο Β τα είδη συνυπάρχουν με τους ανταγωνιστές τους, ο χώρος Μ είναι προσβάσιμος από τα είδη σε δεδομένο χρονικό πλαίσιο βάσει της ικανότητας διασποράς τους. Jo είναι η υπό κατάληψη περιοχή από τα είδη και Jp είναι το κατάλληλο ενδιαιτήμα το οποίο δεν έχει καταληφθεί λόγω περιορισμών

διασποράς του είδους. Εάν το είδος βρίσκεται στο χώρο M ή/και B που εφάπτονται στο A, ή στο A που δεν εφάπτεται στο B τότε βρίσκεται σε δευτερεύον ενδιαίτημα.



Εικόνα 1.3 Το πλαίσιο της εικόνας αναφέρεται περισσότερο στο γεωγραφικό χώρο παρά σε περιβαλλοντικές διαβαθμίσεις. Ο χώρος A είναι η γεωγραφική περιοχή, όπου ο εγγενής ρυθμός ανάπτυξης είναι θετικός. Στο χώρο B τα είδη συνυπάρχουν με τους ανταγωνιστές τους. Ο χώρος M είναι προσβάσιμος από τα είδη σε δεδομένο χρονικό πλαίσιο βάσει της ικανότητας διασποράς τους. J_o είναι η υπό κατάληψη περιοχή από τα είδη και J_p είναι η πιθανή προς κατάληψη περιοχή. Οι μαύροι κύκλοι αφορούν στην παρουσία των ατόμων στα πηγαία ενδιαίτηματα, ενώ οι άσπροι κύκλοι την παρουσία των ατόμων στα δευτερεύοντα ενδιαίτηματα (τροποποιημένο από Franklin, 2009).

Η συζήτηση για τη σχέση μεταξύ της θεωρίας του οικολογικού θώκου και της μοντελοποίησης της κατανομής των ειδών έχει πάρει πολύ μεγάλες διαστάσεις πρόσφατα. Αυτή η συζήτηση συνδέεται άμεσα και με το τι πραγματικά μοντελοποιείται στα ΜΚΕ μεταξύ του θεωρητικού οικολογικού θώκου, του ρεαλιστικού θώκου, της πιθανότητας χρήσης ή της καταλληλότητας του ενδιαίτηματος. Οι περισσότερες μελέτες προσδιορίζουν το ρεαλιστικό θώκο ως αποτέλεσμα των ΜΚΕ (Austin, 2002; Thuiller *et al.*, 2004; Guisan & Thuiller, 2005) επειδή συνήθως κατά τη μοντελοποίηση χρησιμοποιούνται δεδομένα της πραγματικής παρουσίας του είδους, οπότε το μοντέλο εξάγει σε γεωγραφικό χώρο τις συνθήκες εκείνες, που σχετίζονται με την αφθονία ή ύπαρξη του είδους στην περιβαλλοντική υπερδιάσταση.

1.3 Ανάπτυξη των μοντέλων κατανομής ειδών

Κοινές μέθοδοι μοντελοποίησης κατανομής ειδών για δεδομένα μόνο-παρουσίας είναι οι ENFA (ecological niche factor analysis), BIOCLIM (bioclimatic algorithm), DOMAIN, GARP (genetic algorithm of rule-set prediction) και MAXENT (maximum entropy), ενώ οι ANN (artificial neural networks), GLM (generalized linear models), GAM (generalized additive models) και CART (classification and regression trees) απαιτούν ακριβή δεδομένα παρουσίας-απουσίας, προκειμένου να παράγουν στατιστικά αποτελέσματα ή διακριτούς κανόνες, που επιτρέπουν την ταξινόμηση της καταλληλότητας των ενδιαιτημάτων ανάλογα με την κατανομή της παρουσίας και απουσίας των ειδών (Manel *et al.*, 1999; Guisan & Zimmerman, 2000). Κάποιες από τις τελευταίες μεθόδους (GAM, ANN) θα μπορούσαν να χρησιμοποιηθούν και με δεδομένα μόνο-παρουσίας

(Brotans *et al.*, 2004). Άλλες μέθοδοι μοντελοποίησης που δύναται να χρησιμοποιήσουν διωνυμικά δεδομένα είναι οι τεχνικές παρεμβολής και οι τεχνικές προσομοίωσης.

Η εξέλιξη στην επιστήμη των υπολογιστών και τη στατιστική αντικατοπτρίζεται στην εξέλιξη παραδοσιακών ΜΚΕ και την εισαγωγή καινοτόμων μεθόδων. Τα μοντέλα που βασίζονται σε δεδομένα μόνο-παρουσίας (π.χ. Bioclim, Envelope Score, Ecological Niche Factor Analysis-ENFA) αρχικά χρησιμοποιήθηκαν για χερσαία είδη, αξιοποιώντας κυρίως στοιχεία από μουσεία Φυσικής Ιστορίας (Ready *et al.*, 2010). Τα περισσότερα από αυτά βασίζονται στην αρχή του "περιβαλλοντικού φακέλου" (environmental envelope), η οποία είναι συναφής με τη θεωρία του οικολογικού θάκου, και στηρίζονται στον υπολογισμό ενός κατάλληλου για την παρουσία του είδους περιβαλλοντικού εύρους (φάκελος). Σε σύγκριση με τα μοντέλα μόνο-παρουσίας, τα μοντέλα παρουσίας-απουσίας είναι πιο αποτελεσματικά ως προς την ικανότητα πρόβλεψης (Brotans *et al.*, 2004), αφού εκμεταλλεύονται την πρόσθετη πληροφορία σχετικά με τις ακατάλληλες περιβαλλοντικές περιοχές για τα είδη. Τα μοντέλα παλινδρόμησης, όπως τα Γενικευμένα Γραμμικά Μοντέλα (GLMs) και τα Γενικευμένα Αθροιστικά Μοντέλα (GAMs) χρησιμοποιούνται ευρέως με δεδομένα παρουσίας-απουσίας (Olivier & Wotherspoon, 2005; Leathwick *et al.*, 2006b). Διάφορες τεχνικές μοντελοποίησης έχουν αναπτυχθεί πρόσφατα, αξιοποιώντας την εξέλιξη της επιστήμη των υπολογιστών, όπως τα Ενισχυμένα Δέντρα Παλινδρόμησης (Boosted Regression Trees, BRT, Leathwick *et al.*, 2006a) και τα Συνδυαστικά Νευρωνικά Δίκτυα (Associative Neural Networks, ASNN, Tetko, 2002a) τα οποία συνδυάζουν διαφορετικούς αλγόριθμους προκειμένου να βελτιστοποιηθεί η προγνωστική ικανότητα των μοντέλων. Επιπλέον, τα ευρέως χρησιμοποιούμενα μοντέλα, όπως τα GAMs, έχουν αναπτυχθεί περαιτέρω (π.χ. η προσέγγιση BRUTO, Hastie & Tibshirani, 1990) ή τροποποιηθεί (π.χ. MARS, Leathwick *et al.*, 2006a) για να καλύψουν τις ανάγκες που προκύπτουν στην εφαρμογή των ΜΚΕ και στην επιδίωξη βελτίωσης της ακρίβειάς τους.

Η εφαρμογή ενός στατιστικού μοντέλου περιλαμβάνει κάποια βασικά στάδια: την εκτίμηση των παραμέτρων, την επιλογή του βέλτιστου μοντέλου, την εκτίμηση του επιπέδου αβεβαιότητας, την επικύρωση του μοντέλου, και την παραγωγή και έλεγχο των προβλέψεων. Αρκετές εναλλακτικές προσεγγίσεις μπορούν να χρησιμοποιηθούν σε κάθε ένα από τα προαναφερόμενα στάδια. Η επιλογή της κατάλληλης προσέγγισης εξαρτάται από το στόχο της εφαρμογής του ΜΚΕ, από τη φύση των δεδομένων, από την αποτελεσματικότητά της και από πιθανούς περιορισμούς ή από πιθανές παραδοχές που τη συνοδεύουν.

Τα μοντέλα παρουσίας-απουσίας είναι γενικά πιο εύκολο να αναπτυχθούν, εφόσον τα δεδομένα εκπαίδευσης έχουν διωνυμική κατανομή, ενώ τα μοντέλα αφθονίας χαρακτηρίζονται από πιο περίπλοκες κατανομές (π.χ. Poisson, Gaussian) και ως εκ τούτου, οφείλουν να συντάσσονται με αρκετές παραδοχές. Η διαδικασία επικύρωσης και αξιολόγησης των μοντέλων παρουσίας-απουσίας είναι αναλογικά ευκολότερη από την αντίστοιχη για τα μοντέλα αφθονίας. Επισκοπήσεις μεθόδων που έχουν αναπτυχθεί για την αξιολόγηση της ποιότητας των προβλέψεων των μοντέλων περιγράφονται αναλυτικά από τους Fielding & Bell, (1997) και Boyce *et al.* (2002). Τεχνικές, όπως τα χαρακτηριστικά λειτουργίας δέκτη (Receiver Operating Characteristics, ROC), τα Kappa στατιστικά στοιχεία και οι συνεκτικοί πίνακες (confusion matrices) (Fielding & Bell, 1997) σε συνδυασμό με την αξιολόγηση των σφαλμάτων παράλειψης και επιφόρτισης (omission και commission errors) χρησιμοποιούνται ευρέως για την εκτίμηση της επίδοσης μοντέλων παρουσίας-απουσίας και για τη σύγκριση διαφορετικών ΜΚΕ (Elith *et al.*, 2006).

Η εφαρμογή και η εξέλιξη των ΜΚΕ συνοδεύεται από ζητήματα μείζονος σημασίας, που η έρευνα αιχμής στον τομέα των ΜΚΕ καλείται να διαχειριστεί, ώστε να βελτιωθεί η απόδοση και η εφαρμογή τους (Elith & Leathwick, 2009). Ένα κοινό πρόβλημα που αφορά στις προσεγγίσεις παλινδρόμησης με τη χρήση των ελάχιστων τετραγώνων για τη μοντελοποίηση της κατανομής ειδών ως απόκριση στις περιβαλλοντικές μεταβλητές είναι το σφάλμα που εισάγεται από τυχόν μη μετρούμενες μεταβλητές. Δεδομένου ότι μόνο

μερικοί από τους παράγοντες που επηρεάζουν τη χωρική κατανομή των ειδών δύνανται να μετρηθούν και να συμπεριληφθούν σε ΜΚΕ, υπάρχει η πιθανότητα οι επιδράσεις τυχόν μη μετρούμενων παραγόντων να καλύψουν την προγνωστική σχέση μεταξύ κύριας και επεξηγηματικών μεταβλητών. Ακόμη και αν ληφθούν υπόψη όλοι οι παράγοντες, μερικά συστήματα μοντελοποίησης ενσωματώνουν έναν ανεξήγητο στοχαστικό παράγοντα ή εμφανίζουν χαοτική συμπεριφορά. Μια εναλλακτική προσέγγιση είναι η αντιμετώπιση των επεξηγηματικών μεταβλητών ως περιοριστικών παραγόντων και όχι ως συμμεταβλητών. Αυτή η προσέγγιση πηγάζει από μια θεμελιώδη οικολογική αρχή, το νόμου του Liebig για το ελάχιστο. Οι συμβατικές μέθοδοι ανάλυσης που στηρίζονται σε συσχετίσεις ή στην παλινδρόμηση δε βασίζονται στην έννοια των περιοριστικών παραγόντων (Thompson *et al.*, 1996; Cade *et al.*, 1999). Η παλινδρόμηση ποσοστιαίων σημείων από την άλλη, βασίζεται στην αρχή αυτή και ποσοτικοποιεί τους περιοριστικούς παράγοντες με την εφαρμογή καμπυλών παλινδρόμησης στο ποσοστιαίο σημείο κοντά στη μέγιστη απόκριση (Eastwood *et al.*, 2001; Cade & Noon, 2003; Hiddink, 2005). Με αυτόν τον τρόπο η επίδραση άλλων μετρούμενων ή μη παραγόντων υποβαθμίζεται και λαμβάνονται υπόψη μόνο οι παράγοντες που έχουν περιοριστικό αντίκτυπο. Παράλληλα, η παλινδρόμηση ποσοστιαίων σημείων αποκαλύπτει πιθανά καλυμμένα σφάλματα και σημαντικές διεργασίες που δεν αντικατοπτρίζονται επαρκώς από τις μετρούμενες μεταβλητές (Cade *et al.*, 2005). Ωστόσο, ο νόμος του Liebig είναι ανοικτός σε κριτική, εφόσον οι επεξηγηματικές μεταβλητές μπορεί να παρουσιάζουν διαδραστικές επιδράσεις, δηλαδή, ένας παράγοντας μπορεί να έχει περιοριστικό αντίκτυπο μόνο στην παρουσία ή απουσία ενός άλλου (Huisman & Weissing, 2002) ή όλες οι μεταβλητές να περιορίζονται ταυτόχρονα («πολλαπλή υπόθεση περιορισμού», Rubio *et al.*, 2003).

Πέρα από τις δυνατότητες των τεχνικών μοντελοποίησης και των μεθόδων που τις υλοποιούν, η απόδοση και η εφαρμογή των ΜΚΕ εξαρτάται σημαντικά από τη φύση και την ποιότητα των δεδομένων. Η γνώση του κύκλου ζωής των ειδών έχει ιδιαίτερη σημασία κατά την ανάπτυξη και αξιολόγηση του μοντέλου κατανομής ειδών. Υπάρχει πάντα μια ισορροπία μεταξύ της πολυπλοκότητας-υπερπροσαρμογής και της απλότητας-γενικότητας της απόδοσης της πρόβλεψης των ΜΚΕ, η οποία κινείται προς τη μια ή την άλλη κατεύθυνση ανάλογα με τους στόχους της διαδικασίας μοντελοποίησης. Σύμφωνα με τους Guisan & Zimmermann (2000), η φύση είναι πολύπλοκη και ετερογενής για να προβλεφθεί με ακρίβεια σε κάθε πτυχή του χρόνου και του χώρου από ένα ενιαίο και πολύπλοκο μοντέλο. Ο Levins (1966) διατύπωσε την αρχή ότι μόνο δύο από τις τρεις επιθυμητές ιδιότητες του εκάστοτε μοντέλου (γενικότητα-generality, πραγματικότητα-reality, ακρίβεια-precision) μπορούν να βελτιωθούν ταυτόχρονα, ενώ η τρίτη πρέπει να θυσιάσει. Η επιλογή της κατάλληλης τεχνικής μοντελοποίησης θα πρέπει να βασίζεται στον στόχο της κάθε εφαρμογής μοντελοποίησης.

1.4 Εφαρμογή μοντέλων κατανομής ειδών σε υδροακουστικά δεδομένα μικρών πελαγικών ειδών ψαριών

Η μοντελοποίηση της κατανομής των θαλάσσιων ειδών είναι πλέον κοινό εργαλείο των επιστημονικών μελετών και των σχεδίων διαχείρισης. Η γενικότητα των αναπτυγμένων μοντέλων είναι απαραίτητη για ακριβείς προβλέψεις της κατανομής των ειδών σε ευρεία χρονική και χωρική κλίμακα. Από την άλλη πλευρά, η ακρίβεια των μοντέλων είναι απαραίτητη για τη ρεαλιστική απεικόνιση της κατανομής των ειδών και την αναγνώριση των ενδιαιτημάτων των ψαριών. Υπάρχει μια πληθώρα τεχνικών μοντελοποίησης που είναι κατάλληλες για αλιευτικά και υδροακουστικά δεδομένα. Ωστόσο, στη Μεσόγειο Θάλασσα, οι περισσότερες από τις μελέτες πρόβλεψης κατανομής ειδών έχουν χρησιμοποιήσει GAMs (Giannoulaki *et al.*, 2008; Martin *et al.*, 2008; Siapatis *et al.*, 2008), μοντέλα μέγιστης εντροπίας (MAXENT) (Lefkaditou *et al.*, 2008) και Ανάλυση Διαχωριστικών Συναρτήσεων (Discriminant Function Analysis, DFA) (Tsagarakis *et al.*, 2008). Τα κύρια είδη μικρών πελαγικών ψαριών στα οποία αντιστοιχεί η ακουστική πυκνότητα που χρησιμοποιείται στα ΜΚΕ είναι η σαρδέλα (*Sardina pilchardus*) και ο γαύρος (*Engraulis encrasicolus*). Τα συγκεκριμένα είδη ενδείκνυνται για την ανάπτυξη μοντέλων κατανομής ειδικά με τη χρήση

περιβαλλοντικών παραμέτρων, καθώς εμφανίζουν χωρικά πρότυπα στην κατανομή τους τα οποία συσχετίζονται άμεσα ή έμμεσα με τις περιβαλλοντικές μεταβλητές. Αρκετές μελέτες έχουν εστιάσει στη σχέση των δύο ειδών με συγκεκριμένη ή συγκεκριμένες περιβαλλοντικές μεταβλητές (π.χ. Daskalon *et al.*, 2003; Santos *et al.*, 2004; Giannoulaki *et al.*, 2008; Giannoulaki *et al.*, 2011).

Τα υδροακουστικά δεδομένα έχουν χρησιμοποιηθεί σε αρκετές μελέτες κυρίως για τον προσδιορισμό της κατανομής μικρών πελαγικών ειδών (Giannoulaki *et al.*, 2008, Martin *et al.*, 2008; Giannoulaki *et al.*, 2011). Εμπεριέχουν χαρακτηριστικά, που είναι κατάλληλα για την ανάπτυξη ΜΚΕ. Έχει αποδειχθεί ότι η συστηματική δειγματοληψία (όπως αυτή που δημιουργεί τα υδροακουστικά δεδομένα) παρέχει καλύτερα δεδομένα από ότι η τυχαία ή η τυχαία στρωματοποιημένη δειγματοληψία για την μοντελοποίηση της κατανομής ειδών (Hirzel & Guisan, 2002). Επίσης, τα υδροακουστικά δεδομένα χαρακτηρίζονται από μεγάλο αριθμό δειγμάτων που προέρχονται από συνεχείς καταγραφές και καλύπτουν την περιοχή μελέτης σε ένα κανονικό σχέδιο διατομών. Ο αριθμός των δειγμάτων σχετίζεται θετικά με την απόδοση των ΜΚΕ (Cumming, 2000; Hirzel & Guisan, 2002; Reese *et al.*, 2005). Η ιδιότητα τους να φέρουν μεγάλο αριθμό μετρήσεων συστηματικά κατανεμημένων στο χώρο διευκολύνει τη χρήση τους και κατά τη διαδικασία επαλήθευσης των μοντέλων, αφού το σύνολο των δεδομένων δύναται να διαχωριστεί σε σύνολο εκπαίδευσης και σύνολο επαλήθευσης. Επιπλέον, τα υδροακουστικά δεδομένα μπορούν να χρησιμοποιηθούν αυτούσια ως ακουστική πυκνότητα (αφθονία), ή να μετατραπούν σε δεδομένα παρουσίας-απουσίας. Το παραπάνω χαρακτηριστικό τους, δίνει τη δυνατότητα της αξιοποίησης τους με ένα ευρύ σύνολο τεχνικών μοντελοποίησης, που αξιοποιείται πλήρως στην παρούσα εργασία, με την εφαρμογή των περισσότερων ΜΚΕ που περιγράφονται έως σήμερα στη σχετική βιβλιογραφία. Πέρα από τα προφανή χαρακτηριστικά των υδροακουστικών δεδομένων έχουν προσδιοριστεί κάποια άλλα στοιχεία που καθιστούν αυτού του τύπου τα δεδομένα ως πρόκληση για τα ΜΚΕ και ως ευκαιρία για την πλήρη αξιοποίηση των δυνατοτήτων των ΜΚΕ, που είναι απαραίτητη στα πλαίσια σύγκρισης της απόδοσης διαφορετικών τεχνικών μοντελοποίησης. Τα στοιχεία αυτά περιλαμβάνουν τη χωρική αυτοσυσχέτιση που χαρακτηρίζει τα υδροακουστικά δεδομένα, τη μεγάλη διακύμανση και τη μεγάλη τυπική απόκλιση των τιμών. Η χωρική αυτοσυσχέτιση άλλοτε αποτελεί "εργαλείο" όπως όταν χρησιμοποιείται με μεθόδους παρεμβολής και άλλοτε αιτία εισαγωγής σφάλματος στα ΜΚΕ, όταν για παράδειγμα χρησιμοποιείται με μοντέλα παλινδρόμησης χωρίς την κατάλληλη εισαγωγή στη διαδικασία μοντελοποίησης. Τέλος, η ικανοποιητική γνώση του κύκλου ζωής των μικρών πελαγικών ειδών, η γνώση της οικολογίας τους και της σχέσης τους με αβιοτικούς και βιοτικούς παράγοντες (Daskalon *et al.*, 2003; Santos *et al.*, 2004; Ruiz *et al.*, 2006; Planque *et al.*, 2007) συνιστά απαραίτητη προϋπόθεση για να αξιολογηθούν και βιολογικά οι χάρτες πρόβλεψης της κατανομής ειδών, αλλά και να διερευνηθούν ή να επαληθευτούν πιθανές σχέσεις των μικρών πελαγικών ειδών με περιβαλλοντικές παραμέτρους.

1.5 Σκοπός της διατριβής

Η βιβλιογραφία που σχετίζεται με τα ΜΚΕ εμπλουτίζεται με ταχείς ρυθμούς, γεγονός που αντανακλά τη ραγδαία εξέλιξη των ΜΚΕ και τη συμβολή τους στις οικολογικές μελέτες. Κατά την τελευταία δεκαετία έχουν δημοσιευθεί πολυάριθμες επισκοπήσεις που αφορούν ΜΚΕ (π.χ. Guisan & Zimmermann, 2000; Redfern *et al.*, 2006; Richards *et al.*, 2007; Valavanis *et al.*, 2008; Schröder 2008; Elith & Leathwick, 2009). Άλλες μελέτες που ασχολούνται με σημαντικά θέματα της ανάπτυξης των ΜΚΕ περιλαμβάνουν ζητήματα, όπως οι μέθοδοι αξιολόγησης των σφαλμάτων πρόβλεψης (Fielding & Bell, 1997; Boyce *et al.*, 2002), οι επιδράσεις της χωρικής αυτοσυσχέτισης στα ΜΚΕ (Dormann *et al.*, 2007), η οικολογική θεωρία στην εφαρμογή των ΜΚΕ (Guisan & Thuiller, 2005), νέες προσεγγίσεις για τη μοντελοποίηση κατανομής ειδών (Leathwick *et al.*, 2005; Leathwick *et al.*, 2006a) και εκτεταμένες συγκρίσεις ΜΚΕ για να προσδιοριστεί η αποτελεσματικότητα και η δυνατότητα εφαρμογής τους χρησιμοποιώντας διάφορους τύπους δεδομένων

(Elith *et al.*, 2006; Leathwick *et al.*, 2006a; Caruana & Niculescu-Mizil, 2006; MacLeod *et al.*, 2008, Palialexis *et al.*, 2009; Palialexis *et al.*, 2009a; Aertsen *et al.*, 2010).

Η διατριβή αυτή επιχειρεί να συμβάλει κυρίως:

- Στην επιλογή των καταλληλότερων ΜΚΕ όσον αφορά στη δυναμική προσομοίωσης των δειγματοληπτικών δεδομένων και στην αποτελεσματικότητα εξαγωγής ή παρεμβολής των προβλέψεων. Η επιλογή των καταλληλότερων ΜΚΕ πραγματοποιείται μέσα από μια ευρεία σύγκριση τεχνικών, που περιλαμβάνουν τις περισσότερες από τις μεθόδους μοντελοποίησης που συναντώνται στη βιβλιογραφία των ΜΚΕ. Ανάλογης έκτασης σύγκριση και κυρίως μεταξύ μεθόδων που αντιστοιχούν σε διαφορετικές στατιστικές αρχές (μηχανικής μάθησης, μοντέλα παλινδρόμησης, Kriging και άλλα) δεν υπάρχει στη βιβλιογραφία για δεδομένα παρουσίας-απουσίας και αφθονίας. Τα αποτελέσματα της σύγκρισης αποσκοπούν στη δημιουργία ενός οδηγού ικανού να παρέχει το ανάλογο υπόβαθρο για την ορθότερη επιλογή τεχνικής μοντελοποίησης σε δεδομένα παρουσίας-απουσίας, μόνο-παρουσίας και αφθονίας. Επίσης, προτείνονται οι αποτελεσματικότερες μέθοδοι για υδροακουστικά δεδομένα και γενικότερα δεδομένα που έχουν προέλθει από συστηματική δειγματοληψία, όπως και τεχνικές που δύνανται να αναδείξουν σχέσεις ειδών και επεξηγηματικών μεταβλητών (συνήθως περιβαλλοντικών).
- Στην εισαγωγή νέων μεθόδων μοντελοποίησης, ανταγωνιστικών ως προς τις καθιερωμένες εφαρμογές. Εδώ εντάσσεται η χρήση των συνδεδεμένων νευρωνικών δικτύων και των συνόλων νευρωνικών δικτύων τα οποία εισάγονται για πρώτη φορά στα ΜΚΕ. Στα πλαίσια της βελτίωσης της προσομοίωσης δειγματοληπτικών δεδομένων και τη βελτίωση της πρόβλεψης επιχειρείται για πρώτη φορά σε δεδομένα αφθονίας ο συνδυασμός δύο διαφορετικής φύσης τεχνικών (Kriging και νευρωνικά δίκτυα) για τη βελτιστοποίηση των αποτελεσμάτων με άξονα τις ιδιαιτερότητες της κάθε τεχνικής.
- Στην εμφάνιση καίριων ζητημάτων της εφαρμογής και εξέλιξης των ΜΚΕ που τίγονται από μελέτες αιχμής στο χώρο των ΜΚΕ (π.χ. Elith & Leathwick, 2009). Εδώ εντάσσεται η επίδραση της χωρικής αυτοσυσχέτισης στα αποτελέσματα και την ανάπτυξη των μοντέλων, η βέλτιστη επιλογή μεθόδου επικύρωσης των αποτελεσμάτων μοντελοποίησης τόσο σε δεδομένα παρουσίας-απουσίας όσο και αφθονίας, η διαχείριση αλληλεπιδράσεων μεταξύ επεξηγηματικών μεταβλητών και η ερμηνεία των αποτελεσμάτων σε γεωγραφική, περιβαλλοντική και χρονική διάσταση.

Είναι επιβεβλημένη η διάκριση των τεχνικών μοντελοποίησης της κατανομής ειδών μεταξύ αυτών που (α) αναπτύσσονται με δεδομένα αφθονίας, χρησιμοποιώντας την ακουστική πυκνότητα και (β) αυτών που αναπτύσσονται με δεδομένα παρουσίας-απουσίας ή μόνο-παρουσίας. Οι δύο ομάδες τεχνικών μπορεί να διαθέτουν κάποιες κοινές προσεγγίσεις, όμως είναι πολύ διαφορετικός ο τρόπος αξιολόγησης των μοντέλων, η φιλοσοφία ανάπτυξης, η φύση των αποτελεσμάτων, τα προβλήματα και οι προκλήσεις που αντιμετωπίζουν. Οι παραπάνω λόγοι οδηγούν στην πλήρη διάκριση των δύο ομάδων κατά την ανάλυση και έκθεση των αποτελεσμάτων.

Η ακριβής εκτίμηση της κατανομής των ειδών, με τη χρήση δειγματοληπτικών δεδομένων παρουσίας-απουσίας, είναι απαραίτητη για περαιτέρω επιστημονική ανάλυση, όπως η αξιολόγηση των αποθεμάτων, η κατανόηση της διακύμανσης της αφθονίας λόγω της κλιματικής αλλαγής ή της υπεραλίευσης, και η εφαρμογή διαχειριστικών σχεδίων και νομοθετικών πλαισίων. Για την επίτευξη των στόχων της διατριβής πραγματοποιήθηκαν συνοπτικά οι ακόλουθες διαδικασίες:

(α) Αρχικά, πραγματοποιείται μια εκτεταμένη διερεύνηση τόσο των υδροακουστικών όσο και των επεξηγηματικών δεδομένων, για την ανίχνευση χωρικών προτύπων και σχέσεων μεταξύ τους. Για την βελτιστοποίηση της ερμηνείας της κατανομής των δειγματοληπτικών δεδομένων συλλέγονται και δημιουργούνται μεταβλητές στην καλύτερη διαθέσιμη χωρική ανάλυση. Εδώ πρέπει να σημειωθεί ότι

επιχειρείται για πρώτη φορά η εισαγωγή ορισμένων μεταβλητών σε ΜΚΕ, όπως τα θερμικά μέτωπα και οι περιοχές μέγιστης παραγωγικότητας.

(β) Εξετάζονται οι επιδόσεις των μοντέλων παρουσίας-απουσίας από την οπτική της βέλτιστης προσαρμογής σε ένα σύνολο δεδομένων εκπαίδευσης, προκειμένου να προσδιοριστεί η ικανότητα της κάθε προσέγγισης να προσομοιώνει με ακρίβεια τα δεδομένα.

(γ) Στη συνέχεια, χρησιμοποιείται ένα σύνολο δεδομένων επαλήθευσης, για να εκτιμηθεί η ικανότητα πρόβλεψης της κάθε μεθόδου και να αποκτηθεί μια πιο ολοκληρωμένη εικόνα της απόδοσης και εφαρμογής του εκάστοτε ΜΚΕ.

(δ) Τέλος, τα ΜΚΕ εφαρμόζονται σε ανεξάρτητες μεταβλητές υψηλής χωρικής ανάλυσης, προκειμένου να δημιουργήσουν ρεαλιστικούς χάρτες κατανομής των μικρών πελαγικών ειδών, που καλύπτουν μια ευρεία περιοχή.

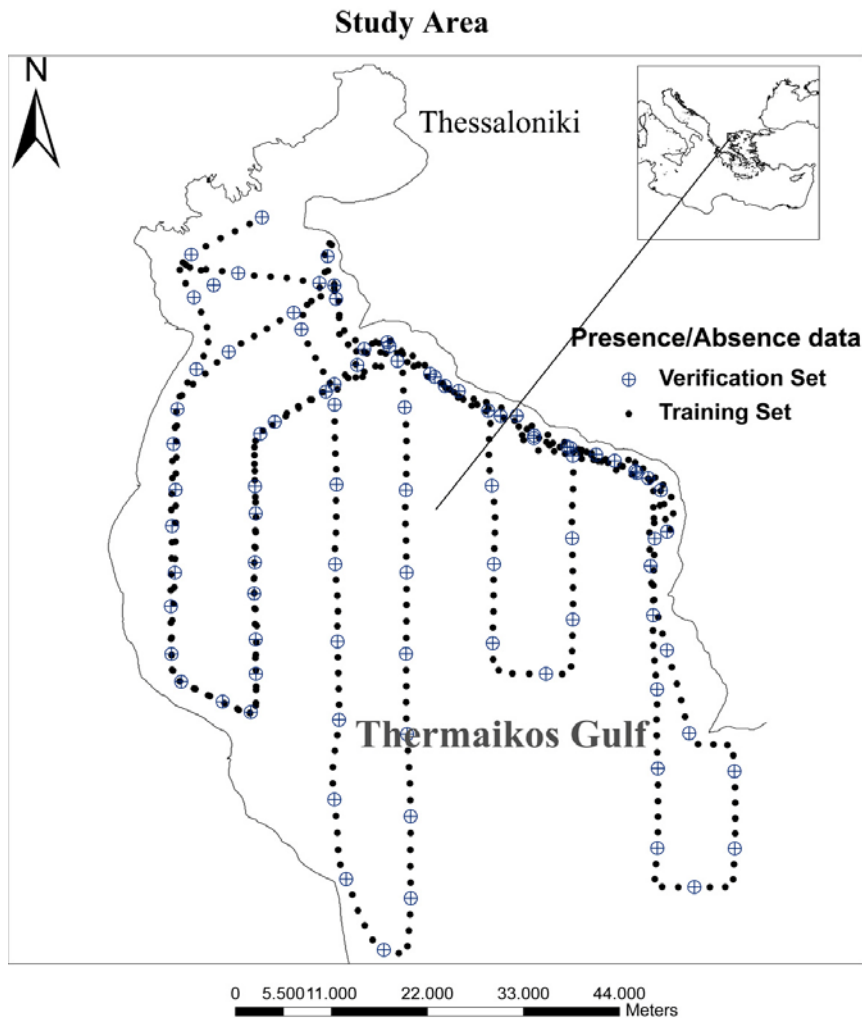
Για την αντικειμενικότερη αξιολόγηση των ΜΚΕ χρησιμοποιήθηκαν πολλά και διαφορετικά κριτήρια, που κυρίως πραγματεύονται τα σφάλματα πρόβλεψης.

Αναφορικά με την ομάδα τεχνικών που χρησιμοποιούν δεδομένα αφθονίας, παρουσιάζονται και αναλύονται τα πλεονεκτήματα και τα μειονεκτήματα της κάθε προσέγγισης μοντελοποίησης ως προς τις στατιστικές τους επιδόσεις. Επιπρόσθετα, γίνεται μια προσπάθεια να βελτιστοποιηθεί η πρόβλεψη της κατανομής των ειδών, συνδυάζοντας τα αποτελέσματα των πλέον αποδοτικών τεχνικών. Τα ΜΚΕ, που καλούνται να συγκριθούν, εφαρμόζονται σε ένα σύνολο εκπαίδευσης, ένα ανεξάρτητο σύνολο επαλήθευσης και ένα ευρύτερο σύνολο δεδομένων το οποίο καλύπτει την περιοχή μελέτης σε πολύ υψηλή χωρική ανάλυση. Η ακρίβεια της πρόβλεψης της κάθε τεχνικής μοντελοποίησης και η ικανότητα εφαρμογής της στα δεδομένα, συγκρίνονται και αναλύονται, όπως επίσης και η καταλληλότητα τους σε ακουστικά δεδομένα και η ικανότητα των μοντέλων να εξάγουν χρήσιμη βιολογική πληροφορία για τα μικρά πελαγικά είδη.

2. Υλικά και μέθοδοι

2.1 Περιοχή μελέτης

Η περιοχή μελέτης της διατριβής είναι ο Θερμαϊκός Κόλπος (Εικ. 2.1). Πρόκειται για μια ημίκλειστη λεκάνη, με υψηλή πρωτογενή παραγωγικότητα συγκριτικά με άλλες περιοχές του Αιγαίου, λόγω των εκροών τεσσάρων μεγάλων ποταμών: Αξιού, Λουδία, Αλιάκμονα και Γαλλικού. Το υπόστρωμα του Θερμαϊκού είναι ομαλό λόγω της συνεχούς απόθεσης των ποταμών. Χαρακτηριστικό επίσης του Κόλπου είναι η αναλογικά πλατιά υφαλοκρηπίδα, που εκτείνεται νότια ως τη λεκάνη των Σποράδων. Η κυκλοφορία των υδάτινων μαζών είναι κυρίως κυκλωνική (Poulos *et al.*, 2000). Οι θαλάσσιες μάζες εισέρχονται από το Αιγαίο από βαθύτερα στρώματα κατά μήκος της ανατολικής ακτής και κινούνται αντίθετα από τη φορά του ρολογιού προς τον κόλπο της Θεσσαλονίκης (Kourafalou και Tsiaras, 2007). Αποτέλεσμα αυτής της κυκλοφορίας είναι η δημιουργία κυκλωνικών υδάτινων δομών (gyres) στην περιοχή, οι οποίες είναι ορατές από δορυφορικές εικόνες και επηρεάζουν τον κύκλο ζωής των μικρών πελαγικών ειδών (Somarakis *et al.*, 2002). Τα νερά που εισρέουν στον Θερμαϊκό από τους ποταμούς κινούνται κυρίως προς νότο κατά μήκος της δυτικής ακτής του κόλπου.



Εικόνα 2.1 Η περιοχή μελέτης και τα σημεία δειγματοληψίας

2.2 Δεδομένα

Οι τεχνικές μοντελοποίησης κατανομής ειδών εξαρτώνται από τον τύπο των δεδομένων και από την ακρίβεια τους. Ανάλογα με τον τύπο των δεδομένων της εξαρτημένης μεταβλητής τα μοντέλα μπορούν να ταξινομηθούν ως: μοντέλα που διαχειρίζονται δεδομένα μόνο-παρουσίας, μοντέλα που διαχειρίζονται δεδομένα παρουσίας-απουσίας και μοντέλα που διαχειρίζονται δεδομένα αφθονίας. Στα μοντέλα παρουσίας τα δείγματα παρέχουν την πληροφορία παρουσίας των ειδών και μόνο, ενώ τα δεδομένα παρουσίας-απουσίας προέρχονται από συστηματικές δειγματοληψίες, όπου υπάρχει και η πληροφορία απουσίας του είδους (1:παρουσία-0:απουσία). Τα δεδομένα αφθονίας μπορούν πολλές φορές να μετατραπούν σε μορφή παρουσίας-απουσίας προκειμένου να αξιοποιηθεί μεγαλύτερο εύρος τεχνικών μοντελοποίησης ή για να εξυπηρετηθούν ανάλογοι σκοποί διαχείρισης και παρακολούθησης (Pollock, 2006). Η πρώτη ομάδα ΜΚΕ αναπτύσσεται χρησιμοποιώντας δεδομένα αφθονίας, ενώ η δεύτερη ομάδα αξιοποιεί τεχνικές που υποστηρίζουν δεδομένα παρουσίας-απουσίας ή μόνο-παρουσίας. Ως ανεξάρτητες μεταβλητές μπορούν να χρησιμοποιηθούν μεταβλητές που άμεσα ή έμμεσα δύναται να ερμηνεύουν μέρος της διακύμανσης της εξαρτημένης μεταβλητής. Υπάρχουν πολλές στρατηγικές επιλογής επεξηγηματικών μεταβλητών, ανάλογα με την εκάστοτε τεχνική μοντελοποίησης. Εδώ έγινε προσπάθεια να συλλεχθούν όλες οι διαθέσιμες παράμετροι και να δημιουργηθούν νέες, ώστε να ενισχυθεί η ρεαλιστικότητα και ακρίβεια των αποτελεσμάτων των ΜΚΕ.

2.2.1 Δεδομένα αφθονίας πελαγικών ειδών

Τα ακουστικά δεδομένα έχουν συλλεχθεί κατά τη διάρκεια του Απριλίου και Μαΐου 1998 με τη χρήση βαθμονομημένου συστήματος 38 kHz SIMRAD EK 500/BI 500 (Bodholt *et al.*, 1989). Ο συντελεστής ακουστικής ανάκλασης ανά μονάδα επιφάνειας (area backscattering coefficient s_A , $m^2 n \text{ mi}^{-2}$, NASC, MacLennan *et al*, 2002) είναι ανάλογος της αφθονίας των πελαγικών ειδών στο δείγμα:

$$\rho_a = s_A \frac{C_E}{\hat{\sigma}}$$

Όπου ρ_a είναι η πυκνότητα ατόμων ανά επιφάνεια (α) στήλης νερού, C_E ο συντελεστής της βαθμονόμησης του όλου συστήματος και $\hat{\sigma}$ η μέση ακουστική διατομή του πελαγικού είδους (Simmonds & MacLennan, 2005). Οι ολοκληρωμένες τιμές s_A με οριζόντια χωρική ανάλυση ενός μιλίου χρησιμοποιήθηκαν, σύμφωνα με τον παραπάνω τύπο, είτε αυτούσιες ως δείκτης αφθονίας (αναφέρονται στην εργασία ως ακουστική πυκνότητα), είτε μετατράπηκαν σε δεδομένα παρουσίας-απουσίας. Δεν προτιμήθηκε η μετατροπή των ακουστικών δεδομένων σε βιομάζα για να αποφευχθεί η εισαγωγή επιπλέον αβεβαιότητας στην εξαρτημένη μεταβλητή, που θα προέκυπτε από την αναγωγή στην ακουστική πληροφορία της κατανομής των ειδών και των μηκών των ψαριών από τις δειγματοληπτικές τράτες (Walline, 2007). Η πλειονότητα των ειδών-στόχων ήταν *Sardina pilchardus* (~55%), *Engraulis encrasicolus* (~25%) and *Trachurus spp* (<10%), όπως προέκυψε από την αναγνώριση των ειδών και των εκφορτώσεων, που έγιναν παράλληλα με τον ηχοβολισμό. Επομένως τα μοντέλα κατανομής ειδών αποτυπώνουν την κατανομή μικρών πελαγικών ειδών στην περιοχή, που κυρίως ανταποκρίνεται στο γάυρο και τη σαρδέλα. Η βιολογία και η οικολογία των παραπάνω ειδών χρησιμοποιήθηκε σε διάφορα στάδια της ανάλυσης, όπως στην επιλογή επεξηγηματικών μεταβλητών, στην ερμηνεία των χαρτών κατανομής και στη σύγκριση των αποτελεσμάτων των μοντέλων με κατανομές άλλων μελετών.

Για την καλύτερη κατανόηση των ακουστικών δεδομένων και για τη διερεύνηση πιθανών μετασχηματισμών πριν τη χρήση τους στα μοντέλα πραγματοποιήθηκαν οι ακόλουθες διεργασίες: γραφήματα Quantile-Quantile για τον έλεγχο πιθανών μετασχηματισμών, boxplots και dotplots για την επισκόπηση της κατανομής των τιμών, βαριόγραμμα και ανάλυση αυτοσυσχέτισης για τον προσδιορισμό των χωρικών συσχετίσεων στα δεδομένα αλλά και τον υπολογισμό της ποσότητας τους (Zuur *et al.*, 2010).

2.2.1.1 Δημιουργία χρονοσειρών σε κάθε σύνολο δεδομένων και κριτήρια σύνθεσης του συνόλου επαλήθευσης.

Ο αριθμός των ακουστικών εγγραφών ήταν αρκετός, ώστε να δημιουργηθούν δύο ομάδες δεδομένων: μια για την εκπαίδευση των μοντέλων και μια για την αξιολόγησή τους. Το πρώτο σύνολο περιλαμβάνει το 80% των αρχικών δεδομένων (μαύρες τελείες στην Εικ. 2.1). Το δεύτερο, που χρησιμοποιείται για την αξιολόγηση της ικανότητας πρόβλεψης των μοντέλων, αποτελείται από το υπόλοιπο 20% των αρχικών τιμών. Ο διαχωρισμός των αρχικών δεδομένων στα δύο σύνολα έγινε με βάση τους Tetko *et al.* (1995), όπου διαδοχικά σημεία διαχωρίζονται για τη δημιουργία του συνόλου εκπαίδευσης και επαλήθευσης. Σε αυτή την εργασία για κάθε πέντε διαδοχικά σημεία τα τέσσερα πρώτα χρησιμοποιήθηκαν για τη σύνθεση του συνόλου εκπαίδευσης ενώ το πέμπτο για το σύνολο επαλήθευσης. Επομένως το σύνολο επαλήθευσης αποτελείται από σημεία δειγματοληψίας που καλύπτουν όλη την περιοχή μελέτης ανά τουλάχιστον ~5 ναυτικά μίλια. Η επιλογή αυτής της απόστασης, που καθορίζει και την αναλογία δεδομένων εκπαίδευσης και επαλήθευσης, έγινε με βάση το γεγονός ότι σε αυτή την απόσταση δεν εντοπίζεται χωρική αυτοσυσχέτιση. Για τον προσδιορισμό πιθανών χωρικών προτύπων στα αρχικά δεδομένα, τα δεδομένα εκπαίδευσης και επαλήθευσης, χρησιμοποιήθηκαν ημι-βαριογράμματα (Matheron, 1971), που δημιουργήθηκαν με τη χρήση του ArcGIS Geostatistical Analyst Software (GAS) της ESRI και γραφήματα, που δημιουργήθηκαν από το Auto-correlation Function Estimation του στατιστικού πακέτου R (R Development Core Team, 2005). Για τον ίδιο σκοπό χρησιμοποιήθηκε και το *I* στατιστικό χωρικής αυτοσυσχέτισης του Moran (Moran, 1950). Πέρα από τα παραπάνω διερευνητικά εργαλεία, μετά την ανάπτυξη των ΜΚΕ εξετάστηκε ο ομοσκεδασμός των σφαλμάτων τους με γραφήματα της εξαρτημένης μεταβλητής και των σφαλμάτων. Για τη σύγκριση των συνόλων εκπαίδευσης και επαλήθευσης χρησιμοποιήθηκαν μη παραμετρικές τεχνικές καθώς κανένα από τα δύο δεν χαρακτηριζόταν από κανονική κατανομή. Συγκεκριμένα χρησιμοποιήθηκαν τα κριτήρια των Mann-Whitney και Kolmogorov-Smirnov. Με το σύνολο των παραπάνω διεργασιών τεκμηριώνεται η καταλληλότητα της χρήσης του συνόλου επαλήθευσης για την αντικειμενική αξιολόγηση των προβλέψεων των ΜΚΕ.

2.2.2 Περιβαλλοντικά και χωροχρονικά δεδομένα

Αρχικά συλλέχθηκαν όλα τα διαθέσιμα δεδομένα, τα οποία θα μπορούσαν να συσχετιστούν με την κατανομή των μικρών πελαγικών ειδών. Τα δεδομένα θα μπορούσαν να ταξινομηθούν ως περιβαλλοντικές παράμετροι (π.χ. αλατότητα, θερμοκρασία κ.α.), χωρικές παράμετροι (π.χ. απόσταση από την ακτή), χρονικές παράμετροι (π.χ. ημέρα και ώρα δειγματοληψίας) και ωκεανογραφικές παράμετροι (π.χ. ρεύματα, σημεία υψηλής παραγωγικότητας όπως περιγράφονται στην εργασία των Valavanis και συνεργατών (2004).

Λεπτομερής διερεύνηση των δεδομένων πραγματοποιήθηκε κατά τη σύνθεση των σειρών δεδομένων με σκοπό τη διερεύνηση χωρικών προτύπων, ακραίων τιμών, συσχετίσεων και αλληλεπιδράσεων μεταξύ των μεταβλητών. Κάποια από τα αποτελέσματα της διερεύνησης των δεδομένων, καθώς και περιγραφές των διαδικασιών που χρησιμοποιήθηκαν παρουσιάζονται στο Παράρτημα II. Η παραπάνω διαδικασία παρέχει σημαντικές πληροφορίες (Zuur *et al.*, 2010), που αξιοποιούνται κατά την επιλογή μεταβλητών στα μοντέλα, ώστε: α) να τηρηθούν οι παραδοχές της κάθε στατιστικής προσέγγισης, β) να επισημανθούν πιθανές μετατροπές μεταβλητών και γ) να αποκτηθεί καλύτερη γνώση των δεδομένων. Η διαδικασία της διερεύνησης των δεδομένων περιλαμβάνει: i) ζευγαρωτά γραφήματα (pair-plots), που προσδιορίζουν τη συγγραμμικότητα των μεταβλητών, η οποία μπορεί να οδηγήσει σε λανθασμένη επιλογή παραμέτρων κατά την ανάπτυξη των μοντέλων (Zuur *et al.*, 2007), ii) dot-plots και box-plots, για τον προσδιορισμό ακραίων τιμών, που μπορεί να επηρεάσουν τις σχέσεις εξαρτημένων-ανεξάρτητων μεταβλητών, iii) γραφήματα Quantile-Quantile (Q-Q plots) και ιστογράμματα, για την επισκόπηση της κατανομής των τιμών των

μεταβλητών και iv) conditional box-plots, για τη διερεύνηση πιθανών σχέσεων και αλληλεπιδράσεων μεταξύ των μεταβλητών.

Η διαδικασία της διερεύνησης (Παράρτημα II) οδήγησε στην επιλογή εννέα κατάλληλων επεξηγηματικών μεταβλητών, που καλύπτουν ενδεχομένως την κατανομή των μικρών πελαγικών ειδών. Οι επεξηγηματικές μεταβλητές είναι οι ακόλουθες: Επιφανειακή θερμοκρασία της θάλασσας (SST), επιφανειακή συγκέντρωση χλωροφύλλης-α (CHL), αλτιμετρία (SLA), διαθέσιμη φωτοσυνθετική ακτινοβολία (PAR), απόσταση από την ακτή (DCoast), κλίση της επιφανειακής θερμοκρασίας της θάλασσας (SSTsl), βάθος (DEP), κλίση του βυθού (DEPsl) και μια κατηγορηματική παράμετρος που χωρίζει τη μέρα σε τέσσερα χρονικά διαστήματα (σούρουπο, μέρα, δύση, βράδυ, DDND). Στο Παράρτημα I απεικονίζονται οι κάρτες (grids) των παραπάνω επεξηγηματικών μεταβλητών. Η παράμετρος DDND επιλέχθηκε με βάση τη συμπεριφορά των μικρών πελαγικών ειδών, ώστε να ανταποκρίνεται στις νυχθήμερες κάθετες μεταναστεύσεις (Giannoulaki *et al.*, 1999) οι οποίες αποτυπώνονται αντίστοιχα στην ηχοβολιστική διαδικασία. Οι μεταβλητές που εξαιρέθηκαν μετά τη διερευνητική διαδικασία είναι οι ακόλουθες: ημέρα δειγματοληψίας, βροχόπτωση, αλατότητα, διεύθυνση και ταχύτητα ρευμάτων, ένταση και κατεύθυνση του ανέμου, περιοχές υψηλής παραγωγικότητας hotspots (Valavanis *et al.*, 2004), θερμικά μέτωπα μέσης κλίμακας (Valavanis *et al.*, 2005) και η ώρα δειγματοληψίας. Οι παραπάνω μεταβλητές παρουσίασαν πολλή υψηλή συσχέτιση με κάποιες από τις μεταβλητές που χρησιμοποιήθηκαν στα μοντέλα, όπως φαίνεται και στα αντίστοιχα pairplots. Επιλέχθηκε η μικρότερη διαθέσιμη χρονική ανάλυση των δορυφορικών παραμέτρων, ώστε να ανταποκρίνεται όσο το δυνατόν περισσότερο στις συνθήκες της δειγματοληψίας. Για τη μετατροπή των δορυφορικών δεδομένων σε κάρτες και την εξαγωγή σημειακών τιμών αντίστοιχων της δειγματοληψίας χρησιμοποιήθηκαν διάφορες τεχνικές Γεωγραφικών Πληροφοριακών Συστημάτων (δες Valavanis, 2002). Στον Πίνακα 2.1 παρουσιάζονται όλα τα δεδομένα που χρησιμοποιήθηκαν και οι πηγές τους.

Πίνακας 2.1 Τα δεδομένα, οι πηγές και η μορφή τους.

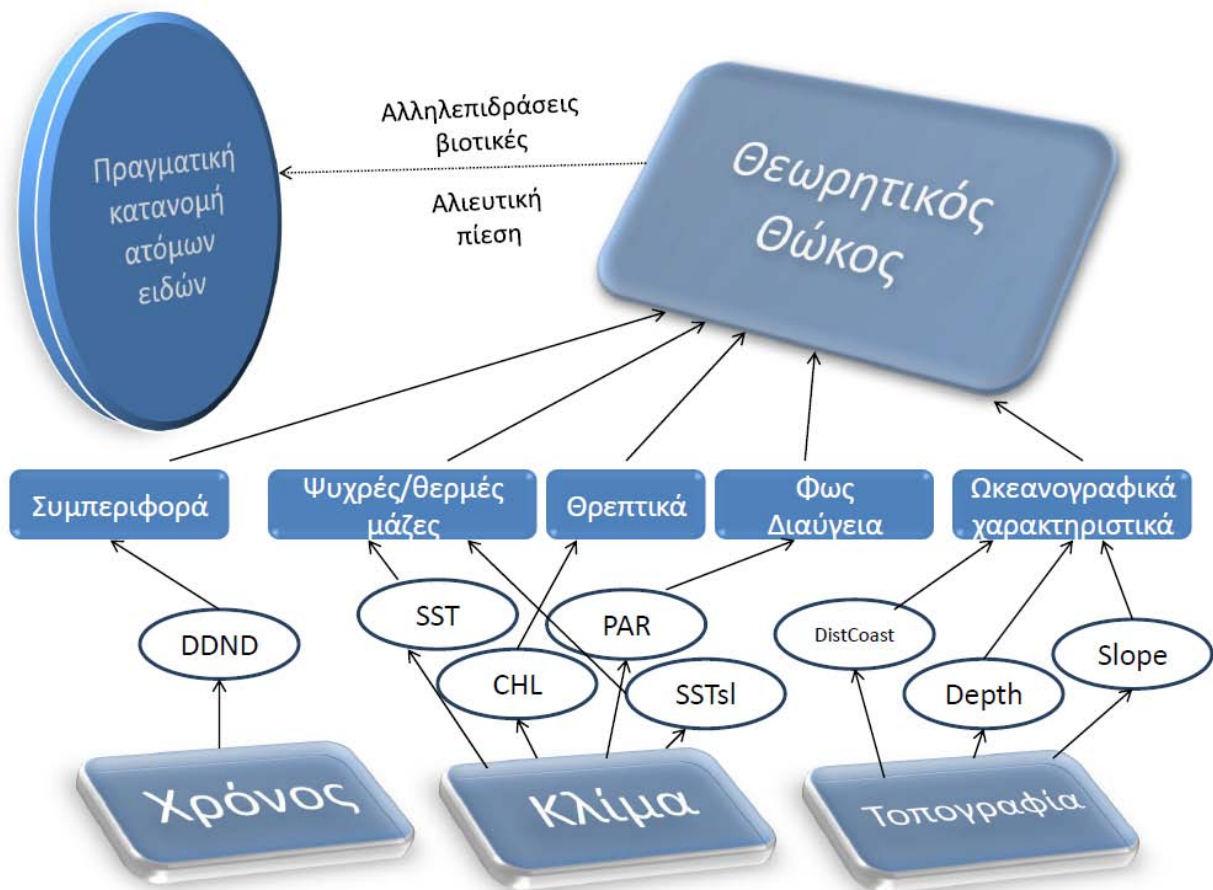
Μεταβλητές	Συντομο/φίες	Τύπος δεδομένων / Αισθητήρας	Πηγή δεδομένων
Ακουστική πυκνότητα	s_A	συντελεστής ακουστικής ανάκλασης ανά μονάδα επιφάνειας s_A ανά ESDU=1nm),	SIMRAD EK500/BI500 σύστημα τον Απρίλη/Μάιο 1998 Θερμαϊκό Κόλπο
Επιφανειακή θερμοκρασία	SST	Κάνναβος / Aqua MODIS	German Aerospace Agency (DLR)
Συγκέντρωση χλωροφύλλης	CHL	Κάνναβος / Aqua MODIS	Distributed Active Archive Center (NASA)
Διαθέσιμη φωτοσυνθετική ακτινοβολία	PAR	Κάνναβος / SeaWiFS	Distributed Active Archive Center (NASA)
Υψομετρία	SLA	Κάνναβος / Merged Jason-1, Envisat, ERS-2, GFO, T/P	AVISO
Βροχόπτωση	PRE	Κάνναβος	Mediterranean Oceanic Database (MODB)
Αλατότητα	SSS	Κάνναβος / CARTON-GIESE SODA and CMA BCC GODAS models	Mercator operational oceanography
Βαθυμετρία	DEP	Κάνναβος / Processed ERS-1, Geostat and historical depth soundings	Laboratory for Satellite Altimetry (NOAA)
Ένταση και κατεύθυνση ανέμου	WS & WD	Κάνναβος & cover	Mercator operational oceanography
Ακτογραμμή	Coast	Cover / Ψηφιοποίηση ναυτικών χαρτών και αεροφωτογραφιών	Υπουργείο Περιβάλλοντος
Απόσταση από ακτή	DCoast	Κάνναβος & cover	Εξάγεται από την ακτογραμμή
Κλίση βάθους	DEPsl	Κάνναβος	Εξάγεται από τον κάβναβο της βαθυμετρίας
Κλίση θερμοκρασίας	SSTsl	Κάνναβος	Εξάγεται από τον κάβναβο SST
Περιοχές αυξημένης παραγωγικότητας	MPH	Κάνναβος	Σύμφωνα με τους Valavanis <i>et al.</i> 2004
Μέσης κλίμακας θερμικά μέτωπα	MTF	Διανυσματικός χάρτης	Σύμφωνα με τους Valavanis <i>et al.</i> 2005
Γεωγραφικό μήκος και πλάτος	LON,LAT	Διανυσματικός χάρτης σε δεκαδικές μοίρες και μέτρα	SIMRAD EK500/BI500 system on April/May 1998 in Thermaikos Gulf
Κατεύθυνση και ταχύτητα ρευμάτων	CURSP & CURDR	Κάβναβος & Διανυσματικός χάρτης / NEMO (OPA9 + LIM)	Mercator operational oceanography
Κατηγορικός παράγοντας χρόνου	DDND	Διανυσματικός χάρτης & Κάβναβος	Βάση της μέρας & ώρας δειγματοληψίας
Ημερομηνία	DT	Διανυσματικός χάρτης	Βάση των δειγματοληπτικών δεδομένων

Για κάθε επεξηγηματική μεταβλητή αλλά και για τα ακουστικά δεδομένα δημιουργήθηκαν τρία σύνολα δεδομένων: Το πρώτο (DS1) ανταποκρίνεται στο σύνολο εκπαίδευσης των μοντέλων, το δεύτερο (DS2) αναπαριστά κάθε σημείο στον κάβναβο της περιοχής μελέτης σε ανάλυση 0.01 δεκαδικές μοίρες (~1km) και το τρίτο (DS3) αποτελεί το σύνολο επαλήθευσης. Η χωρική ανάλυση των επεξηγηματικών μεταβλητών ποικίλει από 0.01 ως 0.04 δεκαδικές μοίρες. Για τους σκοπούς της ανάπτυξης των ΜΚΕ όλα τα σύνολα δεδομένων μετασχηματίστηκαν στη μικρότερη διαθέσιμη χωρική ανάλυση των 0.01 δεκαδικών μοιρών, ώστε να μην απολεσθεί διαθέσιμη πληροφορία. Το σύνολο των επεξηγηματικών μεταβλητών που καλύπτει την περιοχή μελέτης με χωρική ανάλυση 0.01 της μοίρας εφαρμόστηκε στα ΜΚΕ, ώστε να παραχθούν χάρτες πρόβλεψης της κατανομής των ειδών σε πολύ υψηλή ανάλυση και εκτός των περιοχών που προέρχονται τα δειγματοληπτικά δεδομένα. Η μελέτη της εξαρτημένης μεταβλητής οδήγησε στην επιλογή των κατάλληλων ανεξάρτητων μεταβλητών, που να δύνανται να καλύψουν ικανοποιητικά τόσο την περιβαλλοντική όσο και τη γεωγραφική διάσταση του μοντέλου (Elith και Leatwick, 2009). Η πολυδιάστατη

φύση των επεξηγηματικών μεταβλητών γεννά ζητήματα αυτοσυσχέτισης και συγγραμμικότητας, τα οποία ενδέχεται να επηρεάσουν τη λειτουργία των ΜΚΕ και τα οποία αναλύονται παρακάτω.

2.3 Μοντέλα κατανομής ειδών

Τα μοντέλα κατανομής ειδών, όπως αναφέρθηκε και στην εισαγωγή, είναι στενά συνδεδεμένα με τη θεωρία και τις αρχές της Οικολογίας και της Βιογεωγραφίας. Πριν, λοιπόν, την τεχνική παρουσίαση των μεθόδων μοντελοποίησης είναι απαραίτητη η έκθεση του θεμελιώδους μοντέλου (Εικ. 2.2) πάνω στο οποίο στηρίχθηκε η ανάπτυξη των ΜΚΕ. Άλλωστε σύμφωνα με τους Guisan *et al.* (2006) ένα από τα πρώτα βήματα στην κατασκευή ενός μοντέλου πρόβλεψης της κατανομής ατόμων ειδών είναι η θεώρηση ενός θεμελιώδους μοντέλου των αναμενόμενων σχέσεων είδους-περιβάλλοντος. Στόχος της διατριβής είναι, εκτός των άλλων, η σύγκριση της ακρίβειας πρόβλεψης των ΜΚΕ, καθώς και της ικανότητας προσαρμογής τους που σχετίζονται άμεσα με τον προσδιορισμό της πραγματικής κατανομής των ατόμων των ειδών, σε χωρική ανάλυση, που να είναι ικανή να διακρίνει ακόμη και την πιθανή ετερογένεια των ενδιαιτημάτων στην περιοχή μελέτης. Για το λόγο αυτό έγινε προσπάθεια να συλλεχθούν και να αξιοποιηθούν ένας μεγάλος αριθμός επεξηγηματικών μεταβλητών, που να είναι ικανές να προσεγγίσουν την πολυπλοκότητα των σχέσεων κατανομή είδους-περιβαλλοντικές συνθήκες. Στο θεμελιώδες μοντέλο που παρουσιάζεται στην Εικ. 2.2 έχουν χρησιμοποιηθεί οι επεξηγηματικές μεταβλητές, που επελέγησαν τελικά για την ανάπτυξη των ΜΚΕ και όχι όλες όσες αρχικά είχαν προσδιοριστεί. Όπως φαίνεται στο θεμελιώδες μοντέλο έχουν προσδιοριστεί πέντε συστήματα, που επηρεάζουν το θεωρητικό θώκο των μικρών πελαγικών ειδών: (α) Η συμπεριφορά τους, που σχετίζεται με τη βιολογία και τον κύκλο ζωής των ειδών, (β) ψυχρές και θερμές υδάτινες μάζες, (γ) τα διαθέσιμα θρεπτικά συστατικά, (δ) το φως ή η διαύγεια της υδάτινης στήλης και (ε) διάφορα ωκεανογραφικά χαρακτηριστικά. Τα παραπάνω συστήματα περιγράφονται από τις ανάλογες επεξηγηματικές μεταβλητές, που γενικά μπορούν να ταξινομηθούν σε τρεις γενικές κατηγορίες μεταβλητών: χρόνος, κλίμα ή περιβάλλον και τοπογραφία. Παράγοντες όπως η αλιευτική πίεση, που δεν είναι διαθέσιμη στην απαιτούμενη μορφή και οι βιοτικές αλληλεπιδράσεις, που μοντελοποιούνται ως συστατικό της εξαρτημένης μεταβλητής οδηγούν από το θεωρητικό θώκο στην πραγματική κατανομή των ατόμων των ειδών. Θα πρέπει να σημειωθεί ότι η δημιουργία του θεμελιώδους μοντέλου έγινε με τις διαθέσιμες μεταβλητές, γεγονός που όρισε το εύρος της εξέλιξής του, ώστε να αξιοποιηθεί πλήρως η βιολογική γνώση των μικρών πελαγικών ειδών. Από το θεμελιώδες μοντέλο εξάγονται και οι σχέσεις ειδών περιβάλλοντος, που αναλύονται ανά μέθοδο μοντελοποίησης, όπου είναι διαθέσιμες.



Εικόνα 2.2 Θεμελιώδες μοντέλο των περιβαλλοντικών παραμέτρων που ελέγχουν τα κύρια περιβαλλοντικά όρια της κατανομής των μικρών πελαγικών ειδών. Συμπληρώνεται με τις βιοτικές αλληλεπιδράσεις και την αλιευτική πίεση, ώστε να καταλήξει το θεμελιώδες μοντέλο στην πραγματική κατανομή των ατόμων των ειδών (εμπνευσμένο από τους Franklin, 2009, Guissan και Zimmermann, 2000).

Στη συνέχεια περιγράφονται τα μοντέλα που αναπτύχθηκαν στην διατριβή. Κάποια από αυτά είναι πολύ κοινά στο χώρο της μοντελοποίησης κατανομής ειδών, ενώ άλλα χρησιμοποιούνται για πρώτη φορά.

2.3.1 Μοντέλα αφθονίας

2.3.1.1 Μοντέλα παλινδρόμησης

Τα Γενικευμένα Προσθετικά Μοντέλα (GAMs) είναι γενικευμένα μοντέλα, τα οποία ενσωματώνουν μια λειτουργία εξομάλυνσης των μεταβλητών (Hastie και Tibshirani, 1990; Wood, 2008). Τα Γενικευμένα Προσθετικά Μεικτά Μοντέλα (GAMMs) χρησιμοποιούνται συμπληρωματικά με τα GAMs με σκοπό να διαχειριστούν τη χωρική αυτοσυσχέτιση των μεταβλητών. Για την εφαρμογή των παραπάνω μοντέλων η ακουστική πληροφορία μετασχηματίστηκε με τη χρήση λογαρίθμου. Η επιλογή του κατάλληλου τρόπου μετασχηματισμού έγινε με τη χρήση Q-Q plots (Cleveland, 1994). Η επιλογή του εξομαλυντή βασίστηκε στη μέθοδο, που προτάθηκε από τους Wood και Augustin (2002), χρησιμοποιώντας τη βιβλιοθήκη 'mgcv' του στατιστικού πακέτου R (R Development Core Team, 2005). Ο βαθμός εξομάλυνσης επιλέχθηκε βάσει της εξαρτημένης μεταβλητής και της μεθόδου γενικευμένης διασταυρωτής επικύρωσης (Generalized Cross Validation) (Wood, 2006). Η επιλογή του μοντέλου με την καλύτερη εφαρμογή έγινε με τη χρήση του κριτηρίου Akaike's Information Criterion (AIC) (Akaike, 1974), ενώ κατά την ανάπτυξη των μοντέλων ακολουθήθηκε βηματική εμπρόσθια επιλογή για να περιοριστεί η συγγραμμικότητα μεταξύ των εξηγηματικών μεταβλητών. Οι μεταβλητές που δεν είχαν σημαντική συσχέτιση μεταξύ τους χρησιμοποιήθηκαν κατά την ανάπτυξη των μοντέλων ως εξηγηματικές. Με άξονα τα γραφήματα Q-Q

κάποιες από αυτές μετασχηματίστηκαν. Τα μοντέλα GAMMs αναπτύχθηκαν πάνω στο τελικά επιλεγμένο GAM μοντέλο με την προσθήκη μιας επιπλέον παραμέτρου που εισάγει στο μοντέλο μια συγκεκριμένη δομή συσχέτισης μεταξύ των τιμών στα σημεία της δειγματοληψίας. Η προσθήκη αυτή αυξάνει την πολυπλοκότητα των GAMMs σε σχέση με τα GAMs, επιτρέποντας όμως την ενσωμάτωση χωρικών προτύπων στην προσομοίωση των δεδομένων εκπαίδευσης.

2.3.1.2 Συνδεδετικά Νευρωνικά Δίκτυα (Associative Neural Networks)

Τα Συνδεδετικά Νευρωνικά Δίκτυα (ASNN) ανήκουν σε μια ευρύτερη κατηγορία ευέλικτων μη γραμμικών ταξινομητών και παλινδρομικών μοντέλων. Κατά την εφαρμογή τους συνδυάζουν πολύπλοκες μεθόδους, που ανήκουν στις τεχνικές μηχανικής εκμάθησης (machine learning techniques), όπου η διαδικασία ανάπτυξης τους δεν εποπτεύεται απόλυτα από το χρήστη (Michie *et al.*, 1994). Σύμφωνα με τον Haykin (1994), τα νευρωνικά δίκτυα αποτελούν μια ογκώδη, παράλληλα κατανομημένη, επεξεργασία με φυσική κλίση στην αποθήκευση πειραματικής γνώσης, που γίνεται διαθέσιμη για χρήση. Τα νευρωνικά δίκτυα μοιάζουν με τη λειτουργία του εγκεφάλου σε δύο επίπεδα: i) η γνώση αποκτάται από το δίκτυο μέσω μιας λειτουργίας εκμάθησης και ii) οι ενδονευρωνικές συνδέσεις αποδίδουν ενδοσυναπτικά βάρη στις μεταβλητές αποθηκεύοντας έτσι την κεκτημένη γνώση.

Τα ASNN αποτελούν μια μέθοδο με βελτιωμένη ικανότητα πρόβλεψης. Τη μέθοδο αυτή συνθέτουν ένα σύνολο εμπρόσθια-τροφοδοτούμενων νευρωνικών δικτύων και μια τεχνική k-κοντινής γειτνίασης (k-nearest neighbor). Το σύνολο των νευρωνικών δικτύων αποτελεί μια διαδικασία εξαγωγής μέσης πρόβλεψης από διάφορα ανεξάρτητα νευρωνικά δίκτυα και έχει αναπτυχθεί χρησιμοποιώντας ένα κρυμμένο επίπεδο (hidden layer) και τρεις νευρώνες. Ο αριθμός k του αλγορίθμου nearest neighbor και η παράμετρος σ για την παλινδρόμηση Parzen-window αποτελούν τις παραμέτρους εξομάλυνσης των ASNN, ώστε να μειωθεί το σφάλμα πρόβλεψης για το σύνολο των δεδομένων εκπαίδευσης του μοντέλου (Tetko *et al.*, 1995).

Τα ASNNs αρχικά αναπτύχθηκαν και εφαρμόστηκαν στη Χημεία (Tetko *et al.*, 1995) παρέχοντας πιο ακριβείς προβλέψεις από τα απλά νευρωνικά δίκτυα. Η φιλοσοφία και το υπόβαθρο των ASNN αναλύονται εκτεταμένα στις δημοσιεύσεις του Tetko (2002a, b). Λόγω της πρώτης εφαρμογής των ASNN στις θαλάσσιες επιστήμες, εκτός από τα ASNN αναπτύσσεται και το ανάλογο σύνολο νευρωνικών δικτύων ANNE-artificial neural networks ensemble- (που σε συνδυασμό με την τεχνική k-nearest neighbor δίνει τα ASNN) για την πιο ολοκληρωμένη και εποπτική παρακολούθηση και των δύο προσεγγίσεων μοντελοποίησης. Κατά την ανάπτυξη των δύο τύπων νευρωνικών προσεγγίσεων πραγματοποιήθηκαν επιλογές, που έχουν να κάνουν με τον αλγόριθμο εκπαίδευσης, τον αριθμό των νευρώνων, τα κρυμμένα επίπεδα, τις επαναλήψεις του αλγορίθμου και τον αριθμό των συνολικών δικτύων. Η μέθοδος Early Stopping over Ensemble (ESE) χρησιμοποιήθηκε για την εκπαίδευση των νευρωνικών δικτύων (Tetko και Tanchuk, 2002; Bishop, 1995). Τελικά επιλέχθηκαν τα μοντέλα με τη μικρότερη τιμή της ρίζας του μέσου τετραγώνου του σφάλματος πρόβλεψης (RMSE) και το μέσο απόλυτο σφάλμα (MAE) (Tetko *et al.*, 2008). Αυτά στη συνέχεια χρησιμοποιήθηκαν με τον αλγόριθμο δευτέρου βαθμού Levenberg-Marquardt, έχοντας τρεις νευρώνες σε ένα κρυμμένο επίπεδο, 600 δίκτυα σε κάθε σύνολο και 80 επαναλήψεις στην εκπαίδευση κάθε δικτύου. Ο αλγόριθμος Levenberg-Marquardt πλεονεκτεί έναντι άλλων, διότι δεν πέφτει σε τοπικά ελάχιστα (local minima) και παρέχει μειωμένα σφάλματα πρόβλεψης για δεδομένο αριθμό νευρώνων (Shepherd, 1997). Τα ANNE και ASNN αναπτύχθηκαν με τη χρήση του λογισμικού ASNN του VCCLAB (Virtual Computational Chemistry Laboratory, <http://www.vcclab.org>, 2005).

2.3.1.3 Kriging

Η μέθοδος της στάθμισης με αντίστροφη απόσταση (Inverse Distance Weighted, IDW) και η μέθοδος των Splines (π.χ. στα GAMs) χαρακτηρίζονται ως ντετερμινιστικά μοντέλα παρεμβολής, επειδή βασίζονται απευθείας στη μέτρηση γειτονικών τιμών ή σε συγκεκριμένο μαθηματικό τύπο, που καθορίζει το μέγεθος της ομαλότητας στην τελική επιφάνεια. Μια δεύτερη κατηγορία μεθόδων παρεμβολής περιλαμβάνει

γεωστατιστικές προσεγγίσεις όπως το Kriging, οι οποίες βασίζονται σε στατιστικά μοντέλα, που αξιοποιούν την αυτοσυσχέτιση (τη στατιστική συσχέτιση μεταξύ γειτονικών μετρούμενων σημείων). Επομένως οι γεωστατιστικές μέθοδοι εκτός από την ικανότητά τους να παράγουν επιφάνειες πρόβλεψης, μπορούν επίσης να εκτιμούν την ακρίβεια και τη βεβαιότητα της πρόβλεψης (Georgakarakos και Kitsiou, 2008). Ο σκοπός της γεωστατιστικής (Matheron, 1971) είναι η μοντελοποίηση της χωρικής διακύμανσης μιας δεδομένης μεταβλητής και η χρήση του μοντέλου για την εκτίμηση της τιμής της μεταβλητής σε δεδομένα σημεία. Η γεωστατιστική έχει εκτεταμένη εφαρμογή στην εκτίμηση αφθονίας πληθυσμών ψαριών από επιστημονικές δειγματοληψίες (Petitgas, 2001).

Το Kriging θεωρεί ότι οι αποστάσεις και οι κατευθύνσεις μεταξύ των σημείων δειγματοληψίας αντικατοπτρίζουν τη χωρική συσχέτιση, η οποία μπορεί να χρησιμοποιηθεί για να εξηγήσει τη διακύμανση σε μια επιφάνεια. Η μέθοδος του Kriging εφαρμόζει μια μαθηματική λειτουργία σε ένα καθορισμένο αριθμό σημείων ή όλων των σημείων σε καθορισμένη ακτίνα, ώστε να ορίσει τις εξαγόμενες τιμές σε κάθε περιοχή. Είναι μια διαδικασία πολλών σταδίων, που περιλαμβάνει διερευνητικές στατιστικές αναλύσεις των δεδομένων, βαριογράμματα, δημιουργία επιφάνειας πρόβλεψης και (προαιρετικά) διερεύνηση της διακύμανσης της επιφάνειας. Το Kriging είναι κατάλληλο για χρήση όταν υπάρχουν χωρικές συσχετίσεις ή σφάλματα κατεύθυνσης στα δεδομένα. Το Ordinary Kriging αποτελεί την πιο γενική και ευρέως χρησιμοποιούμενη μέθοδο Kriging. Προϋποθέτει ότι ο σταθερός μέσος των δεδομένων είναι άγνωστος, εκτός εάν υπάρχουν επιστημονικές ενδείξεις για την απόρριψη της θεώρησης αυτής.

Σε αυτή την εργασία, για τις γεωστατιστικές αναλύσεις χρησιμοποιήθηκε το ArcGIS Geostatistical Analyst Software (GAS) της ESRI, το οποίο παρέχει ένα ολοκληρωμένο σύνολο εργαλείων για την εφαρμογή διαφόρων μεθόδων Kriging. Η μεθοδολογία του Kriging περιλαμβάνει την επιλογή διαφορετικής βαρύτητας στις εκτιμήσεις, ώστε αυτές να είναι κατά το δυνατό χωρίς σφάλματα και να ελαχιστοποιείται η διασπορά της πρόβλεψης. Το καλύτερα εφαρμοσμένο μοντέλο ή βαριόγραμμα επιλέγεται έπειτα από την εκπαίδευση και επαλήθευση μεγάλου αριθμού μοντέλων, διαδικασία την οποία υποστηρίζει το GAS (Georgakarakos και Kitsiou, 2008).

Τελικά τα μοντέλα Kriging, που αναπτύχθηκαν, αξιολογήθηκαν με ανάλυση των σφαλμάτων τους με leave-one-out διασταυρωτή επικύρωση, επίσης τα στατιστικά τους συγκρίθηκαν μεταξύ τους, όσον αφορά στον έλεγχο των προϋποθέσεων των μοντέλων και την ακρίβεια των τυπικών σφαλμάτων (Isaaks και Srivastava, 1989). Η σύγκριση των μοντέλων έγινε μέσω της ρίζας του μέσου κανονικοποιημένου σφάλματος (RMSE), του μέσου τυπικού σφάλματος (ASE) και του συντελεστή προσδιορισμού (coefficient of determination, R^2). Όταν το ASE είναι κοντά στο RMSE τότε υπάρχει η ένδειξη ορθώς εκτιμώμενης διακύμανσης στην πρόβλεψη. Αν το ASE είναι αρκετά μεγαλύτερο ή μικρότερο από του RMSE η διακύμανση των προβλέψεων υπέρ- ή υποεκτιμάται αντίστοιχα.

Για τους σκοπούς της συγκεκριμένης εργασίας επιλέχθηκαν τελικά δύο μοντέλα Kriging. Το πρώτο (DefKrig) αναπτύχθηκε με σκοπό την ελαχιστοποίηση των σφαλμάτων με τη χρήση των προεπιλεγμένων ρυθμίσεων του λογισμικού ArcGIS Geostatistical. Το δεύτερο μοντέλο (Krig) παραμετροποιήθηκε με βάση τα χαρακτηριστικά των δεδομένων όσον αφορά: στην τάση, στο μετασχηματισμό των δεδομένων, στην επιλογή κοντινής γειτνίασης (near neighbour), αλλά και στην επιλογή του βέλτιστου βαριογράμματος. Επομένως, το πρώτο μοντέλο μπορεί να χαρακτηριστεί ως μοντέλο ελαχιστοποίησης σφαλμάτων και το δεύτερο ως μοντέλο προσαρμοσμένο στα χαρακτηριστικά του συγκεκριμένου συνόλου δεδομένων.

2.3.1.4 Μοντέλο Ιδανικευμένης Πρόβλεψης

Κάθε τεχνική μοντελοποίησης εξηγεί ένα ποσοστό της διακύμανσης των τιμών των παρατηρούμενων δεδομένων από τη δική της οπτική. Τα προσθετικά μοντέλα παλινδρόμησης χρησιμοποιούν έναν εξομαλυντή, τα νευρωνικά δίκτυα κατανέμουν βάρη σε ενδιάμεσες τιμές του συνόλου εκπαίδευσης και

επανεκπαιδεύονται, ενώ η τεχνική Kriging εκτιμά ομοιότητες με βάση την απόσταση των δειγμάτων. Για την αξιοποίηση της διαφορετικής οπτικής κάθε προσέγγισης μοντελοποίησης και με σκοπό την αύξηση της εξήγησης της κατανομής των αρχικών δεδομένων από τα μοντέλα πραγματοποιήθηκε η ακόλουθη διαδικασία: Μέσω ενός νέου μοντέλου GAM συνδυάστηκαν οι προβλέψεις όλων των προσεγγίσεων. Το νέο GAM μοντέλο αναπτύχθηκε με εμπρόσθια επιλογή εξαρτημένων μεταβλητών και με βάση το AIC, όπως περιγράφεται παραπάνω (παράγραφος 2.3.1.1). Στόχος αυτής της διαδικασίας είναι η αξιοποίηση κατά το μέγιστο της ικανότητας πρόβλεψης κάθε μοντέλου για τη δημιουργία ενός νέου, που να εξηγεί μεγαλύτερο ποσοστό της κατανομής των δειγμάτων. Η παραπάνω διαδικασία επομένως μπορεί να ενταχθεί στις διαδικασίες εξαγωγής μέσης τιμής πολλών μοντέλων. Η διαφορετική προέλευση και δυναμική των προβλέψεων, που χρησιμοποιούνται εδώ ως εξαρτημένες μεταβλητές, δεν επιτρέπει την άμεση εξαγωγή μέσης τιμής. Για το λόγο αυτό αξιοποιούνται οι προσθετικές ικανότητες των GAM, ώστε να αποδοθεί η διαφορετική 'απεικόνιση' της κατανομής των τιμών υπό το πρίσμα διαφορετικών προσεγγίσεων. Το παραπάνω μοντέλο αναφέρεται ως AverMod.

2.3.2 Μοντέλα παρουσίας-απουσίας και μόνο-παρουσίας

Τα ΜΚΕ που χρησιμοποιήθηκαν για τα δεδομένα παρουσίας-απουσίας βρίσκονται στον Πίνακα 2.2. Η επιλογή των επεξηγηματικών μεταβλητών σε κάθε μέθοδο βασίστηκε στα εργαλεία διεκπεραίωσης της επιλογής μεταβλητών για τη βελτιστοποίηση του κάθε μοντέλου ή στην επιλογή μεταβλητών με σημαντική στατιστική και βιολογική συνεισφορά στα μοντέλα. Όσον αφορά στα μοντέλα, που αναπτύχθηκαν από την ίδια τεχνική, πραγματοποιήθηκε σύγκριση ως προς την ικανότητα πρόβλεψης τους και ως προς την ισορροπία μεταξύ της κατανομής που εξηγούν και της πολυπλοκότητας τους. Τελικά επελέγησαν τα μοντέλα με την καλύτερη ικανότητα πρόβλεψης, το χαμηλότερο σφάλμα πρόβλεψης και το χαμηλότερο δυνατό ποσό πολυπλοκότητας, ώστε σε κάθε μέθοδο να ελαχιστοποιείται το ενδεχόμενο λιγότερο πολύπλοκα μοντέλα να εμφανίζουν καλύτερη ικανότητα πρόβλεψης. Άλλωστε, αύξηση της πολυπλοκότητας πάνω από ένα όριο οδηγεί στην αύξηση του σφάλματος πρόβλεψης και "τιμωρείται" (penalized) από τις τεχνικές σύγκρισης. Για την εκτίμηση της ισορροπίας μεταξύ της εξηγούμενης κατανομής και της πολυπλοκότητας χρησιμοποιήθηκαν οι ακόλουθες τεχνικές: Κριτήριο πληροφορίας του Akaike (AIC) (Akaike, 1974), διασταυρούμενη επικύρωση, εκτίμηση σφαλμάτων, ανάλογα με τη μέθοδο μοντελοποίησης. Με αυτό τον τρόπο επιχειρείται η μείωση της πολυπλοκότητας στα μοντέλα. Στον Πίνακα 2.2 παρουσιάζονται, επίσης, η σχετική βιβλιογραφία και το λογισμικό που χρησιμοποιήθηκε για κάθε τεχνική μοντελοποίησης. Για όλα σχεδόν τα ΜΚΕ αναπτύχθηκε μεγάλος αριθμός μοντέλων, από όπου και επιλέχθηκε αυτό με την καλύτερη συνολική απόδοση, ώστε να συμμετέχει στη σύγκριση των ΜΚΕ.

Πίνακας 2.2 Τα ΜΚΕ που αναπτύχθηκαν, οι μεταβλητές τους και τα λογισμικά που χρησιμοποιήθηκαν.

Model	Explanatory Variables	Software	Reference
Generalized Additive Models, GAM	SST, SLA, DEP, DDND	R [18], library: mgcv	Wood, 2006; Hastie and Tibshirani, 1990
Generalized Mixed Models, GAMM	SST, SLA, DEP, DDND	R, library: mgcv, geoR, spatstat, spdep	Wood, 2006; Hastie and Tibshirani, 1990
Boosted Regression Trees, BRT	SST, CHL, PAR, SLA, DEP, SSTsl, DCoast, DEPsI, DDND	R, library: gbm	Leathwick <i>et al.</i> , 2006a
Multivariate Analysis and Regression Splines, MARS	SST, CHL, PAR, SLA, DEP, SSTsl, DCoast, DEPsI, DDND	R, library: mda	Leathwick <i>et al.</i> , 2005
Maximum Entropy, MAXENT	SST, CHL, PAR, SLA, DEP, SSTsl, DCoast, DEPsI, DDND	Maxent software for species habitat modeling	Phillips <i>et al.</i> , 2004
Support Vector Machines, SVM and SNM-Nu	SST, CHL, PAR, SLA, DEP, SSTsl, DCoast, DEPsI, DDND	openModeller Desktop	Cristianini and Shawe-Taylor, 2000
Genetic Algorithm for Rule-set Prediction, GARP	SST, CHL, PAR, SLA, DEP, SSTsl, DCoast, DEPsI, DDND	openModeller Desktop	Stockwell, 1999
Envelope Score, EnvScore	SST, CHL, PAR, SLA, DEP, SSTsl, DCoast, DEPsI, DDND	openModeller Desktop	Nix, 1986; Piñeiro <i>et al.</i> , 2007
Bioclim Envelope Model, BIOCLIM	SST, CHL, PAR, SLA, DEP, SSTsl, DCoast, DEPsI, DDND	openModeller Desktop	Nix, 1986
Environmental Distance, EnvDist and EnvDistChe	SST, CHL, PAR, SLA, DEP, SSTsl, DCoast, DEPsI, DDND	openModeller Desktop	Carpenter <i>et al.</i> , 1993
Associative Neural Networks, ASNN	SST, CHL, PAR, SLA, DEP, SSTsl, DCoast, DEPsI, DDND	Associative Neural Network software http://www.vcclab.org	Tetko, 2002a; Tetko, 2002b
Artificial Neural Network Ensemble, ANNE	SST, CHL, PAR, SLA, DEP, SSTsl, DCoast, DEPsI, DDND	Associative Neural Network software http://www.vcclab.org	Tetko, 2002a; Tetko, 2002b

Τα ΜΚΕ που χρησιμοποιούνται σε αυτή την ενότητα μπορούν να ομαδοποιηθούν σε κατηγορίες με βάση τη στατιστική τους προέλευση. Τα GAMs, τα GAMMs και τα Multivariate Analysis and Regression Splines (MARS) ανήκουν στην κατηγορία των μοντέλων παλινδρόμησης. Τα μοντέλα μέγιστης εντροπίας (Maximum Entropy models, MAXENT), τα ενισχυμένα δέντρα παλινδρόμησης (Boosted Regression Trees, BRT), τα ASNN, τα ANNE και οι μηχανές υποστήριξης διανυσμάτων (Support Vector Machines, SVM) ανήκουν στην κατηγορία των μηχανών εκμάθησης. Τέλος τα μοντέλα βιοκλιματικών φακέλων (Bioclim envelope model, BIOCLIM) και Envelope Score (EnvScore) είναι μέθοδοι τύπου φακέλου, που χρησιμοποιούν περιβαλλοντικά δεδομένα για τον καθορισμό βιοκλιματικών φακέλων. Τα μοντέλα περιβαλλοντικών αποστάσεων (Environmental Distance) είναι μοντέλα που βασίζονται στις θεωρητικές αποστάσεις ζευγών τιμών και κάνουν χρήση ενός γενικού αλγορίθμου, που βασίζεται σε περιβαλλοντικές μήτρες διαφορετικότητας. Τέλος, η μέθοδος Genetic Algorithm for Rule-set Prediction (GARP) χρησιμοποιεί ένα γενετικό αλγόριθμο για να δημιουργήσει μοντέλα οικολογικού θώκου για τα είδη.

Τα GAMs και τα GAMMs παρατίθενται παραπάνω (2.3.1.1) και αναπτύσσονται με ανάλογο τρόπο όπως και στη περίπτωση των ΜΚΕ αφθονίας. Η διαφορά στην εφαρμογή τους με τα δεδομένα παρουσίας/απουσίας είναι στο ότι πλέον χρησιμοποιείται διωνυμική κατανομή των σφαλμάτων με λογιστικό παράγοντα σύνδεσης.

Η μέθοδος Multivariate Adaptive Regression Splines (MARS) (Friedman, 1991; Leathwick *et al.*, 2005) είναι μια εναλλακτική μέθοδος παλινδρόμησης για την προσαρμογή μη γραμμικών σχέσεων, που όμως διαφέρει από τα GAMs επειδή χρησιμοποιεί τμηματική γραμμική προσαρμογή αντί για εξομαλυντή. Πιο συγκεκριμένα, στη MARS οι μη γραμμικές σχέσεις μεταξύ του είδους και των περιβαλλοντικών προγνωστικών περιγράφονται από μια σειρά γραμμικών τμημάτων με διαφορετική κλίση, καθένα από τα οποία προσαρμόζεται με τη χρήση μιας βασικής λειτουργίας. Η MARS είναι ικανή να μοντελοποιήσει περίπλοκες μη γραμμικές σχέσεις είδους και περιβάλλοντος, ενώ μία από τις βασικές της δυνατότητες είναι η μοντελοποίηση της σχέσης πολλών ειδών και περιβάλλοντος (Leathwick *et al.*, 2005). Η MARS είναι πολύ πιο γρήγορη από τα GAMs στην ανάπτυξη του μοντέλου, ενώ εύκολα μπορεί να χρησιμοποιηθεί με εφαρμογές GIS για τη δημιουργία χάρτη κατανομής ειδών. Σε αυτή τη μελέτη τα μοντέλα MARS αναπτύχθηκαν με τη χρήση της βιβλιοθήκης 'mda' (Hastie *et al.*, 1994) του R στατιστικού λογισμικού (R Development Core Team, 2005) και της 'MARS public function 3.1' των J. Leathwick και J. Elith (προσωπική επικοινωνία). Η επιλογή των επεξηγηματικών μεταβλητών έγινε με βάση τη συνεισφορά τους στο μοντέλο, που αντικατοπτρίζει και τη γενική απόδοση του μοντέλου. Αρκετά μοντέλα αναπτύχθηκαν και μεταξύ αυτών μοντέλα με διαφορετικές αλληλεπιδράσεις μεταξύ των επεξηγηματικών μεταβλητών. Η τελική επιλογή έγινε με τη χρήση του δείκτη Receiver Operating Characteristics και την Area Under Curve (ROC-AUC).

Η μέθοδος MAXENT υπολογίζει τη ζητούμενη πιθανότητα κατανομής του είδους μέσω της πιθανότητας κατανομής της μέγιστης εντροπίας (δηλαδή το πλέον απομακρυσμένο ή κοντινότερο στην ομοιομορφία), ως υποκείμενο σε ένα σύνολο περιορισμών, που αντιστοιχούν στην πληροφορία για την κατανομή του είδους (Phillips *et al.*, 2006). Η MAXENT είναι μια μηχανή εκμάθησης γενικής χρήσης με μια απλή και ακριβή μαθηματική φόρμουλα και με ειδικές λειτουργίες, που την κάνουν κατάλληλη για την ανάπτυξη μοντέλων κατανομής ειδών. Βασίζεται στην αρχή της μέγιστης εντροπίας, που διατυπώθηκε από τον Jaynes (1957). Για την ανάπτυξη των μοντέλων χρησιμοποιήθηκε η έκδοση 3.3.1 του λογισμικού Maximum Entropy Species Distribution Modelling. Και σε αυτή τη μέθοδο αναπτύχθηκαν πολλά μοντέλα, ενώ τελικά επιλέχθηκε εκείνο με το χαμηλότερο ROC-AUC. Το MAXENT είναι ένα φιλικό προς το χρήστη λογισμικό, που παρέχει εμπειριστικά αποτελέσματα με όλες τις σημαντικές πληροφορίες για κάθε μοντέλο που αναπτύσσεται. Επίσης τα αποτελέσματα μπορούν εύκολα να εισαχθούν σε περιβάλλον GIS για περαιτέρω ανάλυση, αν και το ίδιο λογισμικό έχει τη δυνατότητα να δημιουργεί χάρτες κατανομής ειδών.

Στα BRTs (Leathwick *et al.*, 2006a), κάθε μοντέλο αποτελείται από έναν αριθμό μοντέλων δέντρων ταξινόμησης και παλινδρόμησης (Classification And Regression Tree, CART). Ο αλγόριθμος ενίσχυσης (boosting algorithm) χρησιμοποιεί μια διαδικασία επανάληψης για την ανάπτυξη του τελικού μοντέλου με ένα εμπρόσθιο βηματικό τρόπο, προσθέτοντας διαδοχικά δέντρα και επαναβαθμονομώντας τα δεδομένα με σκοπό να τονίσει τις περιπτώσεις, που έχουν ελάχιστα προβλεφθεί από τα προηγούμενα δέντρα. Τα πλεονεκτήματα που παρέχουν τα BRTs είναι: η ικανότητά τους να διαχειρίζονται διάφορους τύπους επεξηγηματικών μεταβλητών ακόμα και με κενές εγγραφές, η ανοσία τους στις επιδράσεις ακραίων τιμών και μη σχετικών προγνωστικών και η ικανότητα τους να ενσωματώνουν τις αλληλεπιδράσεις μεταξύ των μεταβλητών (Friedman και Meulman, 2003). Τα BRTs αναπτύχθηκαν με τη χρήση της BRT functions έκδοση 2.8 που γράφτηκε από τους J. Leathwick και J. Elith (προσωπική επικοινωνία) για το στατιστικό πακέτο R (R Development Core Team, 2005) και η βιβλιοθήκη 'mda'. Και στην περίπτωση αυτή, η επιλογή του καλύτερου μοντέλου έγινε με άξονα το χαμηλότερο ROC-AUC.

Τα ASNN και ANNE περιγράφονται παραπάνω και χρησιμοποιήθηκαν στα δεδομένα παρουσίας-απουσίας με ανάλογο τρόπο, όπως και στην περίπτωση των δεδομένων αφθονίας. Εδώ η επιλογή του καλύτερου μοντέλου έγινε με βάση το μικρότερο RMSE και MAE (Tetko *et al.*, 2008).

Ο BIOCLIM ανήκει στις μεθόδους "φακέλου" και χρησιμοποιεί έναν αλγόριθμο βιοκλιματικού φακέλου (Nix, 1986; Busby, 1991). Οι βιοκλιματικοί φακέλοι βρίσκονται εννοιολογικά κοντά στο θεώρημα του

θώκου, καθώς προσπαθούν να οριοθετήσουν μια επιφάνεια (ή φάκελο) η οποία εμπεριέχει κατάλληλες συνθήκες μέσα στην υπερ-επιφάνεια του οικολογικού θώκου, όπως ορίζεται από τις περιβαλλοντικές μεταβλητές. Ο αλγόριθμος βρίσκει τις μέσες τιμές και τις τυπικές αποκλίσεις για κάθε περιβαλλοντική μεταβλητή (κανονικά κατανομημένη) που συνδέεται με την παρουσία του είδους στα δειγματοληπτικά δεδομένα. Πέρα από τον 'φάκελο' κάθε μεταβλητή φέρει ανώτερα και κατώτερα όρια που εξάγονται από τη μέγιστη και ελάχιστη τιμή, όπου υπάρχει παρουσία. Σε ένα μοντέλο κάθε σημείο μπορεί να ταξινομηθεί ως: κατάλληλο (όταν όλες οι περιβαλλοντικές μεταβλητές βρίσκονται μέσα στο εύρος του φακέλου), οριακό (όταν μια ή περισσότερες μεταβλητές είναι εκτός του φακέλου αλλά μέσα στα ανώτερα ή κατώτερα όρια) και ακατάλληλο (όταν μια ή περισσότερες μεταβλητές είναι εκτός των ανώτερων και κατώτερων ορίων). Τα κατηγοριοποιημένα αποτελέσματα του BIOCLIM χαρτογραφούνται με πιθανότητες 1, 0.5 και 0 αντίστοιχα. Για την ανάπτυξη του BIOCLIM χρησιμοποιήθηκε το λογισμικό OpenModeller (Muñoz *et al.*, 2009).

Η μέθοδος Envelope Score (EnvScore) είναι ανάλογη του BIOCLIM και χρησιμοποιεί ένα πιο απλό βιοκλιματικό αλγόριθμο. Για κάθε τιμή περιβαλλοντικής μεταβλητής ο αλγόριθμος υπολογίζει την ελάχιστη και μέγιστη τιμή από όλα τα σημεία παρουσίας. Η EnvScore εφαρμόστηκε με το λογισμικό OpenModeller (Muñoz *et al.*, 2009).

Η μέθοδος του κλιματικού χώρου (Climate Space) είναι ένας αλγόριθμος, που βασίζεται στην αρχή των κύριων συνιστωσών (principle components) και έχει αναπτυχθεί από τον Dr. Neil Caithness (<http://openModeller.sf.net>). Η επιλογή των στοιχείων στην εφαρμογή του αλγορίθμου γίνεται με βάση ένα όριο, όπου απορρίπτεται κάθε στοιχείο με ιδιοτιμή μικρότερη από n τυπικές αποκλίσεις ενός τυχαίου δείγματος (Muñoz *et al.*, 2009). Η αρχική εφαρμογή του Climate Space Model είχε δημιουργηθεί από ένα σύνολο λειτουργιών του λογισμικού Matlab.

Η μέθοδος των περιβαλλοντικών αποστάσεων (Environmental Distance, EnvDist) χρησιμοποιεί ένα γενικό αλγόριθμο ο οποίος υπολογίζει μήτρες περιβαλλοντικών ανομοιοτήτων. Όταν χρησιμοποιηθεί με το δείκτη του Gower (Gower και Legendre, 1986) και η μέγιστη απόσταση είναι ίση με 1, τότε ο αλγόριθμος παράγει τα ίδια αποτελέσματα με τον αλγόριθμο που είναι γνωστός ως DOMAIN (Carpenter *et al.*, 1993) και χρησιμοποιείται συχνά στη μοντελοποίηση κατανομής ειδών. Ο DOMAIN είναι ένας αλγόριθμος που βασίζεται στις αποστάσεις και υπολογίζει νέα σημεία στο χώρο ανάλογα με την περιβαλλοντική ομοιότητα τους με γνωστά σημεία παρουσίας του είδους. Επιπλέον του τελικού μοντέλου EnvDist, δημιουργήθηκε με το συγκεκριμένο εργαλείο και ένα παρεμφερές μοντέλο (EnvDistChe), που αναπτύχθηκε με την εφαρμογή της απόστασης Chebyshev αντί του δείκτη Gower. Η απόσταση Chebyshev είναι ένας δείκτης, που ορίζεται σε ένα διανυσματικό χώρο, όπου οι αποστάσεις μεταξύ δύο διανυσμάτων έχουν τις μέγιστες διαφορές κατά μήκος κάθε διεύθυνσης συντεταγμένων. Και τα δύο μοντέλα αναπτύχθηκαν με το λογισμικό OpenModeller (Muñoz *et al.*, 2009).

Οι μηχανές υποστήριξης διανυσμάτων (Support Vector Machines, SVMs) αποτελούνται από ένα σύνολο συσχετιζόμενων συνόλων επιβλεπόμενης εκμάθησης, που ανήκουν στην κατηγορία των γενικών γραμμικών ταξινομητών. Μπορούν να θεωρηθούν ως μια ειδική περίπτωση της κανονικοποίησης του Tychonoff (Tychonoff & Arsenin, 1977). Οι SVMs ελαχιστοποιούν το εμπειρικό ταξινομικό σφάλμα και παράλληλα αυξάνουν τα γεωμετρικά όρια. Τα μοντέλα που αναπτύσσονται με το συγκεκριμένο εργαλείο εξαρτώνται από ένα υποσύνολο των δεδομένων εκπαίδευσης καθώς η λειτουργία ανάπτυξης του μοντέλου δεν λαμβάνει υπόψη σημεία των δεδομένων εκπαίδευσης που βρίσκονται εκτός των ορίων (Vapnik, 1995; Schölkopf *et al.*, 2000). Οι SVMs δύνανται να αναπαραστήσουν μη γραμμικές επιδράσεις και αλληλεπιδράσεις μεταξύ των μεταβλητών προβάλλοντας τις ανεξάρτητες μεταβλητές σε ένα χώρο υψηλότερης διάστασης, όπου η πρόβλεψη επιτυγχάνεται με γραμμική λύση (Moguerza και Muñoz, 2006). Τελικά έγινε η επιλογή δύο SVM μοντέλων (SVM-Nu και SVM) λόγω της πολύ διαφορετικής τους απόδοσης στην ακρίβεια προσαρμογής

τους και την ικανότητα πρόβλεψης. Το SVM-Nu διαφέρει από το SVM στους βαθμούς της λειτουργίας Kernel. Και οι δύο προσεγγίσεις αναπτύχθηκαν με τη χρήση του λογισμικού OpenModeller.

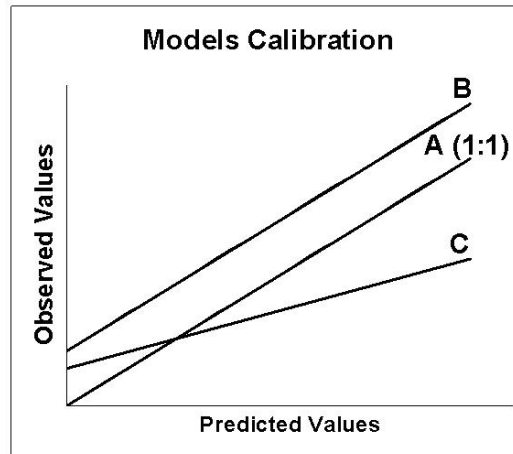
Η μέθοδος Genetic Algorithm for Rule-Set Prediction (GARP) χρησιμοποιεί μια ομάδα από κανόνες (π.χ. προσαρμογή της παλινδρόμησης και καθορισμό του εύρους των τιμών) που περιγράφουν καλύτερα την κατανομή των ειδών (Stockwell και Peters, 1999). Ο γενετικός αλγόριθμος, που χρησιμοποιείται στην GARP, στηρίζεται στη βασική θεωρία που αναπτύχθηκε από τον Holland (1975). Η GARP δημιουργεί μοντέλα οικολογικού θώκου για τα είδη, σε περιοχές, όπου μπορούν να διατηρηθούν πληθυσμοί. Ως δεδομένα εισόδου η GARP χρησιμοποιεί σύνολα σημείων/περιοχών, όπου είναι επιβεβαιωμένη η παρουσία των ειδών και ένα σύνολο στρωμάτων περιβαλλοντικών παραμέτρων, που δύνανται να μειώσουν τις ικανότητες του είδους για επιβίωση. Στην εργασία αυτή ο γενετικός αλγόριθμος εφαρμόστηκε με τη διαδικασία των καλύτερων υποομάδων χρησιμοποιώντας την νέα έκδοση της εφαρμογής του openModeller στο λογισμικό OpenModeller (Muñoz *et al.*, 2009).

2.4 Μέθοδοι επαλήθευσης των μοντέλων κατανομής ειδών

2.4.1 Δεδομένα αφθονίας

Η εφαρμογή ενός εύρους τεχνικών επαλήθευσης μπορεί να αναδείξει διαφορετικές πτυχές της ικανότητας πρόβλεψης των μοντέλων (Potts και Elith, 2006). Ο συντελεστής συσχέτισης του Pearson, r , χρησιμοποιήθηκε μεταξύ των παρατηρούμενων και προβλεπόμενων τιμών ως δείκτης ομοιότητας, με μια πιο ελαστική προσέγγιση, αφού η απόλυτη συσχέτιση και μόνο δεν σημαίνει απαραίτητα και τέλεια πρόβλεψη. Η περίπτωση της τέλει συσχέτισης δεν αποκλείει το ενδεχόμενο οι προβλέψεις να φέρουν σφάλματα με όμοια και σταθερή κατεύθυνση. Ο συντελεστής συσχέτισης βαθμίδων του Spearman, r_s , χρησιμοποιήθηκε ως δείκτης ομοιότητας των βαθμίδων μεταξύ των παρατηρούμενων και προβλεπόμενων τιμών. Η χρήση βαθμίδων ή κλάσεων στις τιμές έχει την έννοια ότι όταν οι αντίστοιχες βαθμίδες των προβλεπόμενων τιμών είναι σωστές τότε ο r_s θα είναι υψηλός (Potts και Elith, 2006). Ο συντελεστής συσχέτισης βαθμίδων του Kendall, τ , είναι αντίστοιχος του Spearman και χρησιμοποιείται ως μέτρο του βαθμού αντιστοιχίας των παρατηρούμενων και προβλεπόμενων τιμών.

Η βαθμονόμηση των μοντέλων πραγματοποιήθηκε όπως περιγράφεται από τους Potts και Elith (2006). Η διαδικασία είναι ανάλογη με αυτή που πρώτοι χρησιμοποίησαν οι Pearce και Ferrier (2000) για διωνυμικά, όμως, δεδομένα. Εφαρμόζεται μια απλή γραμμική παλινδρόμηση μεταξύ των παρατηρούμενων και προβλεπόμενων τιμών (π.χ. παρατηρούμενες = m (προβλεπόμενες) + b) η οποία παρέχει πληροφορίες για το σφάλμα και τη σταθερότητα των προβλέψεων. Σε αυτή την εξίσωση το b αποτελεί ένδειξη του μεγέθους του σφάλματος της πρόβλεψης, ενώ το m πληροφορεί για την έκταση των προβλεπόμενων τιμών συγκριτικά με τις παρατηρούμενες. Ένα απόλυτα βαθμονομημένο μοντέλο θα εφάπτεται της 1:1 γραμμής παλινδρόμησης. Μια παράλληλη γραμμή παλινδρόμησης ($m=1$, $b \neq 0$) στην 1:1 υποδεικνύει σταθερό σφάλμα. Όταν $b \neq 0$ και $m \neq 1$, τότε οι προβλεπόμενες τιμές εκτός από το σταθερό σφάλμα έχουν μικρότερο ή μεγαλύτερο εύρος από τις παρατηρούμενες τιμές (Εικ. 2.3).



Εικόνα 2.3 Παρατηρούμενες προς προβλεπόμενες τιμές. A: μοντέλο με τέλεια βαθμονόμηση (παρατηρούμενες = προβλεπόμενες), B: μοντέλο με σταθερό σφάλμα (δηλ. παρατηρούμενες = προβλεπόμενες + σταθερό σφάλμα), C: μοντέλο με ακανόνιστο σφάλμα (τροποποιημένο από Potts και Elith, 2006).

Επιπλέον των παραπάνω μεθόδων, υπολογίστηκαν κι άλλα περιγραφικά στατιστικά για τη σύγκριση των προβλεπόμενων με τις παρατηρούμενες τιμές. Στόχος της χρήσης αυτών των στατιστικών είναι η απόδοση μιας πιο ολοκληρωμένης εικόνας, όσον αφορά στην κατάταξη των μοντέλων ως προς την ακρίβεια πρόβλεψής τους. Στις προβλέψεις το μέσο σφάλμα (Mean Error, ME) θα πρέπει να είναι κοντά στο 0, όταν πρόκειται για ορθή πρόβλεψη, ενώ η μέση τετραγωνική ρίζα του τετραγώνου των κανονικοποιημένων σφαλμάτων (root-mean-square standardized prediction error, RMSqSt) θα πρέπει να είναι κοντά στο 1, όταν τα τυπικά σφάλματα είναι ακριβή. Η μέση τετραγωνική ρίζα του τετραγώνου των σφαλμάτων (root-mean-square prediction error, RMSE) θα πρέπει να είναι μικρή εφόσον οι προβλέψεις είναι κοντά στις παρατηρούμενες τιμές. Το τετράγωνο του μέσου σφάλματος (Mean Square Error, MSqE) όπως επίσης και το κανονικοποιημένο μέσο σφάλμα (Mean Standardized Error, MStE) θα πρέπει να είναι κοντά στο 0 για ακριβείς προβλέψεις. Το πρώτο (MSqE) γενικά αναφέρεται στην μεταβλητότητα των δεδομένων, ενώ το δεύτερο (MStE) είναι εκτιμητής της αβεβαιότητας του μοντέλου. Το μέσο απόλυτο σφάλμα (Mean Absolute Error, MAE) και το μέσο τυπικό σφάλμα (Average Standard Error, ASE) αποτελούν δείκτη του πόσο κοντά είναι οι προβλέψεις στις αναμενόμενες τιμές (όσο πιο μικρές τιμές τόσο μεγαλύτερη ακρίβεια). Όλα τα παραπάνω στατιστικά υπολογίστηκαν για τα σφάλματα των προβλέψεων των τελικά επιλεγμένων μοντέλων κάθε προσέγγισης. Επιπρόσθετα, υπολογίστηκαν περιγραφικές στατιστικές παράμετροι των παρατηρούμενων και προβλεπόμενων τιμών όπως η μέση τιμή (Ave) και η τυπική απόκλιση (StDev).

2.4.2 Δεδομένα παρουσίας-απουσίας

Για τα μοντέλα που χρησιμοποιούν δεδομένα παρουσίας-απουσίας έχουν αναπτυχθεί πολλές μέθοδοι επαλήθευσης, αξιολόγησης της ποιότητας της πρόβλεψης και σύγκρισης των παρατηρούμενων και προβλεπόμενων τιμών. Χαρακτηριστικές μέθοδοι είναι τα Kappa στατιστικά, ROC-AUC που έχουν ήδη αναφερθεί παραπάνω κατά την επιλογή του καλύτερου μοντέλου σε διάφορες τεχνικές, k-Fold διασταυρούμενη επικύρωση (cross validation), συνεκτικούς πίνακες (confusion matrices) και πίνακες κατάταξης (classification tables) οι οποίες περιγράφονται αναλυτικά στους Boyce *et al.* (2002).

Τα μοντέλα παρουσίας-απουσίας συνήθως αξιολογούνται με τον αριθμό των λάθος προβλέψεων. Υπάρχουν δύο τύποι σφαλμάτων πρόβλεψης: θετικό σφάλμα πρόβλεψης (false positive, FP) και αρνητικό σφάλμα πρόβλεψης (false negative, FN). Το πρώτο αναφέρεται σε πρόβλεψη απουσίας όταν στην πραγματικότητα έχουμε παρουσία και το δεύτερο σε πρόβλεψη παρουσίας όταν στην πραγματικότητα έχουμε απουσία. Η απόδοση των μοντέλων παρουσίας-απουσίας συνοψίζεται περιληπτικά με συνεκτικούς πίνακες (confusion matrix) (Πίνακας 2.2) οι οποίοι παρουσιάζουν ταυτόχρονα τα πρότυπα παρουσίας-απουσίας στις προβλεπόμενες και παρατηρούμενες τιμές. Οι Morrison *et al.* (1992) αναφέρουν τις FP ως σφάλμα τύπου I

και τις FN ως σφάλμα τύπου II. Το FP ή σφάλμα επιφόρτισης (commission error) οδηγεί σε υπερβολικά θετική πρόβλεψη (μεγάλο αριθμό προβλεπόμενων παρουσιών), ενώ το FN ή σφάλμα παράλειψης (omission error) οδηγεί σε μειωμένη θετική πρόβλεψη. Γενικά, το σφάλμα παράλειψης μπορεί να χαρακτηριστεί ως πραγματικό σφάλμα (hard error), ενώ το σφάλμα επιφόρτισης μπορεί και να μην είναι πραγματικό σφάλμα. Το σφάλμα επιφόρτισης αναφέρεται σε (α) ακατάλληλες περιοχές διαβίωσης του είδους (πραγματικό σφάλμα), (β) κατάλληλες περιοχές όπου όμως δεν έχει καταγραφεί το είδος λόγω της δειγματοληψίας (όμως μπορεί να υπάρχει παρουσία) ή (γ) σε κατάλληλες περιοχές, όπου ιστορικοί λόγοι (φράγματα, ικανότητα διασποράς του είδους) ή βιοτικοί λόγοι (διαειδικός ανταγωνισμός, θήρευση) εμπόδισαν το είδος να εποικίσει την περιοχή ή εξαφανίστηκε από εκεί. Ένα ακριβές μοντέλο παρουσίας-απουσίας θα πρέπει να χαρακτηρίζεται υποχρεωτικά από χαμηλό σφάλμα παράλειψης. Όσον αφορά όμως το σφάλμα επιφόρτισης, χαμηλές τιμές υποδηλώνουν ότι το μοντέλο υπερ-προσαρμόζεται στα δεδομένα εκπαίδευσης και υψηλές τιμές υποδηλώνουν ότι το μοντέλο οδηγεί σε υπερβολική θετική πρόβλεψη. Η ειδικότητα και η ευαισθησία αποτελούν ανάλογες και συμπληρωματικές έννοιες με τα σφάλματα επιφόρτισης και παράλειψης και αναφέρονται στις ορθά προβλεπόμενες παρουσίες και απουσίες αντί στα σφάλματα. Η ειδικότητα εκφράζει το ποσοστό των παρατηρούμενων απουσιών που έχουν προβλεφθεί σωστά και αναφέρεται στην ικανότητα του μοντέλου να προβλέπει την απουσία, με δεδομένο ότι το είδος πραγματικά απουσιάζει από μια περιοχή. Η ευαισθησία εκφράζει το ποσοστό των παρατηρούμενων παρουσιών που έχουν προβλεφθεί σωστά και αναφέρεται στην ικανότητα του μοντέλου να προβλέπει σωστά την παρουσία, με δεδομένο ότι το είδος πραγματικά εμφανίζεται στην περιοχή.

Πίνακας 2.2 Συνεκτικός πίνακας, που δίνει περιληπτικά τις σχέσεις των παρατηρούμενων και προβλεπόμενων τιμών παρουσίας-απουσίας.

Συνεκτικός Πίνακας	Προβλεπόμενη Παρουσία	Προβλεπόμενη απουσία
Πραγματική Παρουσία	Ορθή Θετική	Εσφαλμένη θετική (σφάλμα τύπου I)
Πραγματική Απουσία	Εσφαλμένη Αρνητική (σφάλμα τύπου II)	Ορθή Αρνητική

Όπως έχει ήδη αναφερθεί η σύγκριση μεταξύ των MKE πραγματοποιήθηκε χρησιμοποιώντας τον αποδοτικότερο "εκπρόσωπο" κάθε τεχνικής. Το ROC-AUC (Fielding και Bell, 1997) αποτελεί ένα από τα σημαντικότερα κριτήρια αξιολόγησης της απόδοσης των τεχνικών επειδή σε αντίθεση με άλλες μεθόδους επαλήθευσης μοντέλων (Kappa, συνεκτικοί πίνακες και πίνακες ταξινόμησης) αποφεύγει την υποκειμενικότητα της επιλογής τιμής κατωφλίου (threshold) (Lehmann *et al.*, 2002). Τα γραφήματα ROC και η επιφάνεια κάτω από την καμπύλη των Receiver Operating Characteristic μετρούν την ικανότητα του μοντέλου να διαχωρίζει τις περιοχές, όπου το είδος είναι παρόν, με τις περιοχές από όπου απουσιάζει και έχουν ευρέως χρησιμοποιηθεί στη βιβλιογραφία των MKE (Elith *et al.*, 2006). Οι τιμές των ROC-AUC κυμαίνονται από 0 έως 1, όπου το 1 δείχνει τέλεια διάκριση, το 0.5 δείχνει διάκριση των προβλέψεων κοντά στην τυχαιότητα και τιμές μικρότερες του 0.5 δείχνουν απόδοση της πρόβλεψης χειρότερη της τυχαιότητας (Boyce *et al.*, 2002; Elith *et al.*, 2006).

Η συσχέτιση (COR) μεταξύ των παρατηρούμενων τιμών παρουσίας-απουσίας (διχοτόμος μεταβλητή) και των προβλέψεων είναι γνωστή ως σημειακή διαύξουσα συσχέτιση και μπορεί να υπολογιστεί με τον συντελεστή συσχέτισης του Pearson (Zheng και Agresti, 2000). Είναι αντίστοιχη του ROC-AUC και φέρει επιπλέον πληροφορία: αντί να βασίζεται σε διαδικασία ταξινόμησης, λαμβάνει υπόψη τις διαφορές μεταξύ των παρατηρούμενων και προβλεπόμενων τιμών. Αυτή η πληροφορία παρέχει καλύτερη επισκόπηση στην κατανομή των τιμών των προβλέψεων πέρα της ικανότητας διάκρισης των μοντέλων (Murphy και Winkler, 1992).

Το στατιστικό Kappa (Cohen, 1960) συνοψίζει όλες τις πληροφορίες ενός συνεκτικού πίνακα. Μετρά το ποσοστό των ορθά ταξινομημένων μονάδων, αφού υπολογιστεί η πιθανότητα του τυχαίου γεγονότος. Το στατιστικό Kappa συνήθως χρησιμοποιείται σε οικολογικές μελέτες με δεδομένα παρουσίας-απουσίας (Boyce *et al.*, 2002). Απαιτεί όμως την εφαρμογή μιας τιμής ορίου (threshold) στις προβλέψεις με σκοπό να

τις μετατρέψει από πιθανότητες σε παρουσίες-απουσίες. Το Kappa αποτελεί ένα δείκτη που λαμβάνει υπόψη του τα σφάλματα επιφόρτισης και παράλειψης. Σε αυτή την εργασία χρησιμοποιείται το μέγιστο Kappa (maxKAPPA) ως δείκτης απόδοσης των ΜΚΕ, ώστε να αποφευχθεί η υποκειμενική επιλογή του ορίου (Liu *et al.*, 2005). Για τον υπολογισμό του Kappa, όπως και των περισσότερων δεικτών ποιότητας των μοντέλων παρουσίας-απουσίας χρησιμοποιήθηκε η βιβλιοθήκη 'PresenceAbsence' του στατιστικού πακέτου R (R Development Core Team, 2005).

Οι συνεκτικοί πίνακες όλων των ΜΚΕ σχηματίστηκαν για τις προβλέψεις τόσο στο εκπαιδευτικό σύνολο δεδομένων, όσο και στο σύνολο επαλήθευσης. Από αυτές εκτιμήθηκαν τα σφάλματα επιφόρτισης και παράλειψης, οι δείκτες ιδιαιτερότητας και ευαισθησίας και το maxKAPPA και για τα δύο σύνολα δεδομένων. Επίσης, και στα δύο σετ υπολογίστηκαν οι συντελεστές συσχέτισης μεταξύ των προβλεπόμενων και των παρατηρούμενων τιμών, ενώ το ROC-AUC χρησιμοποιήθηκε για την κατάταξη της ακρίβειας πρόβλεψης των ΜΚΕ. Τέλος οι χάρτες πιθανότητας παρουσίας, που δημιουργήθηκαν από κάθε ΜΚΕ, συγκρίθηκαν ως προς τις χωρικές ομοιότητες τους με την εφαρμογή συσχέτισης καμβάδων του λογισμικού ArcInfo της ESRI.

3. Αποτελέσματα

3.1 Μοντέλα Αφθονίας

3.1.1 Προβλέψεις κατανομής ειδών

3.1.1.1 Μοντέλα παλινδρόμησης

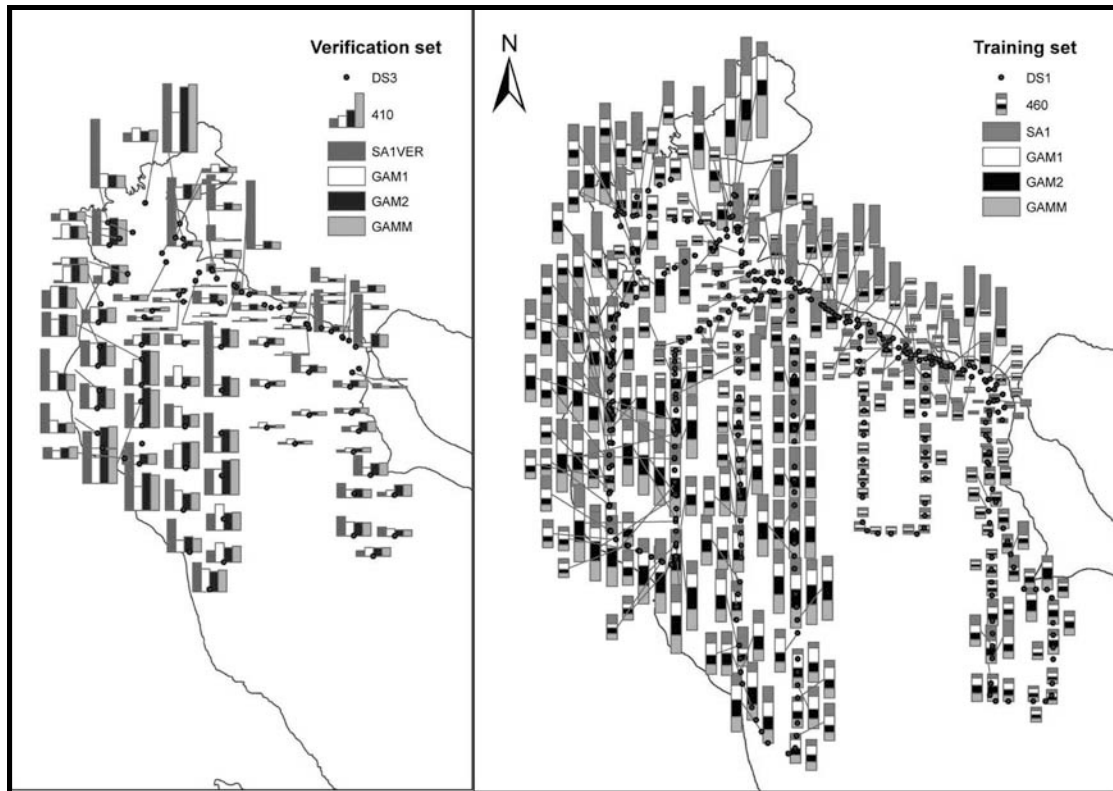
Το GAM1 επιλέχθηκε τελικά ως το καλύτερο μοντέλο στην κατηγορία των GAMs μετά τη διαδικασία της διερεύνησης των δεδομένων, την επιλογή των μεταβλητών και τη βελτιστοποίηση του μοντέλου. Ακόμη ένα μοντέλο (GAM2) της ίδιας κατηγορίας επιλέχθηκε για τη διαδικασία της σύγκρισης των ΜΚΕ. Το GAM1 είναι ένθετο (nested) του GAM2 το οποίο περιλαμβάνει επιπλέον των επεξηγηματικών μεταβλητών του GAM1 ακόμη μια χρονική μεταβλητή. Τα ποιοτικά χαρακτηριστικά του GAM2 είναι βελτιωμένα σε σχέση με το GAM1 (AIC και συντελεστής απόκλισης, Πίνακας 3.1), ωστόσο το GAM2 χρησιμοποιεί περισσότερους βαθμούς ελευθερίας από το GAM1, που δείχνει ότι πρόκειται για πιο σύνθετο μοντέλο. Και τα δύο μοντέλα χρησιμοποιούνται στη σύγκριση των ΜΚΕ με σκοπό να εξεταστεί αν η αυξημένη πολυπλοκότητα του GAM2 βελτιώνει την ακρίβεια πρόβλεψης του μοντέλου, χωρίς να θυσιάζει τη γενικότητά του. Από την ομάδα των GAMMs επιλέχθηκε το GAMM1 το οποίο μαζί με τα δύο προαναφερθέντα GAMs χρησιμοποιήθηκαν στην πρόβλεψη της κατανομής των ειδών και την ποιοτική αξιολόγηση τους ως προς την ακρίβεια, τη γενικότητα και τη βιολογική ερμηνεία των μεταβλητών των μοντέλων.

Πίνακας 3.1. Τα τελικά επιλεγμένα προσθετικά μοντέλα και τα χαρακτηριστικά τους. Διάστημα εμπιστοσύνης: 0.05, το σύμβολο “:” υποδηλώνει αλληλεπίδραση, Dev. Exp. = επεξηγούμενη απόκλιση, Res. d.f = βαθμοί ελευθερίας των υπολοίπων, R_a^2 = διορθωμένο R^2 , AIC = Akaike Information Criterion, p-value (χ^2) = στατιστικά σημαντικές τιμές, s = εξομαλυντής των επεξηγηματικών μεταβλητών.

Μοντέλο	Επεξηγηματικές μεταβλητές	Dev. Exp.	Res. d.f.	R_a^2	AIC	p-value
GAM1	s(SLA) + s(DEP:PAR)	38.9%	23.687	0.345	570.35	<<0.05
	s(SLA)		5.105			<<0.05
	s(DEP:PAR)		18.682			<<0.05
GAM2	s(SLA) + s(DEP:PAR) + as.factor(DDND)	47.9%	29.678	0.458	510.76	<<0.05
	s(SLA)		7.051			<<0.05
	s(DEP:PAR)		22.627			<<0.05
	as.factor(DDND)					<<0.05
GAMM1	s(SLA) + s(DEP:PAR) + Παράγοντας συσχέτισης	NA	24.597	0.444	NA	<<0.05
	s(SLA)		5.561			<<0.05
	s(DEP:PAR)		19.036			<<0.05
	Παράγοντας συσχέτισης					<<0.05

Όπως φαίνεται και στον Πίνακα 3.1 όλα τα μοντέλα είναι στατιστικά σημαντικά. Σύμφωνα με το κριτήριο AIC (όσο μικρότερο τόσο καλύτερο) το GAM2 υπερτερεί του GAM1, όσον αφορά στην προσαρμογή του στα δεδομένα. Επιπλέον, ο συντελεστής απόκλισης (δεν είναι διαθέσιμος για το GAMM1) και το διορθωμένο R^2 υποδηλώνουν ότι το GAM2 εξηγεί μεγαλύτερο ποσοστό της διακύμανσης της ανεξάρτητης μεταβλητής από τα άλλα. Εκτός των άλλων, το GAM2 περιγράφει καλύτερα τις σχέσεις μεταξύ των ακουστικών δεδομένων και των περιβαλλοντικών παραμέτρων. Οι προβλέψεις στο DS1 (Εικ. 3.1) δείχνουν ότι σε συγκεκριμένες περιοχές τα μοντέλα αποτυγχάνουν να αποδώσουν επακριβώς τις μετρούμενες τιμές

(κυρίως βόρεια και βορειοδυτικά). Το ίδιο διαφαίνεται και στις προβλέψεις στο DS3. Οι άλλες περιοχές προβλέπονται ικανοποιητικά και στα δύο σύνολα τιμών. Οι μπάρες στην Εικ. 3.1 δείχνουν ότι το GAM2 και GAMM1 αποδίδουν ανάλογα ως προς τις προβλέψεις τους, ενώ οπτικά διαφέρουν με το GAM1.



Εικόνα 3.1 Συγκριτική απεικόνιση των παρατηρούμενων (SA1VER για το σύνολο επαλήθευσης & SA1 για το σύνολο εκπαίδευσης) και προβλεπόμενων τιμών (GAM1, GAM2, GAMM) που παρήχθησαν από τα GAMs και τα GAMMs. Το ύψος κάθε μπάρας αντιστοιχεί στην ακουστική πυκνότητα του σημείου δειγματοληψίας και στην προβλεπόμενη ακουστική πυκνότητα. Ο χάρτης αριστερά αντιστοιχεί στο σύνολο επαλήθευσης (DS3-μαύρα σημεία) και ο χάρτης δεξιά αντιστοιχεί στο σύνολο εκπαίδευσης (DS1-μαύρα σημεία).

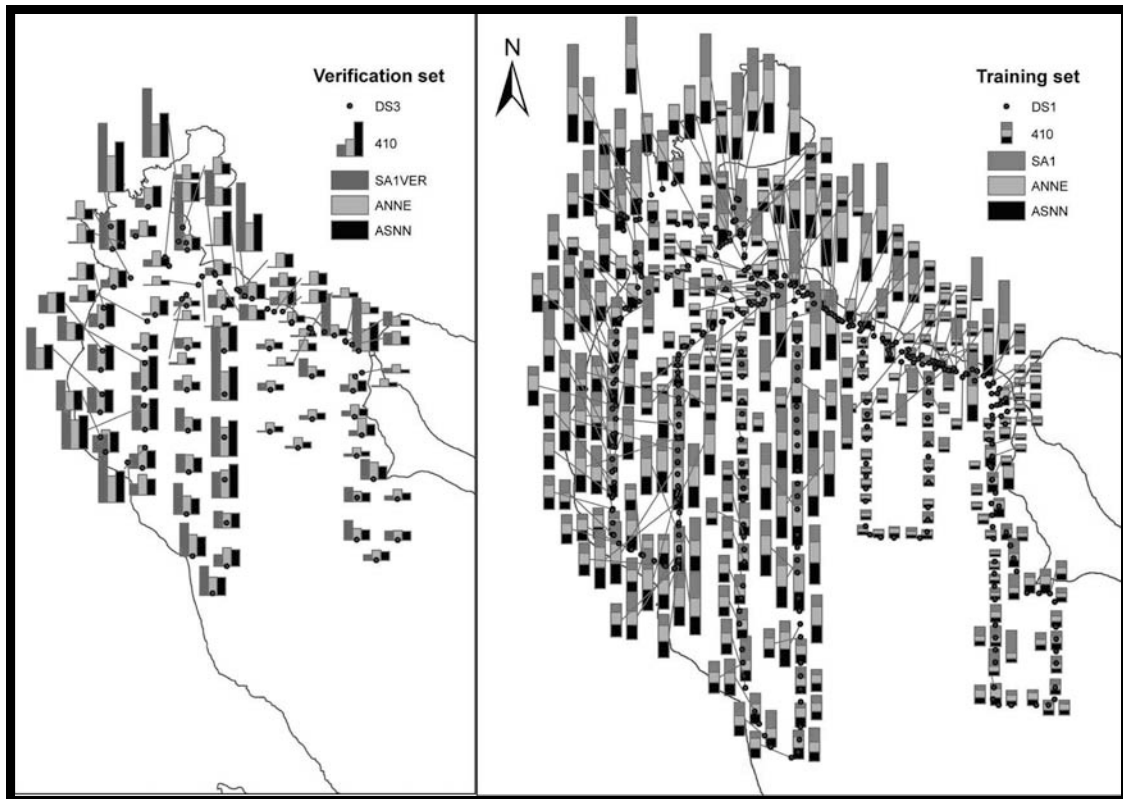
3.1.1.2 Συνδυασμένα νευρωνικά δίκτυα

Τόσο το ANNE όσο και το ASNN αναπτύχθηκαν χρησιμοποιώντας ως επεξηγηματικές μεταβλητές τις: DDND, SLA, CHL, PAR, SST, DEP, DCoast, DEPsI, και SSTsI. Η απόδοση των παραπάνω μοντέλων στις προβλέψεις τους στο σύνολο εκπαίδευσης παρουσιάζονται στον Πίνακα 3.2.

Πίνακας 3.2 Τα ποιοτικά χαρακτηριστικά των ANNE και ASNN. RMSE-τετραγωνική ρίζα του μέσου τετραγώνου του σφάλματος, MAE-μέσο απόλυτο σφάλμα, r^2 - τετράγωνο του συντελεστή συσχέτισης, LOO-Leave One Out, k - k near neighbour, σ -sigma.

Network, k, (σ)	Αποτελέσματα LOO		
	RSME	MAE	r^2
ANNE	191.10	136.64	0.219
ASNN, 54, (0.01)	176.77	116.25	0.325

Στο DS1 (Εικ. 3.2 δεξιά) και τα δύο μοντέλα αποδίδουν σχεδόν το ίδιο καλά. Οπτικά φαίνεται να παρέχουν πολύ ακριβείς προβλέψεις σε όλη την περιοχή, ενώ αποδίδεται ικανοποιητικά και η τοπική διακύμανση των μετρούμενων τιμών. Οι προβλέψεις στο DS3 (Εικ. 3.2, αριστερά) δείχνουν ότι και τα δύο μοντέλα περιγράφουν ικανοποιητικά τη γενική τάση των μετρούμενων τιμών, όμως υποεκτιμούν ή υπερεκτιμούν τις υψηλές και χαμηλές τιμές του DS3 αντίστοιχα.



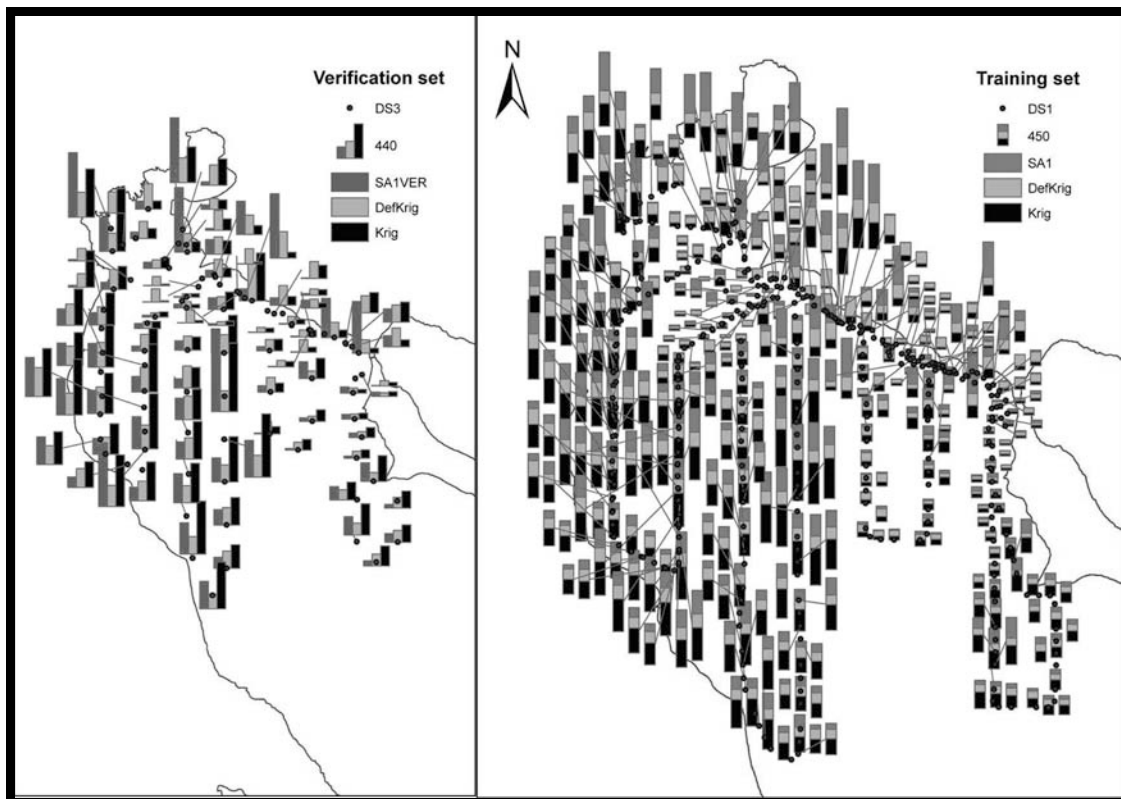
Εικόνα 3.2. Συγκριτική απεικόνιση των παρατηρούμενων (SA1VER για το σύνολο επαλήθευσης & s_{A1} για το σύνολο εκπαίδευσης) και προβλεπόμενων τιμών (ASNN & ANNE) που παρήχθησαν από τα ASNN και τα ANNE. Το ύψος κάθε μπάρας αντιστοιχεί στην ακουστική πυκνότητα του σημείου δειγματοληψίας και στην προβλεπόμενη ακουστική πυκνότητα. Ο χάρτης αριστερά αντιστοιχεί στο σύνολο επαλήθευσης (DS3-μαύρα σημεία) και ο χάρτης δεξιά αντιστοιχεί στο σύνολο εκπαίδευσης (DS1-μαύρα σημεία).

3.1.1.3 Kriging

Τα χαρακτηριστικά των μοντέλων Kriging παρουσιάζονται στον Πίνακα 3.3. Τα δύο τελικά επιλεγμένα μοντέλα διαφέρουν σημαντικά. Το μοντέλο Krig μπορεί να χαρακτηριστεί ως τοπικά εστιασμένο μοντέλο σε σύγκριση με το DefKrig, που είναι ευρύτερα εστιασμένο ως προς τις προβλέψεις. Αυτή η παρατήρηση απεικονίζεται στην Εικ. 3.3, όπου επίσης είναι φανερό οπτικά ότι οι προβλέψεις των δύο μοντέλων διαφέρουν αρκετά. Ο χάρτης δεξιά (Εικ. 3.3) επιβεβαιώνει τις ακριβείς προβλέψεις του Krig στις υψηλές και χαμηλές τιμές των παρατηρούμενων τιμών σε τοπικό επίπεδο, τις οποίες το DefKrig υποεκτιμά. Αντίστοιχη είναι και εικόνα στις προβλέψεις στο DS3 (Εικ. 3.3, αριστερά).

Πίνακας 3.3. Τα χαρακτηριστικά των Ordinary Kriging μοντέλων.

Krig	DefKrig
Αφαίρεση τάσης 1 ^{ου} βαθμού	Καμία αφαίρεση τάσης
<u>Βαριόγραμμα:</u> 15 διαστήματα Μέγεθος διαστήματος 1852 m Nugget 1.184	<u>Βαριόγραμμα:</u> 12 διαστήματα Μέγεθος διαστήματος 6965.1 Nugget 46257.18
Σφαιρικό μοντέλο Εύρος 26571.13 m Περιλαμβάνει ανισοτροπία Ελάχιστο εύρος 5100.6 Κατεύθυνση 3.35°	Σφαιρικό μοντέλο Εύρος 79065.67 Δεν περιλαμβάνει ανισοτροπία



Εικόνα 3.3 Συγκριτική απεικόνιση των παρατηρούμενων (SA1VER για το σύνολο επαλήθευσης & SA1 για το σύνολο εκπαίδευσης) και προβλεπόμενων τιμών (DefKrig & Krig) που παρήχθησαν από τις Ordinary Kriging τεχνικές. Το ύψος κάθε μπάρας αντιστοιχεί στην ακουστική πυκνότητα του σημείου δειγματοληψίας και στην προβλεπόμενη ακουστική πυκνότητα. Ο χάρτης αριστερά αντιστοιχεί στο σύνολο επαλήθευσης (DS3-μαύρα σημεία) και ο χάρτης δεξιά αντιστοιχεί στο σύνολο εκπαίδευσης (DS1-μαύρα σημεία).

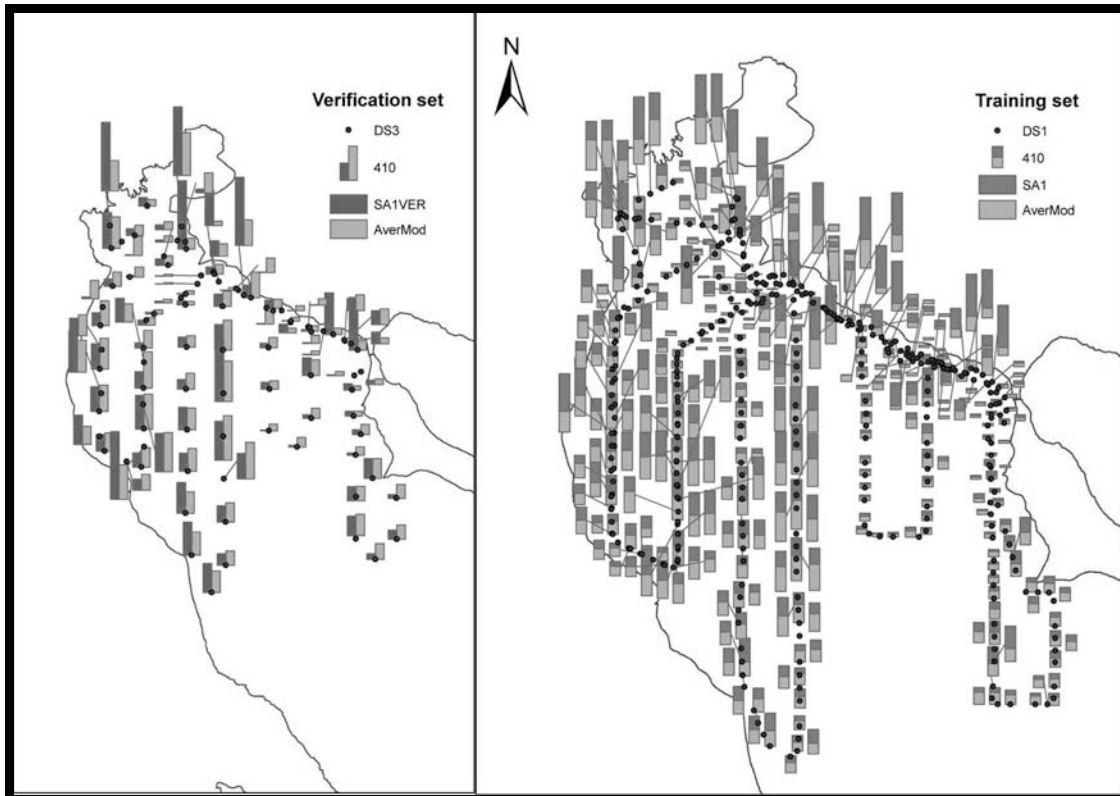
3.1.1.4 Μοντέλο Ιδανικευμένης Πρόβλεψης

Το τελικά επιλεγμένο μοντέλο και τα χαρακτηριστικά του παρουσιάζονται στον Πίνακα 3.4. Το AverMod είναι ένα αναλογικά απλό GAM μοντέλο με δύο ανεξάρτητες μεταβλητές και σχετικά λίγους βαθμούς ελευθερίας. Παρόλα αυτά, η εξηγούμενη από το μοντέλο μεταβλητότητα είναι αναλογικά μεγάλη (45.2%). Οι προβλέψεις του GAM2 δοκιμάστηκαν, επίσης, ως ανεξάρτητη μεταβλητή στη διαδικασία αυτή, όμως δεν ήταν στατιστικά σημαντική η παρουσία τους στο μοντέλο ($p > 0.05$) και απορρίφτηκε στη φάση επιλογής των παραμέτρων. Η Εικ. 3.4 δείχνει ότι οι προβλέψεις του AverMod για το DS1 είναι πολύ ακριβείς για όλη

την έκταση της περιοχής μελέτης (Εικ. 3.4, αριστερά). Οι προβλέψεις στο DS3 έχουν μια τάση να υποεκτιμούν τις υψηλές τιμές και να υπερεκτιμούν τις χαμηλές (Εικ. 3.4, δεξιά).

Πίνακας 3.4. Τα χαρακτηριστικά του μοντέλου GAM που χρησιμοποιήθηκε για τη δημιουργία του μοντέλου ιδανικευμένης πρόβλεψης. Διάστημα εμπιστοσύνης: 0.05, Dev. Exp. = επεξηγούμενη απόκλιση, Res. d.f = βαθμοί ελευθερίας των υπολοίπων, R_a^2 = διορθωμένο R^2 , p-value (χ^2) = στατιστικά σημαντικές τιμές, s = εξομαλυντής των επεξηγηματικών μεταβλητών.

Όνομα μοντέλου	Ανεξάρτητες μεταβλητές	Dev. Exp.	Res. d.f.	R_a^2	p-value
AverMod	s(προβλέψεις Krig) + s(προβλέψεις ASNN)	45.2%	11.093	0.436	<<0.05



Εικόνα 3.4 Συγκριτική απεικόνιση των παρατηρούμενων (SA1VER για το σύνολο επαλήθευσης & SA1 για το σύνολο εκπαίδευσης) και προβλεπόμενων τιμών (AverMod) που παρήχθησαν από το μοντέλο ιδανικευμένης πρόβλεψης. Το ύψος κάθε μπάρας αντιστοιχεί στην ακουστική πυκνότητα του σημείου δειγματοληψίας και στην προβλεπόμενη ακουστική πυκνότητα. Ο χάρτης αριστερά αντιστοιχεί στο σύνολο επαλήθευσης (DS3-μαύρα σημεία) και ο χάρτης δεξιά αντιστοιχεί στο σύνολο εκπαίδευσης (DS1-μαύρα σημεία).

3.1.2. Σύγκριση των μοντέλων αφθονίας

3.1.2.1 Προσαρμογή των μοντέλων στα δεδομένα εκπαίδευσης

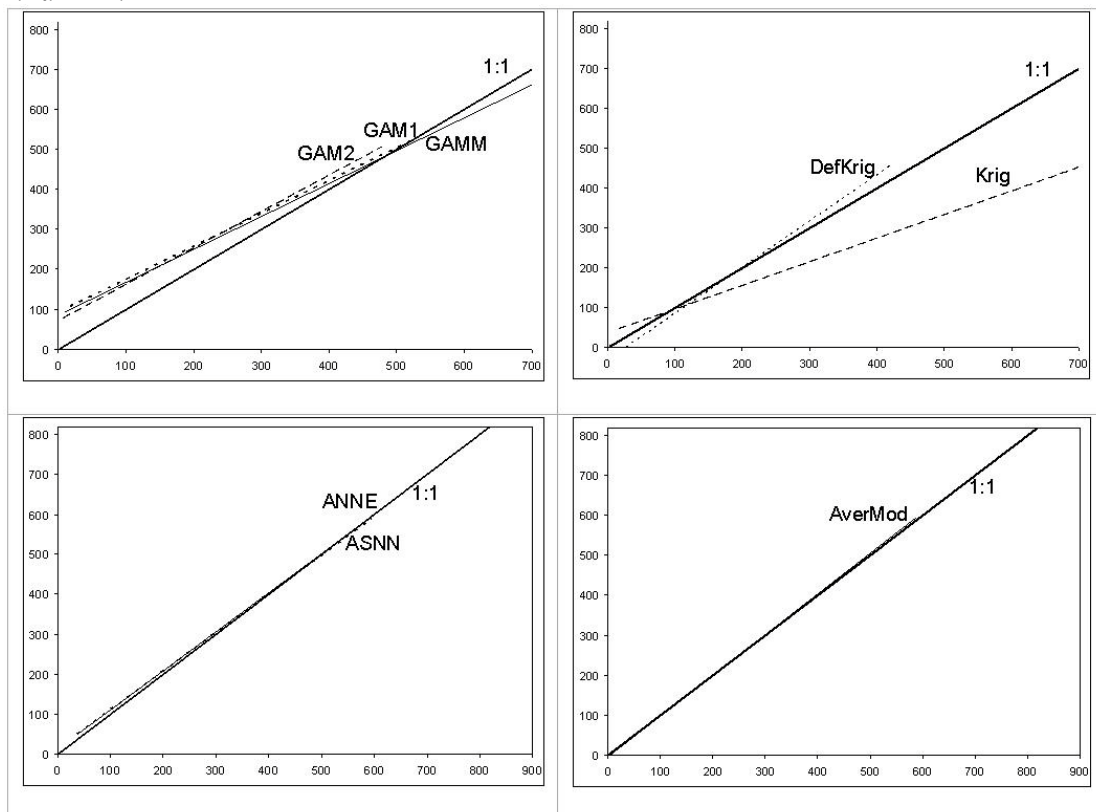
Η σύγκριση της προσαρμογής των μοντέλων στα δεδομένα εκπαίδευσης έγινε με την εφαρμογή των προαναφερθέντων δεικτών στις προβλεπόμενες τιμές και τις τιμές εκπαίδευσης του DS1. Για τον προσδιορισμό του καλύτερα εφαρμοσμένου μοντέλου χρησιμοποιήθηκαν οι δείκτες που παρουσιάζονται στον Πίνακα 3.5.

Πίνακας 3.5. Δείκτες σύγκρισης των μοντέλων για τον προσδιορισμό της ακρίβειας προσαρμογής τους στα δεδομένα εκπαίδευσης

Συσχετίσεις	GAM1	GAM2	GAMM1	Krig	DefKrig	ANNE	ASNN	AverMod
Pearson's r	0.38	0.49	0.45	0.50	0.59	0.57	0.56	0.67
Spearman's r_s	0.62	0.71	0.69	0.54	0.79	0.55	0.56	0.81
Kendall's tau	0.45	0.53	0.51	0.39	0.62	0.39	0.40	0.64
Βαθμονόμηση								
Intercept, m	0.821	0.819	0.823	1.160	0.592	0.980	0.967	1.019
Slope, b	92.89	82.65	85.51	-30.22	37.07	12.31	15.42	-3.67
R ²	0.145	0.224	0.206	0.250	0.351	0.327	0.316	0.453

Οι ομοιότητες μεταξύ των παρατηρούμενων και προβλεπόμενων τιμών, σύμφωνα με το συντελεστή συσχέτισης του Pearson, περιγράφονται καλύτερα από τα μοντέλα AverMod, Krig, ASNN και ANNE. Το GAM2 αποδίδει καλύτερα από τα υπόλοιπα μοντέλα παλινδρόμησης. Από την άλλη οι συντελεστές συσχέτισης των Spearman και Kendall δείχνουν ότι τις μεγαλύτερες συσχετίσεις μεταξύ των τάξεων των παρατηρούμενων και προβλεπόμενων τιμών τις φέρουν τα μοντέλα Krig και GAM2.

Η διαδικασία της βαθμονόμησης των μοντέλων (Εικ. 3.5) έδειξε ότι τα μοντέλα που ανήκουν στα νευρωνικά δίκτυα είναι σχεδόν τέλεια βαθμονομημένα. Ελαχιστοποιούν τα σφάλματα και προβλέπουν τιμές σε ανάλογο εύρος με τις παρατηρούμενες, όπως φανερώνει η επικάλυψη με την 1:1 γραμμή παλινδρόμησης στην Εικ. 3.5. Υπάρχουν σημαντικές ομοιότητες ως προς την απόδοση των προσθετικών μοντέλων, αφού όλα εμφανίζουν το ίδιο σταθερό σφάλμα και των εύρος των προβλεπόμενων τιμών τους είναι ανάλογο. Η γραμμή παλινδρόμησης του μοντέλου Krig διαφέρει πολύ από το DefKrig υποδηλώνοντας ένα υποβαθμονομημένο μοντέλο.



Εικόνα 3.5. Βαθμονόμηση των μοντέλων, όπως προκύπτει από τις γραμμικές παλινδρομήσεις μεταξύ των παρατηρούμενων (άξονας y) και προβλεπόμενων (άξονας x) τιμών του συνόλου δεδομένων εκπαίδευσης (DS1). Η 1:1 γραμμή αντιστοιχεί σε ένα άριστα βαθμονομημένο μοντέλο.

Ανάλυση σφαλμάτων των προβλέψεων στο DS1

Τα περιγραφικά στατιστικά (Ave, StDev) των παρατηρούμενων (s_{A1-DS1}) και προβλεπόμενων τιμών και οι δείκτες σφαλμάτων παρουσιάζονται στον Πίνακα 3.6. Οι μέσες τιμές (Ave) των προβλέψεων των ASNN και DefKrig είναι πιο κοντά στη μέση τιμή των δειγματοληπτικών τιμών. Το ίδιο ισχύει και για τα ASNN και ANNE. Οι μέσες τιμές πρόβλεψης των προσθετικών μοντέλων είναι χαμηλότερα από τις s_{A1-DS1} ενώ οι προβλέψεις του Krig είναι υψηλότερα. Η τυπική απόκλιση των δειγματοληπτικών δεδομένων είναι σχετικά υψηλή γεγονός που γενικά χαρακτηρίζει τη φύση των ακουστικών δεδομένων. Μόνο το Krig προβλέπει τιμές με ανάλογη τυπική απόκλιση με αυτή του s_{A1-DS1} , ενώ τα υπόλοιπα μοντέλα προβλέπουν τιμές με χαμηλότερη τυπική απόκλιση.

Πίνακας 3.6. Δείκτες σφαλμάτων των μοντέλων για τη σύγκριση της ακρίβειας προσαρμογής τους στα δεδομένα εκπαίδευσης

DS1	s_{A1-DS1}	GAM1	GAM2	GAMM1	DefKrig	Krig	ANNE	ASNN	AverMod
Ave	193.414	122.348	133.445	131.089	192.869	263.935	184.856	184.104	193.414
StDev	215.192	99.740	117.158	118.775	92.748	215.353	125.587	125.078	142.058
RMSE		1335.194	1126.701	1170.972	10.232	1324.978	160.789	174.904	~0.000
MSqE		1782742.121	1269454.155	1371174.907	104.696	1755565.923	25853.085	30591.264	~0.000
MAE		117.781	107.436	110.484	130.249	152.782	116.248	116.187	97.469
ASE		0.752	0.728	0.739	0.728	0.742	0.707	0.710	0.672
MStE		-0.356	-0.320	-0.323	-0.003	0.363	-0.048	-0.052	~0.000
RMSqSt		6.684	6.016	6.072	0.055	6.820	0.911	0.982	~0.000
ME		-71.065	-59.968	-62.325	-0.545	70.521	-8.558	-9.309	~0.000

Μια γενική επισκόπηση των δεικτών σφαλμάτων δείχνει ότι τα μοντέλα AverMod, DefKrig, ASNN και ANNE (με φθίνουσα σειρά) παρέχουν πιο ακριβείς και με μικρότερα σφάλματα προβλέψεις, ενώ εξηγούν καλύτερα τη μεταβλητότητα των παρατηρούμενων δεδομένων, σύμφωνα με τα ME, MSqE, MStE, RMSqSt και RMSE. Τα MAE και ASE δείχνουν ότι και το GAM2 προβλέπει τιμές κοντά στις παρατηρούμενες. Σύμφωνα με το ME και το άθροισμα των σφαλμάτων τα AverMod, ASNN, ANNE και DefKrig προσαρμόζονται καλύτερα στα δεδομένα με τα οποία έχουν εκπαιδευτεί σε σχέση με τα υπόλοιπα μοντέλα.

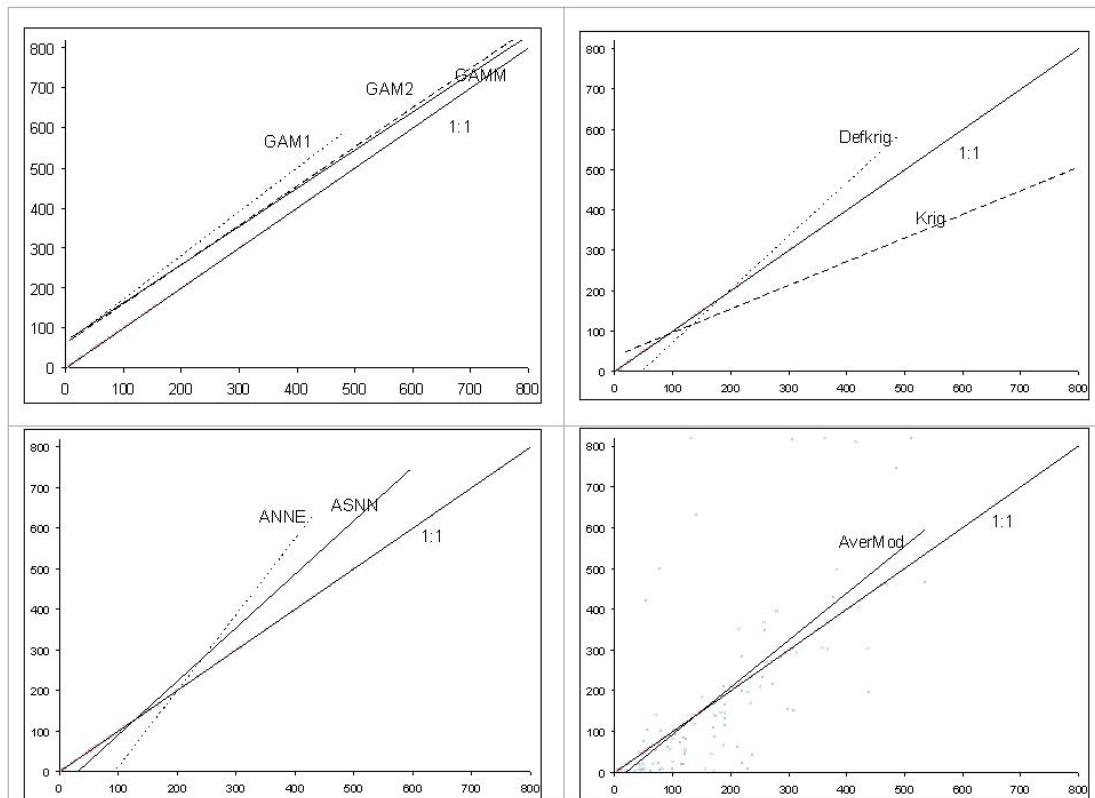
3.1.2.2 Σύγκριση των προβλέψεων του συνόλου επαλήθευσης

Η αξιολόγηση των προβλέψεων των μοντέλων στο σετ επαλήθευσης πραγματοποιείται με την εφαρμογή των παραπάνω εκτιμητών στο DS3. Το μοντέλο με την καλύτερη απόδοση στην πρόβλεψη του συνόλου επαλήθευσης μπορεί να χαρακτηριστεί από γενικότητα και να προβλέπει με μεγαλύτερη ακρίβεια την κατανομή των ειδών σε εκτεταμένη χωρο-χρονική κλίμακα. Στον Πίνακα 3.7 παρουσιάζονται οι διάφοροι δείκτες βάσει των οποίων θα προσδιοριστεί το μοντέλο με την καλύτερη προσαρμογή στο σύνολο επαλήθευσης.

Πίνακας 3.7 Δείκτες σύγκρισης των μοντέλων για τον προσδιορισμό της ακρίβειας προσαρμογής τους στα δεδομένα επαλήθευσης

Συσχετίσεις	GAM1	GAM2	GAMM1	Krig	DefKrig	ANNE	ASNN	AverMod
Pearson's r	0.48	0.58	0.56	0.53	0.58	0.67	0.73	0.67
Spearman's r_s	0.62	0.68	0.67	0.62	0.71	0.62	0.65	0.73
Kendall's tau	0.44	0.50	0.48	0.43	0.56	0.45	0.50	0.57
Βαθμολόγηση								
Intercept, m	1.090	0.981	0.954	1.317	0.586	1.871	1.317	1.154
Slope, b	62.92	61.83	67.97	60.09	37.33	175.28	40.79	22.14
R^2	0.233	0.338	0.315	0.281	0.332	0.454	0.536	0.445

Σύμφωνα με το συντελεστή συσχέτισης του Pearson (Πίνακας 3.7) οι ομοιότητες μεταξύ των προβλεπόμενων και παρατηρούμενων τιμών περιγράφονται καλύτερα από τα μοντέλα που ανήκουν στα νευρωνικά δίκτυα και κυρίως από το ASNN. Το GAM2 αποδίδει καλύτερα από τα άλλα προσθετικά μοντέλα και το Krig καλύτερα από το DefKrig. Από την άλλη, οι συντελεστές συσχέτισης των Spearman και Kendall φανερώνουν ότι η συσχέτιση μεταξύ των βαθμίδων των παρατηρούμενων και προβλεπόμενων τιμών είναι υψηλότερη για το Krig, ενώ είναι πολύ μικρές οι διαφορές μεταξύ των άλλων μοντέλων. Η διαδικασία βαθμονόμησης των μοντέλων (Εικ. 3.6 και Πιν. 3.7) δείχνει ότι τα προσθετικά μοντέλα χαρακτηρίζονται από σταθερό σφάλμα, παρά το γεγονός ότι το εύρος των τιμών που προβλέπουν είναι αντίστοιχο με αυτό των τιμών επαλήθευσης. Οι προβλέψεις από τα νευρωνικά δίκτυα είναι λιγότερο ακριβείς στο σύνολο επαλήθευσης από ότι σε αυτό της εκπαίδευσης, παρουσιάζοντας μεγαλύτερα σφάλματα και εύρος τιμών από ότι το σύνολο επαλήθευσης. Οι προβλέψεις από το Krig δείχνουν σημαντική απόκλιση από την 1:1 γραμμή παλινδρόμησης, ενώ το DefKrig προβλέπει με μεγαλύτερη ακρίβεια από το Krig στις τιμές του συνόλου επαλήθευσης.



Εικόνα 3.6. Βαθμονόμηση των μοντέλων, όπως προκύπτει από τις γραμμικές παλινδρομήσεις μεταξύ των παρατηρούμενων (άξονας y) και προβλεπόμενων (άξονας x) τιμών του συνόλου επαλήθευσης (DS3). Η 1:1 γραμμή αντιστοιχεί σε ένα άριστα βαθμονομημένο μοντέλο.

Ανάλυση των σφαλμάτων των προβλέψεων στο DS3

Τα περιγραφικά στατιστικά (Ave, StDev) των παρατηρούμενων (s_A -DS3) και προβλεπόμενων τιμών, όπως και οι δείκτες σφαλμάτων για τις προβλέψεις, παρουσιάζονται στον Πίνακα 3.8. Η ανάλυση των σφαλμάτων των προβλέψεων στο DS3 έδωσε διαφορετικά αποτελέσματα, από ότι οι προβλέψεις στο DS1, γεγονός που δείχνει τη διαφορετική δυναμική των μοντέλων στην προσαρμογή τους στα δεδομένα εκπαίδευσης ή στη δυνατότητα πρόβλεψης εκτός αυτών. Τα DefKrig και ANNE προβλέπουν με μεγαλύτερη ακρίβεια κοντά στις μέσες παρατηρούμενες τιμές, ενώ αντίστοιχη είναι και η απόδοση των ASNN και AverMod. Και εδώ το Krig είναι το μόνον μοντέλο, που παρέχει προβλέψεις με τυπική απόκλιση κοντά σε αυτή των παρατηρούμενων τιμών. Οι δείκτες σφαλμάτων δείχνουν ότι τα DefKrig και ANNE παρέχουν τις πιο ακριβείς και με μικρότερο σφάλμα προβλέψεις, τα AverMod και ASNN είναι επίσης αρκετά ακριβή, ενώ τα υπόλοιπα μοντέλα υποεκτιμούν την ακουστική συχνότητα. Σύμφωνα με το MAE, τα ASNN, AverMod και GAM2 προβλέπουν τιμές παρόμοιες με τις παρατηρούμενες. Το μέσο σφάλμα και το άθροισμα των

σφαλμάτων δείχνουν ότι το Krig υπερεκτιμά τις παρατηρούμενες τιμές. Η ανάλυση των σφαλμάτων στα νευρωνικά δίκτυα έδειξε ότι το ANNE αποδίδει καλύτερα από το ASNN στην πρόβλεψη του ανεξάρτητου συνόλου DS3. Τα ANNE και DefKrig παρέχουν τις πιο ακριβείς και με το μικρότερο σφάλμα προβλέψεις στο ανεξάρτητο σύνολο DS3, ενώ τα AverMod και ASNN είναι τα πιο αποτελεσματικά στην προσαρμογή τους στο DS1. Και στις δύο περιπτώσεις το GAM2 απέδωσε καλύτερα μεταξύ των προσθετικών μοντέλων.

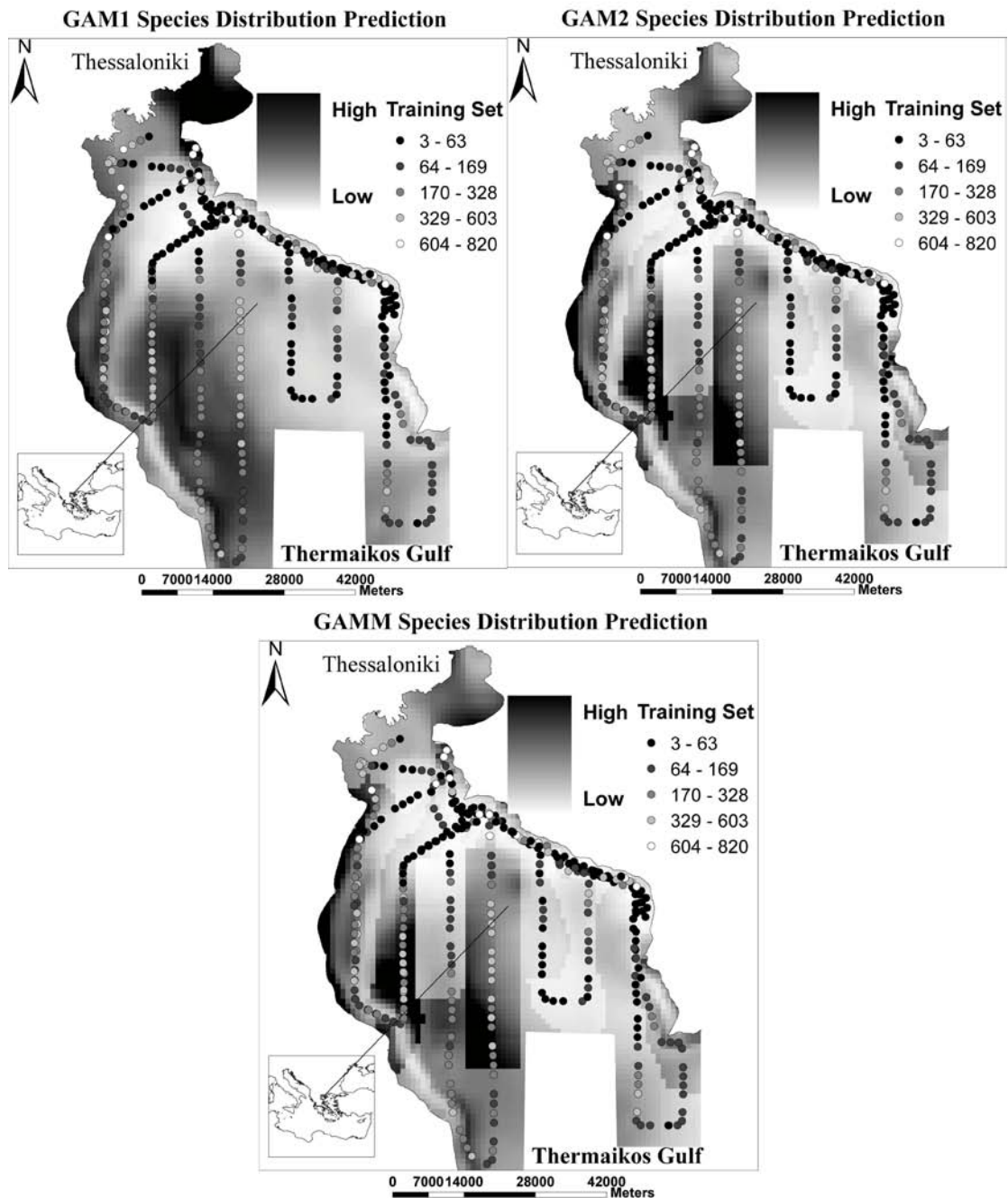
Πίνακας 3.8. Δείκτες σφαλμάτων των μοντέλων για τη σύγκριση της ακρίβειας προσαρμογής τους στα δεδομένα επαλήθευσης.

DS3	Sa1-DS3	GAM1	GAM2	GAMM	DefKrig	Krig	ANNE	ASNN	AverMod
Ave	195.011	121.167	135.830	133.212	193.649	269.270	197.925	179.003	188.242
StDev	218.712	96.831	129.724	128.699	87.956	215.041	78.762	121.495	126.426
RMSE		696.642	558.315	583.008	12.852	700.551	27.485	151.023	63.861
MSqE		485310.678	311715.763	339898.565	165.183	490772.162	755.408	22808.006	4078.247
MAE		119.579	106.025	109.288	134.417	154.808	130.066	104.537	101.902
ASE		1.568	1.468	1.414	1.647	1.452	1.498	1.325	1.315
MSIE		-0.338	-0.309	-0.347	-0.006	0.396	0.015	-0.102	-0.044
RMSqSt		3.185	2.912	3.276	0.053	3.735	0.138	0.967	0.415
ME		-73.844	-59.181	-61.799	-1.362	74.258	2.913	-16.008	-6.769

3.1.2.3 Σύγκριση των καννάβων πρόβλεψης κατανομής των ειδών

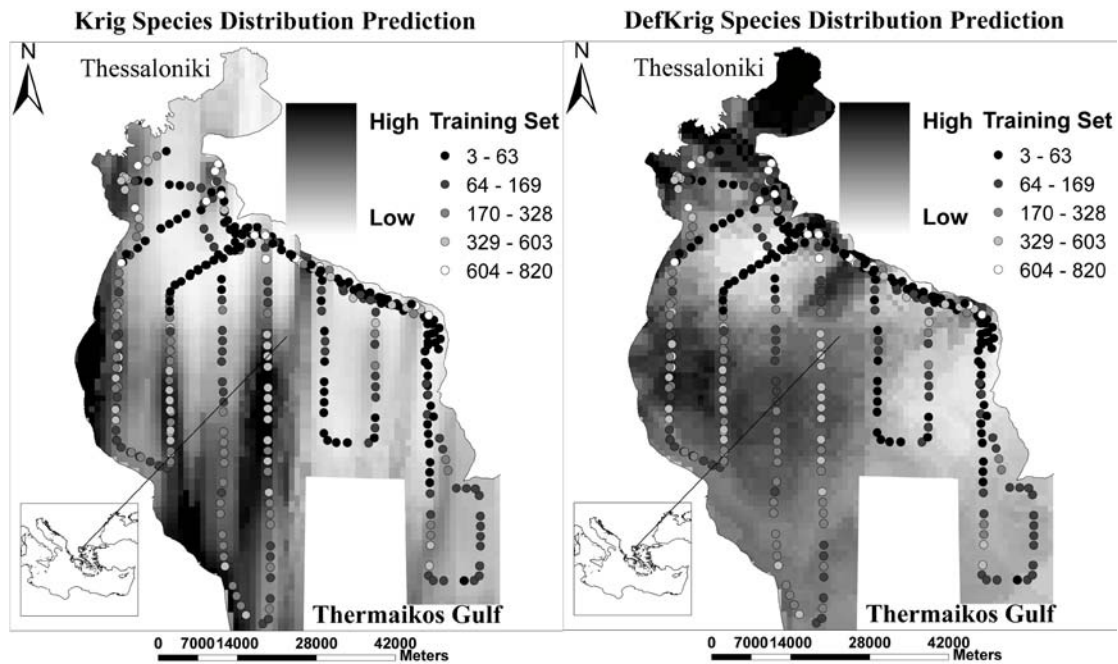
Στις Εικόνες 3.1 ως και 3.4 παρατηρείται μια γενική τάση κατανομής της ακουστικής πυκνότητας, η οποία περιγράφεται από χαμηλές τιμές νότια-νοτιοανατολικά και υψηλές βόρεια-βορειοδυτικά. Τοπικά, κυρίως στο βορειοανατολικό τμήμα της περιοχής μελέτης, παρατηρείται σημαντική διακύμανση των δειγματοληπτικών τιμών. Οι κάνναβοι πρόβλεψης (Εικ. 3.7, 3.8, 3.9, 3.10) δημιουργήθηκαν με σκοπό την αξιολόγηση των προβλέψεων των μοντέλων σε μια εκτεταμένη χωρικά περιοχή και την ανάδειξη των τεχνικών, που δύνανται να αποδώσουν καλύτερα την κατανομή των ειδών σε γενικό και σε τοπικό επίπεδο σε πολύ υψηλή χωρική ανάλυση. Πέρα από την κατανομή των ειδών, τα πλέγματα πρόβλεψης αξιολογούνται ως προς την ικανότητα τους να αποδώσουν λεπτομερώς σημαντικά ενδιαίτηματα. Οι τιμές στους καννάβους (Εικ. 3.7, 3.8, 3.9, 3.10) αντιστοιχούν στην ακουστική πυκνότητα (μαύρο: υψηλές τιμές, άσπρο: χαμηλές τιμές). Οι κύκλοι αντιπροσωπεύουν την ακουστική συχνότητα των τιμών της δειγματοληψίας (s_A -DS1) και έχουν αντίθετη διαβάθμιση χρώματος (μαύροι κύκλοι: χαμηλές τιμές, άσπροι κύκλοι: υψηλές τιμές).

Οι κάνναβοι πρόβλεψης κατανομής των ατόμων των ειδών που δημιουργήθηκαν από τα GAMs παρουσιάζονται στην Εικ. 3.7. Σύμφωνα με τη βαθμονόμηση των μοντέλων, τις συσχετίσεις, τα περιγραφικά στατιστικά και την ανάλυση των σφαλμάτων, το GAM2 αποδίδει καλύτερα από τα άλλα προσθετικά μοντέλα. Το GAMM1 από την άλλη, υπερτερεί του GAM1. Η Εικ. 3.7 παρέχει τη δυνατότητα οπτικής σύγκρισης των καμβάδων πρόβλεψης και των παρατηρούμενων τιμών (τελείες). Και τα τρία μοντέλα χαρακτηρίζονται από ένα κοινό πρότυπο ως προς την προβλεπόμενη κατανομή των ατόμων στο χώρο. Όμως στα GAM2 και GAMM1 είναι χαρακτηριστική η επίδραση του παράγοντα DDND, ο οποίος προκαλεί απότομες αλλαγές στην επιφάνεια της προβλεπόμενης ακουστικής πληροφορίας. Αυτή η παρατήρηση δείχνει την επίδραση της μέρας-σούρουπο-νύχτας-ξημέρωμα στο πρότυπο κατανομής των μικρών πελαγικών ειδών. Οι επεξηγηματικές μεταβλητές έχουν επιλεγεί κοντά στον πραγματικό χρόνο της δειγματοληψίας και επομένως οι κάνναβοι πρόβλεψης αντιστοιχούν στην πιθανή κατανομή των ειδών σε εκείνο το χρονικό διάστημα. Η γενικότητα των μοντέλων εξετάζεται κυρίως από την οπτική του χώρου, παρά του χρόνου με τη χρήση των DS2 και DS3. Το GAM1 αποτυπώνει μια σχετικά ομαλή χωρική κατανομή της ακουστικής πυκνότητας, που είναι σε συμφωνία με το γενικό χωρικό πρότυπο στην περιοχή, αλλά αδυνατεί να αποδώσει τις τοπικές διακυμάνσεις και την ετερογένεια των πιθανών ενδιαιτημάτων, όπως συμβαίνει στα GAMM1 και GAM2. Μεταξύ των τελευταίων δεν υπάρχουν σημαντικές διαφορές, ενώ η ομοιότητα τους είναι σε πολύ υψηλά επίπεδα ($r = 0.98$).



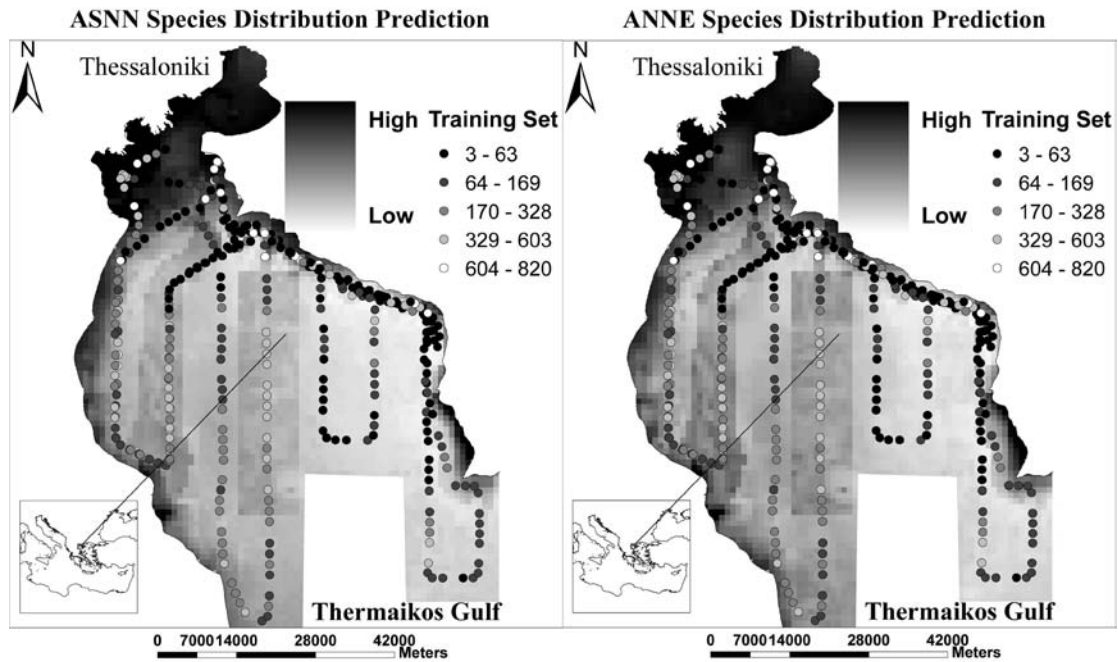
Εικόνα 3.7. Κάνναβι πρόβλεψης για τα GAMs και GAMM που απεικονίζουν την πυκνότητα της ακουστικής πληροφορίας (μαύρο: υψηλές τιμές, άσπρο: χαμηλές τιμές). Οι κύκλοι αντιστοιχούν στη δειγματοληπτική ποσότητα της ακουστικής αφθονίας των ψαριών (s_{A1}) σε αντίθετη χρωματική κλίμακα από τους καννάβους (μαύρο: χαμηλές τιμές, άσπρο: υψηλές τιμές). Άσπροι κύκλοι σε μαύρες περιοχές του χάρτη δείχνουν ακριβή πρόβλεψη και το αντίστροφο.

Οι χάρτες πρόβλεψης που δημιουργήθηκαν από τα μοντέλα DefKrig και Krig (Εικ. 3.8) διαφέρουν σημαντικά ($r = 0.36$) γεγονός που δείχνει τη διαφορετική φιλοσοφία ανάπτυξης των δύο μοντέλων. Η ανάλυση των σφαλμάτων, επίσης, τονίζει τις διαφορές των δύο μοντέλων, που αποτυπώνονται και στην κλίμακα των προβλεπόμενων τιμών στους χάρτες (Εικ. 3.8). Το DefKrig δεν απεικονίζει τη γενική τάση της ακουστικής πυκνότητας σε όλη την περιοχή στα επίπεδα που το επιτυγχάνει το Krig, το οποίο όμως τείνει να αποδίδει υπερβολικά όμοια τις τιμές από τις οποίες έχει εκπαιδευτεί. Το παραπάνω γεγονός είναι ιδιαίτερα εμφανές στο κέντρο της περιοχής μελέτης, όπου οι ακραία μεγάλες προβλεπόμενες τιμές ακουστικής πυκνότητας αντιστοιχούν ακριβώς στις δειγματοληπτικές τιμές.



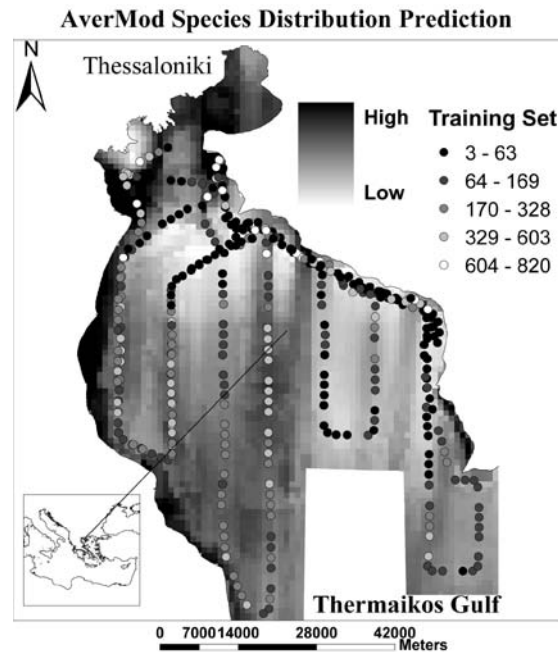
Εικόνα 3.8. Κάνναβοι πρόβλεψης για τα Krig και DefKrig που απεικονίζουν την πυκνότητα της ακουστικής πληροφορίας (μαύρο: υψηλές τιμές, άσπρο: χαμηλές τιμές). Οι κύκλοι αντιστοιχούν στη δειγματοληπτική ποσότητα της ακουστικής αφθονίας των ψαριών (s_A1) σε αντίθετη χρωματική κλίμακα από τους καννάβους (μαύρο: χαμηλές τιμές, άσπρο: υψηλές τιμές). Άσπροι κύκλοι σε μαύρες περιοχές του χάρτη δείχνουν ακριβή πρόβλεψη και το αντίστροφο.

Οι χάρτες πρόβλεψης που δημιουργήθηκαν με τα ASN και ANNE παρουσιάζονται στην Εικόνα 3.9 και εμφανίζουν στατιστικά σημαντικές ομοιότητες ($r = 0.98$). Το πρότυπο που δημιουργεί ο παράγοντας DDND είναι λιγότερο εμφανές από ότι στα GAMM και GAM2. Γενικά και τα δύο πλέγματα παρέχουν μια ομαλή και χωρίς συσσωματώματα επιφάνεια πιθανής κατανομής αποδεικνύοντας έτσι τη γενικότητα τους ως μοντέλα, όπως επίσης έχει επιβεβαιωθεί και από την ακρίβεια των μοντέλων στις προβλέψεις τους στο DS3. Το γενικό χωρικό πρότυπο κατανομής αποδίδεται ικανοποιητικά, ιδιαίτερα στις παράκτιες περιοχές. Ωστόσο το κέντρο της περιοχής μελέτης εμφανίζει γενικά μικρότερες τιμές από τις παρατηρούμενες, γεγονός που μπορεί να αποδοθεί στη γενική ομαλή διαβάθμιση των προβλέψεων.



Εικόνα 3.9. Κάρτες πρόβλεψης για τα ASNN και ANNE που απεικονίζουν την πυκνότητα της ακουστικής πληροφορίας (μαύρο: υψηλές τιμές, άσπρο: χαμηλές τιμές). Οι κύκλοι αντιστοιχούν στη δειγματοληπτική ποσότητα της ακουστικής αφθονίας των ψαριών (s_{A1}) σε αντίθετη χρωματική κλίμακα από τους κανάβους (μαύρο: χαμηλές τιμές, άσπρο: υψηλές τιμές). Άσπροι κύκλοι σε μαύρες περιοχές του χάρτη δείχνουν ακριβή πρόβλεψη και το αντίστροφο.

Οι χάρτες πρόβλεψης που δημιουργήθηκαν με το AveriMod παρουσιάζονται στην Εικόνα 3.10. Η διαδικασία βαθμονόμησης των μοντέλων, η ανάλυση των σφαλμάτων και η ανάλυση των συσχετίσεων δείχνει ότι το AveriMod χαρακτηρίζεται από μεγαλύτερη ικανότητα πρόβλεψης συγκριτικά με τις άλλες προσεγγίσεις. Ο παραγόμενος χάρτης πρόβλεψης της κατανομής των ατόμων των ειδών συνδυάζει τη γενικότητα στην πρόβλεψη του ASNN με την τοπική ικανότητα πρόβλεψης του Krig με ένα προσθετικό τρόπο, που οφείλεται στη λογική του GAM. Ο κάρναβος διατηρεί τη γενική χωρική τάση της κατανομής της αφθονίας των ειδών, ενώ παράλληλα εντοπίζονται σημεία υψηλής ή χαμηλής αφθονίας, που αναδεικνύουν την ετερογένεια της κατανομής της αφθονίας.

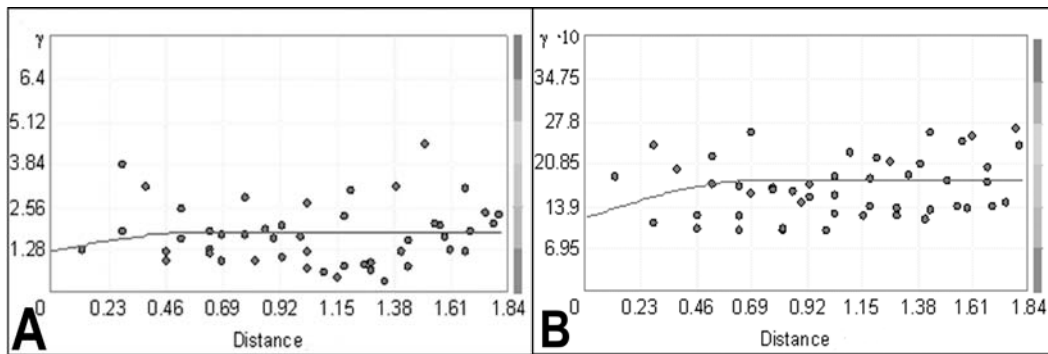


Εικόνα 3.10. Κάνναβος πρόβλεψης για το AverMod που απεικονίζει την πυκνότητα της ακουστικής πληροφορίας (μαύρο: υψηλές τιμές, άσπρο: χαμηλές τιμές). Οι κύκλοι αντιστοιχούν στη δειγματοληπτική ποσότητα της ακουστικής αφθονίας των ψαριών (S_{A1}) σε αντίθετη χρωματική κλίμακα από τους καννάβους (μαύρο: χαμηλές τιμές, άσπρο: υψηλές τιμές). Άσπροι κύκλοι σε μαύρες περιοχές του χάρτη δείχνουν ακριβή πρόβλεψη και το αντίστροφο.

3.2 Μοντέλα παρουσίας-απουσίας

3.2.1 Χωρικά πρότυπα και σύγκριση μεταξύ συνόλου δεδομένων εκπαίδευσης και επαλήθευσης

Τα σύνολα δεδομένων επαλήθευσης δημιουργήθηκαν από τα αρχικά δείγματα, από όπου επελέγησαν τιμές σε απόσταση h (όπου h τουλάχιστον $5 n$ mi). Η απόσταση επιλέχτηκε σύμφωνα με τα εμπειρικά βαριογράμματα (Εικ. 3.11) τόσο των τιμών επαλήθευσης όσο και των τιμών εκπαίδευσης των μοντέλων, ώστε να ελαχιστοποιηθεί η χωρική αυτοσυσχέτιση η οποία υπερ-εκτιμά την απόδοση των μοντέλων στη διαδικασία επαλήθευσης. Τα εμπειρικά βαριογράμματα και τα γραφήματα που δημιουργήθηκαν από τη λειτουργία αυτο-συσχέτισης έδειξαν χαμηλή αυτοσυσχέτιση των δεδομένων ακόμα και σε αποστάσεις μικρότερες των $5 n$ mi. Στα βαριογράμματα το nugget ήταν μικρότερο από το 65% από το κατώφλι (sill) και το εύρος μεγαλύτερο από 10 km ($5.4 n$ mi). Το σύνολο επαλήθευσης δεν παρουσιάζει καμία χωρική αυτοσυσχέτιση, όπως φαίνεται από την Εικ. 3.11A. Ο δείκτης χωρικής αυτοσυσχέτισης Moran για το σύνολο επαλήθευσης υπολογίστηκε σε 0.03 ενώ για το σύνολο εκπαίδευσης ήταν 0.14. Ερμηνεύοντας τις παραπάνω τιμές φαίνεται ότι το σύνολο επαλήθευσης εμφανίζει χαμηλή συσσωμάτωση τιμών του δείγματος, που μπορεί να αποδοθεί σε τυχαίο γεγονός. Από την άλλη, στο σύνολο εκπαίδευσης η συσσωμάτωση όμοιων τιμών δεν μπορεί να αποδοθεί στη τύχη για διάστημα εμπιστοσύνης 99%. Τα σφάλματα των ΜΚΕ ελέγχθηκαν επίσης για χωρικά πρότυπα, χωρίς να παρατηρηθεί κάποιο σημαντικό, ενδεχομένως λόγω των σχετικά πολλών επεξηγηματικών μεταβλητών που χρησιμοποιήθηκαν και της ικανοποιητικής προσαρμογής των μοντέλων (Elith & Leathwick, 2009). Τα παραπάνω αποτελέσματα επιβεβαιώνουν ότι το σύνολο επαλήθευσης δεν παρουσιάζει χωρική αυτοσυσχέτιση, σε αντίθεση με το σύνολο εκπαίδευσης των μοντέλων.



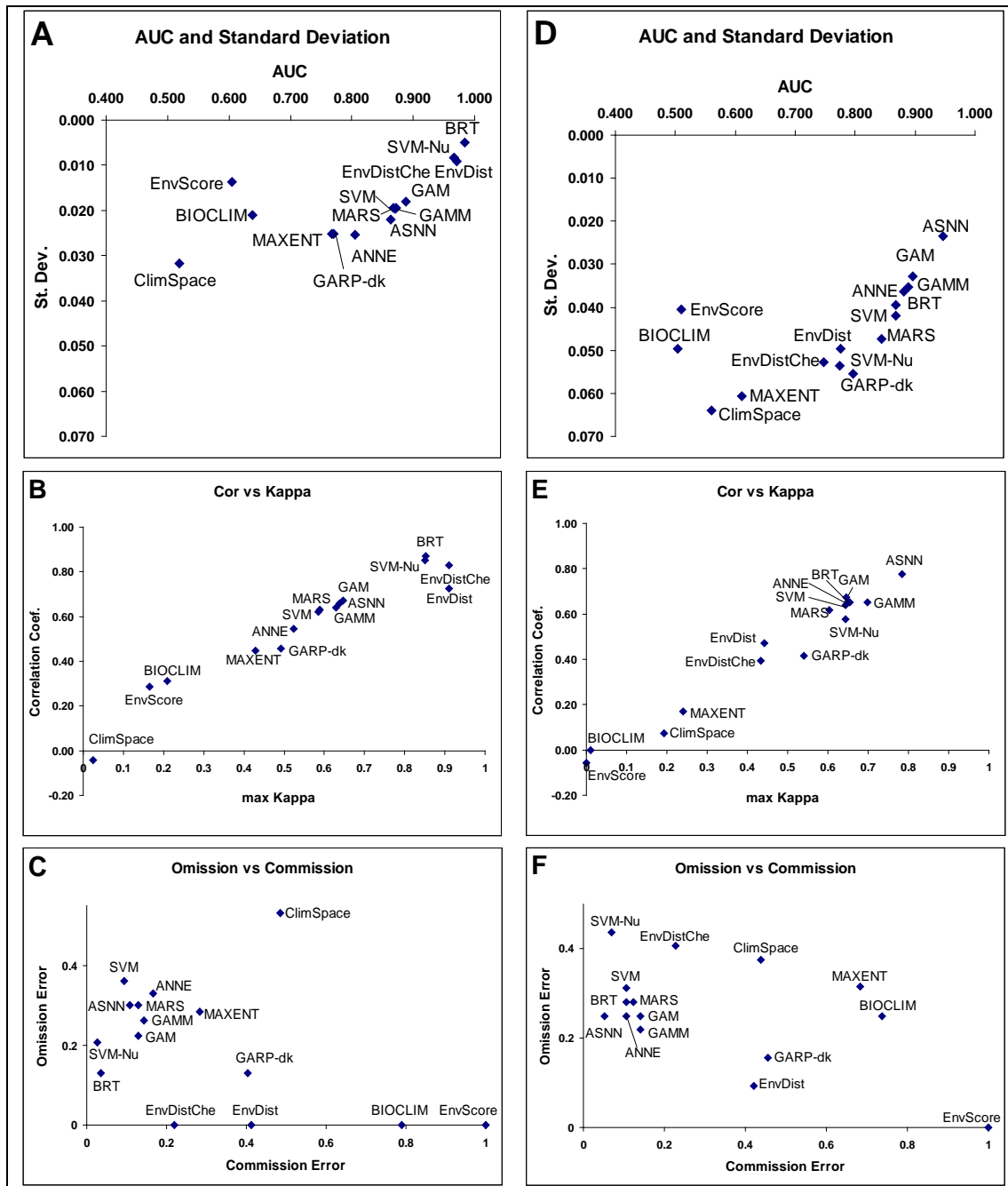
Εικόνα 3.11. Εμπειρικά βαριογράμματα της ακουστικής πυκνότητας. (Α) Για το σύνολο επαλήθευσης και (Β) για το σύνολο εκπαίδευσης

Τα σύνολα εκπαίδευσης και επαλήθευσης διαφέρουν ως προς το μέγεθος αυτο-συσχέτισης που παρουσιάζουν. Επιπλέον τα στατιστικά των κατανομών τους δεν μας επιτρέπουν να αμφισβητήσουμε την κοινή προέλευσή τους από τον ίδιο πληθυσμό. Σύμφωνα με το Mann-Whitney τεστ η μηδενική υπόθεση δεν απορρίπτεται ($U = 1550$, $p\text{-value} = 0.48$) οπότε βάση του ορισμού του κριτηρίου τα δύο σύνολα ανήκουν στον ίδιο "πληθυσμό", καθώς επίσης το Kolmogorov-Smirnov τεστ δύο δειγμάτων έδειξε ότι έχουν την ίδια κατανομή τιμών, αφού η μηδενική υπόθεση δεν απορρίπτεται ($Z=0.56$, $p\text{-value} = 0.87$). Η μέση τιμή και η τυπική απόκλιση του συνόλου επαλήθευσης (196 s_A και 220) και του συνόλου εκπαίδευσης (193 s_A και 215 αντίστοιχα) δεν διαφέρουν στατιστικά μεταξύ τους. Παρά τις ομοιότητες των δύο συνόλων δεν μπορούν να χαρακτηριστούν ως όμοια αφού τα διανύσματα των τιμών τους δεν συσχετίζονται σημαντικά (συντελεστής συσχέτισης Pearson = -0.202).

Οι παραπάνω συγκρίσεις και τα κριτήρια επιβεβαιώνουν την καταλληλότητα του συνόλου δεδομένων επαλήθευσης για το σκοπό αυτό, καθώς δεν εμφανίζει χωρική αυτοσυσχέτιση και συσχέτιση με το σύνολο εκπαίδευσης, ενώ από την άλλη παρουσιάζει ομοιότητες στην κατανομή τιμών με το σύνολο εκπαίδευσης, ώστε να αποφευχθεί η επαλήθευση με ένα "ξένο" σύνολο. Είτε η μια, είτε η άλλη περίπτωση θα οδηγούσαν σε υπερεκτίμηση ή υποεκτίμηση της απόδοσης των ΜΚΕ. Θα πρέπει να σημειωθεί ωστόσο ότι το σύνολο επαλήθευσης είναι το ίδιο για όλα τα ΜΚΕ, ώστε να επιτευχθεί η πλέον αντικειμενική σύγκριση τους.

3.2.2 Απόδοση Εφαρμογής

Τα ROC-AUC και οι ανάλογες τυπικές αποκλίσεις όλων των ΜΚΕ παρουσιάζονται στην Εικ. 3.12Α. Τα ΜΚΕ με το υψηλότερο ROC-AUC και τη χαμηλότερη τυπική συσχέτιση παρέχουν την καλύτερη εφαρμογή στα δεδομένα εκπαίδευσης. Αυτά παρουσιάζονται στην Εικ. 3.12Α στη πάνω δεξιά περιοχή του γραφήματος. Τα BRTs, EnvDist, EnvDistChe και SVM-Nu αποδίδουν καλύτερα από τις άλλες προσεγγίσεις επιτυγχάνοντας ROC-AUC μεγαλύτερο από 0.9. Τα μοντέλα παλινδρόμησης (GAM, GAMM, MARS) όπως και τα ASNN και SVM επίσης επιτυγχάνουν υψηλά ROC-AUC (0.86-0.9). Τα ANNE, GARP και MAXENT έχουν AUCs σε εύρος τιμών από 0.81-0.76. Τα BIOCLIM, EnvScore και ClimSpace δεν αποδίδουν ικανοποιητικά, επιτυγχάνοντας ROC-AUC λιγότερο από 0.64, ενώ το αντίστοιχο του ClimSpace είναι 0.52.



Εικόνα 3.12. Σύγκριση της απόδοσης εφαρμογής (A, B, C) και της ικανότητας πρόβλεψης (D, E, F) μεταξύ των ΜΚΕ. Απεικόνιση των ROC-AUC και των αντίστοιχων τυπικών αποκλίσεων των προσεγγίσεων μοντελοποίησης για το σύνολο εκπαίδευσης (A) και επαλήθευσης (D). Συντελεστές συσχέτισης και MaxKappa των ΜΚΕ για το σύνολο εκπαίδευσης (B) και επαλήθευσης (E). Σφάλματα παράλειψης και επιφόρτισης των ΜΚΕ για το σύνολο εκπαίδευσης (C) και επαλήθευσης (F).

Οι συσχετίσεις COR, που αποτελούν δείκτη ομοιότητας μεταξύ των παρατηρούμενων και προβλεπόμενων τιμών, και το MaxKappa παρουσιάζονται στην Εικ. 3.12B. Οι ομάδες των ΜΚΕ που φαίνονται στην Εικ. 3.12B είναι ανάλογες με αυτές που αναδεικνύονται από τα αποτελέσματα του ROC-AUC, αν και υπάρχουν μικροδιαφορές όσον αφορά την ταξινόμηση τους από το MaxKappa και COR. Οι προβλέψεις των BRTs και SVM-Nu παρουσιάζουν τις μέγιστες συσχετίσεις με το σύνολο εκπαίδευσης (0.87 και 0.85, αντίστοιχα). Από την άλλη τα EnvDist και EnvDistChe εμφανίζουν τα μεγαλύτερα MaxKappa (0.91). Τα ASNN, GAM και GAMM αποδίδουν σχεδόν όμοια (COR: 0.67-0.64 και MaxKappa: 0.65-0.62). Το MARS αποδίδει ανάλογα με το SVM (COR: 0.63-0.62 και MaxKappa: 0.59-0.58). Το ANNE επιτυγχάνει COR ίσο με 0.55 και MaxKappa 0.52, ενώ το MAXENT και το GARP παρουσιάζουν όμοια COR (0.45 και 0.46, αντίστοιχα), αλλά το MaxKappa είναι υψηλότερο για το GARP από ότι για το MAXENT (0.49 έναντι 0.43). Τα Bioclim,

EnvScore και κυρίως το ClimSpace αποτυγχάνουν να προσομοιάσουν τα δεδομένα εκπαίδευσης παρουσιάζοντας COR ίσο με 0.31, 0.29 και 0.04, αντίστοιχα και MaxKappa ίσο με 0.2, 0.16 και 0.02.

Τα EnvDist και EnvDistChe αποδίδουν αναλογικά ικανοποιητικά στην προσομοίωση των δεδομένων εκπαίδευσης (ROC-AUC 0.97 και για τα δύο). Μόνο τα BRT και SVM-Nu επιτυγχάνουν μεγαλύτερα ROC-AUC, ενώ τα δύο πρώτα έχουν μεγαλύτερα MaxKappa (0.91) και αναλογικά μεγαλύτερες τιμές COR (0.73 και 0.83, αντίστοιχα). Και τα δύο μοντέλα χαρακτηρίζονται από μηδενικά σφάλματα παράλειψης (omission error) και το EnvDistChe παρουσιάζει σχεδόν το μισό του σφάλματος επιφόρτισης (commission error) από το EnvDist (0.22 και 0.41). Το μηδενικό σφάλμα παράλειψης όταν συνδυάζεται με χαμηλό σφάλμα επιφόρτισης, όπως συμβαίνει στο EnvDistChe, αποτελεί σαφή ένδειξη ότι το μοντέλο μιμείται ακριβώς τα δεδομένα εκπαίδευσης. Αυτό επιβεβαιώνει και ο χάρτης πρόβλεψης του EnvDistChe, όπου οι υψηλές προβλεπόμενες τιμές συγκεντρώνονται γύρω από τις υψηλές τιμές των δειγμάτων και αντίστοιχα για τις χαμηλές τιμές. Το EnvDist εμφανίζει επίσης τάση για υπερ-προσαρμογή των δεδομένων εκπαίδευσης, σε μικρότερο όμως βαθμό από το EnvDistChe. Ακόμη ένα στοιχείο που επιβεβαιώνει τα παραπάνω είναι η μειωμένη ικανότητα πρόβλεψης του EnvDist. Αφού προσομοιάζει απόλυτα τα δεδομένα εκπαίδευσης, αδυνατεί να πραγματοποιήσει επιτυχημένη πρόβλεψη στο σύνολο επαλήθευσης παρουσιάζοντας χαμηλότερα ROC-AUC, COR και MaxKappa από το EnvDist. Συγκριτικά με τις άλλες προσεγγίσεις και τα δύο μοντέλα είναι λιγότερο αποτελεσματικά στην ικανότητα πρόβλεψης, τόσο σε σχέση με τα μοντέλα παλινδρόμησης όσο και με τις περισσότερες από τις τεχνικές μηχανικής μάθησης. Μεταξύ των ΜΚΕ που δεν υποστηρίζουν δεδομένα απουσίας τα EnvDist και EnvDistChe αποδίδουν σχετικά ικανοποιητικά, όχι όμως στο επίπεδο του SVM.

Όπως αναφέρθηκε και προηγουμένως, τα σφάλματα επιφόρτισης και παράλειψης αντικατοπτρίζουν την ποιότητα των προβλεπόμενων τιμών, όσον αφορά την υπερβολική ή μειωμένη πρόβλεψη τιμών και την υπερ-προσαρμογή στα δεδομένα εκπαίδευσης. Η Εικ. 3.12C αποτυπώνει τα σφάλματα επιφόρτισης και παράλειψης των προσεγγίσεων που εφαρμόστηκαν. Υψηλές τιμές σφαλμάτων παράλειψης δείχνουν ελλιπή απόδοση εφαρμογής (π.χ. ClimSpace). Μηδενικό σφάλμα επιφόρτισης σε συνδυασμό με υψηλές τιμές σφάλματος παράλειψης δείχνουν υπερβολική πρόβλεψη κατανομής ειδών (κυρίως EnvScore και BIOCLIM). Μηδενικό σφάλμα επιφόρτισης σε συνδυασμό με μηδενικές τιμές σφάλματος παράλειψης δείχνουν ότι οι προβλεπόμενες τιμές υπερ-προσαρμόζονται στο σετ εκπαίδευσης (π.χ. EnvDistChe). Οι χάρτες πρόβλεψης (αναλυτικά σε επόμενη παράγραφο) οι οποίοι παράγονται από μοντέλα με τα προαναφερθέντα σφάλματα δίνουν μια οπτική εικόνα των εσφαλμένων προβλεπόμενων χωρικά κατανομών.

3.2.3 Ικανότητα πρόβλεψης

Ένα ακόμη διαγνωστικό μέσο της απόδοσης των ΜΚΕ είναι η ικανότητα τους να προβλέπουν ικανοποιητικά τιμές, όταν εφαρμόζονται σε δεδομένα διαφορετικά από αυτά με τα οποία έχουν εκπαιδευτεί. Αυξημένη ικανότητα πρόβλεψης υποδηλώνει ΜΚΕ, που χαρακτηρίζονται από γενικότητα στην εφαρμογή τους διατηρώντας τα πρότυπα από τα οποία έχουν δημιουργηθεί. Σε αυτή την περίπτωση όλα τα ΜΚΕ δοκιμάζονται ως προς την ικανότητα πρόβλεψης ενός 'άγνωστου' για αυτά συνόλου δεδομένων. Η συνεκτίμηση των αποτελεσμάτων σε συνδυασμό με την απόδοση εφαρμογής τους και τα διαγνωστικά τους θα δώσουν μια πιο ολοκληρωμένη εικόνα για τη συμπεριφορά και καταλληλότητα των ΜΚΕ σε αλιευτικά δεδομένα. Τα ROC-AUC και οι ανάλογες τυπικές αποκλίσεις από την εφαρμογή των ΜΚΕ στα δεδομένα επαλήθευσης παρουσιάζονται στην Εικ. 3.12D. Τα ΜΚΕ με το υψηλότερο ROC-AUC και τη χαμηλότερη τυπική απόκλιση αποτελούν τις προσεγγίσεις με την υψηλότερη ικανότητα πρόβλεψης (Εικ. 3.12D, περιοχή πάνω δεξιά του γραφήματος). Οι τεχνικές που αποδίδουν σχετικά αποτελεσματικά στην πρόβλεψη του συνόλου εκπαίδευσης, αλλά αποτυγχάνουν να προβλέψουν ικανοποιητικά τα δεδομένα επαλήθευσης, πιθανόν να υπερ-προσαρμόζονται στο σετ εκπαίδευσης και τελικά να παρουσιάζουν μειωμένη γενικότητα στην εφαρμογή τους. Σύμφωνα με την ταξινόμηση των ΜΚΕ στην Εικ. 3.12D το ASNN αποδίδει σαφώς

καλύτερα από τις άλλες προσεγγίσεις επιτυγχάνοντας ROC-AUC κοντά στο 0.96. Τα μοντέλα παλινδρόμησης (GAM, GAMM, MARS) όπως και τα BRT, SVM και ANNE επιτυγχάνουν επίσης υψηλά ROC-AUC (0.84-0.9). Τα SVM-Nu, EnvDist, EnvDistChe και GARP έχουν ROC-AUC από 0.75 ως 0.8. Τα MAXENT και ClimSpace επιτυγχάνουν τιμές ROC-AUC 0.61 και 0.56 αντίστοιχα. Τα BIOCLIM και EnvScore δεν αποδίδουν ικανοποιητικά με τιμές AUC 0.50 και 0.51, αντίστοιχα.

Οι συσχετίσεις COR, που είναι δείκτες ομοιότητας μεταξύ των παρατηρούμενων τιμών στο σύνολο επαλήθευσης και των προβλεπόμενων τιμών και οι τιμές του MaxKappa παρουσιάζονται στην Εικ. 3.12E. Το γενικό πρότυπο της ταξινόμησης των ΜΚΕ από το γράφημα των ROC-AUC είναι σχεδόν ίδιο και στο γράφημα των COR. Το ASNN δείχνει τη μεγαλύτερη ικανότητα πρόβλεψης (0.78 COR, 0.78 MaxKappa). Ακολουθεί μια διακριτή ομάδα ΜΚΕ με τιμές COR μεταξύ 0.58-0.67 και MaxKappa 0.6-0.7 που περιλαμβάνει τα GAM, GAMM, ANNE, BRT, SVM, SVM-Nu και MARS. Το GARP επιτυγχάνει COR ίσο με 0.42 και MaxKappa ίσο με 0.54. Τα EnvDist και EnvDistChe έχουν το ίδιο MaxKappa (0.44) αλλά διαφέρουν ως προς τη συσχέτιση παρατηρούμενων και προβλεπόμενων τιμών (0.47 και 0.39, αντίστοιχα). Τα MAXENT, ClimSpace, BIOCLIM και EnvScore παρουσιάζουν μικρότερη ικανότητα πρόβλεψης επιτυγχάνοντας τιμές COR από 0.17 ως 0.00 και MaxKappa από 0.24 ως 0.00.

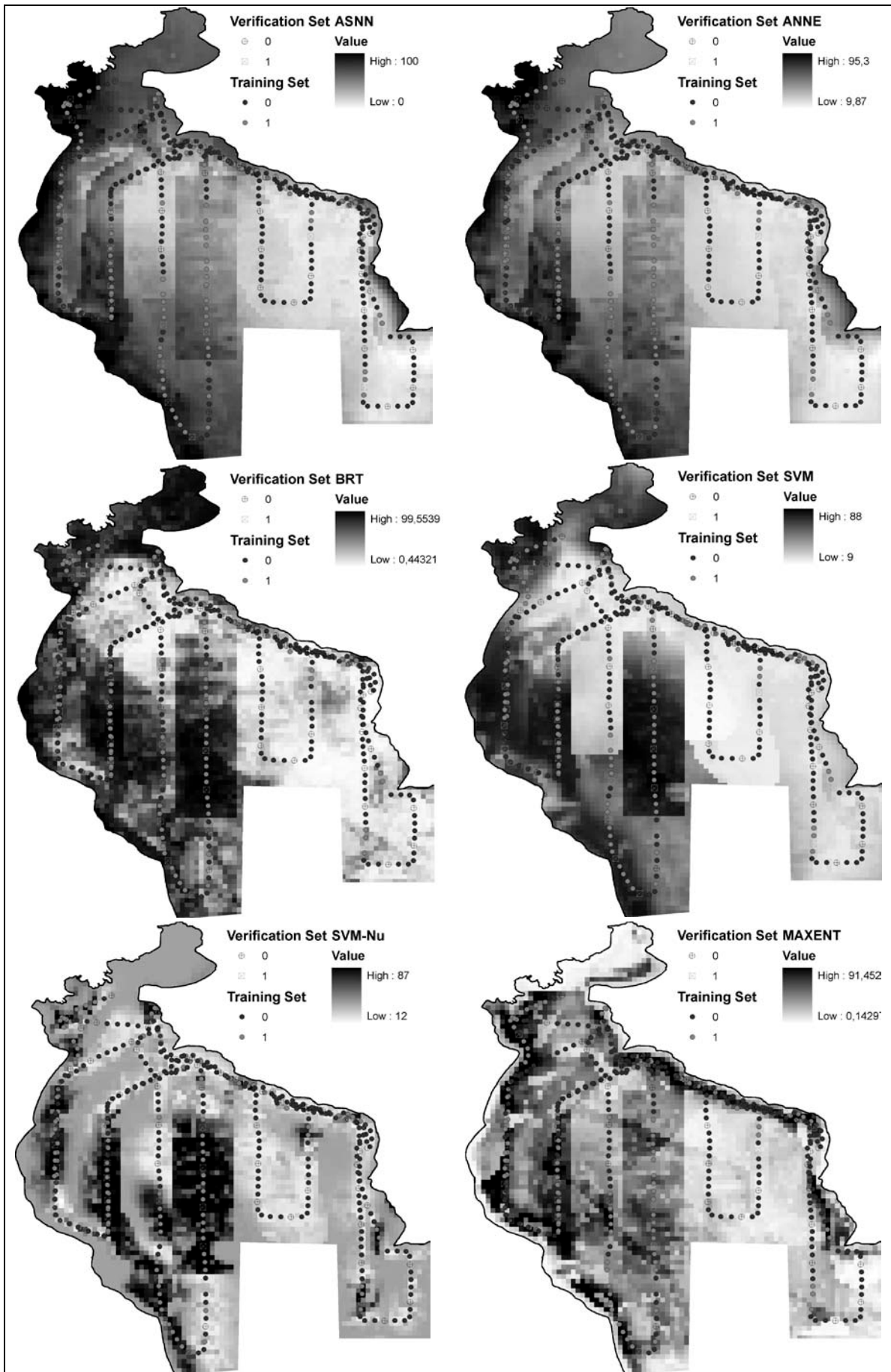
Σε αντίθεση με την Fig. 3.12C (που αναφέρεται στην εκπαίδευση των μοντέλων), η Fig. 3.12F δίνει την ικανότητα πρόβλεψης των ΜΚΕ σε σχέση με τα δεδομένα επαλήθευσης, όσον αφορά τα παραγόμενα σφάλματα επιφόρτισης και παράλειψης. Υψηλό σφάλμα παράλειψης δείχνει την αδυναμία του ΜΚΕ να προσδιορίσει την παρουσία των ειδών, ενώ υψηλό σφάλμα επιφόρτισης δείχνει την αδυναμία του ΜΚΕ να διακρίνει ακατάλληλα ενδιαιτήματα. Τα ASNN, GAM, GAMM, ANNE, BRT, και SVM αποτελούν τα ΜΚΕ με τα μικρότερα σφάλματα, όσον αφορά την πρόβλεψη που κάνουν στο σύνολο επαλήθευσης. Τα MAXENT, BIOCLIM και EnvScore είναι τα ΜΚΕ με τα μεγαλύτερα σφάλματα επιφόρτισης, ενώ τα SVM-Nu, EnvDistChe και ClimSpace έχουν τα υψηλότερα σφάλματα παράλειψης.

3.2.4 Χάρτες Πιθανοτήτων

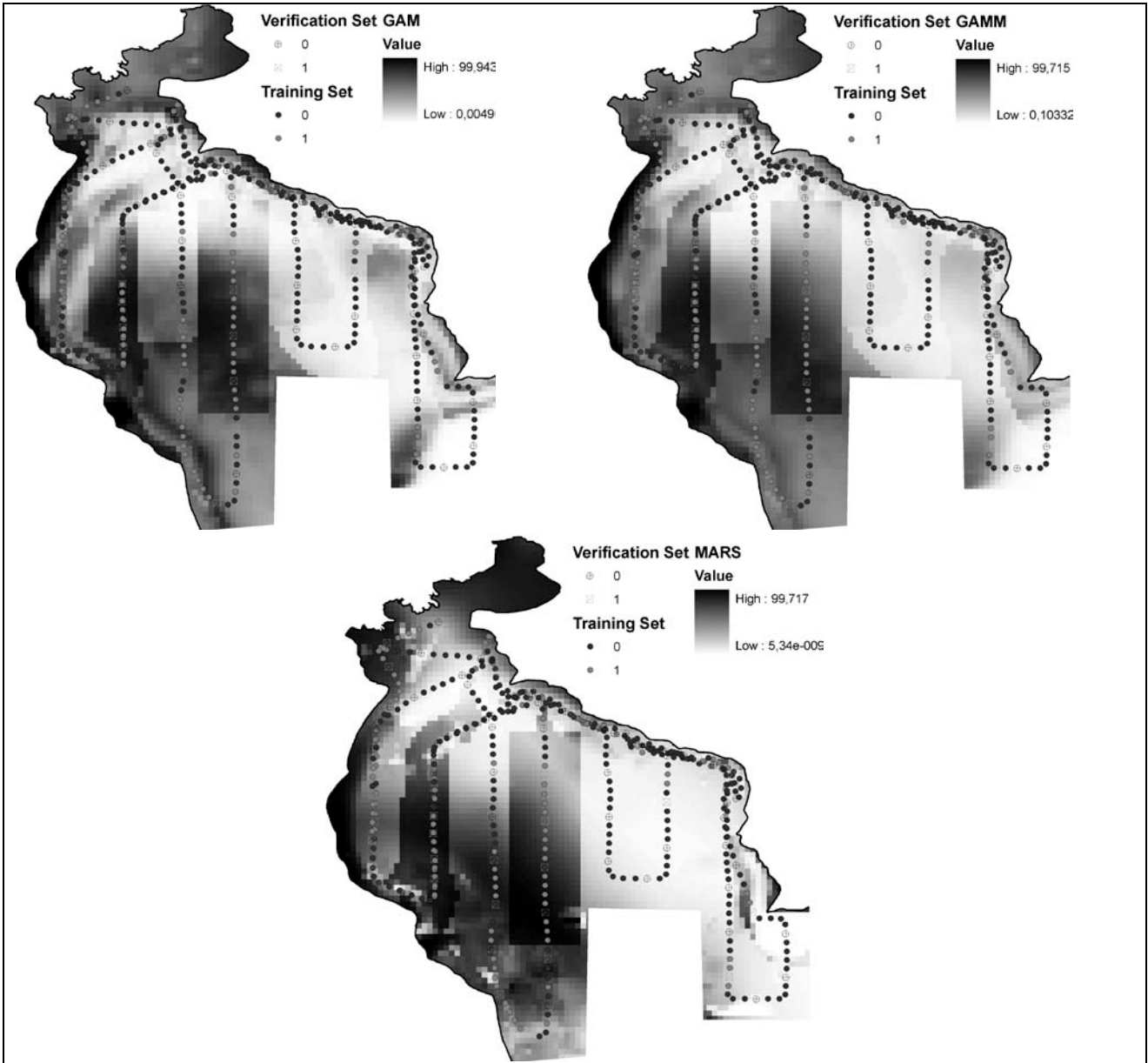
Τα μοντέλα που αναπτύχθηκαν από κάθε προσέγγιση εφαρμόστηκαν στους καννάβους των επεξηγηματικών μεταβλητών, ώστε να παραχθούν οι αντίστοιχοι χάρτες πρόβλεψης της κατανομής των ατόμων των ειδών, που καλύπτουν σε υψηλή ανάλυση (0.01 δεκαδικές μοίρες) την περιοχή μελέτης. Τα δεδομένα παρουσίας-απουσίας παράγουν χάρτες πιθανοτήτων, που εκφράζουν την πιθανότητα παρουσίας ατόμων των ειδών. Η παρουσία αντιστοιχεί σε υψηλή ακουστική πυκνότητα και υποδηλώνει κατάλληλα ενδιαιτήματα, ενώ η απουσία σε χαμηλή ή μηδενική ακουστική πυκνότητα και απουσία ατόμων. Πέρα από την ικανότητα πρόβλεψης της παρουσίας των ατόμων των ειδών σε λεπτομερή χωρική διάσταση, οι παραγόμενοι χάρτες αναδεικνύουν την πιθανή ετερογένεια των κατάλληλων ενδιαιτημάτων. Η Εικ. 3.13 δείχνει τους χάρτες πιθανοτήτων που έχουν προκύψει από τις μεθόδους που ανήκουν στις τεχνικές μηχανών μάθησης (BRT, ASNN, ANNE, MAXENT, SVM). Οι χάρτες πιθανοτήτων που προέρχονται από τις μεθόδους παλινδρόμησης παρουσιάζονται στην Εικ. 3.14, ενώ αυτοί που αντιστοιχούν στις μεθόδους τύπου φακέλου είναι στην Εικ. 3.15. Οι EnvScore και BIOCLIM δεν υποστηρίζουν δεδομένα απουσίας και έχουν αναπτυχθεί μόνο με δεδομένα παρουσίας. Τέλος οι χάρτες που έχουν δημιουργηθεί από τις τεχνικές EnvDist, EnvDistChe, ClimSpace και GARP αποτυπώνονται στην Εικ. 3.16. Μεταξύ των προαναφερθέντων μεθόδων μόνο η GARP υποστηρίζει δεδομένα παρουσίας-απουσίας, ενώ οι άλλες αναπτύχθηκαν με δεδομένα μόνο παρουσίας. Ο Πίνακας 3.9 παρουσιάζει τους συντελεστές συσχέτισης του Pearson μεταξύ των ΜΚΕ. Η πάνω αριστερή πλευρά του Πίνακα 3.9 αντιστοιχεί στις συσχετίσεις μεταξύ προβλέψεων στα δεδομένα επαλήθευσης, ενώ η κάτω αριστερή πλευρά του Πίνακα 3.9 αντιστοιχεί στις συσχετίσεις μεταξύ των καννάβων πρόβλεψης της κατανομής των ατόμων, όπως υπολογίστηκε από το εργαλείο συσχέτισης καννάβων του λογισμικού ArcInfo της ESRI.

Πίνακας 3.9 Συντελεστής συσχέτισης του Pearson μεταξύ των προβλέψεων των ΜΚΕ για το σύνολο επαλήθευσης (πάνω δεξιά) και για τους καννάβους πρόβλεψης (κάτω αριστερά).

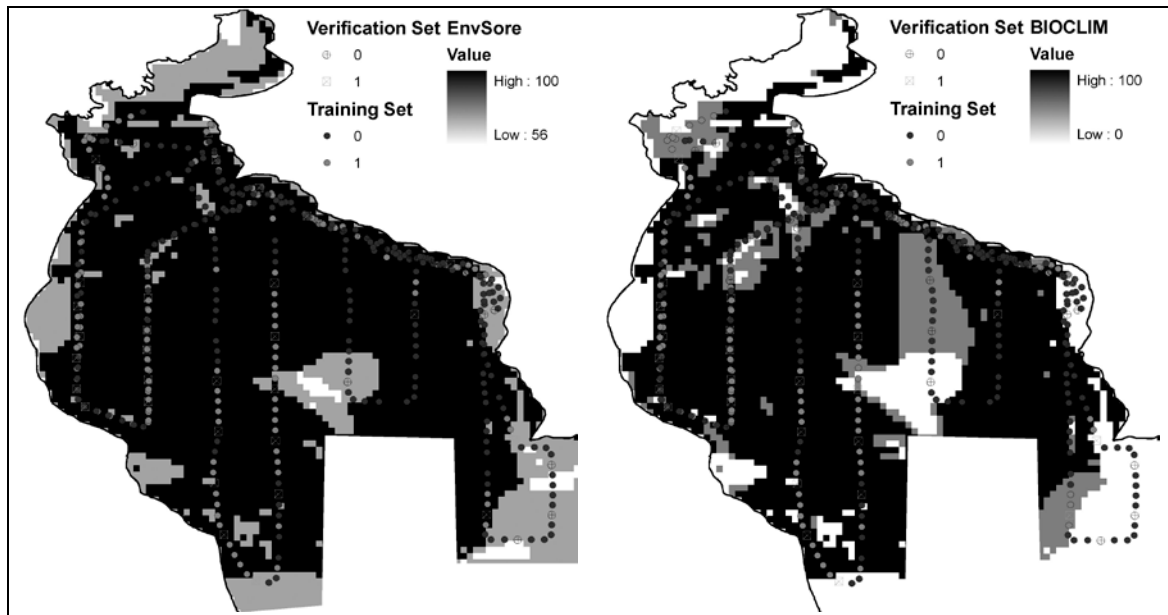
Pearson R	GAM	GAMM	BRT	MARS	BIO CLIM	Clim Space	Env Dist	Env DistChe	Env Score	GARP -dk	SVM	SVM -Nu	MAX ENT	ANNE	ASNN
GAM		0.98	0.87	0.90	0.17	-0.04	0.64	0.53	0.10	0.63	0.82	0.70	0.49	0.90	0.85
GAMM	0.97		0.87	0.93	0.19	-0.02	0.61	0.49	0.12	0.69	0.85	0.70	0.48	0.93	0.85
BRT	0.84	0.87		0.85	0.18	-0.03	0.66	0.60	0.14	0.63	0.84	0.83	0.48	0.85	0.88
MARS	0.87	0.91	0.86		0.21	-0.04	0.60	0.47	0.14	0.65	0.80	0.64	0.47	0.89	0.80
BIOCLIM	-0.05	-0.02	-0.01	0.00		0.38	0.40	0.32	0.87	0.14	0.10	0.07	0.46	0.17	0.10
ClimSpace	-0.23	-0.20	-0.24	-0.26	0.41		-0.08	-0.07	0.30	0.17	0.04	0.08	-0.08	0.01	0.00
EnvDist	0.41	0.43	0.43	0.42	0.34	-0.06		0.87	0.24	0.28	0.49	0.54	0.69	0.60	0.69
EnvDistChe	0.30	0.29	0.32	0.28	0.26	-0.05	0.80		0.21	0.10	0.40	0.57	0.57	0.45	0.60
EnvScore	-0.11	-0.10	-0.10	-0.11	0.85	0.40	0.26	0.21		0.11	0.10	0.09	0.30	0.13	0.02
GARP-dk	0.55	0.60	0.56	0.55	0.16	-0.05	0.39	0.24	0.05		0.69	0.58	0.30	0.73	0.56
SVM	0.78	0.82	0.82	0.80	0.00	-0.23	0.50	0.34	-0.10	0.65		0.80	0.33	0.91	0.83
SVM-Nu	0.57	0.55	0.59	0.51	0.15	-0.14	0.63	0.65	0.09	0.46	0.67		0.20	0.75	0.82
MAXENT	0.31	0.34	0.36	0.33	0.43	-0.06	0.65	0.55	0.38	0.39	0.38	0.44		0.43	0.41
ANNE	0.73	0.79	0.71	0.76	-0.05	-0.16	0.39	0.23	-0.13	0.72	0.78	0.39	0.35		0.89
ASNN	0.73	0.78	0.72	0.76	-0.93	-0.20	0.36	0.21	-0.16	0.66	0.75	0.35	0.33	0.97	



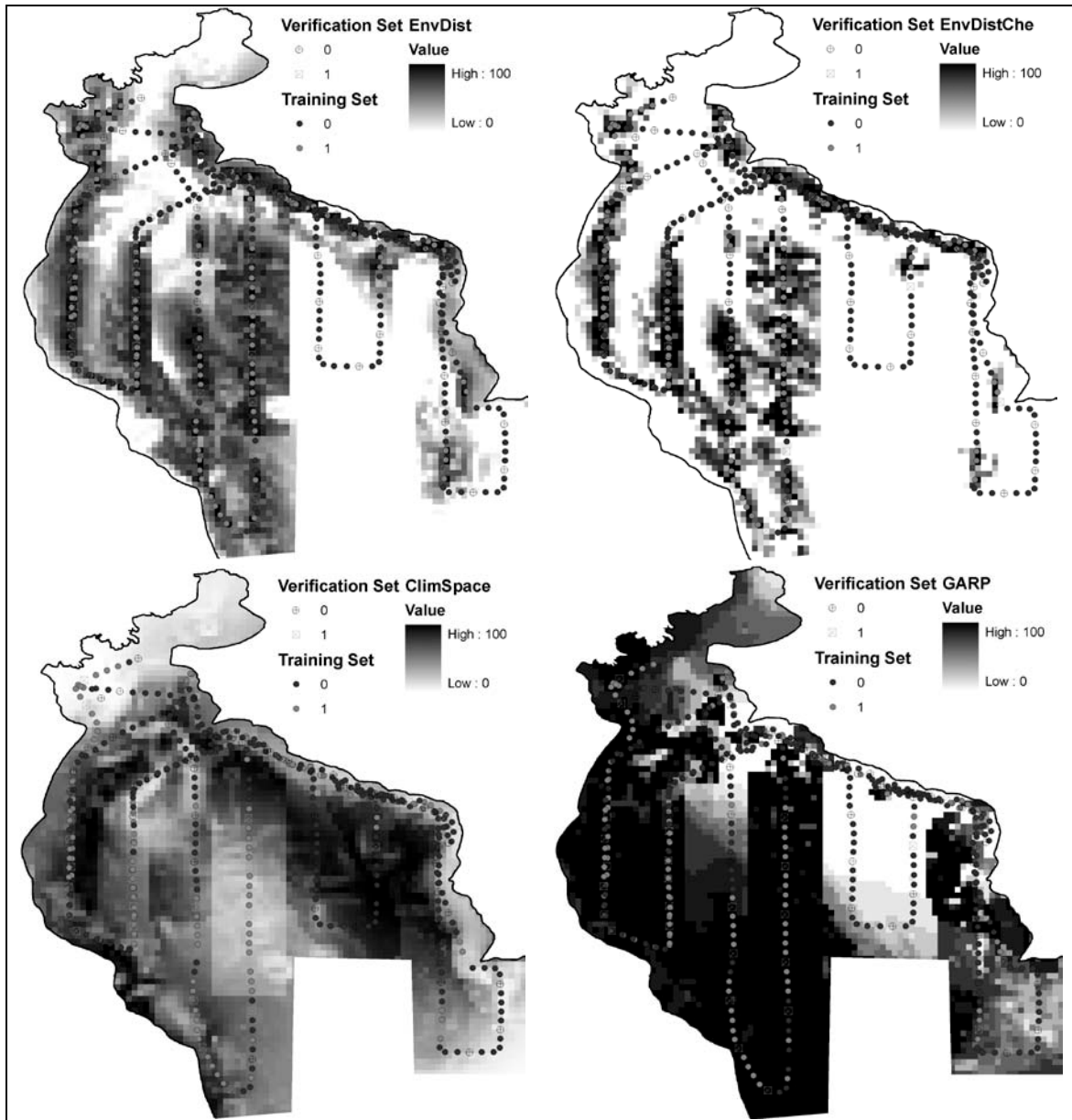
Εικόνα 3.13 Χάρτες πιθανοτήτων της κατανομής των ατόμων ειδών από τις τεχνικές μηχανικής μάθησης.



Εικόνα 3.14. Χάρτες πιθανοτήτων της κατανομής των ατόμων ειδών από τις τεχνικές παλινδρόμησης.



Εικόνα 3.15 Χάρτες πιθανοτήτων της κατανομής των ατόμων ειδών από τις μεθόδους τύπου "φακέλου".



Εικόνα 3.16. Χάρτες πιθανότητας της κατανομής των απόμων ειδών από τις μεθόδους: κλιματικές αποστάσεις, GARP και κλιματικό χώρο.

4. Συζήτηση

Ένας από τους κύριους σκοπούς της διατριβής είναι η επιλογή των καταλληλότερων Μοντέλων Κατανομής Ειδών (ΜΚΕ) όσον αφορά στη δυναμική προσομοίωση των δειγματοληπτικών δεδομένων και στην αποτελεσματικότητα εξαγωγής ή παρεμβολής των προβλέψεων. Στα πλαίσια της επιλογής των ΜΚΕ παρατίθεται σε ένα πρώτο επίπεδο η σύγκριση και αξιολόγηση των μοντέλων που αναπτύχθηκαν με την ίδια μέθοδο, όπου παρουσιάζονται τα ιδιαίτερα χαρακτηριστικά κάθε προσέγγισης και οι ιδιαιτερότητες κατά την εφαρμογή τους. Σε ένα δεύτερο επίπεδο, επιχειρείται η σύγκριση συνολικά των μοντέλων που προέρχονται από διαφορετικές προσεγγίσεις, ώστε να επιτευχθεί η τελική αξιολόγηση ως προς τη δυναμική κάθε προσέγγισης. Στα πλαίσια των παραπάνω αξιολογήσεων και συγκρίσεων έχουν ενσωματωθεί και οι νεοεισαχθείσες στα ΜΚΕ προσεγγίσεις (συνδεδετικά νευρωνικά δίκτυα, μοντέλο βελτιωμένης πρόβλεψης), που αποτελούν εναλλακτικές και πολλά υποσχόμενες τεχνικές στο χώρο της μοντελοποίησης κατανομής ειδών.

Μετά τη σύγκριση και αξιολόγηση των ΜΚΕ, αναλύονται καίρια ζητήματα για την ορθή εφαρμογή των ΜΚΕ, όπως η ύπαρξη χωρικών προτύπων και χωρικών αυτοσυσχετίσεων στα δεδομένα και η επιλογή της κατάλληλης μεθόδου επαλήθευσης και επικύρωσης των μοντέλων. Τέλος επιχειρείται η ερμηνεία των χαρτών πρόβλεψης, που συνδέεται στενά με την ποιότητα της απόδοσης των μοντέλων, από τη διάσταση της βιολογίας των ειδών. Επίσης, αναλύονται τα ΜΚΕ για την ανάδειξη σχέσεων είδους-περιβάλλοντος.

Υδροακουστικά δεδομένα

Η εφαρμογή των υδροακουστικών δεδομένων σε είκοσι προσεγγίσεις μοντελοποίησης, που ανήκουν σε δεκατέσσερις διαφορετικές μεθόδους επιβεβαιώνει τη δυναμική των δεδομένων αυτού του τύπου, καθώς και τη δυνατότητα να μετασχηματίζονται από ακουστική αφθονία σε παρουσία-παρουσία αξιοποιώντας μεγάλο εύρος εφαρμογών. Οι μετασχηματισμοί πραγματοποιούνται χωρίς να υφίσταται σημαντική απώλεια πληροφορίας, καθώς οι χάρτες πρόβλεψης που προέρχονται από μεθόδους αφθονίας είναι ανάλογοι με τους χάρτες πιθανοτήτων που δημιουργούνται με δεδομένα παρουσίας-απουσίας (π.χ. σύγκριση Εικ. 3.7 και Εικ. 3.14). Η μη απώλεια πληροφορίας κατά το μετασχηματισμό μπορεί να αποδοθεί στο γεγονός ότι τα δεδομένα αποτελούν συνεχείς καταγραφές, σε ένα κανονικό σχέδιο διατομών, που καλύπτουν όλη την περιοχή μελέτης, δηλαδή στον τύπο της δειγματοληψίας και την πυκνότητα των δειγμάτων. Άλλωστε, σύμφωνα με τους Hirzel & Guisan (2002) η συστηματική δειγματοληψία (όπως αυτή που δημιουργεί τα υδροακουστικά δεδομένα) παρέχει καλύτερα δεδομένα από ότι η τυχαία ή η τυχαία στρωματοποιημένη δειγματοληψία για τη μοντελοποίηση της κατανομής ειδών. Η παραπάνω θεώρηση επιβεβαιώνεται και από τη σύγκριση των αποτελεσμάτων της διατριβής με εργασίες μοντελοποίησης που χρησιμοποιούν αλιευτικά δεδομένα από τυχαία στρωματοποιημένη δειγματοληψία (π.χ. Leukaditou *et al.* 2008, δεδομένα από το πρόγραμμα Δειγματοληψίες Τράτας στη Μεσόγειο-MEDITS; Damalas, αδημοσίευτα δεδομένα, on-board δεδομένα σε αλιευτικά σκάφη διάφορων εργαλείων). Η πυκνότητα των δειγμάτων σε συνδυασμό με τα πολύ υψηλής χωρικής ανάλυσης περιβαλλοντικά δεδομένα επιτρέπει την πλήρη αξιοποίηση των πρώτων, ώστε οι χάρτες πρόβλεψης των πλέον αποδοτικών προσεγγίσεων, να μπορούν να διακρίνουν τη ρεαλιστική-πραγματική κατανομή των ειδών και να αναδείξουν την προβλεπόμενη ετερογένεια των ενδιαιτημάτων.

Η υψηλή αντιπροσώπευση της σαρδέλας στη σύνθεση ειδών και η έντονη ανάμειξη της με τις συγκεντρώσεις γαύρου στα βιολογικά δείγματα δεν επέτρεψαν τη χρήση της σύνθεσης των ειδών ως ανεξάρτητη μεταβλητή στην ανάπτυξη των μοντέλων. Για τον ίδιο λόγο χρησιμοποιήθηκε η ακουστική πυκνότητα ως εξαρτημένη μεταβλητή αντί της βιομάζας, ώστε να αποφευχθεί η διάδοση της μεταβλητότητας από την ακουστική πυκνότητα στην εκτίμηση της βιομάζας λόγω της δειγματοληψίας. Ως αποτέλεσμα οι χάρτες πρόβλεψης αντανακλούν την κατανομή του συνόλου των μικρών πελαγικών ειδών στην περιοχή και κυρίως του γαύρου και της σαρδέλας. Η ανάπτυξη των ΜΚΕ, η επιλογή των ανεξάρτητων μεταβλητών και η δημιουργία των χαρτών πρόβλεψης της κατανομής έγιναν με άξονα την πολυ-ειδική φύση των δεδομένων με συνέπεια την κοινή χρήση χαρακτηριστικών του κύκλου ζωής των ειδών στις αναλύσεις, όπως είναι για παράδειγμα η συσχέτιση τους με αναβλύσεις. Σύμφωνα με τους Stergiou & Lascaratos (1997) η κατανομή των προαναφερθέντων ειδών επηρεάζεται από περιβαλλοντικές παραμέτρους, την αλιευτική πίεση και τον δια-ειδικό και ενδο-ειδικό ανταγωνισμό. Η χρήση των μικρών πελαγικών ως σύνολο αντί του κάθε είδους χωριστά, εμπεριέχει την πληροφορία του δια-ειδικού ανταγωνισμού μειώνοντας έτσι τους βιοτικούς παράγοντες που επηρεάζουν την κατανομή των ειδών. Από την άλλη, τα ΜΚΕ και οι χάρτες κατανομής που εξάγονται δεν ενδείκνυται για τον προσδιορισμό σχέσεων συγκεκριμένων ειδών με περιβαλλοντικούς παράγοντες, αλλά αξιοποιούνται τα κοινά χαρακτηριστικά των μικρών πελαγικών ειδών, όπως η ικανότητα τους να σχηματίζουν χωρικά πρότυπα που σχετίζονται με συγκεκριμένες περιβαλλοντικές μεταβλητές και αντανακλούν ωκεανογραφικά χαρακτηριστικά, όπως αναβλύσεις και κυκλώνες.

4.1 Απόδοση μοντέλων αφθονίας

Μοντέλα παλινδρόμησης

Στην κατηγορία των Γενικευμένων Προσθετικών Μοντέλων επελέγησαν τρία μοντέλα για τη διαδικασία της σύγκρισης. Τα τρία μοντέλα είναι ένθετα, δηλαδή εμπεριέχουν τις ίδιες παραμέτρους και διαφοροποιούνται ως προς μια παράμετρο σε σχέση με το αρχικό (GAM1). Η επιπλέον μεταβλητή του GAM2 είναι ο χρονικός παράγοντας (DDND), που χωρίζει τα δειγματοληπτικά δεδομένα σε μέρα-σούρουπο-νύχτα-αυγή. Πάνω στο GAM2 αναπτύχθηκε το GAMM1 που περιέχει επιπλέον ένα παράγοντα για τη μοντελοποίηση της χωρικής αυτοσυσχέτισης των δεδομένων. Η σύγκριση έδειξε ότι το GAM2 αποδίδει καλύτερα από τα GAM1 και GAMM1 σε όλα τα επίπεδα. Το GAM2 παρουσίασε τα καλύτερα ποιοτικά χαρακτηριστικά (χαμηλότερο AIC, μεγαλύτερη εξηγούμενη απόκλιση) από τη διαδικασία επιλογής των μοντέλων, γεγονός που τελικά αντικατοπτρίζεται και στην ικανότητα πρόβλεψης του GAM2 σε όλα τα σύνολα δεδομένων που δοκιμάστηκε. Τελικά το GAM2 είναι το καταλληλότερο μεταξύ των γενικευμένων προσθετικών μοντέλων ως προς την προσαρμογή του στα δεδομένα εκπαίδευσης, την πρόβλεψη ενός ανεξάρτητου σετ δεδομένων καθώς και την πρόβλεψη ενός σετ που καλύπτει όλη την περιοχή μελέτης. Το GAMM1 αποδίδει σχεδόν όμοια με το GAM2 όμως η αύξηση της πολυπλοκότητας του με την ενσωμάτωση ενός παράγοντα αυτοσυσχέτισης δεν πέτυχε να βελτιώσει την απόδοση του GAMM1. Αυτό μπορεί να αποδοθεί στην πολύ χαμηλή αυτοσυσχέτιση των αρχικών δεδομένων, όπως έδειξαν τα ανάλογα τεστ, την οποία δεν μπόρεσε να εξηγήσει το GAMM1. Τέλος το GAM1 που επιλέχθηκε ως το απλούστερο μοντέλο που διατηρεί την ιδιότητα της γενικότητας, δεν απέδωσε ανάλογα στην πρόβλεψη του ανεξάρτητου συνόλου.

Η γενικότητα, ο ρεαλισμός και η ακρίβεια (generality, reality, precision) είναι οι ιδιότητες που οφείλει να έχει ένα ιδανικό στατιστικό μοντέλο, όμως μόνο δύο από τις τρεις μπορούν να επιτευχθούν κάθε φορά (Levins, 1966). Αν και το GAM1 είναι το μοντέλο που εστιάζει στη γενικότητα εις βάρος της ακρίβειας, ώστε να μπορεί να εφαρμοστεί σε μεγαλύτερο εύρος χωρικών και χρονικών προβλέψεων, δεν κατάφερε να αποδώσει ανάλογα με το GAM2. Το τελευταίο αποδείχθηκε πιο ικανό στην εξήγηση της διακύμανσης των τιμών των ακουστικών δεδομένων επιδεικνύοντας μεγαλύτερο ρεαλισμό και ακρίβεια. Από την άλλη το GAMM1 σχεδιάστηκε ώστε να διαχειριστεί την υφιστάμενη χωρική αυτοσυσχέτιση των δεδομένων (Keitt *et al.*, 2002), η οποία μπορεί να οδηγήσει σε σφάλματα κατά τη μοντελοποίηση ακουστικών δεδομένων, χωρίς όμως να καταφέρει να έχει την ακρίβεια πρόβλεψης και την απόδοση εφαρμογής του GAM2.

Τα GAMs είναι από τις πλέον διαδεδομένες τεχνικές μοντελοποίησης της κατανομής ειδών. Οι Giannoulaki *et al.* (2011) εφάρμοσαν τα GAMs σε τρεις περιοχές της Μεσογείου για να προβλέψουν την κατανομή των ανώριμων ατόμων της σαρδέλας από ακουστικά δεδομένα παρουσίας-απουσίας. Τα μοντέλα που αναπτύχθηκαν εμφάνισαν υψηλές τιμές ως προς την ικανότητα ορθής πρόβλεψης. Οι Maravelias *et al.* (2007) χρησιμοποίησαν τα GAMs για να προσδιορίσουν τα πιθανά ενδιαίτηματα του μπαλά (*Dentex maroccanus*), αλλά και να διερευνήσουν σχέσεις περιβαλλοντικές και εποχιακές συγκεντρώσεις του είδους με την κατάληψη ενδιαιτημάτων. Οι Maravelias & Reid (1997) ήταν από τους πρώτους που εφάρμοσαν τα Γενικευμένα προσθετικά μοντέλα στη θαλάσσια βιολογία για να διερευνήσουν τις σχέσεις ωκεανογραφικών δεδομένων και ζωοπλαγκτού με την αφθονία της ρέγγας. Οι Redfern *et al.* (2006) χρησιμοποίησαν γενικευμένα προσθετικά μεικτά μοντέλα για τον προσδιορισμό των ενδιαιτημάτων θαλάσσιων θηλαστικών, στα πλαίσια μιας ευρύτερης σύγκρισης μεθόδων κατάλληλων για την πρόβλεψη της κατανομής τους, με σκοπό να διαχειριστούν τη χωρική αυτοσυσχέτιση των δεδομένων.

Kriging

Οι γεωστατιστικές και άλλες μέθοδοι παρεμβολής αν και έχουν χρησιμοποιηθεί ευρέως σε άλλες επιστήμες, έχουν περιορισμένες εφαρμογές στο χώρο της μοντελοποίησης κατανομής ειδών (Franklin, 2009; Segurado & Araujo, 2004). Η θεωρία των αλγορίθμων των γεωστατιστικών μεθόδων βασίζεται στις χωρικές σχέσεις των δειγματοληπτικών τιμών και δεν μπορεί να αξιοποιήσει τα πρότυπα των περιβαλλοντικών παραμέτρων για την παρεμβολή των προβλέψεων. Παράλληλα, οι γεωστατιστικές μέθοδοι είναι οι μόνες μεταξύ αυτών που αναπτύχθηκαν στην διατριβή, που δεν βασίζονται στο θεμελιώδες θεώρημα (Εικ. 2.2) και δεν δύνανται να αναδείξουν σχέσεις ειδών-περιβάλλοντος. Ωστόσο, η ικανότητάς τους να δημιουργούν χάρτες κατανομής από δειγματοληπτικά δεδομένα επιβάλλει τη σύγκρισή τους με άλλες τεχνικές, ώστε να αξιολογηθούν οι εφαρμογές τους στο χώρο των ΜΚΕ. Τα μοντέλα που αναπτύχθηκαν με αλγόριθμους kriging παρουσιάζουν μεγάλες διαφορές μεταξύ τους. Το DefKrig, το οποίο αναπτύχθηκε ώστε να ελαχιστοποιούνται τα σφάλματα σε σχέση με τα δεδομένα εκπαίδευσης, αποδίδει καλύτερα στη διαδικασία βαθμονόμησης των μοντέλων και στην ανάλυση των σφαλμάτων. Από την άλλη το Krig το οποίο προβλέπει τιμές κοντά στη διασπορά και την τυπική απόκλιση των αρχικών ακουστικών δεδομένων, αναγεννά ένα πιο ρεαλιστικό χάρτη πιθανοτήτων από το DefKrig. Ο παραγόμενος χάρτης παρουσιάζει σημαντική ετερογένεια στον προσδιορισμό των ενδιαιτημάτων, αν και εμφανίζει σαφή τάση να υπερ-προσαρμόζεται στα δεδομένα εκπαίδευσης.

Από τις πρώτες εφαρμογές γεωστατιστικών μεθόδων σε ακουστικά δεδομένα είναι αυτή των Maravelias *et al.* (1996) για τη ρέγγα στη Β. Θάλασσα, που αποσκοπεί στη χαρτογράφηση της κατανομής του είδους. Οι Georgakarakos & Kitsiou (2008) χρησιμοποίησαν γεωστατιστικές μεθόδους για τη χαρτογράφηση της

κατανομής αφθονίας μικρών πελαγικών ειδών, αξιοποιώντας όμως μια γεωστατιστική προσέγγιση που λαμβάνει υπόψη περιβαλλοντικές παραμέτρους (Co-Kriging). Τα αποτελέσματα της εργασίας τους είναι ανάλογα με τα αποτελέσματα του μοντέλου DefKrig.

Νευρωνικά Δίκτυα

Τα συνδεδετικά νευρωνικά δίκτυα αποτελούν μια εξελιγμένη προσέγγιση κλασικών νευρωνικών δικτύων, όπου συνδυάζονται δύο αλγόριθμοι για τη βελτιστοποίηση της πρόβλεψης. Η βελτιωμένη ικανότητα πρόβλεψης τους σε εφαρμογή τους στη Χημεία για τον προσδιορισμό χημικών δεσμών, οδήγησε στη δοκιμή της μεθόδου στο χώρο της μοντελοποίησης κατανομής ειδών. Το ANNE είναι ένα μοντέλο που δημιουργείται από τη στάθμιση ενός συνόλου νευρωνικών δικτύων, ενώ το ASNN το οποίο ανήκει στα συνδεδετικά νευρωνικά δίκτυα συνδυάζει το ANNE με έναν επιπλέον αλγόριθμο για την περαιτέρω ταξινόμηση των αποτελεσμάτων του ANNE. Τα ASNN και ANNE αποδίδουν σχεδόν το ίδιο σε όλες τις δοκιμές. Ωστόσο τα ASNN εμφανίζουν λίγο καλύτερη προσαρμογή στα δεδομένα εκπαίδευσης, ενώ τα ANNE παρέχουν καλύτερες προβλέψεις στο σύνολο επαλήθευσης, όσον αφορά τα δεδομένα αφθονίας. Στη συγκεκριμένη εφαρμογή και σε αντίθεση με τις αρχικές προσδοκίες η αυξημένη ικανότητα πρόβλεψης των ASNN σε σχέση με τα ANNE δεν επαληθεύτηκε, γεγονός που θα μπορούσε να αποδοθεί στο μεγάλο εύρος διασποράς των ακουστικών δεδομένων.

Τα χαρακτηριστικά των μοντέλων δείχνουν ότι το ASNN είναι αποτελεσματικότερο στην ακρίβεια εφαρμογής των δεδομένων από το ANNE. Ένα ιδιαίτερο χαρακτηριστικό των ASNN είναι η ικανότητα τους να βελτιώνουν την εφαρμογή των ANNE αναλύοντας περαιτέρω τα αποτελέσματα τους και ενσωματώνοντας σε αυτά τις συσχετίσεις μεταξύ των τιμών στη διάσταση πρόβλεψης του μοντέλου (Tetko, 2002b). Αυτή η καινοτόμος προσέγγιση παρέχει μια πιο ολοκληρωμένη προσομοίωση της σχέσης περιβάλλοντος-ειδών, αφού αποτυπώνει τόσο τις τοπικές όσο και τις γενικές τάσεις των δεδομένων εκπαίδευσης (Tetko, 2002a). Όταν νέα δεδομένα είναι διαθέσιμα (όπως τα δεδομένα επαλήθευσης στη συγκεκριμένη περίπτωση) τα ASNN μπορούν να βελτιώσουν την ικανότητα πρόβλεψης τους και να παρέχουν μια λογική εκτίμηση του "άγνωστου" συνόλου χωρίς να απαιτείται επανεκπαίδευση των νευρωνικών (Tetko, 2002a). Στη συγκεκριμένη διατριβή και εφόσον δεν υπάρχουν ενδείξεις υπερ- ή υπό-εκτίμησης των δεδομένων εκπαίδευσης, τα ASNN μπορούν να χαρακτηριστούν από γενικότητα ενώ διατηρούν την ικανότητα εντοπισμού της τοπικής διακύμανσης των δεδομένων. Αυτή η ικανότητα αποδίδεται στο συνδυασμό ενός τεχνητού εμπρόσθια τροφοδοτούμενου νευρωνικού δικτύου, χωρίς ικανότητα απομνημόνευσης, και ενός αλγόριθμου k-nearest neighbours εφαρμοσμένου με έναν τύπο παλινδρόμησης (Parzen-window), που έχει δυνατότητα απομνημόνευσης (Tetko 2002a). Τα νευρωνικά δίκτυα χαρακτηρίζονται ως σφαιρικά (global) μοντέλα ενώ οι άλλες δύο προσεγγίσεις χαρακτηρίζονται ως τοπικά (local) μοντέλα (Lawrence *et al.*, 1996).

Γενικότερα, τα νευρωνικά δίκτυα χαρακτηρίζονται ως "μαύρα κουτιά", λόγω της αδυναμίας του χρήστη να επιτηρεί τις ενδιάμεσες λειτουργίες των νευρωνικών ως το τελικό αποτέλεσμα. Τα παράγωγα των ASNN, σε συνδυασμό με προσεγγίσεις όπως η "evaluation strip" που έχει προταθεί από τους Potts & Elith (2006), επιτρέπουν στον χρήστη την επιθεώρηση των επιδράσεων των εξαρτημένων μεταβλητών στην ανεξάρτητη. Με τον τρόπο αυτό, τα ASNN αποκτούν περισσότερο πληροφοριακό ρόλο και λιγότερο ρόλο "μαύρου κουτιού". Παρά τη δυναμική που εμφανίζουν τα νευρωνικά δίκτυα στη μοντελοποίηση βιολογικών και περιβαλλοντικών δεδομένων, οι εφαρμογές τους είναι περιορισμένες. Αυτό μπορεί να αποδοθεί στις υψηλές απαιτήσεις χειρισμού που απαιτούν κατά την ανάπτυξη τους, ώστε να αξιοποιηθούν πλήρως οι δυνατότητες τους (Franklin, 2009) και στη δυσκολία ανάδειξης σχέσεων ειδών-περιβάλλοντος ή γενικότερα τη δυσκολία ερμηνείας των

αποτελεσμάτων τους. Χαρακτηριστική εφαρμογή των νευρωνικών δικτύων στο χώρο της οικολογικής μοντελοποίησης είναι η εργασία των Brosse *et al.* (1999) για την εκτίμηση της αφθονίας και κατανομής ψαριών στην παράκτια περιοχή λιμναίου οικοσυστήματος.

Μοντέλο βελτιωμένης πρόβλεψης

Το σύνθετο μοντέλο βελτιωμένης πρόβλεψης (AverMod) δημιουργήθηκε αξιοποιώντας τις προβλέψεις των παραπάνω προσεγγίσεων και αφομοιώνοντας αυτές με προσθετικό τρόπο μέσω ενός νέου GAM. Η λογική δημιουργίας του είναι ανάλογη με τα ASNN αφού αξιοποιεί την ικανότητα ενός μοντέλου για τοπική πρόβλεψη (Krig) και ενός με πολύ δυνατή πρόβλεψη σε τοπικό αλλά και γενικό επίπεδο (ASNN). Το AverMod είναι η προσέγγιση με την καλύτερη βαθμονόμηση και έχει σημαντικές ομοιότητες με τα δεδομένα εκπαίδευσης καθώς και την καλύτερη ικανότητα πρόβλεψης, χωρίς να εμφανίζει τάσεις υπερ-προσαρμογής στα δεδομένα εκπαίδευσης. Κάθε προσέγγιση μοντελοποίησης στοχεύει στην ερμηνεία ενός ποσοστού της διακύμανσης των δειγματοληπτικών δεδομένων. Ο συνδυασμός των αποτελεσμάτων δύο προσεγγίσεων μοντελοποίησης δύναται να παρέχει ένα μοντέλο με βελτιωμένη ικανότητα ερμηνείας της διακύμανσης των δεδομένων. Επιπρόσθετα, το AverMod δείχνει ανάλογη αύξηση της ακρίβειας πρόβλεψης. Τα πρότυπα κατανομής των ειδών, όπως αποτυπώνονται στους καννάβους πρόβλεψης, επιβεβαιώνονται και από άλλες δημοσιεύσεις που αφορούν την ίδια περιοχή, από ωκεανογραφικά χαρακτηριστικά που σχετίζονται με την παρουσία των ειδών καθώς και από την προτίμηση ενδιαιτημάτων βάση του κύκλου-ζωής τους.

Οι Elith & Leathwick (2009) έδειξαν ότι η πραγματική κατανομή ειδών τοποθετείται τόσο στον περιβαλλοντικό όσο και στο γεωγραφικό χώρο. Σε αυτή τη διατριβή, το μοντέλο Krig φαίνεται να εξηγεί τη διακύμανση των δειγματοληπτικών δεδομένων βασίζόμενο σχεδόν αποκλειστικά στο γεωγραφικό χώρο, ενώ το ASNN αναπτύχθηκε αξιοποιώντας επί το πλείστον την περιβαλλοντική διάσταση. Με άλλα λόγια, στο Kriging χρησιμοποιείται η υπάρχουσα αυτοσυσχέτιση, ενώ στα άλλα μοντέλα επιπλέον με την αυτοσυσχέτιση χρησιμοποιείται και η ετεροσυσχέτιση με περιβαλλοντικές παραμέτρους. Λαμβάνοντας υπόψη τα παραπάνω, η αυξημένη ακρίβεια του AverMod θα μπορούσε να αποδοθεί στο συνδυασμό ενός "γεωγραφικού" (Krig) και ενός "περιβαλλοντικού" (ASNN) μοντέλου. Ο συνδυασμός των δύο οδήγησε στη γένεση ενός υβριδικού μοντέλου, του οποίου ο χώρος πρόβλεψης εμπεριέχει ταυτόχρονα μια περιβαλλοντική και μια γεωγραφική διάσταση.

4.1.1 Σύγκριση και αξιολόγηση των τεχνικών μοντελοποίησης δεδομένων αφθονίας

Το εύρος των τεχνικών επαλήθευσης και αξιολόγησης των προβλέψεων κατανομής ειδών οριοθετείται από τη χρήση τελείως ανεξάρτητων δεδομένων έως τις τεχνικές διασταυρωτής επικύρωσης (cross-validation) και επαναδειγματοληψίας (bootstrapping) που χρησιμοποιούν τα ίδια δεδομένα βάσει των οποίων έχει γίνει η εκπαίδευση των μοντέλων. Οι βιβλιογραφικές αναφορές στον τομέα των τεχνικών επαλήθευσης και επικύρωσης δεν φαίνεται να καταλήγουν σε μια αποκλειστικά ενδεδειγμένη λύση, αλλά τεκμηριώνουν την κάθε τεχνική κατά περίπτωση. Σύμφωνα με τους Elith *et al.* (2006) ένα βήμα για τη βελτίωση της αξιολόγησης των προβλέψεων κατανομής ειδών είναι η χρήση καλά δομημένων ανεξάρτητων δεδομένων. Από την άλλη, οι Lehmann *et al.* (2002) υποστηρίζουν ότι η χρήση τελείως ανεξάρτητων δεδομένων εμπεριέχει τον κίνδυνο της σύγκρισης διαφορετικών δειγματοληπτικών στρατηγικών αντί της αξιολόγησης των προβλέψεων. Σύμφωνα με τους Boyce *et al.* (2002) ο καλύτερος τρόπος αξιολόγησης δεδομένων παρουσίας είναι η δέσμευση δεδομένων από το αρχικό σύνολο (καταμερισμός k-Fold) για τον έλεγχο των προβλέψεων του μοντέλου ή για τη σύγκριση

της πρόβλεψης από μοντέλα που έχουν αναπτυχθεί σε διαφορετικές περιόδους ή περιοχές (δειγματοληψία προοπτικής). Σύμφωνα με τους Lehmann *et al.* (2002) οι τεχνικές cross-validation και bootstrapping είναι πιο πρακτικές γιατί δημιουργούν σχετικά ανεξάρτητα, τυχαία σύνολα δεδομένων και επιτρέπουν τη χρήση όλων των διαθέσιμων πληροφοριών τόσο στην εκπαίδευση των μοντέλων όσο και στην αξιολόγηση τους. Ως αντίλογος στην παραπάνω πρόταση εμφανίζεται η αναφορά ότι οι τεχνικές διασταυρούμενης επικύρωσης και επαναδειγματοληψίας υπερεκτιμούν την απόδοση των υπό αξιολόγηση μοντέλων, ειδικότερα στην περίπτωση που τα δεδομένα εμπεριέχουν χωρικά πρότυπα και χωρική αυτοσυσχέτιση. Οι απόψεις περί των τεχνικών αξιολόγησης των εκπαιδευμένων μοντέλων είναι τόσο συγκεχυμένες ώστε οι Elith και Leathwick (2009) στην επισκόπηση τους για τα μοντέλα κατανομής ειδών τονίζουν την έλλειψη μια συγκεκριμένης και αποτελεσματικής μεθοδολογίας. Στη συγκεκριμένη διατριβή η μέθοδος αξιολόγησης επιλέχθηκε, όπως έχει περιγραφεί προηγουμένως, με στόχο την αμερόληπτη και ουσιαστική κρίση των ΜΚΕ. Τέλος, οι διάφορες προσεγγίσεις μοντελοποίησης της κατανομής των ειδών αξιολογούνται με τον ίδιο ακριβώς τρόπο, ώστε να διασφαλιστεί η αντικειμενικότητα της σύγκρισης.

Συγκεκριμένα, το σύνολο επαλήθευσης δημιουργήθηκε από τα αρχικά δειγματοληπτικά δεδομένα. Αυτή η προσέγγιση παρουσιάζει ομοιότητες με τη διασταυρωτή επικύρωση, αλλά προκειμένου να μην υπερ-εκτιμηθούν τα ΜΚΕ η επιλογή του καταλληλότερου συνόλου επαλήθευσης πραγματοποιήθηκε μετά από μια σειρά διερευνητικών διαδικασιών στα αρχικά δεδομένα, οι οποίες παρουσιάζονται αναλυτικά στη συνέχεια. Τελικά το σύνολο δεδομένων επαλήθευσης χρησιμοποιήθηκε ως ανεξάρτητο σύνολο, άγνωστο στα ΜΚΕ αφού δε συμμετείχε στην εκπαίδευση τους. Η επιλογή του συνόλου επαλήθευσης περιέχει διαδοχικές μετρήσεις, οι οποίες αποτελούν μέρος των αρχικών δεδομένων, όπου δύο σημεία έχουν μεγαλύτερη απόσταση από το εύρος αυτοσυσχέτισης. Η απόσταση αυτή επιλέχθηκε σύμφωνα με εμπειρικά βαριογράμματα τόσο του συνόλου επαλήθευσης όσο και των σφαλμάτων των προβλέψεων, που έδειξαν πολύ χαμηλή αυτοσυσχέτιση ακόμα και σε αποστάσεις μικρότερες των 5 nm. Ανάλογες χωρικές δομές εμφανίστηκαν και σε προηγούμενες δειγματοληψίες στη περιοχή μελέτης (Οκτώβρης 1996, Μάιος 1997) παρά τις διαφορετικές εποχιακές συνθήκες (Georgakarakos & Kitsiou, 2008). Το σύνολο επαλήθευσης DS3 τελικά αποτελεί ένα ανεξάρτητο, "άγνωστο" για τα μοντέλα, σύνολο τιμών, που από βιολογικής άποψης διατηρεί την δειγματοληπτική πληροφορία.

Ως γενικό συμπέρασμα θα μπορούσε να αναφερθεί ότι τα ASNN είναι πιο αποτελεσματικά από τις άλλες τεχνικές στο να μοντελοποιούν τα ακουστικά δεδομένα ενώ τα GAMs και κυρίως το GAM2 είναι πιο ευέλικτα και ακριβή στην πρόβλεψη του ανεξάρτητου συνόλου επαλήθευσης. Οι μέθοδοι Kriging από την άλλη, αποτελούν ένα χρήσιμο εργαλείο της πρόβλεψης κατανομής των ειδών. Θα πρέπει να σημειωθεί ότι η πρόβλεψη της κατανομής των ειδών με τη χρήση χωρικών συσχετίσεων των δειγμάτων φαίνεται να μην είναι τόσο ισχυρή όσο η χρήση περιβαλλοντικών παραμέτρων, άμεσα συσχετιζόμενων με την βιολογία των ειδών. Άλλωστε, οι Petitgas *et al.* (2001) σε μια συγκριτική μελέτη με ακουστικά δεδομένα από πέντε διαφορετικές περιοχές της Ευρώπης έδειξαν ότι η χωρική οργάνωση των αποθεμάτων εξαρτάται περισσότερο από περιβαλλοντικές παραμέτρους παρά από την αφθονία τους. Σε κάθε περίπτωση, τα ΜΚΕ είναι "ευαίσθητα" στην ποιότητα και χωρική ανάλυση των δεδομένων τα οποία καθορίζουν και την ποιότητα των τελικών αποτελεσμάτων. Μεταξύ των βιολογικών δεδομένων, τα δεδομένα ακουστικής συχνότητας όταν χρησιμοποιούνται ως δεδομένα αφθονίας αποτελούν ιδιαίτερη πρόκληση στις μελέτες μοντελοποίησης κατανομής ειδών και στα μοντέλα ειδικότερα, καθώς χαρακτηρίζονται από σημαντική διασπορά τιμών και μεγάλες τυπικές αποκλίσεις. Οι περισσότερες τεχνικές μοντελοποίησης χρησιμοποιούν εξομαλυντές για την προσομοίωση των δεδομένων (Spline για GAMs, Parzen-window για ASNN) και για το λόγο αυτό παρουσιάζουν την τάση να υπο-εκτιμούν τις υψηλές τιμές και

να υπερ-εκτιμούν τις χαμηλές τιμές των ακουστικών δεδομένων. Ως αποτέλεσμα το συνολικό σφάλμα πρόβλεψης μειώνεται υποδηλώνοντας πιο ακριβή πρόβλεψη, αλλά η ακρίβεια της πρόβλεψης αναφορικά με τις τοπικές διακυμάνσεις υστερεί. Η χρήση υψηλής ανάλυσης πληροφοριακών επεξηγηματικών μεταβλητών δύναται εν μέρει να ξεπεράσει το παραπάνω πρόβλημα και να αναδείξει πιο ακριβή διακύμανση της κατανομής των ειδών και κατά συνέπεια την πιθανή ετερογένεια των ενδιαιτημάτων.

Τα ASNN και ANNE έδειξαν ιδιαίτερη δυναμική στην ακριβή μοντελοποίηση δεδομένων, που προέρχονται από κανονική δειγματοληψία και καλύπτουν την περιοχή μελέτης σε ένα κανονικό σχέδιο διατομών. Και οι δύο μέθοδοι αποδίδουν ρεαλιστικά τα δειγματοληπτικά δεδομένα και έχουν την ικανότητα να διακρίνουν τόσο την τοπική όσο και τη γενική διακύμανση των δεδομένων, χωρίς να υπερ-προσαρμόζονται στα δεδομένα εκπαίδευσης. Τα χαρακτηριστικά και η απόδοση των ASNN και ANNE τα καθιστούν ικανά για εφαρμογές που απαιτούν λεπτομερή και ρεαλιστική απεικόνιση της κατανομής ειδών που προέρχεται από δειγματοληπτικά δεδομένα, όπως η διαχείριση, ο σχεδιασμός προστατευόμενων περιοχών και η επιλογή κατάλληλης στρατηγικής δειγματοληψίας. Τα αποτελέσματα, επίσης, αναδεικνύουν την ικανότητα πρόβλεψης των GAMs, ειδικά όταν αυτή απαιτείται σε εκτεταμένη χωρική κλίμακα, πέρα από την περιοχή δειγματοληψίας. Επομένως, τα GAMs ενδείκνυται περισσότερο στις εφαρμογές, όπου απαιτείται ο προσδιορισμός πιθανών ενδιαιτημάτων ειδών σε ευρύτερη της δειγματοληψίας κλίμακα, παρά στην ταυτοποίηση της ετερογένειας των ενδιαιτημάτων σε πολύ υψηλή χωρική ανάλυση. Επιπλέον, τα GAMs, με την εποπτική παρουσίαση των αποτελεσμάτων τους, μπορούν να αναδείξουν καλύτερα τις σχέσεις ειδών-περιβάλλοντος από τις υπόλοιπες τεχνικές. Εν συνεχεία, διευκολύνεται η ανίχνευση και ερμηνεία των σχέσεων ειδών-περιβάλλοντος.

Το AverMod αποτέλεσε το πλέον ακριβές μοντέλο πρόβλεψης της κατανομής και θα μπορούσε να είναι η ιδανικότερη προσέγγιση για τις εφαρμογές που αναφέρθηκαν παραπάνω. Το μειονέκτημα αυτής της προσέγγισης έγκειται στο γεγονός ότι προαπαιτεί την ανάπτυξη διαφόρων άλλων προσεγγίσεων πριν το τελικό συνδυασμό τους για τον πιο ολοκληρωμένο προσδιορισμό της κατανομής των ειδών. Επιπλέον ο συνδυασμός των αποτελεσμάτων διαφορετικών μοντέλων πρέπει σε κάθε περίπτωση να εξετάζεται ως προς τα αρχικά δεδομένα και τις αρχικές προβλέψεις εφόσον δεν αποτελεί μονόδρομη διαδικασία και κρύβει κινδύνους υπερ-προσαρμογής των δεδομένων ή ενίσχυσης των σφαλμάτων των επιμέρους μοντέλων. Σε αυτή την εφαρμογή ο συνδυασμός των προβλέψεων πραγματοποιήθηκε προσθετικά με τη χρήση ενός GAM μοντέλου. Και άλλες μέθοδοι θα μπορούσαν να χρησιμοποιηθούν για αυτό το σκοπό, όπως τα ASNN που αποδεδειγμένα προσομοιώνουν ικανοποιητικά τα δεδομένα εκπαίδευσης. Σε κάθε περίπτωση η βελτιστοποίηση της πρόβλεψης κατανομής ειδών με τη διαδικασία που προτείνεται φαίνεται από τα αποτελέσματα να είναι πολλά υποσχόμενη.

4.2 Απόδοση μοντέλων που εκπαιδεύονται με δεδομένα παρουσίας-απουσίας ή μόνο-παρουσίας

Όπως έχει αναφερθεί οι γενικές κατηγορίες τύπων δεδομένων στα οποία εφαρμόζονται τα MKE μπορούν να διακριθούν σε αφθονίας, παρουσίας-απουσίας και μόνο-παρουσίας. MKE, όπως τα GAMs και τα ASNN μπορούν να αξιοποιήσουν όλους τους τύπους δεδομένων. Άλλα MKE, όπως τα BRT, τα GARP και τα MARS αξιοποιούν δεδομένα παρουσίας-απουσίας και άλλα όπως οι βιοκλιματικοί φάκελοι και τα SVMs χρησιμοποιούν δεδομένα μόνο-παρουσίας. Από την οπτική της εφαρμογής των MKE τα μοντέλα παρουσίας-απουσίας και μόνο-παρουσίας εμφανίζουν ουσιαστικές ομοιότητες στην ανάπτυξη τους, την επαλήθευση των μοντέλων, την αξιολόγηση τους και τις απεικονίσεις των αποτελεσμάτων τους. Για τους παραπάνω λόγους, τα

ΜΚΕ που εκπαιδεύονται με δεδομένα παρουσίας-απουσίας και μόνο-παρουσίας εξετάστηκαν μαζί και ενσωματώθηκαν σε μια ενιαία διαδικασία σύγκρισης. Οι δύο παραπάνω κατηγορίες ΜΚΕ, που βασίζονται σε δεδομένα διωνυμικών κατανομών διαφέρουν σε όλες τις διαδικασίες ανάπτυξης, επαλήθευσης και επικύρωσης από τα ΜΚΕ που εκπαιδεύονται από δεδομένα αφθονίας και βασίζονται σε πιο πολύπλοκες κατανομές. Οι διαφορές αυτές καθιστούν τα αποτελέσματα των ΜΚΕ αφθονίας μη ανάλογα με τις άλλες κατηγορίες και επομένως μη συγκρίσιμα.

4.2.1 Σύγκριση των ΜΚΕ για τα δεδομένα παρουσίας-απουσίας και μόνο-παρουσίας

Μεταξύ των τεχνικών μηχανικής εκπαίδευσης, τα SVM και MAXENT αξιοποιούν μόνο τα δεδομένα παρουσίας. Οι Brotons *et al.* (2004) έδειξαν ότι οι προβλέψεις που βασίζονται σε δεδομένα παρουσίας-απουσίας γενικά αποδίδουν καλύτερα σε σχέση με αυτές που βασίζονται σε δεδομένα μόνο-παρουσίας. Σύμφωνα με τους MacLeod *et al.*, (2008) τα μοντέλα μόνο-παρουσίας μπορούν να αποδώσουν ανάλογα με τα παρουσίας-απουσίας όταν η κάλυψη της δειγματοληψίας είναι οργανωμένη σε κανονικά διαστήματα και πραγματοποιείται σε αναλογικά μεγάλη έκταση. Όμως τα μοντέλα μόνο-παρουσίας δεν εμπεριέχουν μηχανισμούς για τον έλεγχο εσφαλμένων δειγμάτων. Στη συγκεκριμένη διατριβή το MAXENT αποδίδει λιγότερο καλά από τα άλλα μοντέλα μηχανικής εκμάθησης, ενώ το SVM έχει ανάλογη απόδοση με κάποια από τα μοντέλα παρουσίας-απουσίας.

Οι χάρτες πιθανοτήτων, που αντιστοιχούν στις τεχνικές μηχανικής μάθησης, παρουσιάζουν αξιοσημείωτες ομοιότητες, εντοπίζοντας υψηλές πιθανότητες παρουσίας των ειδών κοντά στις ακτές, κυρίως τις δυτικές καθώς και στην κεντρική και νότια περιοχή του χώρου μελέτης. Τα SVM-Nu και MAXENT όμως εμφανίζουν αισθητά μικρότερες πιθανότητες παρουσίας κατά μήκος των δυτικών ακτών συγκριτικά με τα άλλα μοντέλα. Επιπλέον, οι υψηλές πιθανότητες φαίνεται να επικαλύπτουν τις περιοχές δειγματοληψίας, γεγονός που υποδηλώνει πιθανή υπερ-προσαρμογή των δύο μοντέλων στα δεδομένα εκπαίδευσης. Όσον αφορά τα σφάλματα παράλειψης, τα δύο μοντέλα εμφανίζουν υψηλές τιμές γεγονός που δείχνει την αδυναμία τους να προβλέψουν ικανοποιητικά τις αρχικές τιμές. Το SVM-Nu αν και μοντέλο μόνο-παρουσίας εμφανίζει αξιοσημείωτη ακρίβεια προσαρμογής έχοντας το δεύτερο καλύτερο ROC-AUC και COR (μετά το BRT). Όμως η ακρίβεια πρόβλεψης του είναι σχετικά χαμηλή, όπως έδειξε το μέτριο ROC-AUC στο σετ επαλήθευσης και το υψηλό σφάλμα παράλειψης, αποδεικνύοντας ότι το SVM-Nu υπερ-προσαρμόζεται στα δεδομένα. Το SVM από την άλλη δεν εμφανίζει τάσεις υπερ-προσαρμογής των δεδομένων και χαρακτηρίζεται ως το πλέον αποδοτικό μεταξύ των τεχνικών που χρησιμοποιούν δεδομένα μόνο-παρουσίας, με απόδοση ανάλογη με τα μοντέλα παλινδρόμησης.

Τα BRT, ASNN και ANNE αποτελούν τα πιο αποτελεσματικά μοντέλα στα δεδομένα παρουσίας-απουσίας. Συγκεκριμένα, το BRT παρουσιάζει την καλύτερη απόδοση εφαρμογής, ενώ η ικανότητα πρόβλεψης του είναι σχετικά υψηλή σε σύγκριση με το σύνολο των μοντέλων. Το ASNN παρουσιάζει την υψηλότερη ικανότητα πρόβλεψης και αποδίδει ικανοποιητικά ως προς την απόδοση εφαρμογής (ROC-AUC 0.86). Ανάλογη είναι και η αποτελεσματικότητα του ANNE που ακολουθεί το ASNN, όπως αναμενόταν, εφόσον το ASNN αποτελεί μια βελτιωμένη έκδοση του ANNE (Tetko, 2002). Τέλος το ASNN επιτυγχάνει τιμές ROC-AUC, COR και MaxKappa κατά τι μεγαλύτερες από το δεύτερο καλύτερο μοντέλο που είναι το GAM.

Αναφορικά με τα μοντέλα παλινδρόμησης, τα GAM, GAMM και MARS παράγουν σχεδόν όμοιους χάρτες πιθανοτήτων, δεδομένων των ομοιοτήτων που παρουσιάζουν στην απόδοση εφαρμογής και στην ικανότητα πρόβλεψης. Οι ομοιότητες θα μπορούσαν να αποδοθούν στην κοινή στατιστική τους προέλευση. Η σύγκριση μεταξύ των μοντέλων παλινδρόμησης έδειξε ότι το GAM αποδίδει συνολικά καλύτερα από το GAMM, ενώ το

τελευταίο αποδίδει καλύτερα από το MARS. Συγκριτικά με τις άλλες μεθόδους μοντελοποίησης τα μοντέλα παλινδρόμησης μπορεί να μην έχουν επιτύχει τις υψηλότερες τιμές, όμως κατατάσσονται αρκετά υψηλά τόσο στην εφαρμογή τους στο σετ εκπαίδευσης όσο και στο σετ επαλήθευσης. Οι επιδόσεις τους δικαιολογούν την ευρεία χρήση των μοντέλων παλινδρόμησης έναντι των μοντέλων τύπου φακέλου, του GARP και MAXENT τα τελευταία χρόνια, καθώς εμφανίζουν σταθερότητα, ευελιξία και γενικότητα στην εφαρμογή τους.

Τα μοντέλα τύπου φακέλου απέτυχαν να προβλέψουν ικανοποιητικά την κατανομή των ειδών, επιτυγχάνοντας τις χαμηλότερες τιμές στα ROC-AUC, COR και MaxKappa. Χαρακτηρίζονται από υψηλές τιμές σφαλμάτων επιφόρτισης (1 για το EnvScore και 0.79 για το BIOCLIM) και όπως έδειξαν οι χάρτες πιθανοτήτων υπερ-προσαρμόζονται στα δεδομένα εκπαίδευσης. Αρχικά, οι προσεγγίσεις τύπου φακέλου αναπτύχθηκαν για να μοντελοποιήσουν την κατανομή χερσαίων ειδών από δεδομένα μουσείων Φυσικής Ιστορίας και πιθανόν δεν ενδεικνύονται για τη μοντελοποίηση ειδών με παρουσία σε υψηλή ανάλυση και συχνότητα, αλλά ούτε και για τον προσδιορισμό της ετερογένειας ενδιαιτημάτων.

Τα EnvDist και EnvDistChe αποδίδουν με σχετική ακρίβεια στην εφαρμογή των δεδομένων εκπαίδευσης (ROC-AUC 0.97 και για τα δύο). Υστερούν έναντι των BRT και SVM-Nu στις ROC-AUC τιμές, ενώ επιτυγχάνουν τις υψηλότερες τιμές MaxKappa (0.91) και από τις υψηλότερες COR τιμές (0.73 και 0.83, αντίστοιχα). Και τα δύο μοντέλα χαρακτηρίζονται από μηδενικό σφάλμα παράλειψης και το EnvDistChe έχει σχεδόν το μισό σφάλμα επιφόρτισης από το EnvDist (0.22 και 0.41, αντίστοιχα). Το μηδενικό σφάλμα παράλειψης βέβαια σε συνδυασμό με το χαμηλό σφάλμα επιφόρτισης υποδηλώνουν ότι το μοντέλο υπερ-προσαρμόζεται στα δεδομένα εκπαίδευσης. Η θεώρηση αυτή αποδεικνύεται και από τους χάρτες πρόβλεψης, όπου οι περιοχές υψηλής πιθανότητας συγκεντρώνονται γύρω από τις υψηλές τιμές του δείγματος. Και το EnvDist παρουσιάζει την τάση να υπερ-προσαρμόζεται στα δεδομένα αλλά σε μικρότερο βαθμό από EnvDistChe, γεγονός που επιβεβαιώνεται και από την ικανότητα πρόβλεψης των δύο μοντέλων. Δεδομένου ότι το EnvDistChe υπερ-προσαρμόζεται στα δεδομένα εκπαίδευσης, αδυνατεί να επιτύχει ακριβή πρόβλεψη στο ανεξάρτητο σετ δεδομένων, γεγονός που οδηγεί σε χαμηλότερες τιμές ROC -AUC, COR και MaxKappa από το EnvDist. Τα δύο προαναφερόμενα μοντέλα είναι λιγότερο αποδοτικά στην ικανότητα πρόβλεψης τους συγκριτικά με τα μοντέλα παλινδρόμησης καθώς και με τα περισσότερα μοντέλα μηχανικής εκμάθησης. Παρόλα αυτά τα σε σύγκριση με τις άλλες μεθόδους μοντελοποίησης και τα δύο μοντέλα από τα και από τα EnvDist και EnvDistChe αποδίδουν αναλογικά ικανοποιητικά συγκριτικά με τις υπόλοιπες μεθόδους που δεν υποστηρίζουν δεδομένα απουσίας, όχι όμως όσο το SVM.

Το ClimSpace αποτυγχάνει να προσαρμοστεί στα δεδομένα εκπαίδευσης ή να προβλέψει καλά το ανεξάρτητο σετ επαλήθευσης, ενώ έχει τις χειρότερες τιμές ROC-AUC, COR και MaxKappa. Κατά την εφαρμογή της τεχνικής αναπτύχθηκε ένας μεγάλος αριθμός μοντέλων εκ των οποίων το καλύτερο ποιοτικά επιλέχθηκε για τις διαδικασίες της σύγκρισης. Τελικά, αποδεικνύεται ότι το ClimSpace είναι ακατάλληλο για την πρόβλεψη κατανομής ειδών από ακουστικά δεδομένα.

Το GARP αποδίδει μέτρια τόσο στην προσαρμογή του στα δεδομένα εκπαίδευσης όσο και στην ικανότητα πρόβλεψης. Τα αποτελέσματα της μελέτης αλλά και η ανάλυση των σφαλμάτων δείχνουν ότι το GARP ανταποκρίνεται καλά στις επεξηγηματικές μεταβλητές που χρησιμοποιήθηκαν και που αντανακλούν τις περιοχές όπου οι περιβαλλοντικές συνθήκες είναι κατάλληλες για τη διατήρηση πληθυσμών. Ο καμβάς πρόβλεψης ωστόσο είναι αρκετά αδρός σε σύγκριση με τα άλλα μοντέλα. Η αδυναμία του GARP να παράγει περισσότερο λεπτομερείς χάρτες κατανομής το καθιστούν λιγότερο αποτελεσματικό σε σχέση με τις προσεγγίσεις που

υποστηρίζουν δεδομένα παρουσίας-απουσίας. Παρόλα αυτά τα αποτελέσματα του GARP συνολικά ταυτίζονται με τα χωρικά πρότυπα υψηλών συχνοτήτων που αναδεικνύουν τα πλέον ακριβή ΜΚΕ.

Η σύνδεση των ΜΚΕ με την αντίστοιχη οικολογική θεωρία του θώκου των ειδών (στην οποία στηρίζεται και το θεμελιώδες μοντέλο που περιγράφεται στην παράγραφο 2.3) οφείλει να είναι όσο το δυνατό εκτενέστερη στην (α) επιλογή επεξηγηματικών μεταβλητών, (β) την επισκόπηση αλληλεπιδράσεων μεταξύ των μεταβλητών, (γ) τη λειτουργία της εξαρτημένης μεταβλητής, (δ) την επιλογή του κατάλληλου τύπου μοντέλου και (ε) την ερμηνεία των χαρτών πρόβλεψης. Για παράδειγμα, έχει προταθεί ότι τα μοντέλα τύπου περιβαλλοντικού φακέλου (EnvScore και BIOCLIM σε αυτή τη μελέτη) που χρησιμοποιούν δεδομένα μόνο παρουσίας, τείνουν να απεικονίζουν την πιθανή κατανομή (κατάλληλο ενδιαίτημα), ενώ πιο σύνθετα μοντέλα (ASNN, BRT, GAM) που αξιοποιούν δεδομένα παρουσίας-απουσίας τείνουν να απεικονίζουν την πραγματική-ρεαλιστική κατανομή (κατελιμμένο ενδιαίτημα), άρα είναι πιο ικανά για παρεμβολή (Jiménez-Valverde *et al.*, 2008; Hirzel & Le Lay, 2008). Η παραπάνω διευκρίνιση πραγματοποιείται για να τονιστεί ότι οι μέθοδοι που αποτυγχάνουν να αποδώσουν την κατανομή των ατόμων των ειδών (π.χ. EnvScore και BIOCLIM) όσο άλλες, ενδέχεται να προορίζονται για άλλο σκοπό (π.χ. την περιγραφή του θεωρητικού θώκου) χωρίς να δύνανται να φτάσουν στον τελικό στόχο, δηλαδή την πραγματική κατανομή των ατόμων των ειδών, όπως φαίνεται στην Εικ. 2.2 της παραγράφου 2.3.

Στους κάρναβους πρόβλεψης που προκύπτουν από τα ΜΚΕ που διαχειρίζονται δεδομένα παρουσίας-απουσίας ή μόνο-παρουσίας εντοπίζονται δύο διακριτές περιοχές συγκέντρωσης πληθυσμών μικρών πελαγικών ειδών: (α) η δυτική ακτή από βορρά προς νότο, που χαρακτηρίζονται από την παρουσία μεγάλων ποταμών με σημαντικές εκροές και η ανατολική ακτή του Θερμαϊκού κόλπου και (β) η κεντρική περιοχή μελέτης, που σχετίζεται με την παρουσία κυκλωνικών σχηματισμών (Somarakis *et al.*, 2002). Οι παραπάνω περιοχές είναι πλούσιες σε θρεπτικά συστατικά και ευνοούν τη συγκέντρωση μικρών πελαγικών ειδών. Αντίστοιχα ήταν και τα αποτελέσματα των ΜΚΕ με δεδομένα αφθονίας. Όπως αναφέρθηκε και προηγουμένως, οι περιοχές αυτές είναι σε αντιστοιχία με τα αποτελέσματα άλλων μελετών για μικρά πελαγικά είδη στην περιοχή (Somarakis *et al.*, 2002; Giannoulaki *et al.*, 2008; Tsagarakis *et al.*, 2008). Οι συσχετίσεις μεταξύ των καννάβων πρόβλεψης της κατανομής των ειδών είναι σε συμφωνία με τις συσχετίσεις των προβλέψεων των ΜΚΕ στο σύνολο επαλήθευσης. Συγκεκριμένα, ΜΚΕ που αποδίδουν σχεδόν ανάλογα στην ικανότητα πρόβλεψης αναγεννού χάρτες πρόβλεψης που συσχετίζονται σημαντικά, όπως συμβαίνει π.χ. στο MARS και GAMM. Επίσης, οι κάρναβοι που προέρχονται από ΜΚΕ με κοινή στατιστική προέλευση, όπως τα μοντέλα παλινδρόμησης ή τα νευρωνικά δίκτυα παρουσιάζουν σημαντικές συσχετίσεις.

4.2.2 Αξιολόγηση των ΜΚΕ για τα δεδομένα παρουσίας-απουσίας και μόνο-παρουσίας

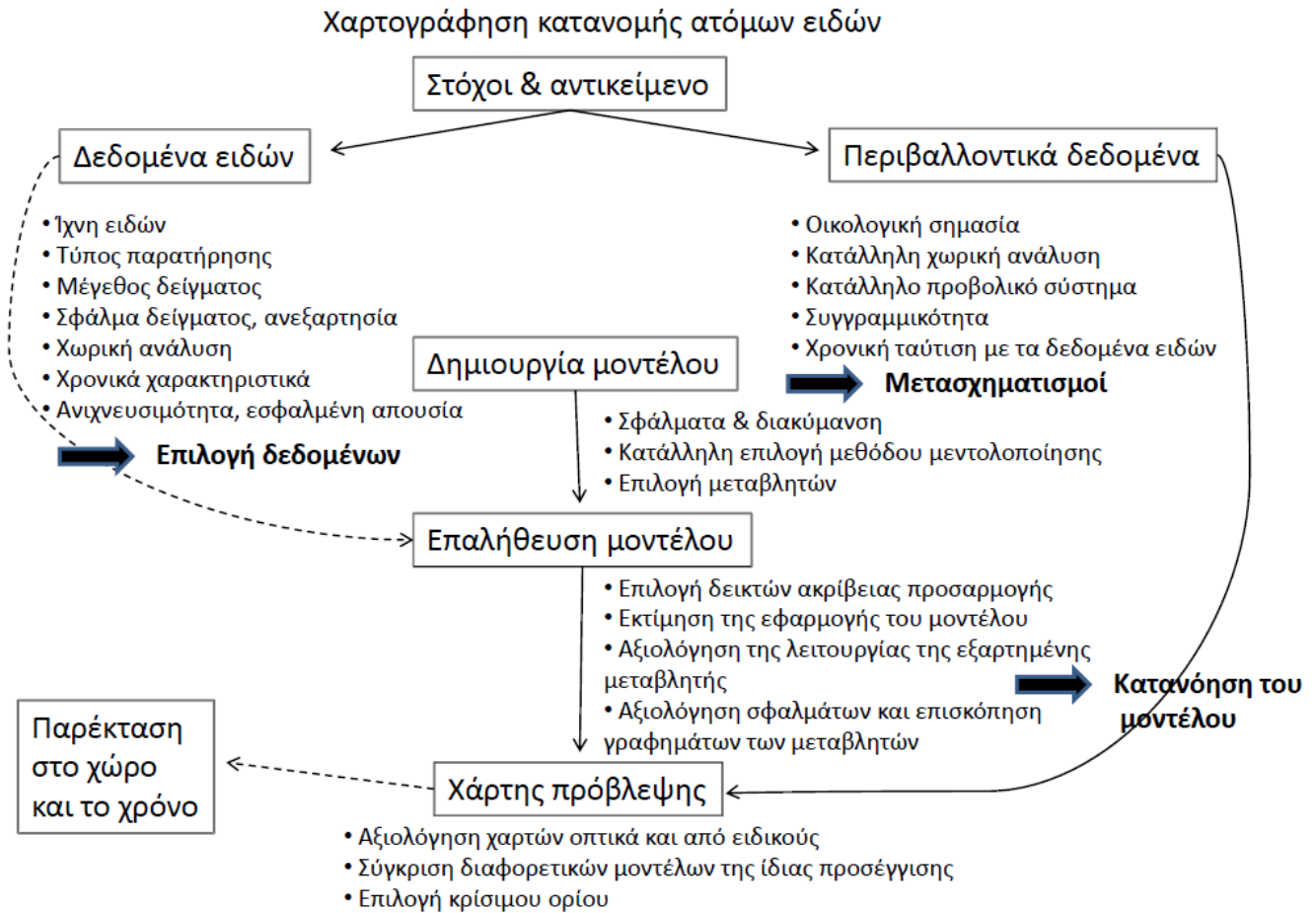
Η ποιότητα των μοντέλων κατανομής ειδών είναι ανάλογη των δεδομένων που χρησιμοποιούνται (Hirzel & Guisan, 2002). Επιπλέον, η απόδοση των ΜΚΕ εξαρτάται από τον αριθμό των δειγμάτων που χρησιμοποιούνται για την εκπαίδευση του μοντέλου. Διαφορετικού τύπου δεδομένα (π.χ. αφθονίας, παρουσίας-απουσίας, μόνο-παρουσίας) παράγουν ΜΚΕ σε διαφορετική κλίμακα απόδοσης (Elith *et al.*, 2006). Γενικά οι προβλέψεις που παράγονται από μοντέλα που χρησιμοποιούν δεδομένα παρουσίας-απουσίας είναι πιο ακριβείς από αυτές που παράγονται από μοντέλα που χρησιμοποιούν δεδομένα μόνο-παρουσίας (Brotans *et al.*, 2004). Αντιστοίχως, τα μοντέλα παρουσίας-απουσίας γενικά αποδίδουν καλύτερα από τα μοντέλα αφθονίας (Francis *et al.*, 2005). Τα μοντέλα μόνο-παρουσίας μπορούν να αποδώσουν εξίσου καλά με τα παρουσίας-απουσίας μόνο στην περίπτωση που η δειγματοληψία είναι κανονική και κατανεμημένη σε μεγάλο εύρος (MacLeod *et al.*, 2008). Θεωρητικά, τα

μοντέλα αφθονίας θα έπρεπε να είναι περισσότερο πληροφοριακά, όμως η φτώχη τους απόδοση στην πράξη σχετίζεται με το γεγονός ότι τα πραγματικά δεδομένα αφθονίας σπάνια χαρακτηρίζονται από κανονική κατανομή, παραβιάζοντας μια βασική προϋπόθεση μοντελοποίησης. Στην περίπτωση των δεδομένων παρουσίας-απουσίας, τα οποία έχουν διωνυμική κατανομή, οι προϋποθέσεις μοντελοποίησης ικανοποιούνται ευκολότερα. Επιπλέον, οι διαδικασίες αξιολόγησης των μοντέλων παρουσίας-απουσίας (ROC-AUC, Kappa, Confusion matrix) είναι καλύτερα ανεπτυγμένες και παρέχουν περισσότερη πληροφορία συγκριτικά με τις διαδικασίες αξιολόγησης των μοντέλων αφθονίας (k-fold cross-validation, models calibration, correlation). Αυτό αποδίδεται στην ευκολότερη ανάλυση των αποτελεσμάτων των πρώτων, που έχουν διωνυμική κατανομή, σε σύγκριση με τα μοντέλα αφθονίας που έχουν πιο σύνθετες κατανομές (Gaussian, Poisson, κ.α.).

Μελέτες μοντελοποίησης δεδομένων παρουσίας-απουσίας έδειξαν ότι οι μη γραμμικές τεχνικές (π.χ. GAM, ANN και MARS) είναι συγκρίσιμες μεταξύ τους στην προσαρμογή των δεδομένων, ενώ στο σύνολο τους είναι ανώτερες στην ικανότητα πρόβλεψης από τις τεχνικές δενδροειδών αποφάσεων (Ferrier & Watson, 1997; Elith & Burgman, 2002, Moisen & Frescino, 2002, Muñoz & Felicísimo, 2004; Segurado & Araujo, 2004). Στη συγκεκριμένη διατριβή επιβεβαιώνονται τα παραπάνω με την σχεδόν όμοια απόδοση των GAM και MARS. Αντιθέτως, τα ANNE και πολύ περισσότερο τα ASNN, έδειξαν υψηλότερη ικανότητα πρόβλεψης όχι μόνο σε σχέση με τα παραδοσιακά νευρωνικά δίκτυα, αλλά ακόμα και σε σχέση με άλλες ευρέως διαδεδομένες τεχνικές (π.χ. GAMs). Οι Elith *et al.* (2006) αξιολόγησαν την ικανότητα πρόβλεψης έντεκα τεχνικών μοντελοποίησης και δεκαέξι προσεγγίσεων χρησιμοποιώντας δεδομένα μόνο-παρουσίας. Κατηγοριοποίησαν τα μοντέλα σε τρεις κατηγορίες απόδοσης. Η πρώτη, που εμπεριέχει τα μοντέλα με την καλύτερη απόδοση, περιλαμβάνει τα MARS, BRT, γενικευμένης ανομοιότητας (GDM και GDM-SS) και μέγιστης εντροπίας (MAXENT και MAXENT-T). Στην επόμενη κατηγορία κατατάσσονται οι τεχνικές παλινδρόμησης (GAM/BRUTO, GLM, MARS και GARP) και η τρίτη κατηγορία με λιγότερο αποτελεσματικά μοντέλα περιλαμβάνει τα BIOCLIM, DOMAIN (μια κατηγορία του Environmental Distance, EnvDist) και LIVES. Τα αποτελέσματα της διατριβής επιβεβαιώνουν την ικανότητα πρόβλεψης του BRT και την χαμηλή απόδοση του Bioclim, παρόλο που χρησιμοποιούνται δεδομένα παρουσίας-απουσίας. Το Environmental Distance αποδίδει καλύτερα στην παρούσα μελέτη από ότι στην εργασία των Elith *et al.* (2006) αν και από ότι αποδείχθηκε υπερ-προσαρμόζεται στα δεδομένα εκπαίδευσης. Η μικρή διαφορά στο ROC-AUC μεταξύ MARS και GAMs που παρατηρείται εδώ, επιβεβαιώνεται και από άλλες εργασίες. Συγκεκριμένα, οι Leathwick *et al.* (2006a, b) εφάρμοσαν τα GAMs και MARS για να μοντελοποιήσουν την κατανομή δεκαπέντε ειδών γλυκού νερού σε σχέση με το περιβάλλον τους και οι τιμές ROC-AUC παρουσίασαν πολύ μικρές αποκλίσεις μεταξύ των δύο μοντέλων. Η υψηλή ικανότητα πρόβλεψης των ASNN και ANNE επιβεβαιώθηκε και στη μοντελοποίηση των δεδομένων αφθονίας.

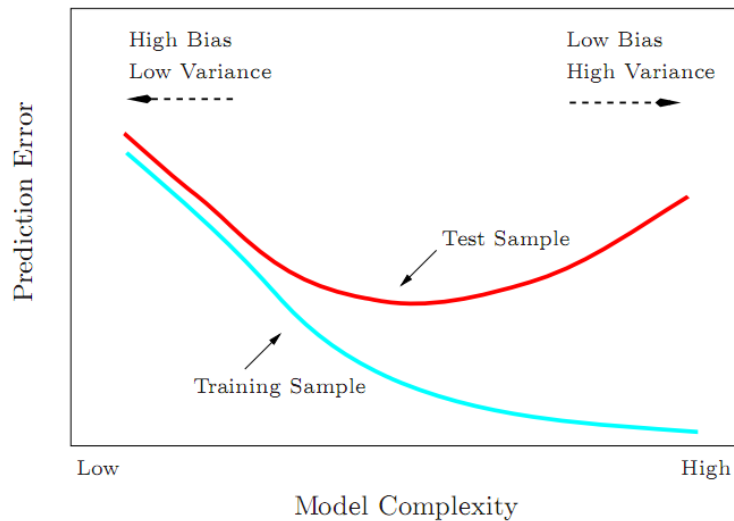
Η διαχείριση της αβεβαιότητας που συνοδεύει τις προβλέψεις των ΜΚΕ απαιτεί ιδιαίτερη προσοχή, ιδίως όταν τα αποτελέσματα των ΜΚΕ προορίζονται για λήψη αποφάσεων ή σκοπούς διαχείρισης. Η αβεβαιότητα στα ΜΚΕ προκύπτει τόσο από τη διακύμανση των δεδομένων όσο και από τις αποκλίσεις στην εφαρμογή των δεδομένων από τα μοντέλα (Elith & Leathwick, 2009). Τα προβλήματα που σχετίζονται με την αβεβαιότητα των προβλέψεων συνήθως ηθελημένα αγνοούνται διότι είναι δύσκολη η διαχείριση τους. Όμως η αβεβαιότητα μπορεί να περιοριστεί με (α) την επιλογή μεταβλητών που εξηγούν ικανοποιητικά τη διακύμανση της εξαρτημένης μεταβλητής τόσο στον περιβαλλοντικό όσο και το γεωγραφικό χώρο και (β) με την επιλογή ΜΚΕ που δύνανται να διαχειριστούν πολύπλοκες σχέσεις ειδών-περιβάλλοντος και αλληλεπιδράσεων μεταξύ των ανεξάρτητων μεταβλητών. Η Εικ. 4.1 παρουσιάζει εποπτικά τα σημεία κατά την ανάπτυξη ενός μοντέλου

κατανομής ειδών που αποτελούν πηγές αβεβαιότητας καθώς και τα σημεία που πρέπει να ληφθούν αποφάσεις, ώστε η διαδικασία μοντελοποίησης να επιτύχει το σκοπό για τον οποίο προορίζεται.



Εικόνα 4.1 Σημεία κατά τη μοντελοποίησης της κατανομής ειδών που αποτελούν πηγές αβεβαιότητας και σημεία αποφάσεων στην επιλογή δεδομένων και μεθόδου ώστε να επιτευχθεί ο σκοπός της μοντελοποίησης.

Η φύση τύπου "μαύρου κουτιού" (black box) που χαρακτηρίζει τις μεθόδους μηχανικής εκμάθησης δεν μπορεί να χαρακτηριστεί ως ιδιαίτερα πληροφοριακή για τέτοιες αλληλεπιδράσεις, η απόδοση τους όμως τις κατατάσσει υψηλότερα από τις άλλες μεθόδους. Πάντα κατά την εφαρμογή ΜΚΕ υπάρχει μια αλληλεπίδραση μεταξύ της πολυπλοκότητας και της αβεβαιότητας των προβλέψεων, η οποία φαίνεται ευκρινώς στην Εικ. 4.2. Αύξηση της πολυπλοκότητας του ΜΚΕ μπορεί να ελαχιστοποιεί τα σφάλματα στο σύνολο εκπαίδευσης. Στο σύνολο επαλήθευσης από την άλλη, από κάποιο ποσό πολυπλοκότητας και έπειτα, τα σφάλματα πρόβλεψης αρχίζουν να αυξάνουν. Κατά συνέπεια η αύξηση της πολυπλοκότητας πέρα από ένα οριακό σημείο δεν αντιστοιχεί στην αύξηση της ικανότητας πρόβλεψης και η διαδικασία επιλογής του βέλτιστου μοντέλου αποβλέπει στην αναζήτηση του οριακού σημείου.



Εικόνα 4.2 Σχέσεις πολυπλοκότητας ενός ΜΚΕ και σφάλματος πρόβλεψης (από Hastie *et al.* 2009)

Η χρήση βιοτικών αλληλεπιδράσεων (π.χ. σχέση θηρευτή-θηράματος, αλιευτική θνησιμότητα, κ.ά.), που σχετίζονται με τη χωρική κατανομή ειδών σε διάφορα στάδια του κύκλου ζωής τους, όταν χρησιμοποιηθούν ως επεξηγηματικές μεταβλητές στα ΜΚΕ μπορούν να βελτιώσουν την εξηγούμενη διακύμανση της εξαρτημένης μεταβλητής. Όπως αναφέρεται από τους Guisan & Thuiller (2005), πολύ λίγες μελέτες περιλαμβάνουν μεταβλητές που περιγράφουν βιολογικές αλληλεπιδράσεις. Οι Elith & Leathwick (2009) έδειξαν τις δυσκολίες της χρήσης βιολογικών επιδράσεων ως επεξηγηματικών μεταβλητών. Η χρήση τους όμως θα μπορούσε να ολοκληρώσει την εξηγούμενη διακύμανση στην περιβαλλοντική διάσταση, αλλά και να προσδιορίσει πιο σύνθετες σχέσεις στην οικολογική διάσταση. Στην πράξη, η διαδικασία επιλογής των ανεξάρτητων μεταβλητών εξαρτάται από: (α) τη διαθεσιμότητα και ποιότητα των δεδομένων, (β) την ικανότητα των δεδομένων να ερμηνεύσουν ένα σημαντικό ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής, και (γ) τις προϋποθέσεις των εκάστοτε τεχνικών ΜΚΕ. Το τελευταίο σημείο ενδέχεται να εξαιρέσει σημαντικές μεταβλητές από βιολογική άποψη. Στην περίπτωση αυτή θα μπορούσε να χρησιμοποιηθεί μια εναλλακτική μέθοδος μοντελοποίησης, ώστε να αξιοποιηθεί η διαθέσιμη βιολογική πληροφορία. Στη συγκεκριμένη διατριβή προτιμήθηκε η χρήση γνωστών, καλά μελετημένων επεξηγηματικών μεταβλητών, εφόσον ο βασικός στόχος είναι η παράλληλη σύγκριση των ΜΚΕ. Επιπλέον, η επιλογή των μεταβλητών έγινε και με άξονα τις απαιτήσεις των λογισμικών με τα οποία αναπτύχθηκαν τα ΜΚΕ και τη δυνατότητα να αποτυπώνονται σε καμβάδες, ώστε να εξασφαλιστεί η εφαρμοσιμότητά τους στο DS3. Η αλιευτική πίεση, ο δια-ειδικός ανταγωνισμός και οι σχέσεις θηράματος-θηρευτή φαίνεται να επηρεάζουν την κατανομή μικρών πελαγικών ειδών (π.χ. Sabates *et al.*, 2006; Ramzi *et al.*, 2006). Αυτή η πληροφορία μπορεί να εξηγήσει μέρος της διακύμανσης της κατανομής των ειδών, η οποία δεν καλύπτεται από την αποκλειστική χρήση αβιοτικών παραμέτρων, ειδικά στις περιπτώσεις όπου στοχεύουν στον προσδιορισμό αλληλεπιδράσεων ειδών ή στο χαρακτηρισμό κατανομής ειδών που προσεγγίζει τα πραγματικά ενδιαιτήματα (Planque *et al.*, 2007).

Τα περισσότερα ΜΚΕ κατάφεραν να απεικονίσουν τα βασικά πρότυπα κατανομής ειδών, τα οποία επιβεβαιώνονται και από αντίστοιχες μελέτες στην ίδια περιοχή. Οι πλέον καινοτόμες προσεγγίσεις παρέχουν πιο λεπτομερή αποτελέσματα και αναδεικνύουν την ετερογένεια των ενδιαιτημάτων σε πολύ υψηλή χωρική ανάλυση. Στα ΜΚΕ με παρόμοια απόδοση υπάρχουν διαφορές στις τελικά επιλεγμένες μεταβλητές, καθώς και στη σημαντικότητα της συνδρομής τους στο τελικό μοντέλο. Κάθε ΜΚΕ εξηγεί ένα συγκεκριμένο ποσοστό της διακύμανσης της ακουστικής πληροφορίας. Ακόμα και στην περίπτωση που δύο ΜΚΕ ερμηνεύουν το ίδιο ποσό διακύμανσης, ενδέχεται να διαφοροποιείται η διάσταση της διακύμανσης που εξηγούν, γεγονός που φαίνεται από τη διαφορετική σημαντικότητα και συνεισφορά των επεξηγηματικών μεταβλητών. Το γεγονός αυτό πρέπει να αντιμετωπίζεται με ιδιαίτερη προσοχή, ειδικά στην περίπτωση που τα ΜΚΕ χρησιμοποιούνται για τη μελέτη οικολογικών σχέσεων ή για πλάνα διαχείρισης.

4.3 Διαδικασία επαλήθευσης ΜΚΕ

Η χρήση ενός ανεξάρτητου καλά δομημένου συνόλου παρουσίας-απουσίας ως σύνολο επαλήθευσης προτείνεται ως η βέλτιστη μέθοδος αξιολόγησης της απόδοσης πρόβλεψης των ΜΚΕ (Elith *et al.*, 2006). Από την άλλη, η χρήση ενός εντελώς ανεξάρτητου συνόλου δεδομένων φέρει τον κίνδυνο σύγκρισης διαφορετικών δειγματοληπτικών στρατηγικών αντί της επαλήθευσης μοντέλων (Lehmann *et al.*, 2002). Εναλλακτικά, τεχνικές όπως η διασταυρωτή-επαλήθευση (cross-validation) (Jaberg & Guisan, 2001) και η jackknife (Lehmann *et al.*, 2002) επίσης χρησιμοποιούνται για την αξιολόγηση ΜΚΕ, ειδικά στις περιπτώσεις όπου δεν υπάρχουν επαρκή δεδομένα, ικανά να δημιουργήσουν διαφορετικά σύνολα εκπαίδευσης και επαλήθευσης. Σύμφωνα με τους Lehmann *et al.* (2002) και Jaberg & Guisan (2001), οι προαναφερθείσες τεχνικές έχουν μεγάλη πρακτική αξία, επειδή δημιουργούν σχετικά ανεξάρτητες, τυχαίες, ομάδες δεδομένων και επιτρέπουν τη συμμετοχή όλων των διαθέσιμων δεδομένων στην εκπαίδευση των μοντέλων. Ως αντίλογο στα παραπάνω, οι Fu *et al.* (2005) και Simon *et al.* (2003) παρατήρησαν ότι η διασταυρωτή-επαλήθευση και κυρίως η leave-one-out cross-validation μπορεί να οδηγήσει στην υποεκτίμηση των σφαλμάτων πρόβλεψης. Το σύνολο επαλήθευσης στα δεδομένα παρουσίας-απουσίας δημιουργήθηκε με τον τρόπο που παρουσιάστηκε παραπάνω, εφόσον τα δειγματοληπτικά δεδομένα ήταν αρκετά ώστε να δημιουργήσουν τα σύνολα εκπαίδευσης και επαλήθευσης. Η συγκεκριμένη επιλογή του συνόλου επαλήθευσης αποτρέπει το πρόβλημα της υποεκτίμησης των σφαλμάτων πρόβλεψης, το οποίο γίνεται ακόμη πιο έντονο όταν τα ακουστικά δεδομένα είναι χωρικά αυτοσυσχετιζόμενα (Hastie *et al.*, 2009). Επιπλέον, εξαλείφεται και ο κίνδυνος σύγκρισης διαφορετικών δειγματοληπτικών στρατηγικών αφού το σύνολο επαλήθευσης αποτελεί μέρος των αρχικών δεδομένων. Γενικά, η διαδικασία επαλήθευσης και αξιολόγησης των ΜΚΕ αποτελεί ένα από τα σημαντικότερα ζητήματα μεταξύ των ερευνητών που ασχολούνται με τη μοντελοποίηση της κατανομής ειδών (Elith & Leathwick, 2009). Η έλλειψη κοινής ενδεδειγμένης λύσης για όλες τις περιπτώσεις, καθιστά απαραίτητη την αξιοποίηση χρήσιμων τεχνικών από άλλους τομείς, όπως η μετεωρολογία (Elith & Leathwick, 2009).

Το κριτήριο πληροφορίας του Akaike (AIC) που ήδη χρησιμοποιήθηκε κατά την επιλογή του βέλτιστου μοντέλου μέσα σε κάποιες τεχνικές, θα μπορούσε να χρησιμοποιηθεί και για τη σύγκριση των διαφορετικών τεχνικών. Για μη γραμμικές τεχνικές ο υπολογισμός του αριθμού των δραστικών παραμέτρων και συνάμα του AIC είναι πολύ δύσκολος (Hastie *et al.*, 2009). Για το λόγο αυτό οι σημαντικές συγκρίσεις βασίστηκαν κυρίως σε τεχνικές cross-validation και ROC-AUC. Το AIC χρησιμοποιήθηκε κυρίως για τη σύγκριση ένθετων μοντέλων της ίδιας τεχνικής και για τον έλεγχο της ισορροπίας πολυπλοκότητας και ακρίβειας πρόβλεψης των

μοντέλων. Οι Hastie *et al.* (2009) χρησιμοποιώντας πειραματικές προσομοιώσεις διαπίστωσαν ότι το AIC υπερεκτιμά το σφάλμα πρόβλεψης (>30%) σε σχέση με το cross-validation, γεγονός που μπορεί να αποδοθεί στην "τιμωρία" της πολυπλοκότητας. Από την άλλη, οι τεχνικές cross-validation δεν λειτουργούν σωστά όταν υπάρχει αυτοσυσχέτιση στα δεδομένα, οδηγώντας σε υποεκτίμηση των σφαλμάτων πρόβλεψης και κατά συνέπεια σε εσφαλμένη επιλογή μοντέλου (Hastie *et al.*, 2009). Τα υδροακουστικά δεδομένα φέρουν χωροχρονικές αυτοσυσχετίσεις, οι οποίες αποδίδονται περισσότερο στη δομή της βιομάζας και λιγότερο στις μετρήσεις κατά τη δειγματοληψία (Simmonds και MacLennan, 2005). Η ύπαρξη χωρικών προτύπων στα σφάλματα μπορεί να προκαλέσει, εκτός των άλλων, υποεκτίμηση των τυπικών σφαλμάτων της κλίσης των μοντέλων παλινδρόμησης και αδυναμίες στον έλεγχο υποθέσεων και τις προβλέψεις (Ostrom, 1990).

4.4 Χωρικά πρότυπα στα δεδομένα

Τα πρότυπα χωρικής αυτοσυσχέτισης είναι συνήθη στα δεδομένα αφθονίας και βιομάζας οργανισμών ή σε ανάλογες οικολογικές εγγραφές (Legendre, 1993). Κατά συνέπεια, τα συνήθη στατιστικά μοντέλα όταν διαχειρίζονται τέτοιου τύπου δεδομένα ενδέχεται να παραβιάζουν τη βασική προϋπόθεση της ανεξαρτησίας των σφαλμάτων. Οι πιθανοί λόγοι που δημιουργούν τη χωρική αυτοσυσχέτιση στα δεδομένα μπορούν να ομαδοποιηθούν σε τρεις κατηγορίες: (α) στη φύση των βιολογικών διεργασιών που εμπλέκονται στα δεδομένα, (β) στην απουσία σημαντικών επεξηγηματικών μεταβλητών των μοντέλων, ικανών να εξηγήσουν τη διασπορά της εξαρτημένης μεταβλητής και (γ) στη γραμμική μοντελοποίηση διεργασιών μη γραμμικών (Legendre & Legendre, 1998). Μεταξύ των κοινών μεθόδων που χρησιμοποιούνται για τη διαχείριση των χωρικών προτύπων των σφαλμάτων των μοντέλων είναι και οι ακόλουθες: (α) η προσθήκη συμμεταβλητών, ικανών να απορροφήσουν τα αυτοσυσχετιζόμενα σφάλματα (Elith & Leathwick, 2009), (β) η επιλογή κατάλληλης στοιχειώδους δειγματοληπτικής μονάδας (ESDU), (γ) η εφαρμογή μεθόδων βασισμένων σε Ανάλυση Κυματιδίων (Wavelet analysis) για την απομάκρυνση των φαινομένων αυτοσυσχέτισης (Gudrun & Kühn, 2008), (δ) η αναγωγή του μοντέλου σε πιο σύνθετο, ώστε να συμπεριλάβει τη χωρική αυτοσυσχέτιση (π.χ. από GAM σε GAMM) (Dray *et al.*, 2006) και (ε) η αξιοποίηση αυτής καθ' αυτής της αυτοσυσχέτισης για σκοπούς παρεμβολής (Rossi *et al.*, 1992; Simmonds & MacLennan, 2005).

Διάφορες τροποποιήσεις της τυπικής διασταυρούμενης επικύρωσης (typical cross-validation) έχουν δημοσιευτεί κατά καιρούς για να διαχειριστούν το θέμα εξάρτησης μεταξύ των δεδομένων εκπαίδευσης και επαλήθευσης (Burman *et al.*, 1994; Racine, 2000). Στην παρούσα μελέτη, το σύνολο επαλήθευσης περιλαμβάνει μετρήσεις των αρχικών δειγματοληπτικών σημείων σε βήμα h μονάδων (h ίσο με τουλάχιστον 5 n mi), το οποίο h έχει επιλεγεί με βάση τα εμπειρικά βαριογράμματα των συνόλων επαλήθευσης και των σφαλμάτων των προβλέψεων. Τα εμπειρικά βαριογράμματα και τα γραφήματα αυτοσυσχέτισης των δεδομένων έδειξαν την ύπαρξη σχετικά χαμηλής αυτοσυσχέτισης (σχεδόν αμελητέα) σε αποστάσεις μικρότερες από 5 n mi. Το σύνολο επαλήθευσης παρουσιάζει μηδενική αυτοσυσχέτιση, όπως φαίνεται και από την Εικόνα 3.11. Ανάλογη χωρική δομή παρατηρήθηκε σε αντίστοιχες δειγματοληψίες στην περιοχή που πραγματοποιήθηκαν στο παρελθόν, σε άλλες όμως εποχές (Οκτώβρης 1996 και Νοέμβρης 1997) (Georgakarakos & Kitsiou, 2008). Τα αποτελέσματα από μια συγκριτική μελέτη ακουστικών δεδομένων προερχόμενα από πέντε διαφορετικές περιοχές της Ευρώπης έδειξαν ότι η χωρική οργάνωση των αποθεμάτων εξαρτάται περισσότερο από περιβαλλοντικές παραμέτρους παρά από την αφθονία των ψαριών (Petitgas *et al.*, 2001). Τα αποτελέσματα αυτά έρχονται σε αντίθεση με τη γενική θεώρηση που συσχετίζει το μέγεθος του αποθέματος με τη χωρική του οργάνωση, τουλάχιστον για

υψηλές τιμές αφθονίας (MacCall, 1990). Σε ανάλογη μελέτη τα χαρακτηριστικά των σμηνών και των συστάδων τους (cluster) στα μικρά πελαγικά είδη (π.χ. ενδιάμεσες αποστάσεις, αριθμός σμηνών) έχουν συσχετιστεί με το συνολικό αριθμό των σμηνών και όχι με τη βιομάζα του πληθυσμού (Munio *et al.*, 2003). Τα χαρακτηριστικά αυτοσυσχέτισης της βιομάζας συμφωνούν με την τάση ομαδοποίησης των σμηνών σε προηγούμενες δειγματοληψίες (Οκτώβρης 1996 και Νοέμβρης 1997), γεγονός που έχει υπολογιστεί με βάση την απόσταση μεταξύ δύο κοπαδιών μέσα σε μια ομάδα κοπαδιών (Petitgas *et al.*, 2001). Τα περιγραφικά χαρακτηριστικά των εκτιμώμενων κοπαδιών, αλλά και των συστάδων τους από αυτές τις δειγματοληψίες (μέσος αριθμός κοπαδιών ανά km, μέσο κλάσμα μήκους κοπαδιού προς μήκος ομάδας κοπαδιών, μέγιστη απόσταση μεταξύ δύο κοπαδιών στην ομάδα) δείχνουν μικρή κλίμακα συγκέντρωσης στη χωρική δομή της βιομάζας.

4.5 Περιβαλλοντικοί παράγοντες που επηρεάζουν την κατανομή των μικρών πελαγικών ψαριών

Αξιολογώντας τα μικρά πελαγικά είδη, ως μέσο μοντελοποίησης της κατανομής τους, διαφαίνεται ότι η οικολογία τους ευνοεί σημαντικά τη ρεαλιστική προσομοίωση της κατανομής τους στο χώρο. Τα είδη αυτά χαρακτηρίζονται από τη δημιουργία χωρικών προτύπων, που σχετίζονται άμεσα με περιβαλλοντικούς και ωκεανογραφικούς παράγοντες. Επίσης, η εκτεταμένη μελέτη του κύκλου ζωής τους επιτρέπει τη δημιουργία ενός θεμελιώδους μοντέλου πάνω στο οποίο θα στηριχθεί η εφαρμογή των ΜΚΕ. Όπως έδειξαν οι χάρτες πρόβλεψης, τα χωρικά πρότυπα των μικρών πελαγικών ειδών απεικονίζονται λεπτομερώς. Παράλληλα, οι συσχετίσεις μεταξύ των περιβαλλοντικών παραμέτρων με τα δειγματοληπτικά δεδομένα των μικρών πελαγικών ειδών και η απόδοση των ΜΚΕ δείχνουν ότι ερμηνεύεται μεγάλο ποσοστό της κατανομής τους από τις επιλεγμένες επεξηγηματικές μεταβλητές και επικυρώνεται το θεμελιώδες μοντέλο των περιβαλλοντικών παραμέτρων που ελέγχουν τα κύρια περιβαλλοντικά όρια της κατανομής των μικρών πελαγικών ειδών (Εικ. 2.2). Τα μικρά πελαγικά είδη και κυρίως ο γαύρος και η σαρδέλα δέχονται έντονη αλιευτική πίεση, ενώ χαρακτηρίζονται από ευρεία διακύμανση της αφθονίας τους και της χωρικής κατανομής τους, που εξαρτώνται σημαντικά από τις περιβαλλοντικές διακυμάνσεις. Διάφορα ωκεανογραφικά χαρακτηριστικά, όπως θερμοκρασιακά μέτωπα, δίνες, κυκλωνικά φαινόμενα, και αναβλύσεις έχουν κατά καιρούς συσχετιστεί με τη συγκέντρωση ψαριών (Laurs *et al.*, 1984; Fiedler & Bernard, 1987; Bakun, 2001; Garcia και Palomera, 1996; Chen *et al.*, 2005). Με βάση τον κύκλο ζωής των μικρών πελαγικών οι περισσότερες μεταβλητές που χρησιμοποιήθηκαν σχετίζονται με την κατανομή τους (Daskalov *et al.*, 2003; Santos *et al.*, 2004; Ruiz *et al.*, 2006; Planque *et al.*, 2007). Οι Zwolinski *et al.* (2010) έδειξαν ότι τα χωρικά πρότυπα της σαρδέλας σχετίζονται με υδάτινες μάζες που χαρακτηρίζονται από υψηλή συγκέντρωση χλωροφύλλης, χαμηλή θερμοκρασία και χαμηλή αλατότητα. Σύμφωνα με τους Georgakarakos & Kitsiou (2008) θερμές, φτωχές σε θρεπτικά συστατικά υδάτινες μάζες χαρακτηρίζονται από χαμηλές συγκεντρώσεις χλωροφύλλης-α, ενώ κρύες, πλούσιες σε θρεπτικά υδάτινες μάζες χαρακτηρίζονται από υψηλές συγκεντρώσεις. Είδη όπως ο γαύρος και η σαρδέλα δείχνουν προτίμηση σε περιοχές υψηλής συγκέντρωσης θρεπτικών. Σε ορισμένες περιόδους του κύκλου ζωής τους μπορεί να εμφανίζουν ισχυρή συσχέτιση με περιοχές υψηλής συγκέντρωσης χλωροφύλλης-α, συγκριτικά με είδη που ανήκουν σε υψηλότερα τροφικά επίπεδα. Οι περιοχές υψηλής συγκέντρωσης χλωροφύλλης-α αποτυπώνονται εύκολα σε δορυφορικές εικόνες (Ware & Thomson, 2005; Valavanis, 2002). Σημαντική συσχέτιση υπάρχει επίσης μεταξύ των εκροών ποταμών και της αφθονίας γαύρου όπως έχει εκτιμηθεί στον Βισκαϊκό κόλπο, την Αδριατική, τον κόλπο των Λεόντων (Motos *et al.*, 1996; Agostini & Bakun, 2002), τις ακτές της Τυνησίας και τη θάλασσα της Καταλονίας (Palomera *et al.*, 2007; Sabates *et al.*, 2007).

Οι κάνναβοι πρόβλεψης για τον κόλπο του Θερμαϊκού αναδεικνύουν δύο διακριτές περιοχές συγκέντρωσης της βιομάζας: (α) τη δυτική ακτογραμμή από Βορρά προς Νότο, που χαρακτηρίζεται από την παρουσία εκβολών μεγάλων ποταμών και (β) την περιοχή στο κέντρο του Θερμαϊκού κόλπου, που σχετίζεται με τη δημιουργία κυκλωνικών σχηματισμών (Somarakis *et al.*, 2002). Και οι δύο αυτές περιοχές μπορούν να χαρακτηριστούν ως εύτροφες.

Διάφορες προσεγγίσεις μοντελοποίησης έχουν εφαρμοστεί στον Θερμαϊκό κόλπο όσον αφορά μικρά πελαγικά είδη. Οι Georgakarakos & Kitsiou (2008) εφάρμοσαν μεθόδους Kriging και Co-Kriging σε δεδομένα ακουστικής πυκνότητας και τα αποτελέσματα τους ως προς την προβλεπόμενη κατανομή των ειδών είναι σε συμφωνία με τα αποτελέσματα της διατριβής, ειδικά στον προσδιορισμό της κεντρικής περιοχής μελέτης. Οι Schismenou *et al.* (2008) χαρακτήρισαν τον Θερμαϊκό ως γνωστό αλιευτικό πεδίο για γούρο και σαρδέλα, όμως τα αποτελέσματα της μελέτης τους αποτυπώνονται σε χαμηλότερη χωρική ανάλυση και δεν μπορούν να συγκριθούν άμεσα με τα αποτελέσματα αυτής της διατριβής. Το ίδιο συμβαίνει και με τη μελέτη των Tsagarakis *et al.* (2008) που εφάρμοσαν Discriminant Function Analysis σε δεδομένα ακουστικής πυκνότητας της σαρδέλας. Παρά τη χαμηλή χωρική ανάλυση των χαρτών, που δεν επιτρέπει τη διάκριση της ετερογένειας των ενδιαιτημάτων, τα αποτελέσματα είναι γενικά σε συμφωνία με αυτά της παρούσας διατριβής. Οι Giannoulaki *et al.* (2008) ανέπτυξαν GAMs για τον προσδιορισμό ενδιαιτημάτων του γούρου σε ένα διαφορετικό χρονικό πλαίσιο από αυτό που χρησιμοποιήθηκε εδώ, και κατέληξαν σε ανάλογες τάσεις κατανομής ειδών με αυτές που έχουν χαρτογραφηθεί από τα πιο αποτελεσματικά μοντέλα στην παρούσα διατριβή. Αυτή η ταύτιση επιβεβαιώνει τη μόνιμη ή μακρά παρουσία ορισμένων ωκεανογραφικών χαρακτηριστικών στην περιοχή, όπως κυκλωνικοί σχηματισμοί και αναβλύσεις, που ευνοούν τη συγκέντρωση μικρών πελαγικών ειδών.

Έχει αναφερθεί σε διάφορα σημεία η ικανότητα των ΜΚΕ να αποδίδουν σχέσεις ειδών και περιβάλλοντος. Είναι χαρακτηριστική η ικανότητα των Γενικευμένων Προσθετικών Μοντέλων να διακρίνουν σχέσεις μεταξύ της εξαρτημένης και των ανεξάρτητων μεταβλητών. Στη συγκεκριμένη εφαρμογή με τα δεδομένα αφθονίας, η υψηλή ακουστική πυκνότητα παρατηρείται σε περιοχές που χαρακτηρίζονται από υψηλές ή χαμηλές τιμές SLA, σε συνδυασμό με μικρό βάθος και μεγάλου εύρους PAR ή με μεγάλο βάθος και μεγάλη PAR. Το υποδεικνυόμενο εύρος της SLA, που επηρεάζεται από τους ανέμους και τα ρεύματα, εμφανίζεται στις ανατολικές και δυτικές ακτές της περιοχής μελέτης. Αυτές οι περιοχές είναι γενικά πλούσιες σε θρεπτικά συστατικά λόγω των αναβλύσεων και των εκροών μεγάλων ποταμών, οπότε δημιουργούν κατάλληλες συνθήκες για τη συγκέντρωση μικρών πελαγικών ειδών. Επίσης, οι ίδιες περιοχές υποδεικνύονται και από τον παράγοντα αλληλεπίδρασης μεταξύ PAR και DEP. Με βάση τον παράγοντα αυτό εξηγούνται και οι περιοχές που εντοπίζονται από τα GAMs στα ανοιχτά της περιοχής μελέτης, όπου παρατηρούνται ωκεανογραφικά χαρακτηριστικά, π.χ. δίνες, που σχετίζονται με τον κύκλο ζωής των μικρών πελαγικών ειδών (Valavanis *et al.*, 2005). Όσον αφορά τα μοντέλα παρουσίας-απουσίας θα πρέπει να σημειωθεί ότι δεν παρουσιάζουν όλες οι τεχνικές λεπτομερής σχέσης της κατανομής των ειδών με τις επεξηγηματικές μεταβλητές. Στο Παράρτημα III παρουσιάζονται στατιστικά και γραφικά οι σχέσεις των ανεξάρτητων μεταβλητών με την εξαρτημένη μεταβλητή, όπως εξάγονται από κάποιες προσεγγίσεις μοντελοποίησης της κατανομής ατόμων ειδών με δεδομένα παρουσίας-απουσίας. Οι προσεγγίσεις μοντελοποίησης που παρέχουν προκαθορισμένα την παραπάνω πληροφορία είναι οι: GAMs, GAMMs, BRTs, MARS και MAXENT. Για τα δεδομένα παρουσίας-απουσίας είναι εντυπωσιακή η ταύτιση των αποτελεσμάτων όσον αφορά τις σχέσεις ειδών περιβάλλοντος μεταξύ των προαναφερθέντων μεθόδων, πλην του MAXENT. Φαίνεται, λοιπόν, ότι η ALT είναι ο παράγοντας που εξηγεί το μεγαλύτερο μέρος της διακύμανσης της παρουσίας μικρών πελαγικών, ενώ οι χαμηλές τιμές της αλτιμετρίας

είναι οι κατάλληλες για παρουσία των μικρών πελαγικών. Ακόμη, αποδεικνύεται ότι υπάρχει ισχυρή αλληλεπίδραση μεταξύ DEP και SST, που επιβεβαιώνεται από διαφορετικές προσεγγίσεις (GAM, GAMM, BRT) και σύμφωνα με την οποία η παρουσία των μικρών πελαγικών ειδών είναι συνυφασμένη με περιοχές που χαρακτηρίζονται από ψυχρές υδάτινες μάζες σε αναλογικά ρηγά νερά. Οι περιοχές αυτές αντιστοιχούν στις εκβολές των μεγάλων ποταμών της περιοχής, αλλά και σε περιοχές με αναβλύσεις, που όπως έχει αναφερθεί σε ερευνητικές μελέτες σχετίζονται άμεσα με την παρουσία γαύρου και σαρδέλας (Motos *et al.*, 1996; Agostini & Bakun, 2002). Ταύτιση μεταξύ των προαναφερθέντων μεθόδων υπάρχει επίσης ως προς τη σημαντικότητα του παράγοντα DDND που ερμηνεύει σημαντικό μέρος της διακύμανσης της παρουσίας των ειδών, αλλά και ως προς τη μικρή συνεισφορά της απόστασης από την ακτή στα μοντέλα. Το MAXENT, που όπως έχει αποδειχθεί από την σύγκριση των μοντέλων δεν επιτυγχάνει ακριβείς προβλέψεις έχει ιεραρχήσει διαφορετικά τις μεταβλητές, ως προς τη σημαντικότητά τους. Παρά το ότι υποδεικνύει την αλτιμετρία ως τη βασική παράμετρο για να αυξηθεί η ποιότητα του μοντέλου, θεωρεί την απόσταση από την ακτή ως την μεταβλητή με την σημαντικότερη πληροφορία για την κατανόηση της κατανομής των ατόμων των μικρών πελαγικών ειδών. Όπως αναμένεται τα αποτελέσματα του MAXENT για τις σχέσεις περιβαλλοντικών μεταβλητών και παρουσίας ειδών δεν μπορούν να έχουν την ίδια βαρύτητα με τα αντίστοιχα των BRTs και GAMs, αφού τα δεύτερα υπερτερούν σημαντικά σε όλα τα επίπεδα των συγκρίσεων. Η χρήση περισσότερων της μιας μεθόδων μοντελοποίησης για τα ίδια δεδομένα είναι πολύ χρήσιμη, όπως αναφέρει και η Franklin (2009), για τη βαθύτερη διερεύνηση των προτύπων κατανομής των ειδών, αλλά και για την καλύτερη κατανόηση της δομής και των υποθέσεων των μοντέλων. Επομένως, οι διαφορετικές προσεγγίσεις μοντελοποίησης που αναπτύχθηκαν στο ίδιο σύνολο δεδομένων παρέχουν συμπληρωματικά μια ευρύτερη-συμπληρωματική εικόνα της κατανομής των ειδών.

5. Συμπεράσματα

5.1 Μοντέλα αφθονίας

Τα Γενικευμένα Προσθετικά Μοντέλα είναι πολύ ευέλικτες τεχνικές μοντελοποίησης, που αποδίδουν πολύ καλά με ακουστικά δεδομένα, ενώ χαρακτηρίζονται από την ικανότητα να προσδιορίζουν εύκολα σχέσεις μεταξύ της εξαρτημένης και των ανεξάρτητων μεταβλητών. Η μέθοδος Kriging από την άλλη, στηριζόμενη στη χωρική συσχέτιση των ακουστικών δεδομένων, δεν έδωσε ανάλογα αποτελέσματα με τις τεχνικές που χρησιμοποίησαν περιβαλλοντικές και άλλες παραμέτρους που σχετίζονται με την κατανομή ειδών. Τα ASNN και ANNE αποδεικνύονται ως τα πλέον ακριβή ΜΚΕ. Και οι δύο μέθοδοι προσομοιώνουν τα ακουστικά δεδομένα πιο ρεαλιστικά από τις υπόλοιπες και έχουν τη δυναμική να εντοπίζουν τη διακύμανση των δεδομένων τόσο σε τοπική όσο και σε συνολική διάσταση, χωρίς να εμφανίζουν δείγματα υπερ-προσαρμογής στα δειγματοληπτικά δεδομένα εκπαίδευσης. Η μέθοδος βελτιστοποίησης πρόβλεψης (AverMod) έδωσε τα καλύτερα αποτελέσματα και υπερτερούσε έναντι των άλλων μεθόδων σε όλες τις συγκρίσεις. Η αποτελεσματικότητα της οφείλεται στον προσθετικό συνδυασμό ενός "γεωγραφικού" μοντέλου (Kriging) με ένα "περιβαλλοντικό" μοντέλο (ASNN) που οδήγησαν στη δημιουργία ενός υβριδικού μοντέλου ικανού να ερμηνεύει τη διακύμανση της ακουστικής πληροφορίας τόσο στην περιβαλλοντική όσο και τη γεωγραφική διάσταση.

5.2 Μοντέλα παρουσίας-απουσίας

Η σύγκριση δεκαπέντε διαφορετικών στατιστικών προσεγγίσεων, που ανήκουν σε δεκατρία Μοντέλα Κατανομής Ειδών, έδειξαν ότι οι τεχνικές μηχανικής εκμάθησης είναι γενικά πιο ακριβείς στην πρόβλεψη κατανομής ατόμων ειδών. Στην παρούσα μελέτη χρησιμοποιήθηκαν για τη σύγκριση των ΜΚΕ δεδομένα παρουσίας-απουσίας, που προέρχονται από προκαθορισμένη συστηματική δειγματοληψία και εκπαιδεύονται με ένα σημαντικό αριθμό ανεξάρτητων μεταβλητών υψηλής ανάλυσης. Πιο συγκεκριμένα, το BRT αποδίδει καλύτερα από τις άλλες τεχνικές στην προσαρμογή των δεδομένων εκπαίδευσης, ενώ το ASNN παρουσιάζει αξιοσημείωτη ικανότητα πρόβλεψης. Το SVM αποτελεί την πιο αξιόπιστη προσέγγιση μεταξύ των τεχνικών που δεν υποστηρίζουν δεδομένα απουσίας. Όλα τα προαναφερθέντα ΜΚΕ δεν εμφανίζουν τάση υπερ-προσαρμογής των δεδομένων εκπαίδευσης. Οι τεχνικές μηχανικής μάθησης στηρίζονται σε αρχές στατιστικής που επιφορτίζονται με την ανάπτυξη και τον σχεδιασμό αλγορίθμων ικανών να αλλάζουν συμπεριφορά και να προσαρμόζονται ανάλογα με τα δεδομένα. Η εξέλιξη στην επιστήμη υπολογιστών στηρίζει πλέον πιο σύνθετες προσομοιώσεις δεδομένων και συνδυασμούς αλγορίθμων που αποδίδουν με μεγαλύτερη ακρίβεια. Στην κατηγορία αυτή εντάσσονται τα ASNN, που αποτελούν συνδυασμό ANNE και ενός αλγορίθμου k-nearest neighbour, καθώς και τα BRTs, που αποτελούν συνδυασμό ενός ενισχυτικού αλγορίθμου και δέντρων παλινδρόμησης. Η εξέλιξη και η εφαρμογή ανάλογων τεχνικών στην πρόβλεψη κατανομής ειδών δίνει νέες προοπτικές για πιο ρεαλιστικά και εφαρμόσιμα αποτελέσματα, που η απόδοση τους ξεπερνά τις συμβατικές μεθόδους μοντελοποίησης (Elith & Leathwick, 2009).

Τα μοντέλα παλινδρόμησης κατατάσσονται αναλογικά υψηλά στην απόδοση τους, όσον αφορά στην προσαρμογή των δεδομένων εκπαίδευσης και την ικανότητα πρόβλεψης και αποτελούν από τις πλέον ευέλικτες μεθόδους στη μοντελοποίηση διαφόρων τύπων δεδομένων. Τα GAMs, MARS και GAMMs αποδίδουν σχεδόν ανάλογα, αν και το GAM υπερτερεί κατά τι των άλλων. Το GAMM φέρει την ικανότητα να διαχειρίζεται τη χωρική συσχέτιση των δεδομένων, που χαρακτηρίζει τα δεδομένα εκπαίδευσης, χωρίς όμως να επιτύχει καλύτερες προβλέψεις από το GAM. Το γεγονός αυτό μπορεί να αποδοθεί στη σχετικά περιορισμένη χωρική αυτοσυσχέτιση των δεδομένων εκπαίδευσης, όπως έδειξαν τα εμπειρικά βαριογράμματα, ή/και στην επαρκή εξήγηση της αυτοσυσχέτισης από τις γεωγραφικές και περιβαλλοντικές ανεξάρτητες μεταβλητές (Elith & Leathwick, 2009). Θα πρέπει να σημειωθεί ότι το GAMM δεν απέδωσε ικανοποιητικά στο συγκεκριμένο σύνολο δεδομένων εκπαίδευσης, όμως δεν πρέπει να υποτιμηθεί η ικανότητα πρόβλεψης και διαχείρισης της χωρικής αυτοσυσχέτισης ή ανάλογων πολύπλοκων σχέσεων, ως εξέλιξη του GAM. Εδώ βέβαια υπεισέρχεται και το ζήτημα της αύξησης της πολυπλοκότητας των μοντέλων, η οποία δε συνεπάγεται αυξημένη ακρίβεια πρόβλεψης, αλλά θα πρέπει πάντα να αναζητείται η χρυσή τομή μεταξύ πολυπλοκότητας και ακρίβειας. Συμπερασματικά, από τις τεχνικές παλινδρόμησης προτείνεται το GAM ως το πιο αποτελεσματικό για τη μοντελοποίηση κατανομής ειδών και εναλλακτικά το MARS ως μια πιο φιλική στη χρήση εφαρμογή, σχεδόν εξίσου αποτελεσματική.

Τα μοντέλα μόνο-παρουσίας δεν μπορούν να ανταγωνιστούν τα παρουσίας-απουσίας καθώς η μη χρήση των δεδομένων απουσίας αντιστοιχεί στην απώλεια σημαντικής πληροφορίας. Οι συγκρίσεις έδειξαν ότι τεχνικές όπως BIOCLIM, EnvScore, ClimSpace και MAXENT αποτυγχάνουν να παράγουν ανταγωνιστικά αποτελέσματα. Μόνο το EnvDist αποτέλεσε εξαίρεση και απέδωσε σχετικά ικανοποιητικά, αν και υπάρχουν ενδείξεις ότι αυτή η μέθοδος υπερ-προσαρμόζεται στα δεδομένα εκπαίδευσης και χρειάζεται προσεκτικό χειρισμό για να μην οδηγήσει σε σφάλματα.

Γενικά η ακρίβεια προσαρμογής και η ικανότητα πρόβλεψης που χαρακτηρίζει ένα ΜΚΕ εξαρτάται σημαντικά από την ποιότητα των δεδομένων εκπαίδευσης, όσον αφορά στη ρεαλιστική και ακριβή απεικόνιση της κατανομής των ατόμων των ειδών. Η χρήση ακουστικών δεδομένων από δειγματοληψία σε προκαθορισμένες τομές που μοντελοποιήθηκε με υψηλής ανάλυσης περιβαλλοντικά και γεωγραφικά δεδομένα έδειξε ότι τα BRTs και ASNN αποτελούν τις πλέον ενδεδειγμένες προσεγγίσεις. Οι προσεγγίσεις, που ανήκουν στις μεθόδους μηχανικής μάθησης, με τις εκτεταμένες αναλυτικές ικανότητες, αποτελούν χρήσιμα εργαλεία στην πρόβλεψη της κατανομής ειδών. Θα πρέπει να σημειωθεί κλείνοντας, ότι η διαφορετική περιοχή μελέτης, η διαθεσιμότητα επεξηγηματικών μεταβλητών και δεδομένων εκπαίδευσης ενδέχεται να απαιτούν προσεγγίσεις διαφορετικές από αυτές που προτείνονται σε αυτή τη μελέτη.

6. Βιβλιογραφία

- Aertsen, W., V. Kint, J. van Orshoven, K. Özkan & B. Muys, 2010. Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecological Modelling*, 221: 1119–1130.
- Agostini, V. N. & A. Bakun, 2002. ‘Ocean triads’ in the Mediterranean Sea: physical mechanisms potentially structuring reproductive habitat suitability (with example application to European anchovy, *Engraulis encrasicolus*). *Fisheries Oceanography*, 11: 129–142.
- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–723.
- Amara, R., K. Mahé, O. LePape & N. Desroy, 2004. Growth, feeding and distribution of the solenette *Buglossidium luteum* with particular reference to its habitat preference. *Journal of Sea Research* 51: 211-217.
- Andersen, M. C., Adams, H., Hope, B., & Powell, M., 2004. Risk analysis for invasive species: general framework and research needs. *Risk Analysis*, 24, 893–900.
- Araújo, M. B. & Guisan, A. (2006) Five (or so) challenges for species distribution modeling. *Journal of Biogeography*, 33, 1677–1688.
- Austin, M. P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, 157, 101–118.
- Austin, M. P. & Smith, T. M., 1989. A new model for the continuum concept. *Vegetatio*, 83, 35–47.
- Bakun, A., 2001. ‘School-mix feedback’: a different way to think about low frequency variability in large mobile fish populations. *Progress in Oceanography*, 49: 485–511.
- Bellis, L. M., Pidgeon, A. M., Radeloff, V. C., St-Louis, V., Navarro, J. L., & Martella, M. B., 2008. Modeling habitat suitability for greater rheas based on satellite image texture. *Ecological Applications*, 18, 1956–1966.
- Bishop, M., 1995. *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford.
- Bodholt, H., H. Nes & H. Solli, 1989. A new echo sounder system. *Proceedings of the Institute of Acoustics (UK)* 11(3): 123-130.
- Box, E., 1981. *Macroclimate and Plant Forms: An Introduction to Predictive Modeling in Phytogeography*. The Hague: Dr. W. Junk.
- Boyce, M. S., P. R. Vernier, S. E. Nielsen & F. K. A. Schmiegelow, 2002. Evaluating resource selection functions. *Ecological Modelling*, 157: 281–300.
- Brosse, S., J.F. Guegan, J. Tourenq, S. Lek, 1999. The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. *Ecological Modelling*, 120 (1999) 299–311

- Brotons, L., W. Thuiller, M. B. Araujo & A. H. Hirzel, 2004. Presence–absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography* 27: 437–448.
- Burman, P., E. Chow & D. Nolan, 1994. A Cross-Validatory Method for Dependent Data. *Biometrika* 81(2): 351–358.
- Busby, J. R., 1991. BIOCLIM – A bioclimate analysis and prediction system. In Margules, C. R. & M. P. Austin (eds), *Nature Conservation: Cost effective biological surveys and data analysis*. CSIRO, Australia: 64–68.
- Cade, B. S. & B. R. Noon, 2003. A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, 1: 412–420.
- Cade, B. S., B. R. Noon & C. H. Flather, 2005. Quantile regression reveals hidden bias and uncertainty in habitat models. *Ecology*, 86: 786–800.
- Cade, B. S., J. W. Terrell & R. L. Schroeder, 1999. Estimating effects of limiting factors with regression quantiles. *Ecology*, 80: 311–323.
- Carroll, C., Phillips, M. K., Schumaker, N. H., & Smith, D. W., 2003. Impacts of landscape change on wolf restoration success: Planning a reintroduction program based on static and dynamic spatial models. *Conservation Biology*, 17, 536–548.
- Carpenter, G., A. N. Gillison & J. Winter, 1993. DOMAIN: A flexible modeling procedure for mapping potential distributions of animals and plants. *Biodiversity and Conservation*, 2: 667–680.
- Caruana, R. & A. Niculescu-Mizil, 2006. An empirical comparison of supervised learning algorithms. *Proceedings of International Conference on Machine Learning*, 23rd, Pittsburgh, PA.
- Chen, I. C., P. F. Lee & W. N. Tzeng, 2005. Distribution of albacore (*Thunnus alalunga*) in the Indian Ocean and its relation to environmental factors. *Fisheries Oceanography*, 14: 71–80.
- Cleveland, W.S., 1994. *The Elements of Graphing Data*, Hobart Press ISBN 0-9634884-1-4
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20: 37–46.
- Cristianini, N. & J. Shawe-Taylor, 2000. *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press: London.
- Cumming, G. S., 2000. Using habitat models to map diversity: pan-African species richness of ticks (Acari: Ixodida). *Journal of Biogeography*, 27, 425–440.
- Daskalov, G. M, D. C. Boyer & J. P. Roux, 2003. Relating sardine *Sardinops sagax* abundance to environmental indices in northern Benguela. *Progress in Oceanography*, 59: 257–274.
- Dormann, C.F., J. M. McPherson, M. B. Araujo, R. Bivand & J. Bolliger, 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30: 609–628.
- Dray, S., P. Legendre & P.R. Peres-Neto, 2006. Spatial modeling: a comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM). *Ecological Modelling*, 196: 483–493.

- Dunk, J. R., Zielinski, W. J., & Preisler, H. K., 2004. Predicting the occurrence of rare mollusks in northern California forests. *Ecological Applications*, 14, 713–729.
- Eastwood, P. D., G. J. Meaden & A. Grioche, 2001. Modelling spatial variations in spawning habitat suitability for the sole *Solea solea* using regression quantiles and GIS procedures. *Marine Ecology Progress Series*, 224: 251–266.
- Edwards, T. C., Jr., Cutler, D. R., Zimmermann, N. E., Geiser, L., & Moisen, G. G., 2006. Effects of sample survey design on the accuracy of classification tree models in species distribution models. *Ecological Modelling*, 199, 132–141.
- Elith, J. & J. R. Leathwick, 2009. Species Distribution Models: Ecological Explanation and Prediction across Space and Time. *Annual Review of Ecology Evolution and Systematics* 40: 677-697.
- Elith, J. & M. A. Burgman, 2002. Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. In Scott, J. M. (ed), *Predicting species occurrences: issues of accuracy and scale*. Island Press: 303-314.
- Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. Mc C. Overton, A. T. Peterson, S. J. Phillips, K. S. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz & N. E. Zimmermann, 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151.
- Elton, C., 1927. *Animal Ecology*. Chicago: University of Chicago Press.
- Ferrier, M. D., G. Manion & G. Watson, 2002. Extended statistical approaches to modelling spatial pattern in biodiversity: the north-east New SouthWales experience. I. Species-level modelling. *Biodiversity and Conservation*, 11: 2275-2307.
- Ferrier, S. & G. Watson, 1997. An evaluation of the effectiveness of environmental surrogates and modelling techniques in predicting the distribution of biological diversity. *Environment Australia*, Canberra: <http://www.deh.gov.au/biodiversity/publications/technical/surrogates/>
- Fiedler, P. C. & H. J. Bernard, 1987. Tuna aggregation and feeding near fronts observed in satellite imagery. *Continental Shelf Research*, 7: 871–881
- Fielding, A. H. & J. F. Bell, 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24: 38–49.
- Francis, M. P., M. A. Morrison, J. Leathwick, C. Walsh & C. Middleton, 2005. Predictive models of small fish presence and abundance in northern New Zealand harbours. *Estuarine, Coastal and Shelf Science*, 64: 419–435.
- Franklin J. 2009. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge, UK: Cambridge Univ. Press.
- Franklin, J., 1995. Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, 19, 474–499.

- Friedman, J. H. & J. J. Meulman, 2003. Multiple adaptive regression trees with application in epidemiology. *Statistics in Medicine*, 22: 1365–1381.
- Friedman, J. H., 1991. Multivariate adaptive regression splines. *Ann. Stat.* 19, 1–141.
- Fu, J. W., R. J. Carroll & S. Wang, 2005. Estimating misclassification error with small samples via bootstrap cross-validation. *Bioinformatics*, 21: 1979–1986.
- García, A. & I. Palomera, 1996. Anchovy early life history and its relation to its surrounding environment in the Western Mediterranean basin. *Scientia Marina*, 60(2): 155-166.
- Georgakarakos, S. & D. Kitsiou, 2008. Mapping abundance distribution of small pelagic species applying hydroacoustics and Co-Kriging techniques. *Hydrobiologia* 612(1): 155-169.
- Giannoulaki M., Pyrounaki M.M., Liorzou B., Leonori I., Valavanis V.D., Tsagarakis K., Bigot J.L., Roos D., De Felice A., Campanella F., Somarakis S., Arneri E, Machias A., 2011. Habitat suitability modelling for sardine juveniles in the Mediterranean Sea. *Fisheries Oceanography*, 20(5), 367–382.
- Giannoulaki, M., A. Machias & N. Tsimenides, 1999. Ambient luminance and vertical migration of the sardine *Sardina pilchardus*. *Marine Ecology Progress Series* 178: 29-38.
- Giannoulaki, M., V. D. Valavanis, A. Palialexis, K. Tsagarakis, A. Machias, S. Somarakis & C. Papaconstantinou, 2008. Modelling the presence of anchovy *Engraulis encrasicolus* in the Aegean Sea during early summer, based on satellite environmental data. *Hydrobiologia*, 612(1): 225-240.
- Gibson, L. A., Wilson, B. A., Cahill, D. M., & Hill, J., 2004. Modelling habitat suitability of the swamp antechinus (*Antechinus minimus maritimus*) in the coastal heathlands of southern Victoria, Australia. *Biological Conservation*, 117, 143–150.
- Gower, J.C. & P. Legendre, 1986. Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification*, 3(1): 5-48.
- Graham, C. H., C. Moritz, & S. E. Williams, 2006. Habitat history improves prediction of biodiversity in a rainforest fauna. *Proceedings of the National Academy of Sciences USA*, 103: 632-636.
- Gudrun, C. & I. Kühn, 2008 Analyzing spatial ecological data using linear regression and wavelet analysis. *Stoch. Environ. Res. Risk Assess.* 22: 315–324.
- Guisan, A., Lehmann, A., Ferrier, S. *et al.*, 2006. Making better biogeographical predictions of species distributions. *Journal of Applied Ecology*, 43, 386–392.
- Guisan, A & W. Thuiller, 2005. Predicting species distribution: offering more than simple habitat models. *Ecological Letters* 8: 993–1009.
- Guisan, A. & N. E. Zimmermann, 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, 135: 147–186.
- Guisan, A., J. Edwards, C. Thomas & T. Hastie, 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 157: 89–100.
- Hanski, I., 1999. *Metapopulation Ecology*. Oxford, UK: Oxford University Press.

- Hastie, T. & R. Tibshirani, 1990. Generalized additive models. Chapman & Hall, London.
- Hastie, T. J., R. Tibshirani & A. Buja, 1994. Flexible Discriminant Analysis by Optimal Scoring. *JASA*, 89: 1255-1270.
- Hastie, T., R. Tibshirani & J. Friedman, 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics).
- Haykin, S., 1994. Neural Networks: A Comprehensive Foundation, NY: Macmillan.
- Heikkinen, R. K., Luoto, M., Araujo, M. B., Virkkala, R., Thuiller, W., & Sykes, M. T., 2006. Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, 30, 751–777.
- Hiddink, J. G., 2005. Implications of Liebig’s law of the minimum for the use of ecological indicators based on abundance. *Ecography*, 28: 264–271.
- Hirzel, A. H. & A. Guisan, 2002. Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, 157: 331–341.
- Hirzel, A. H., Hausser, J., Chessel, D., & Perrin, N., 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology*, 83, 2027–2036.
- Hirzel, A. H., Posse, B., Oggier, P. A., Crettenand, Y., Glenz, C., & Arlettaz, R., 2004. Ecological requirements of reintroduced species and the implications for release policy: the case of the bearded vulture. *Journal of Applied Ecology*, 41, 1103–1116.
- Hirzel, A. H. & Le Lay, G., 2008. Habitat suitability modelling and niche theory. *Journal of Applied Ecology*, 45, 1372–1381.
- Holland, J. H., 1975, *Adaptation in Natural and Artificial Systems* (Ann Arbor: University of Michigan Press).
- Huisman, J. & F. J. Weissing, 2002. Oscillations and chaos generated by competition for interactively essential resources. *Ecological Research*, 17: 175–181.
- Hutchinson, M. F., 1987. Methods for generation of weather sequences. In Bunting, A. H. (Ed.) *Agricultural environments: characterisation, classification and mapping*. Wallingford: CAB, International, 149–157.
- Isaaks, E. H. & R. M. Srivastava, 1989. *Applied Geostatistics*. Oxford University Press, New York.
- Jaberg, C. & A. Guisan, 2001. Modelling the distribution of bats in relation to landscape structure in a temperate mountain environment. *Journal of Applied Ecology*, 38: 1169–1181.
- Jaynes, E.T., 1957. Information theory and statistical mechanics. *Physics Revisions* 106: 620–630.
- Jiménez-Valverde, A., Lobo, J. M., & Hortal, J. (2008) Not as good as they seem: the importance of concepts in species distribution modeling. *Diversity and Distributions*, 14, 885–890.
- Keating, K. A. & Cherry, S., 2004. Use and interpretation of logistic regression in habitat selection studies. *Journal of Wildlife Management*, 68, 774–789.

- Kearney, M., 2006. Habitat, environment and niche: what are we modelling? *Oikos*, 115, 186–191.
- Keitt, T. H., O. N. Bjornstad, P. M. Dixon & S. Citron-Pousty, 2002. Accounting for spatial pattern when modelling organism–environment interactions. *Ecography*, 25: 616–625.
- Kourafalou, V. & K. Tsiaras, 2007. A nested circulation model for the North Aegean Sea. *Ocean Science* 3: 1–16.
- Lawler, J. J., Shafer, S. L., White, D., 2009. Projected climate-induced faunal changes in the Western Hemisphere. *Ecology*, 90, 588–597.
- Laurs, R. M., P. C. Fiedler & D. R. Montgomery, 1984. Albacore tuna catch distributions relative to environmental features observed from satellites. *Deep-Sea Research*, 31: 1085–1099.
- Lawrence, S., A. C. Tsoi & A. D. Back, 1996. Function Approximation with Neural Networks and Local Methods: Bias, Variance and Smoothness. Australian Conference on Neural Networks; Australian National University: Australian National University, pp 16-21.
- Leathwick, J. R., D. Rowe, J. Richardson, J. Elith & T. Hastie, 2005. Using multivariate adaptive regression splines to predict the distributions of New Zealand’s freshwater diadromous fish. *Freshwater Biology*, 50: 2034–2052.
- Leathwick, J. R., J. Elith & T. Hastie, 2006b. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological Modelling*, 199: 188–196.
- Leathwick, J. R., J. Elith, M. P. Francis, T. Hastie & P. Taylor, 2006a. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series*, 321: 267–281.
- Lefkaditou, E., C-Y. Politou, A. Palialexis, J. Dokos, P. Cosmopoulos & V. D. Valavanis, 2008. Influences of environmental variability on the population structure and distribution patterns of the short-fin squid *Illex coindetii* (Cephalopoda: Ommastrephidae) in the Eastern Ionian Sea. *Hydrobiologia*, 612 (1): 71–90.
- Legendre, P. & L. Legendre, 1998. Numerical ecology. 2nd English edition. Elsevier Science BV, Amsterdam.
- Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74:1659–1673.
- Lehmann, A., C. Overton & J. R. Leathwick, 2002. GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling*, 157: 189–207.
- Lenihan, J. M., 1993. Ecological response surfaces for North American boreal tree species and their use in forest classification. *Journal of Vegetation Science*, 4, 667–680.
- Levins, R., 1966. The strategy of model building in population ecology. *American Scientist* 54: 421–431.
- Liu, C., P. M. Berry, T. P. Dawson & R. G. Pearson, 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28: 385–393.

- MacCall, A. D., 1990. Dynamic geography of marine fish populations. University of Washington Press, Seattle. 153p.
- MacLennan, D. N., Fernandes P. G., Dalen J., 2002. A consistent approach to definitions and symbols in fisheries acoustics. *ICES Journal of Marine Science*. 59: 365–369.
- MacLeod D., C. L. Mandleberg, C. Schweder, S. M. Bannon & G. J. Pierce, 2008. A comparison of approaches for modelling the occurrence of marine animals. *Hydrobiologia*, 612(1): 21-32.
- Manel, S., Dias, J. M., & Ormerod, S. J., 1999b. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling*, 120, 337–347.
- Manly, B. F. J., McDonald, L. L., Thomas, D. L., McDonald, T. L., & Erickson, W., 2002. *Resource Selection by Animals: Statistical Design and Analysis for Field Studies*. New York, NY, USA: Kluwer Press.
- Maravelias, C. D., E. V. Tsitsika & C. Papaconstantinou, 2007. Evidence of Morocco dentex (*Dentex maroccanus*) distribution in the NE Mediterranean and relationships with environmental factors determined by Generalized Additive Modelling. *Fisheries Oceanography*, 16: 294–302.
- Maravelias C.D. & D.G. Reid, 1997. Identifying the effects of oceanographic features and zooplankton on prespawning herring abundance using generalized additive models. *Marine Ecology Progress Series*, 147: 1-9.
- Maravelias, C. D., D. G. Reid, E. J. Simmonds & J. Haralabous, 1996. Spatial Analysis and Mapping of Acoustic Survey Data in the presence of High Local Variability: Geostatistical Application to North Sea Herring (*Clupea harengus*). *Canadian Journal of Fisheries and Aquatic Sciences*, 53: 1497–1505.
- Martin, P., N. Bahamon, A. Sabates, F. Maynou, P. Sanchez & M. Demestre, 2008. European anchovy (*Engraulis encrasicolus*) landings and environmental conditions on the Catalan Coast (NW Mediterranean) during 2000-2005. *Hydrobiologia*, 612(1): 185-199.
- Martínez-Meyer, E., Peterson, A. T., Servin, J. I., & Kiff, L. F. (2006) Ecological niche modelling and prioritizing areas for species reintroductions. *Oryx*, 40, 411–418.
- Matheron, G., 1971. The theory of regionalized variables and its applications. Ecole Nationale Supérieure des Mines de Paris, Fontainebleau.
- Meentemeyer, R. K., Moody, A., & Franklin, J., 2001. Landscape-scale patterns of shrub-species abundance in California chaparral: the role of topographically mediated resource gradients. *Plant Ecology*, 156, 19–41.
- Michie, D. J. & C. C. Spiegelhalter, 1994. *Machine Learning, Neural and Statistical Classification*, Taylor.
- Moguerza, J. & A. Muñoz, 2006. Support vector machines with applications. *Statistical Science* 21(3): 322-336.
- Moisen, G. G. & T. S. Frescino, 2002. Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling*, 157: 209–225.
- Moran, P. A. P., 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37: 17–23.

- Morrison, M. L., B. G. Marcot & R. W. Mannan, 1992. *Wildlife Habitat Relationships. Concepts and Applications*. Madison, WI, USA: University Wisconsin Press: 341p.
- Motos, L., A. Uriarte & V. València, 1996. The spawning environment of the Bay Biscay anchovy (*Engraulis encrasicolus* L.). *Scientia Marina*, 60: 117–140.
- Muiño, R., P. Carrera, P. Petitgas, D. J. Beare, S. Georgakarakos, J. Haralambous, M. Iglesias, B. Liorzou, J. Masse & D. G. Reid, 2003. Consistency in the correlation of school parameters across years and stocks. *ICES Journal of Marine Science*, 60: 164–175.
- Muñoz, J. & A. M. Fellicisimo, 2004. Comparison of statistical methods commonly used in predictive modeling. *Journal of Vegetation Science*, 15: 285-292.
- Muñoz, M. E. S., R. Giovanni, M. F. Siqueira, T. Sutton, P. Brewer, R. S. Pereira, D. A. L. Canhos & V. P. Canhos, 2009. openModeller: a generic approach to species' potential distribution modelling. *GeoInformatica*. DOI: 10.1007/s10707-009-0090-7.
- Murphy, A. H. & R. L. Winkler, 1992. Diagnostic verification of probability forecasts. *International Journal of Forecasting*, 7: 435-455.
- Nix, H. A., 1986. A biogeographic analysis of Australian elapid snakes. In Longmore, R. (ed), *Atlas of Elapid Snakes of Australia*. pp. 4-15. Australian Flora and Fauna Series 7. Australian Government Publishing Service: Canberra.
- Olivier, F. & S. J. Wotherspoon, 2005. GIS-based application of resource selection functions to the prediction of snow petrel distribution and abundance in East Antarctica: comparing models at multiple scales. *Ecological Modelling*, 189: 105–129.
- Ostrom, E., 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. New York: Cambridge University Press.
- Palialexis, A., S. Georgakarakos, I. Karakassis, K. Lika and V. D. Valavanis, 2011. Fish distribution predictions from different points of view: Comparing Associative Neural Networks, Geostatistics and Regression Models. *Hydrobiologia*, 670, 275–287.
- Palialexis, A., S. Georgakarakos, K. Lika & V. D. Valavanis, 2009. Use of GIS, remote sensing and regression models for the identification and forecast of small pelagic fish distribution. *Proceedings of the Second International Conference on Environmental Management, Engineering, Planning and Economics (CEMEPE 09)*. June 21-26, Mykonos, Greece.
- Palialexis, A., S. Georgakarakos, K. Lika & V. D. Valavanis, 2009a. Comparing novel approaches used for prediction of species distribution from presence/absence acoustic data. *Proceedings of the Second International Conference on Environmental Management, Engineering, Planning and Economics (CEMEPE 09)*. June 21-26, 2009, Mykonos, Greece.
- Palomera, I., M. P. Olivar, J. Salat, A. Sabates, M. Coll, A. Garcia & B. Morales-Nin, 2007. Small pelagic in the NW Mediterranean Sea: An ecological review. *Progress in Oceanography*, 74: 377–396.
- Pearce, J. & S. Ferrier, 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, 133: 225–245.

- Pearce, J. & Lindenmayer, D., 1998. Bioclimatic analysis to enhance reintroduction biology of the endangered helmeted honeyeater (*Lichenostomus melanops cassidix*) in southeastern Australia. *Restoration Ecology*, 6, 238–243.
- Pearson, R. G., Thuiller, W., Araújo, M. B. *et al.*, 2006. Model-based uncertainty in species range prediction. *Journal of Biogeography*, 33, 1704–1711.
- Peterson, A. T. & Vieglais, D. A., 2001. Predicting species invasions using ecological niche modeling: New approaches from bioinformatics attack a pressing problem. *Bioscience*, 51, 363–371.
- Peterson, A. T., 2003. Predicting the geography of species invasions via ecological niche modeling. *Quarterly Review of Biology*, 78, 419–433.
- Peterson, A. T., Papes, M., & Soberón, J., 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, 213, 63–72.
- Petitgas, P., 2001. Geostatistics in fisheries survey design and stock assessment: models, variances and applications. *Fish and Fisheries*, 2: 231–249.
- Petitgas, P., D. Reid, P. Carrera, M. Iglesias, S. Georgakarakos, B. Liorzou & J. Masse, 2001. On the relation between schools, clusters of schools, and abundance in pelagic fish stocks. *ICES Journal of Marine Research*, 58: 1150–1160.
- Phillips, S. J., M. Dudik & R. E. Schapire, 2004. A maximum entropy approach to species distribution modeling. In *Proceedings of the Twenty-First International Conference on Machine Learning*: 655–662.
- Phillips, S. J., R. P. Anderson & R. E. Schapire, 2006. Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, 190: 231–259.
- Piñeiro, R., J. F. Aguilar, D. D. Munt & G. N. Feliner, 2007. Ecology matters: Atlantic-Mediterranean disjunction in the sand-dune shrub *Armeria pungens* (Plumbaginaceae). *Molecular Ecology*, 16: 2155–2171.
- Planque, B., E. Bellier & P. Lazure, 2007. Modelling potential spawning habitat of sardine (*Sardina pilchardus*) and anchovy (*Engraulis encrasicolus*) in the Bay of Biscay. *Fisheries Oceanography* 16 (1): 16–30.
- Pollock, J. F., 2006. Detecting population declines over large areas with presence–absence, time-to-encounter, and count survey methods. *Conservation Biology*, 20: 882–892.
- Potts, J. M. & J. Elith, 2006. Comparing species abundance models. *Ecological Modelling*, 199: 153–163.
- Poulos, S. E., G. T. Chronis, M. B. Collins & V. Lykousis, 2000. Thermaikos Gulf Coastal System, NW Aegean Sea: an overview of water/sediment fluxes in relation to airland-ocean interactions and human activities. *Journal of Marine Systems*, 25: 47–76.
- Pulliam, H. R., 1988. Sources, sinks, and population regulation. *American Naturalist*, 132, 652–661.
- Pulliam, H. R., 2000. On the relationship between niche and distribution. *Ecology Letters*, 3, 349–361.
- R Development Core Team, 2005. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. <http://www.Rproject.org>.

- Racine, J., 2000. Consistent Cross-Validatory Model-Selection for Dependent Data: hv-Block Cross-Validation. *Journal of Econometrics*, 99: 39–61.
- Ramzi, A., My L. Hbid & O. Ettahiri, 2006. Larval dynamics and recruitment modelling of the Moroccan Atlantic coast sardine (*Sardina pilchardus*). *Ecological Modelling*, 197: 296–302.
- Ready, J., K. Kaschner, A. B. South, P. D. Eastwood, T. Rees, J. Rius, E. Agbayani, S. Kullander & R. Froese, 2010. Predicting the distributions of marine organisms at the global scale, *Ecological Modelling*, 221(3): 467-478.
- Redfern, J.V., M. C. Ferguson, E. A. Becker, K. D. Hyrenbach, C. Good, J. Barlow, K. Kaschner, M. F. Baumgartner, K. A. Forney, L. T. Ballance, P. Fauchald, P. Halpin, T. Hamazaki, A. J. Pershing, S. S. Qian, A. Read, S. B. Reilly, L. Torres & F. Werner, 2006. Techniques for cetacean–habitat modeling: a review. *Marine Ecology Progress Series*, 310: 271–295.
- Reese, G. C., Wilson, K. R., Hoeting, J. A., & Flather, C. H., 2005. Factors affecting species distribution predictions: a simulation modeling experiment. *Ecological Applications*, 15, 554–564.
- Richards, C. L., B. C. Carstens & L. Knowles, 2007. Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *Journal of Biogeography*, 34: 1833–1845.
- Rodriguèz, J. P., Brotons, L., Bustmante, J., & Seoane, J., 2007. The application of predictive modelling of species distribution to biodiversity conservation. *Diversity and Distributions*, 13, 243–251.
- Rossi, R. E., D. J. Mula, A. G. Journel & E. H. Franz, 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs*, 2: 277–314.
- Rotenberry, J. T., Preston, K. L., & Knick, S. T., 2006. GIS-based niche modeling for mapping species habitat. *Ecology*, 87, 1458–1464.
- Rubio, G., J. Zhu & J. P. Lynch, 2003. A critical test of the two prevailing theories of plant response to nutrient availability. *American Journal of Botany*, 90: 143–152.
- Ruiz, J., E. Garcia-Isarch, I. E. Huertas, L. Prieto, A. Juárez, J. L. Munõz, A. Sánchez-Lamadrid, S. Rodríguez-Gálvez, J. M. Naranjo & F. Baldó, 2006. Meteorological and oceanographic factors influencing *Engraulis encrasicolus* early life stages and catches in the Gulf of Cádiz. *Deep-Sea Research II*, 53: 1363–1376.
- Rushton, S. P., Ormerod, S. J., & Kerby, G., 2004. New paradigms for modelling species distributions? *Journal of Applied Ecology*, 41, 193–200.
- Sabates, A., M. P. Olivar, J. Salat, I. Palomera & F. Alemany, 2007. Physical and biological processes controlling the distribution of fish larvae in the NW Mediterranean. *Progress in Oceanography*, 74: 355–376.
- Sabatés, A., P. Martìn, J. Lloret & V. Raya, 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.

- Santos, A. M. P., A. Peliz, J. Dubert, P. B. Oliveira, M. M. Angélico & P. Ré, 2004. Impact of a winter upwelling event on the distribution and transport of sardine (*Sardina pilchardus*) eggs and larvae off western Iberia: a retention mechanism. *Continental Shelf Research*, 24: 149–165.
- Schadt, S., Revilla, E., Wiegand, T. *et al.*, 2002. Assessing the suitability of central European landscapes for the reintroduction of Eurasian lynx. *Journal of Applied Ecology*, 39, 189–203.
- Schismenou, E., M. Giannoulaki, V. D. Valavanis & S. Somarakis, 2008. Modeling and predicting potential spawning habitat of anchovy (*Engraulis encrasicolus*) and round sardinella (*Sardinella aurita*) based on satellite environmental information. *Hydrobiologia*, 612(1): 201-214.
- Schölkopf, B., A. Smola, R. Williamson & P. L. Bartlett, 2000. New support vector algorithms. *Neural Computation*, 12: 1207-1245.
- Schröder, B. 2008. Challenges of species distribution modeling belowground. *Journal of Plant Nutrition and Soil Science*, 171: 325–337.
- Scott, J. M., Heglund, P. J., Morrison, M. L. *et al.* (Eds.), 2002. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Covelo, CA: Island Press.
- Segurado, P. & M. B. Araujo, 2004. An evaluation of methods for modelling species distributions. *Journal of Biogeography*, 31: 1555-1568.
- Shepherd, A. J., 1997. *Second-Order Methods for Neural Networks*. Springer-Verlag, London, p 145.
- Siapatis, A., M. Giannoulaki, V. D. Valavanis, A. Palialexis, E. Schismenou, A. Machias & S. Somarakis, 2008. Modelling potential habitat of the invasive ctenophore *Mnemiopsis leidyi* in Aegean Sea. *Hydrobiologia*, 612 (1): 281-295.
- Simmonds, E. J. & D. N. MacLennan, 2005. *Fisheries Acoustics: Theory and Practice*. Blackwell Science Ltd, Oxford.
- Simon, R., M. D. Radmacher, K. Dobbin & L. M. McShane, 2003. Pitfalls in the Use of DNA Microarray Data for Diagnostic and Prognostic Classification. *Journal of the National Cancer Institute*, 95(1): 14-18.
- Soberón, J., 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, 10, 1115–1123.
- Soberón, J. & Peterson, A. T. (2004) Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 359, 689–698.
- Somarakis, S., P. Drakopoulos & V. Filippou, 2002. Distribution and abundance of larval fishes in the northern Aegean Sea -Eastern Mediterranean- in relation to early summer oceanographic conditions. *Journal of Plankton Research*, 24: 339–357.
- Stergiou, I. K. & A. Lascaratos, 1997. Climatic variability and the anchovy/sardine ratio in Hellenic waters. *GeoJournal*, 41(3): 245–254.

- Stockwell, D. & D. Peters, 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, 13: 143–158.
- Stockwell, D. R. B., 1999. Genetic algorithms II. In Fielding A. H. (ed). *Machine learning methods for ecological applications*. Kluwer Academic Publishers, Boston: 123-144.
- Tetko, I. V. & V. Y. Tanchuk, 2002. Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. *Journal of Chemical Information in Computer Science*, 42: 1136-1145.
- Tetko, I. V., 2002a. Associative neural network, *Neural Processing Letters* 16: 187-199.
- Tetko, I. V., 2002b. Neural network studies. 4. Introduction to associative neural networks. *Journal of Chemical Information in Computer Science*, 42: 717-28.
- Tetko, I. V., D. J. Livingstone & A. I. Luik, 1995. Neural Network Studies. 1. Comparison of Overfitting and Overtraining. *Journal of Chemical Information in Computer Science*, 35: 826-833.
- Tetko, I. V., I. Sushko, A. K. Pandey, H. Zhu, A. Tropsha, E. Papa, T. Oberg, R. Todeschini, D. Fourches & A. Varnek, 2008. Critical assessment of QSAR models of environmental toxicity against *Tetrahymena pyriformis*: focusing on applicability domain and overfitting by variable selection *Journal of Chemical Information and Modeling*, 48(9): 1733-1746.
- Thogmartin, W. E., Sauer, J. R., & Knutson, M. G., 2004. A hierarchical spatial model of avian abundance with application to Cerulean Warblers. *Ecological Applications*, 14, 1766–1779.
- Thomson, J., Weiblen, G., Thomson, B., Alfaro, S., & Legendre, P., 1996. Untangling multiple factors in spatial distributions: lilies, gophers and rocks. *Ecology*, 77, 1698–1715.
- Thuiller, W., Lavorel, S., Sykes, M. T., & Araújo, M. B., 2006. Using niche-based modelling to assess the impact of climate change on tree functional diversity in Europe. *Diversity and Distributions*, 12, 49–60.
- Thuiller, W., Brotons, L., Araújo, M. B., & Lavorel, S., 2004. Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, 27, 165–172.
- Tsagarakis, K., A. Machias, S. Somarakis, M. Giannoulaki, A. Pali Alexis & V. D. Valavanis, 2008. Habitat discrimination of juvenile sardines in the Aegean Sea using remotely sensed environmental data. *Hydrobiologia*, 612(1): 215-223.
- Tychonoff, A. N. & V. Y. Arsenin, 1977. *Solution of Ill-posed Problems*. Washington: Winston and Sons. ISBN 0-470-99124-0.
- Valavanis, V. D., 2002. *Geographic Information Systems in Oceanography and Fisheries*. Taylor & Francis, London, p 240.
- Valavanis, V. D., I. Katara & A. Pali Alexis, 2005. Marine GIS: identification of mesoscale oceanic thermal fronts. *International Journal of Geographical Information Science*, 19: 1131–1147.
- Valavanis, V. D., G. J. Pierce, A. F. Zuur, A. Pali Alexis, A. Saveliev, I. Katara & J. Wang., 2008. Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. *Hydrobiologia*, 612(1): 5-20.

- Valavanis, V. D., S. Georgakarakos, A. Kapantagakis, A. Palialexis & I. Katara, 2004. A GIS environmental modelling approach to Essential Fish Habitat Designation. *Ecological Modelling*, 178: 417–427.
- Vapnik, V., 1995. *The Nature of Statistical Learning Theory*. SpringerVerlag.
- Walline, P. D., 2007. Geostatistical simulations of eastern Bering Sea walleye pollock spatial distributions, to estimate sampling precision. *ICES Journal of Marine Science*, 64: 559–569.
- Ware, D. M. & R. E. Thomson, 2005. Bottom-Up Ecosystem Trophic Dynamics Determine Fish Production in the Northeast Pacific. *Science*, 308: 1280–1285.
- Wisz, M. S., Dendoncker, N., Madsen, J. *et al.*, 2008. Modelling pink-footed goose (*Anser brachyrhynchus*) wintering distributions for the year 2050: potential effects of land-use change in Europe. *Diversity and Distributions*, 14, 721–731.
- Whittaker, R. H., 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, 30, 279–338.
- Whittaker, R. H., Levin, S. A., & Root, R. B., 1973. Niche, habitat and ecotope. *The American Naturalist*, 107, 321–338.
- Wood, S. N. & N. H. Augustin, 2002. GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling*, 157: 157–177.
- Wood, S. N., 2006. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC Press.
- Wood, S. N., 2008. Fast stable direct fitting and smoothness selection for generalized additive models. *Journal of the Royal Statistical Society: Series B*, 70(3): 495-518.
- Wu, X. B. & Smeins, F. E., 2000. Multiple-scale habitat modeling approach for rare plant conservation. *Landscape and Urban Planning*, 51, 11–28.
- Zaniewski, A. E., A. Lehman & J. Overton, 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, 157: 261–280.
- Zheng, B. & A. Agresti, 2000. Summarizing the predictive power of a generalized linear model. *Statistics in Medicine*, 19: 1771-1781.
- Zuur, A. F., E. N. Ieno & C. S. Elphick, 2010. A protocol for data exploration to avoiding common statistical problems. *Methods in Ecology and Evolution*, 1: 3-14.
- Zuur, A. F., E. N. Ieno & G. M. Smith, 2007. *Analysing Ecological Data*. Springer Series: Statistics for Biology and Health. Zuur, A. F., E. N. Ieno & G. M. Smith, 2007. *Analysing Ecological Data*. Springer Series: Statistics for Biology and Health.
- Zwolinski, J. P., Oliveira, P. B., Quintino, V., and Stratoudakis, Y. 2010. Sardine potential habitat and environmental forcing off western Portugal. – *ICES Journal of Marine Science*, 67: 1553–1564.

Περίληψη

Η ανάγκη για επιστημονικά τεκμηριωμένες μελέτες, σχετικά με το χαρακτηρισμό των θαλάσσιων οικοσυστημάτων, τη διακύμανση της βιοποικιλότητας και αφθονίας των ειδών και τις επιπτώσεις της αλλαγής του κλίματος στα θαλάσσια οικοσυστήματα, έχει οδηγήσει στην αύξηση των εργασιών που αφορούν στην πρόβλεψη της κατανομής των ειδών. Η πρόβλεψη της κατανομής των ειδών μπορεί να διεξαχθεί με τη χρήση πολλών και διαφορετικών προσεγγίσεων για τη μοντελοποίηση των ενδιαιτημάτων και την αναγνώριση της ετερογένειας τους. Η επιλογή της κατάλληλης προσέγγισης μοντελοποίησης κατανομής ειδών έχει ιδιαίτερη σημασία στην ποιότητα των αποτελεσμάτων. Στα πλαίσια της παρούσας διατριβής, σε ένα πρώτο επίπεδο, πραγματοποιήθηκε μια ευρεία σύγκριση τεχνικών με μοντέλα που εκπαιδεύονται με δεδομένα αφθονίας. Κατά τη σύγκριση αξιολογήθηκαν τα μοντέλα κατανομής ειδών (MKE) ως προς την ικανότητα πρόβλεψης χρησιμοποιώντας δεδομένα ακουστικής πυκνότητας μικρών πελαγικών ειδών. Ειδικότερα με τα δεδομένα αφθονίας αναπτύχθηκαν γενικευμένα προσθετικά μοντέλα (Generalized Additive Models) και μεικτά μοντέλα (Generalized Additive Mixed Models) στα πλαίσια των μοντέλων παλινδρόμησης, συνδεδετικά νευρωνικά δίκτυα (Associative Neural Networks) και σύνολα τεχνιτών νευρωνικών δικτύων (Artificial Neural Networks Ensemble) στα πλαίσια των νευρωνικών δικτύων και Kriging στα πλαίσια των γεωστατιστικών προσεγγίσεων. Τα νευρωνικά δίκτυα απέδωσαν ικανοποιητικότερα από τις άλλες τεχνικές στην προσομοίωση των δειγματοληπτικών δεδομένων, ενώ τα γενικευμένα προσθετικά μοντέλα ξεχώρισαν στην πρόβλεψη του "άγνωστου" συνόλου επαλήθευσης. Σε ένα δεύτερο επίπεδο της διατριβής πραγματοποιήθηκε μια ευρεία σύγκριση τεχνικών MKE που αναπτύσσονται με δεδομένα παρουσίας-απουσίας από δειγματοληπτικά δεδομένα ακουστικής πυκνότητας. Εδώ πραγματοποιήθηκε η αξιολόγηση των MKE ως προς την ικανότητα προσαρμογής τους στα δειγματοληπτικά δεδομένα, η σύγκριση των χαρτών πρόβλεψης σε πολύ υψηλή χωρική ανάλυση και ελέγχθηκε η ικανότητα πρόβλεψης τους με τη βοήθεια ενός νέου συνόλου δεδομένων αξιολόγησης. Για τις διαδικασίες σύγκρισης των MKE χρησιμοποιήθηκαν τεχνικές όπως τα Receiver Operating Characteristics (ROC), τα στατιστικά Kappa, οι συντελεστές συσχέτισης και οι συνεκτικοί πίνακες. Τα ενισχυμένα δέντρα παλινδρόμησης (Boosted Regression Trees) και τα συνδεδετικά νευρωνικά δίκτυα, που αμφότερα ανήκουν στις τεχνικές μηχανικής εκμάθησης, ξεχώρισαν ως προς την απόδοσή τους μεταξύ των MKE που εφαρμόστηκαν. Στα στάδια ανάπτυξης και σύγκρισης των MKE αναλύονται σημαντικά ζητήματα που άπτονται της ορθής εφαρμογής των MKE, όπως η επιλογή των ανεξάρτητων μεταβλητών, η χωρική αυτοσυσχέτιση στα δεδομένα, η διαχείριση και εισαγωγή της αβεβαιότητας στα μοντέλα, η επεξήγηση της διακύμανσης των δειγματοληπτικών δεδομένων στη γεωγραφική, περιβαλλοντική και χρονική διάσταση.

Λέξεις κλειδιά: Μοντέλα Κατανομής Ειδών, μοντελοποίηση ενδιαιτημάτων, μέθοδοι σύγκρισης μοντέλων, χάρτες ενδιαιτημάτων, χάρτες πρόβλεψης, μικρά πελαγικά είδη.

Summary

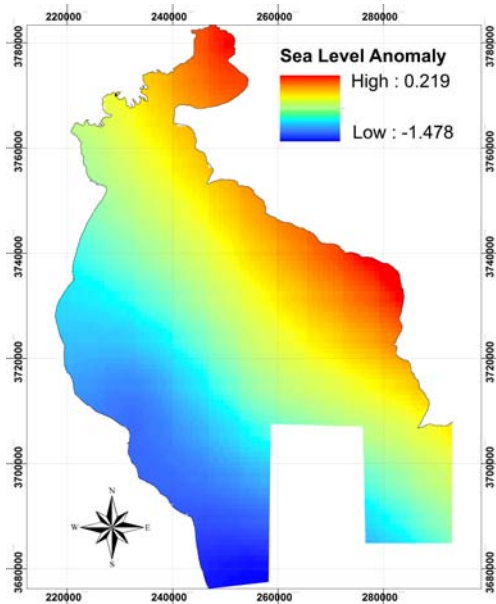
Accurate prediction of species distributions based on sampling and environmental data is essential for further scientific analysis, such as stock assessment, detection of abundance fluctuation due to climate change or overexploitation, and to underpin management and legislation processes. The evolution of computer science and statistics has allowed the development of sophisticated and well-established modelling techniques as well as a variety of promising innovative approaches for modelling species distribution. The appropriate selection of

modelling approach is crucial to the quality of predictions about species distribution. In this study, modelling techniques based on different approaches are compared and evaluated in relation to their predictive performance, utilizing fish density acoustic data. Generalized additive models and mixed models amongst the regression models, associative neural networks (ANNs) and artificial neural networks ensemble amongst the artificial neural networks and ordinary Kriging amongst the geostatistical techniques are applied and evaluated. A verification dataset is used for estimating the predictive performance of these models. A combination of outputs from the different models is applied for prediction optimization to exploit the ability of each model to explain certain aspects of variation in species acoustic density. Neural networks and especially ANNs appear to provide more accurate results in fitting the training dataset while generalized additive models appear more flexible in predicting the verification dataset. The efficiency of each technique in relation to certain sampling and output strategies is also discussed. The accurate representation of species distribution derived from sampled data is essential for management purposes and to underpin population modelling. Additionally, the prediction of species distribution for an expanded area, beyond the sampling area can reduce sampling costs. Here, several well-established and recently developed habitat modelling techniques are investigated in order to identify the most suitable approach to use with presence–absence acoustic data. The fitting efficiency of the modelling techniques are initially tested on the training dataset while their predictive capacity is evaluated using a verification set. For the comparison among models, Receiver Operating Characteristics (ROC), Kappa statistics, correlation and confusion matrices are used. Boosted Regression Trees (BRT) and Associative Neural Networks (ASNN), which are both within the machine learning category, outperformed the other modelling approaches tested.

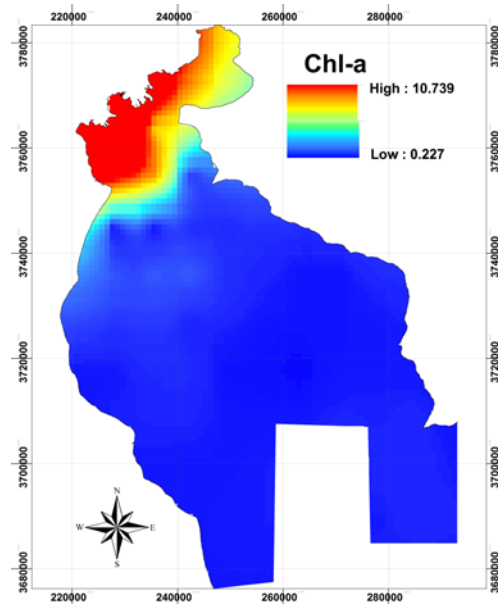
Keywords: Species distribution models, Species distribution predictions, Habitat modelling, Models comparison, Habitat maps, Spatial autocorrelation, small pelagic species.

7. Παράρτημα Ι

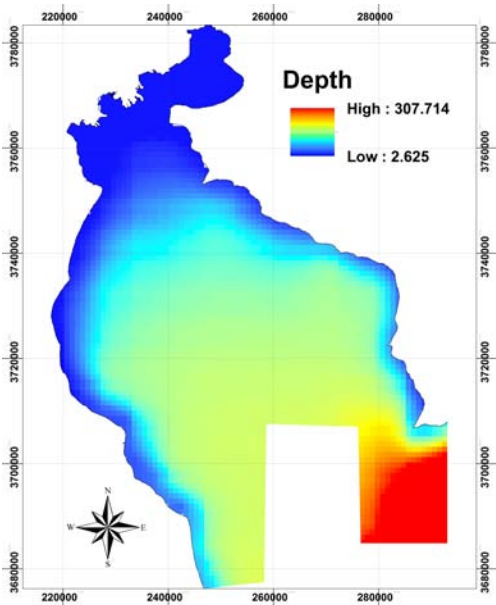
Οι καμβάδες των παραμέτρων που τελικά επελέγησαν για την εκπαίδευση των μοντέλων, όπως απεικονίζονται στους καμβάδες.



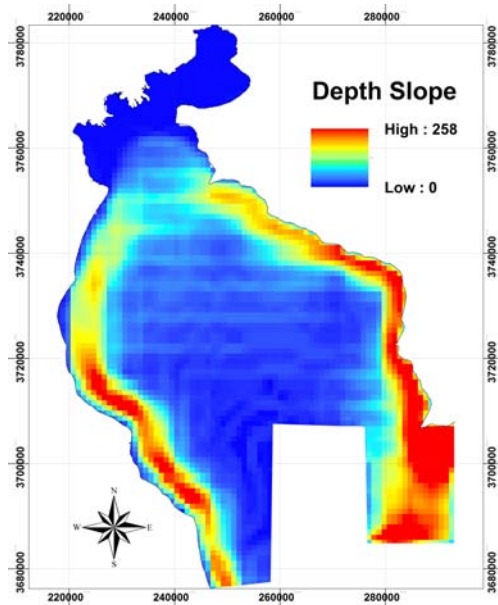
Αλιμετρία



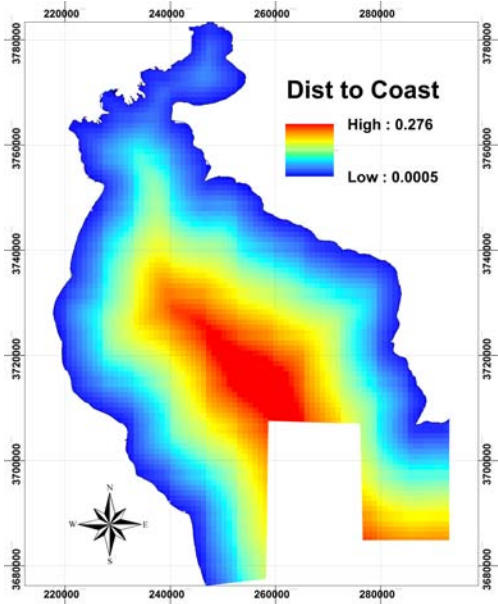
Επιφανειακή συγκέντρωση χλωροφύλλης-α



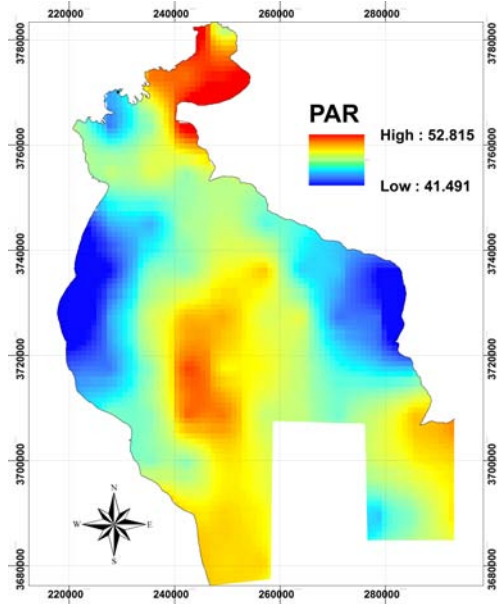
Βάθος



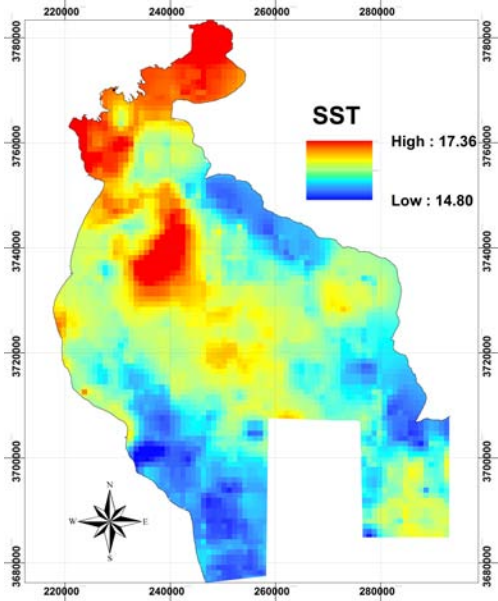
Κλίση του βυθού



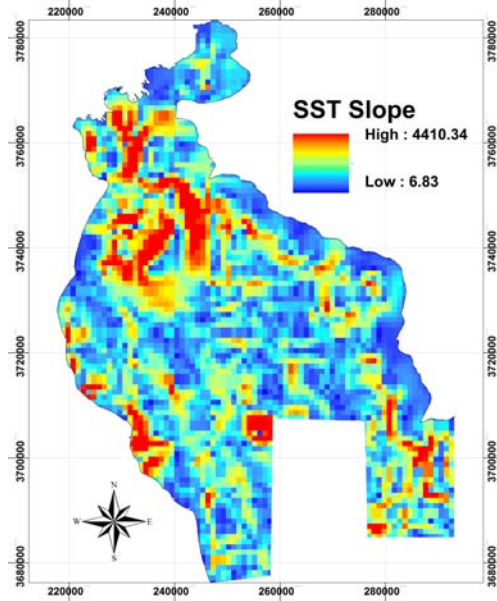
Απόσταση από την ακτή



Διαθέσιμη φωτοσυνθετική ακτινοβολία



Επιφανειακή θερμοκρασία της θάλασσας



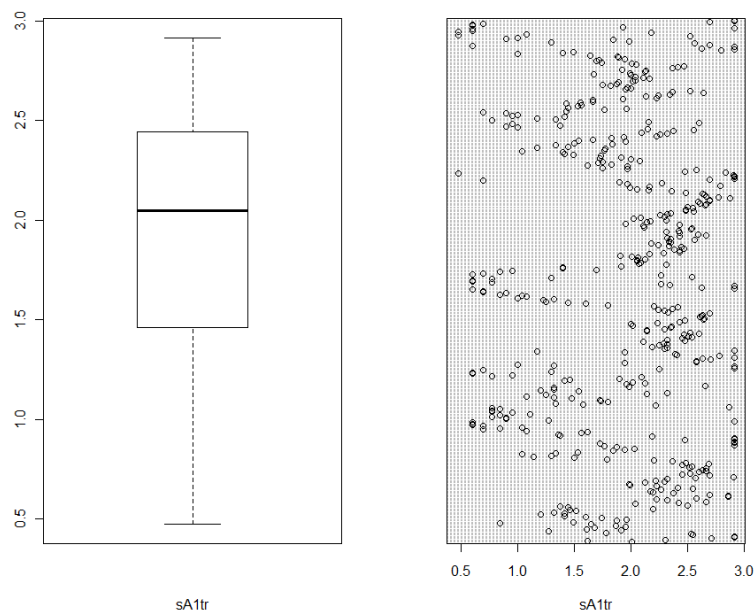
Κλίση της επιφανειακής θερμοκρασίας της θάλασσας

8. Παράρτημα II

Διερεύνηση των δεδομένων

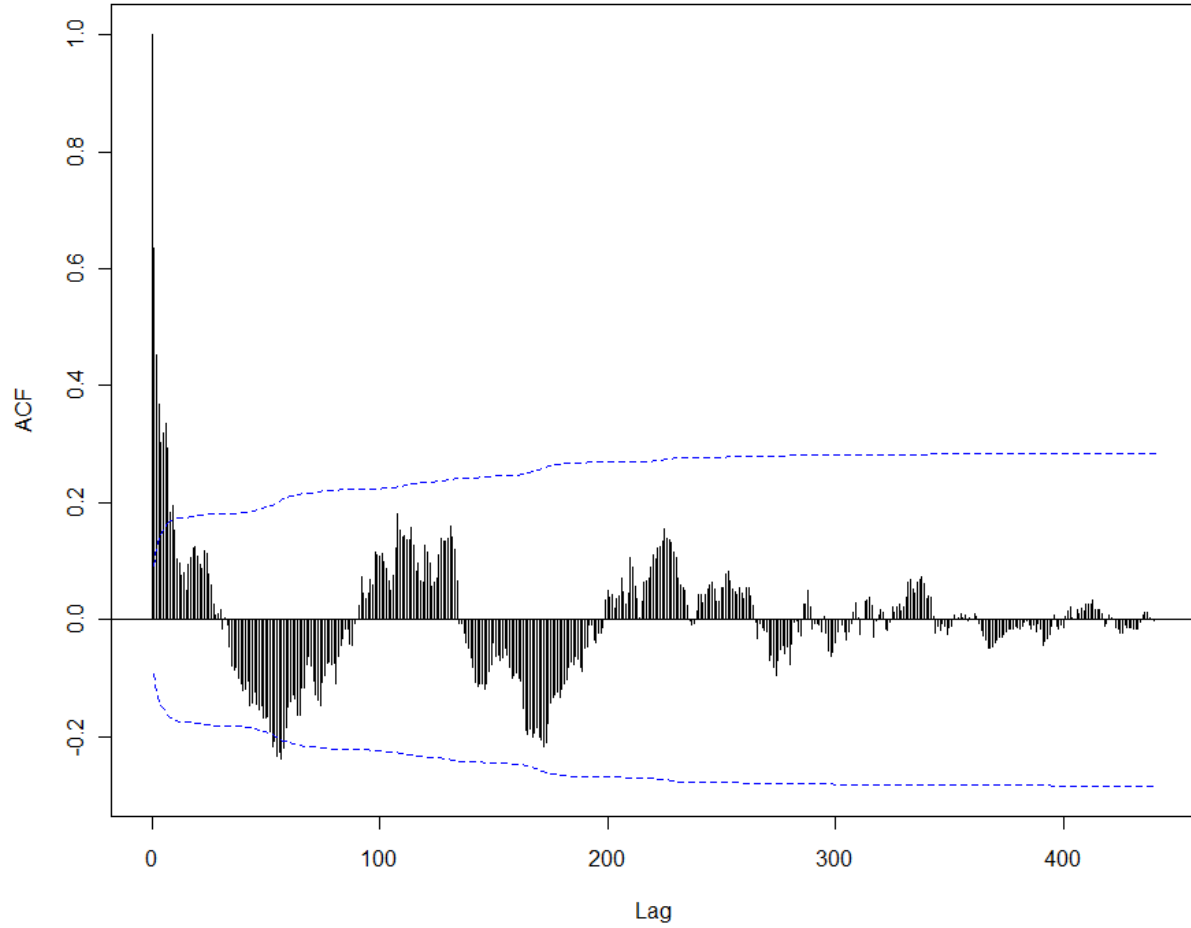
Περιγραφικά χαρακτηριστικά της ακουστικής πληροφορίας

Ελάχιστο	1 ^ο τεταρτημόριο	Διάμεσος	μέσος	3 ^ο τεταρτημόριο	μέγιστο
3.0	29.0	111.5	193.7	276.5	820.0

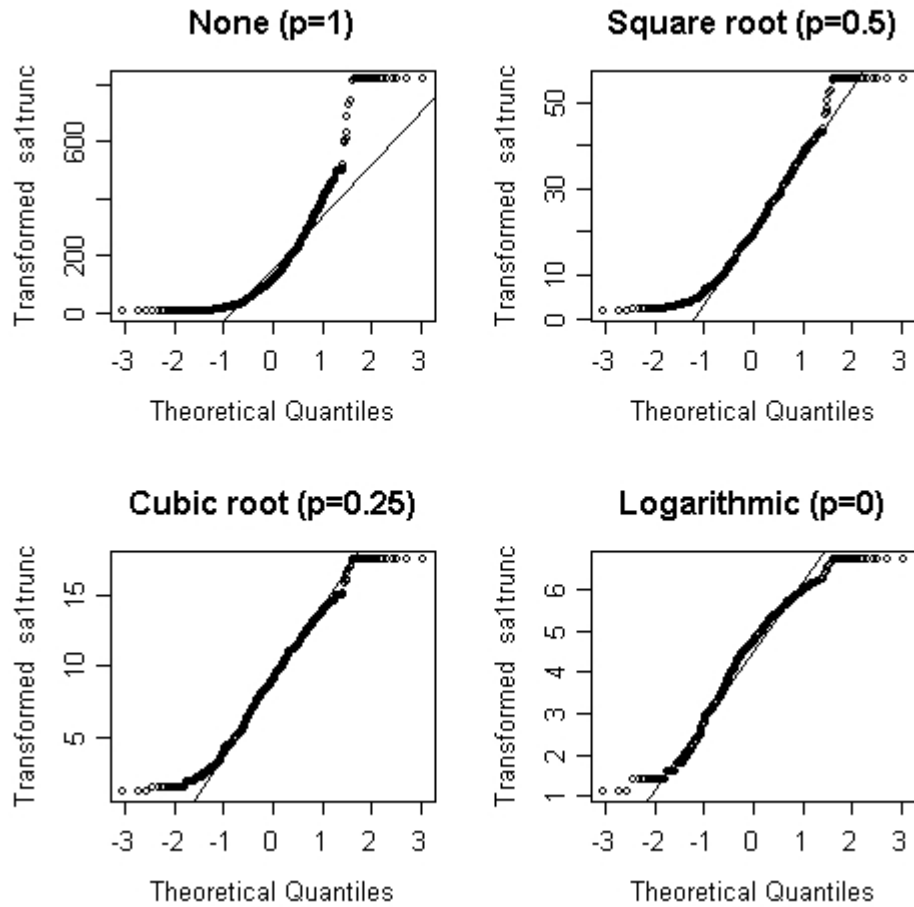


Boxplot και Dotplot της s_A για την ανίχνευση ακραίων τιμών και για την επισκόπηση της κατανομής των τιμών

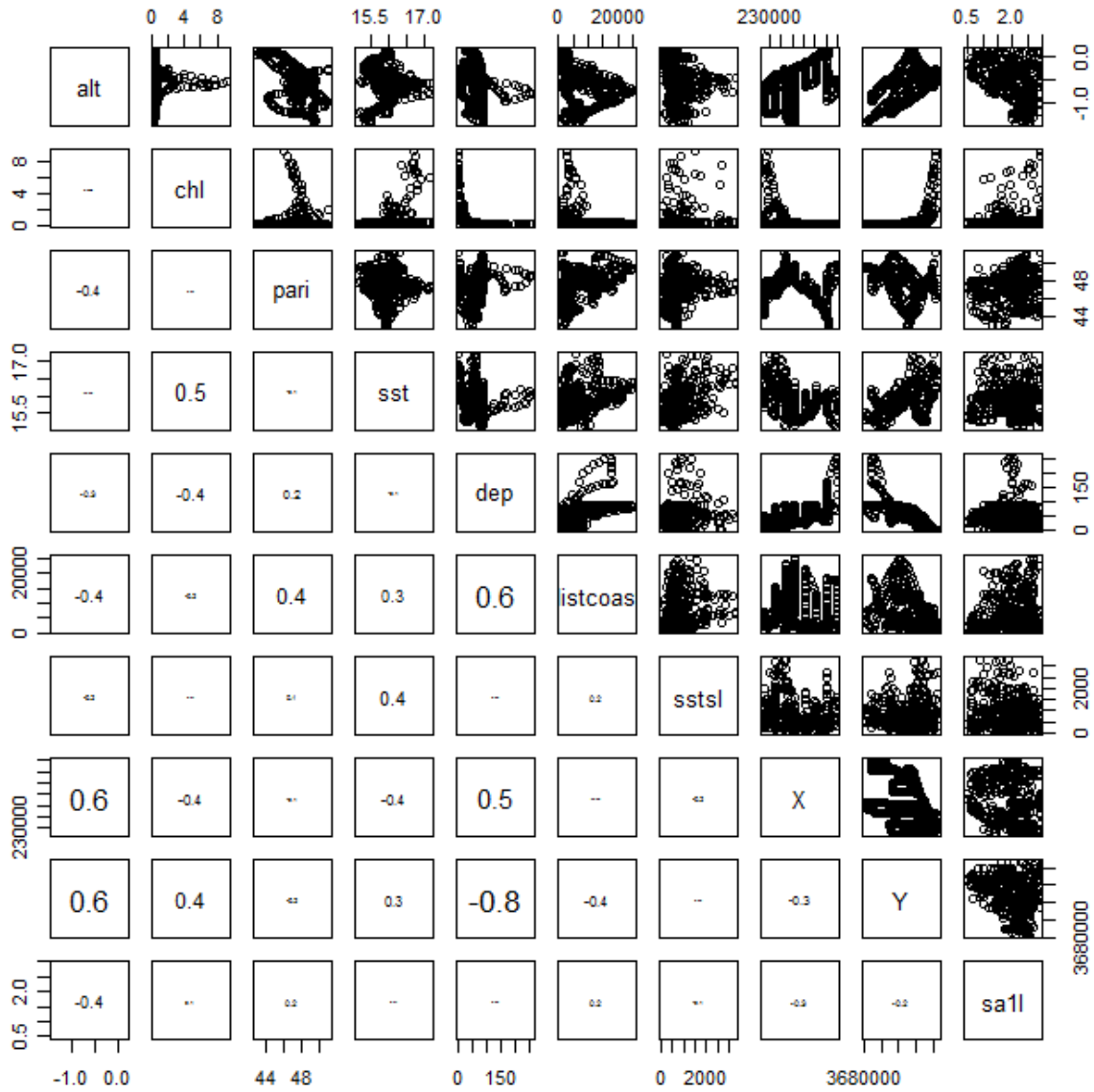
Series sA11



Γράφημα λειτουργίας αυτοσυσχέτισης για το s_A

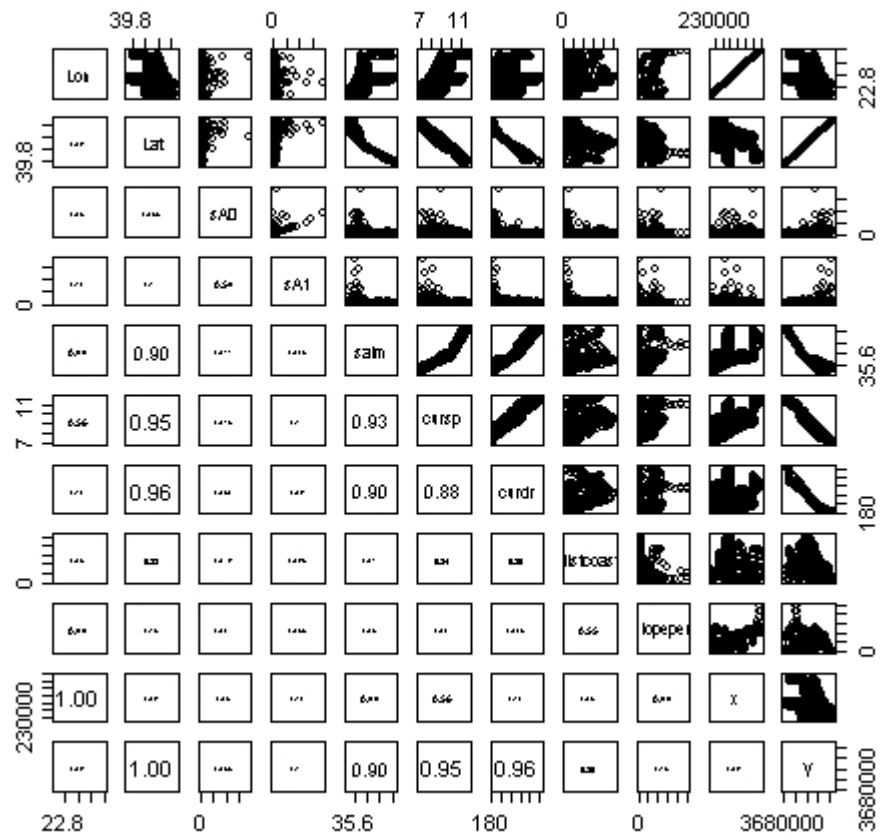


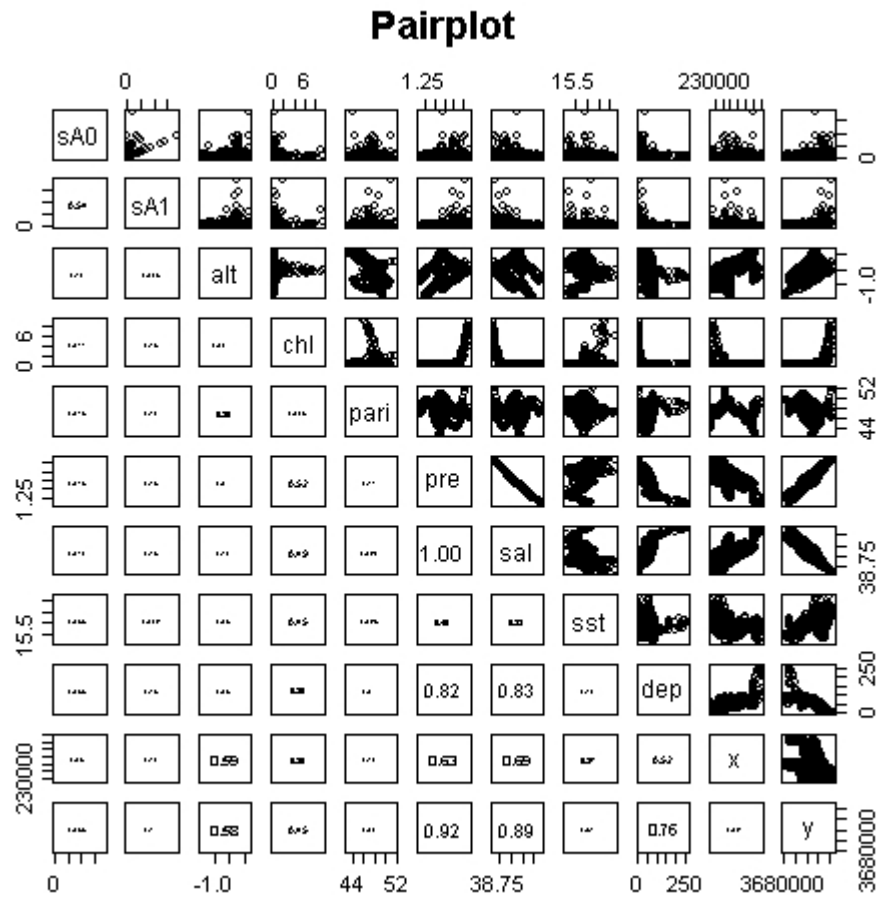
Παράλληλη έκθεση της s_A με τους αντίστοιχους μετασχηματισμούς που φαίνονται πάνω από κάθε γράφημα, ώστε να βρεθεί ο ενδεδειγμένος μετασχηματισμός που κάνει τα δεδομένα να βρίσκονται κοντά σε κανονική κατανομή



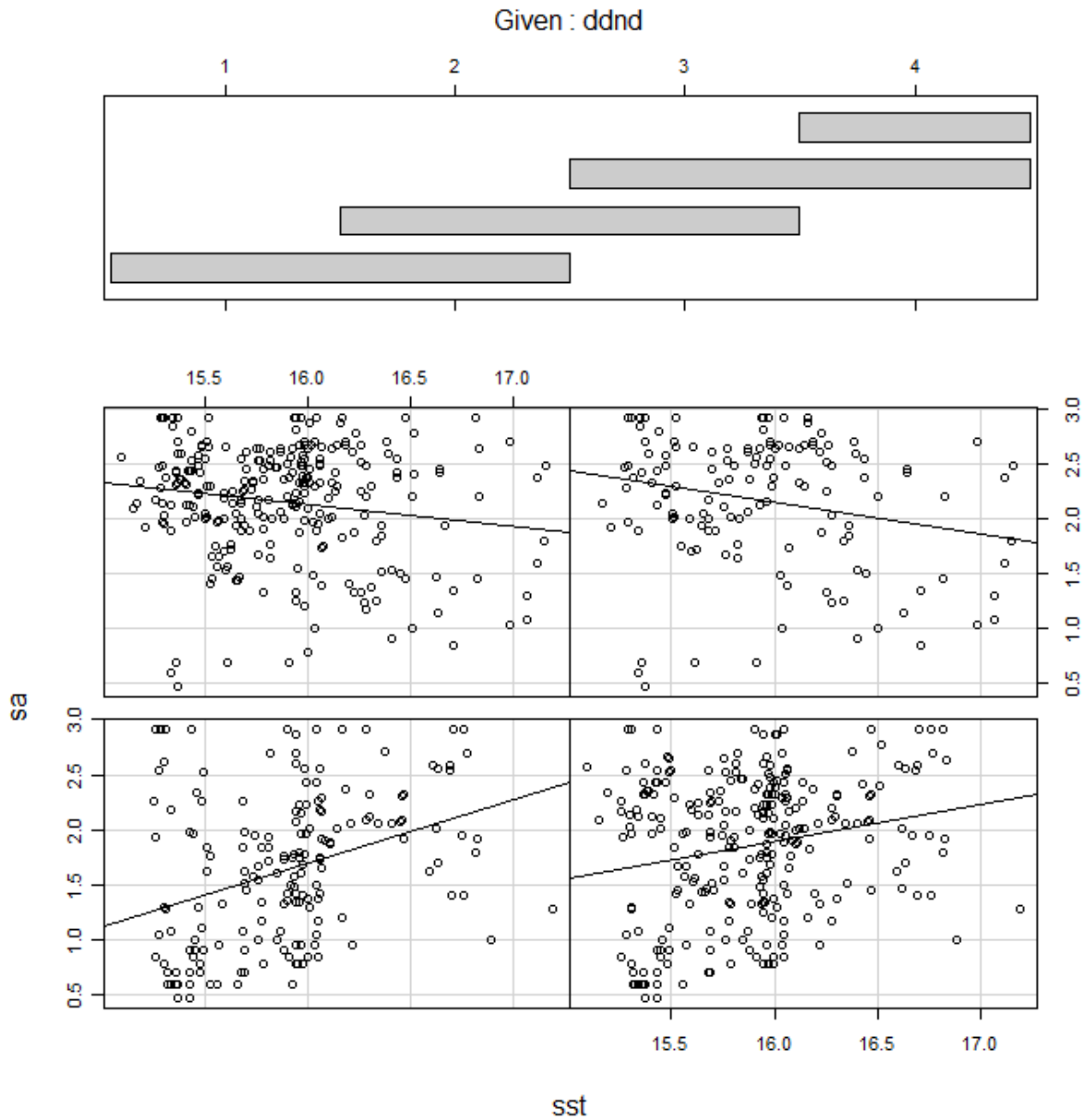
Pairplot που δείχνει τις συσχετίσεις μεταξύ των επεξηγηματικών παραμέτρων που χρησιμοποιήθηκαν και της εξαρτημένης μεταβλητής

Pairplot

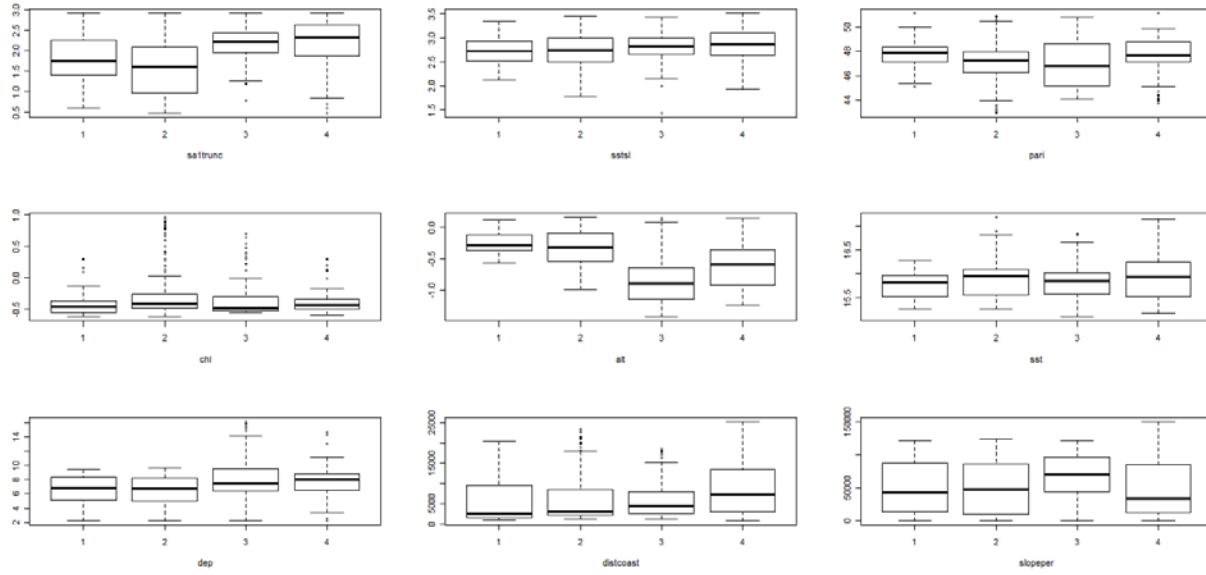




Pairplots όλων των μεταβλητών που χρησιμοποιήθηκαν για να προσδιορίσουν τη συγγραμμικότητα τους.



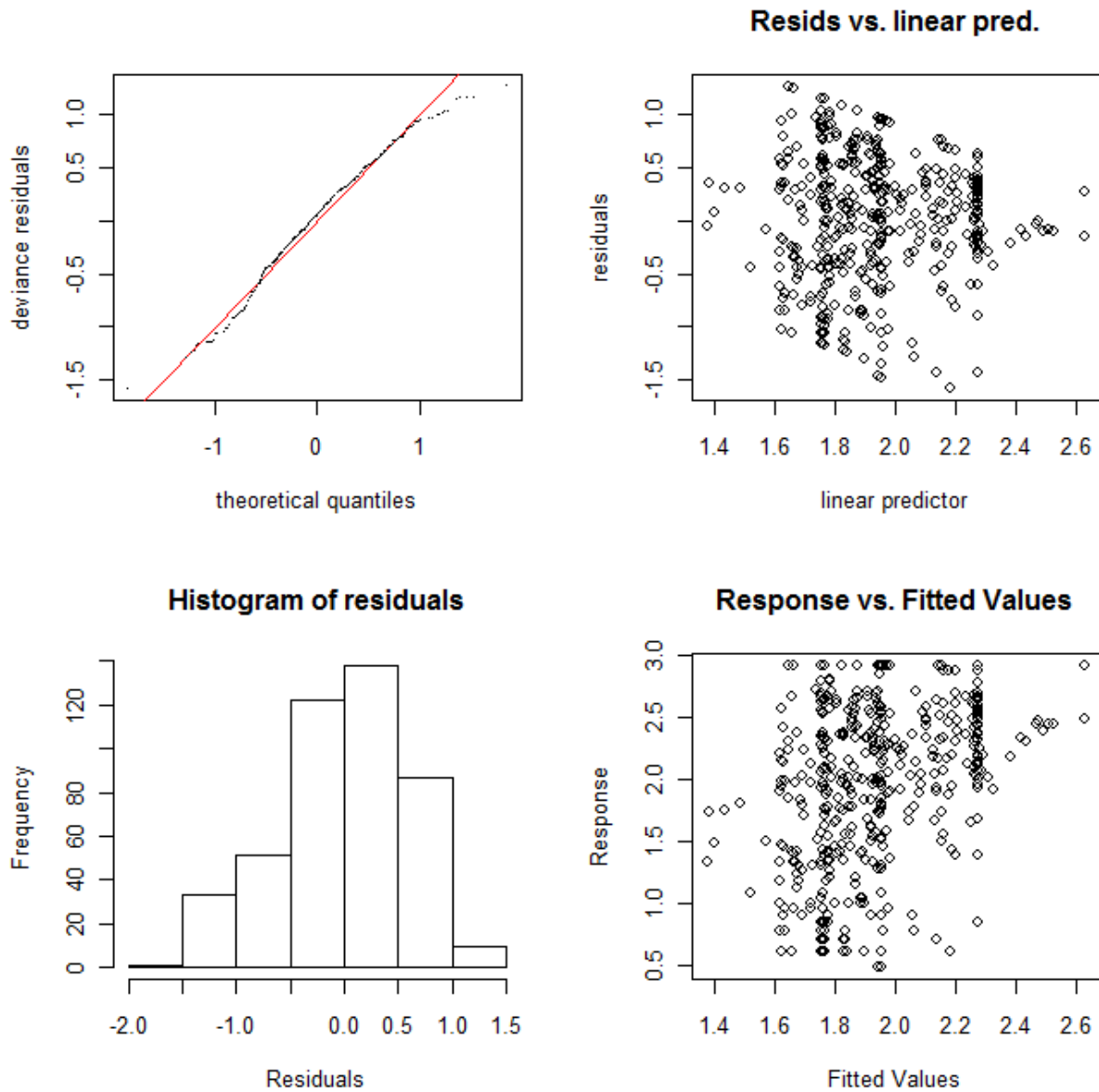
Conditional co-plot μεταξύ επιφανειακής θερμοκρασίας, ακουστικής πληροφορίας και του παράγοντα DDND, που δείχνει ότι υπάρχει αλληλεπίδραση μεταξύ θερμοκρασίας και DDND ως προς την κατανομή των τιμών της ακουστικής πληροφορίας.

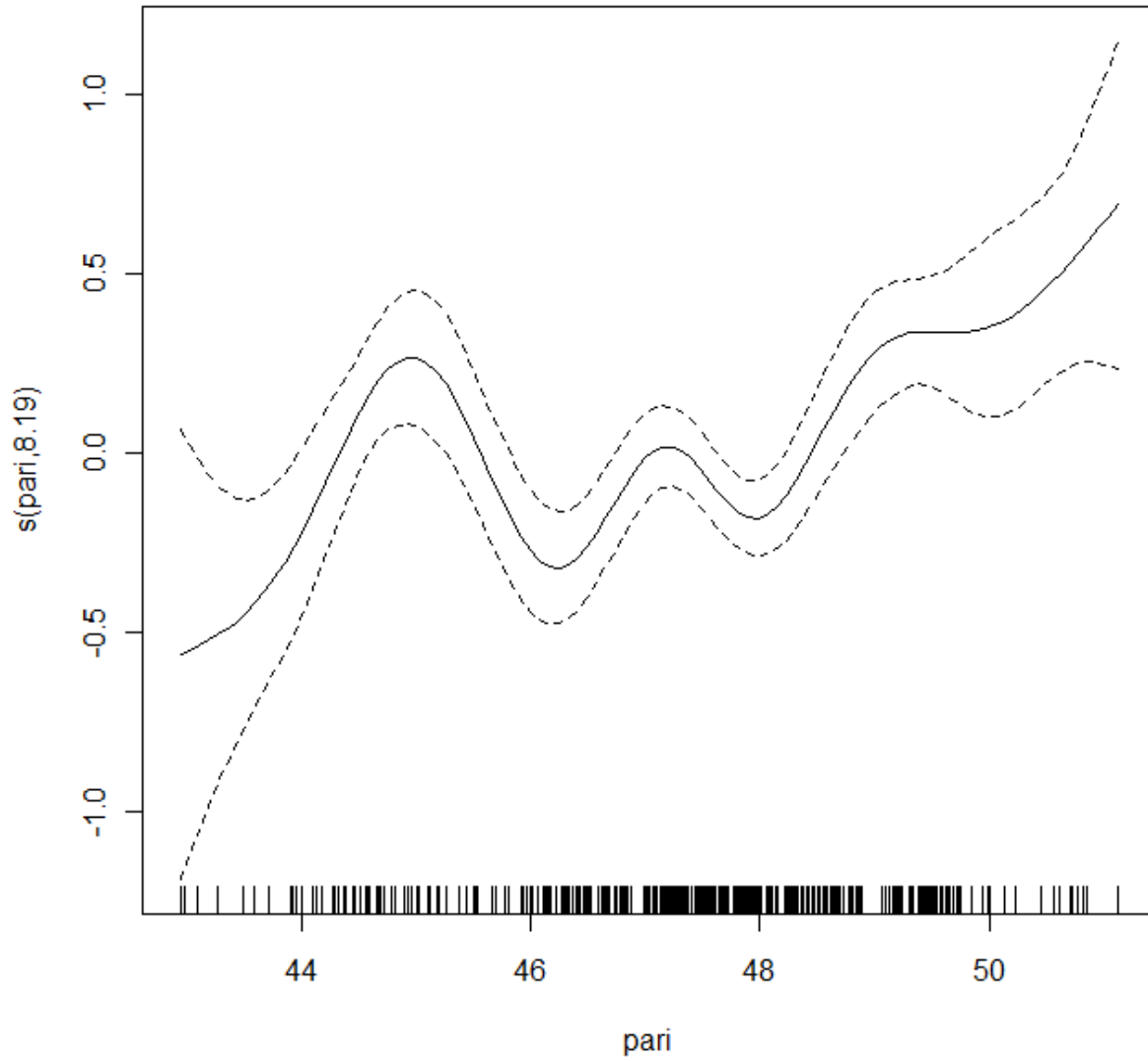


Conditional boxplots που δείχνουν τις σχέσεις του παράγοντα DDND με τις μεταβλητές που επιλέγηκαν τελικά για την ανάπτυξη των μοντέλων

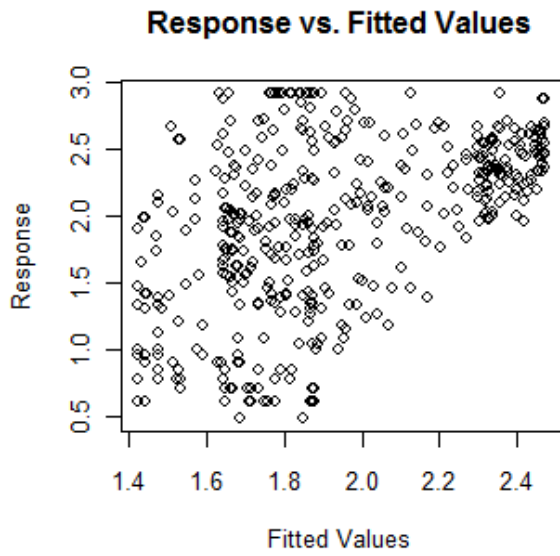
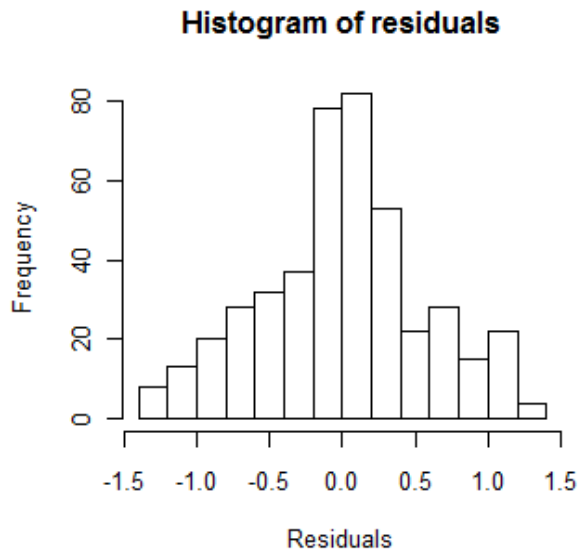
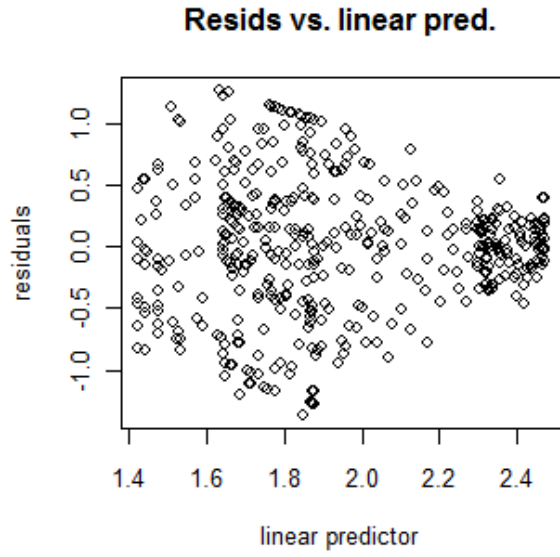
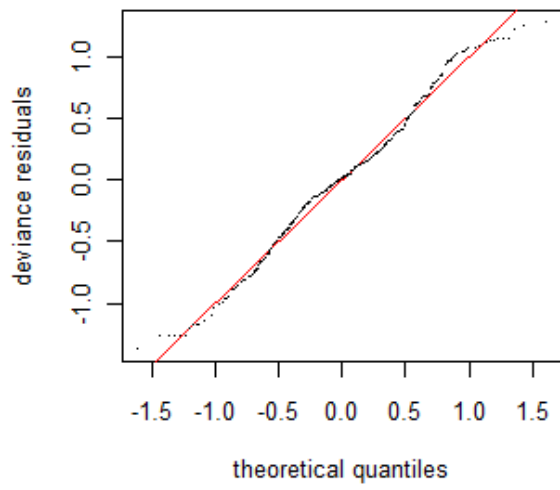
Μονοπαραμετρικά GAMs που χρησιμοποιούνται ως διαγνωστικό εργαλείο για τη σχέση της εξαρτημένης μεταβλητής με την εκάστοτε ανεξάρτητη και ως δείκτης της σημαντικότητας που έχει η κάθε ανεξάρτητη μεταβλητή στην εξηγούμενη μεταβλητότητα της εξαρτημένης.

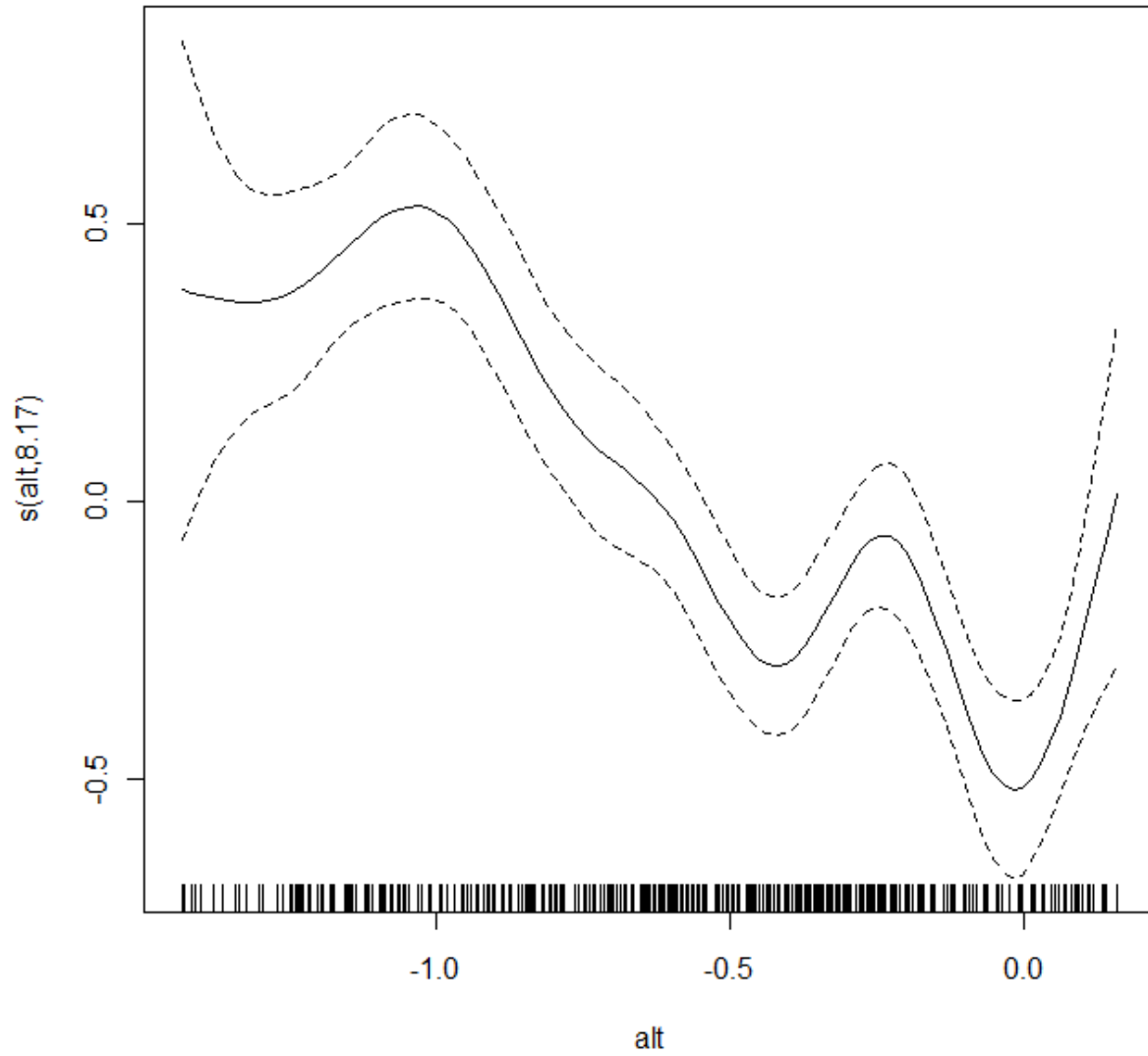
PAR: Deviance explained = 13.5%, p-value: 1.30e-09, edf=8.188



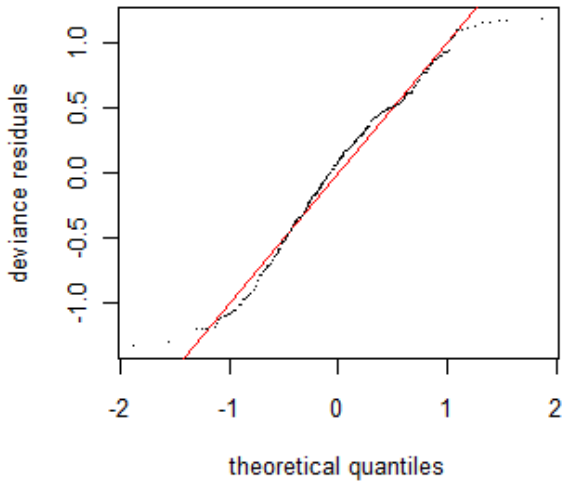


ALT: Deviance explained = 23.9%, p-value: 2e-16, edf=8.173

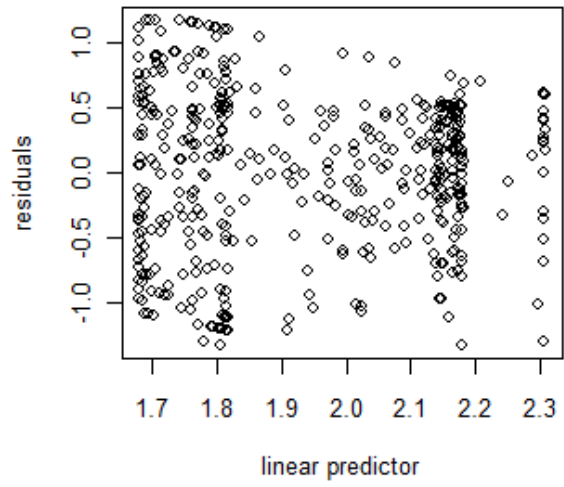




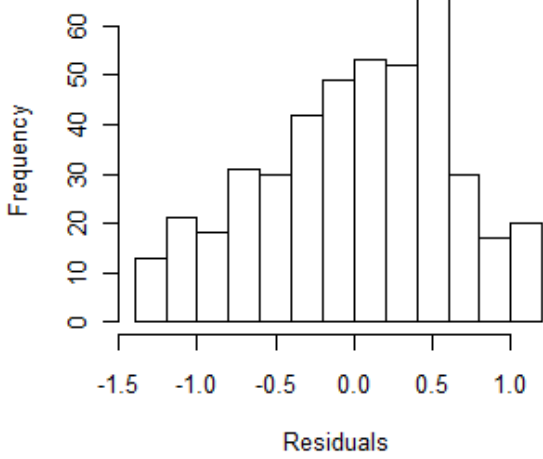
DEP: Deviance explained = 10.9%, p-value: 1.29e-07, edf=7.633



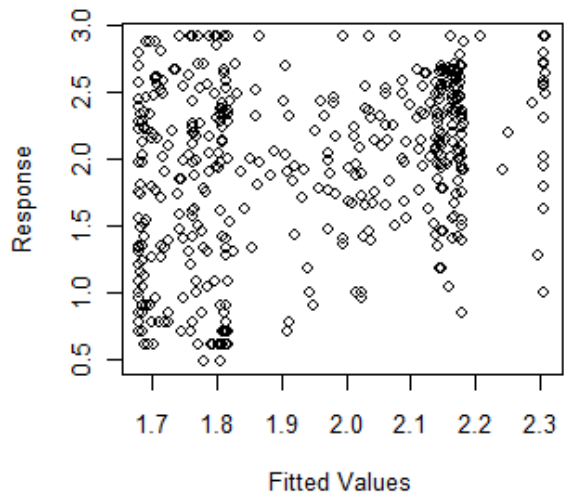
Resids vs. linear pred.

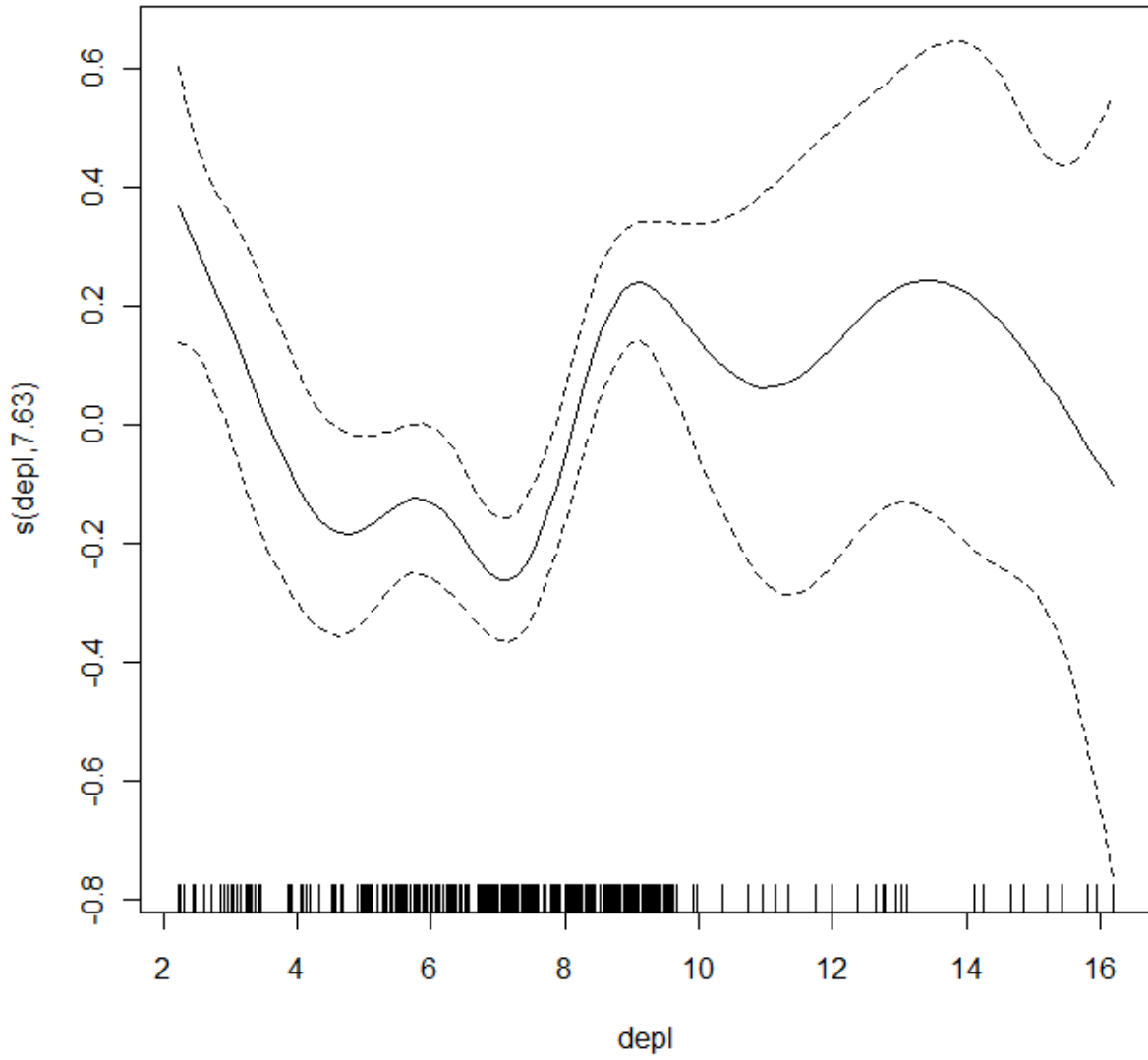


Histogram of residuals

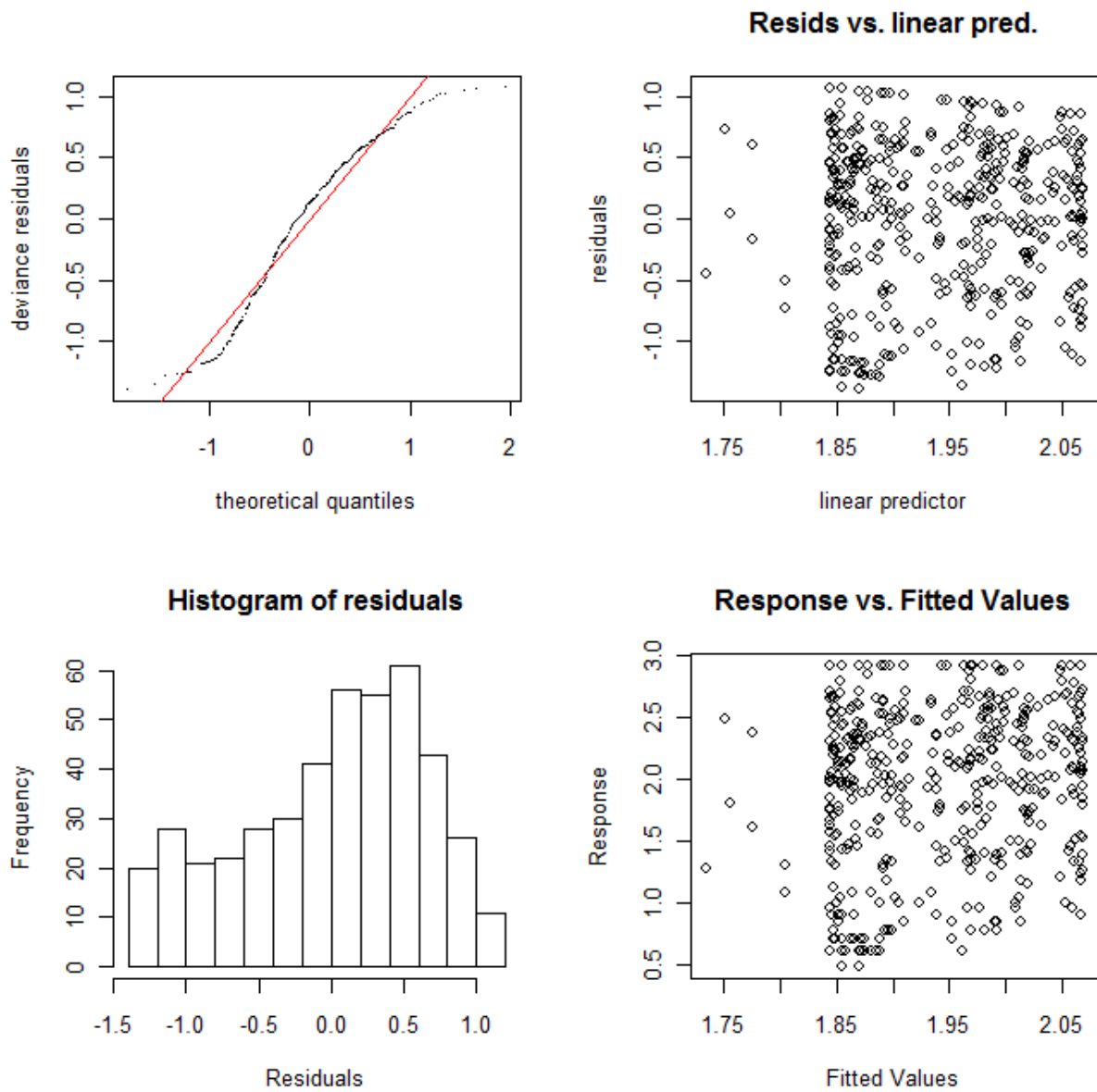


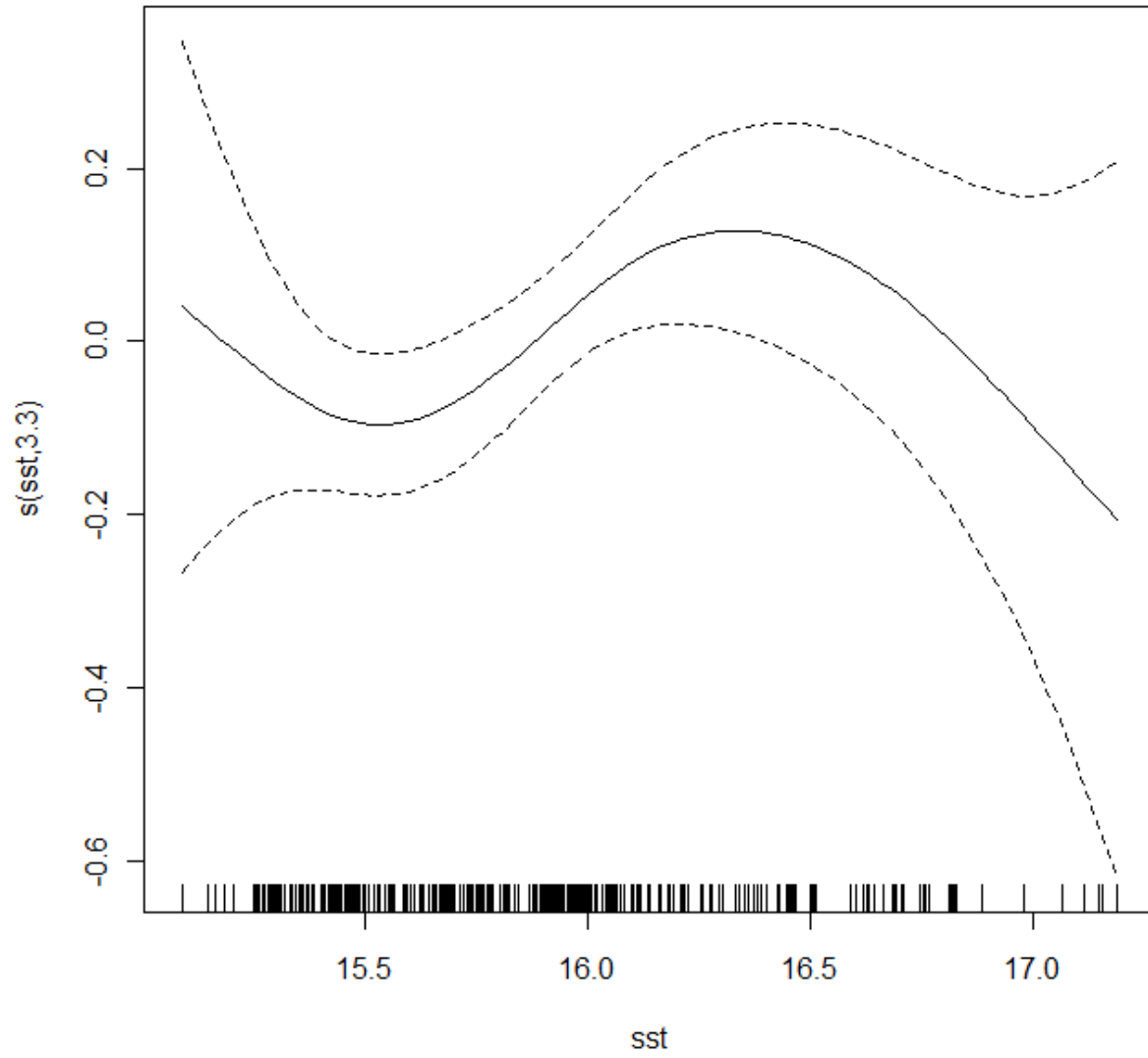
Response vs. Fitted Values



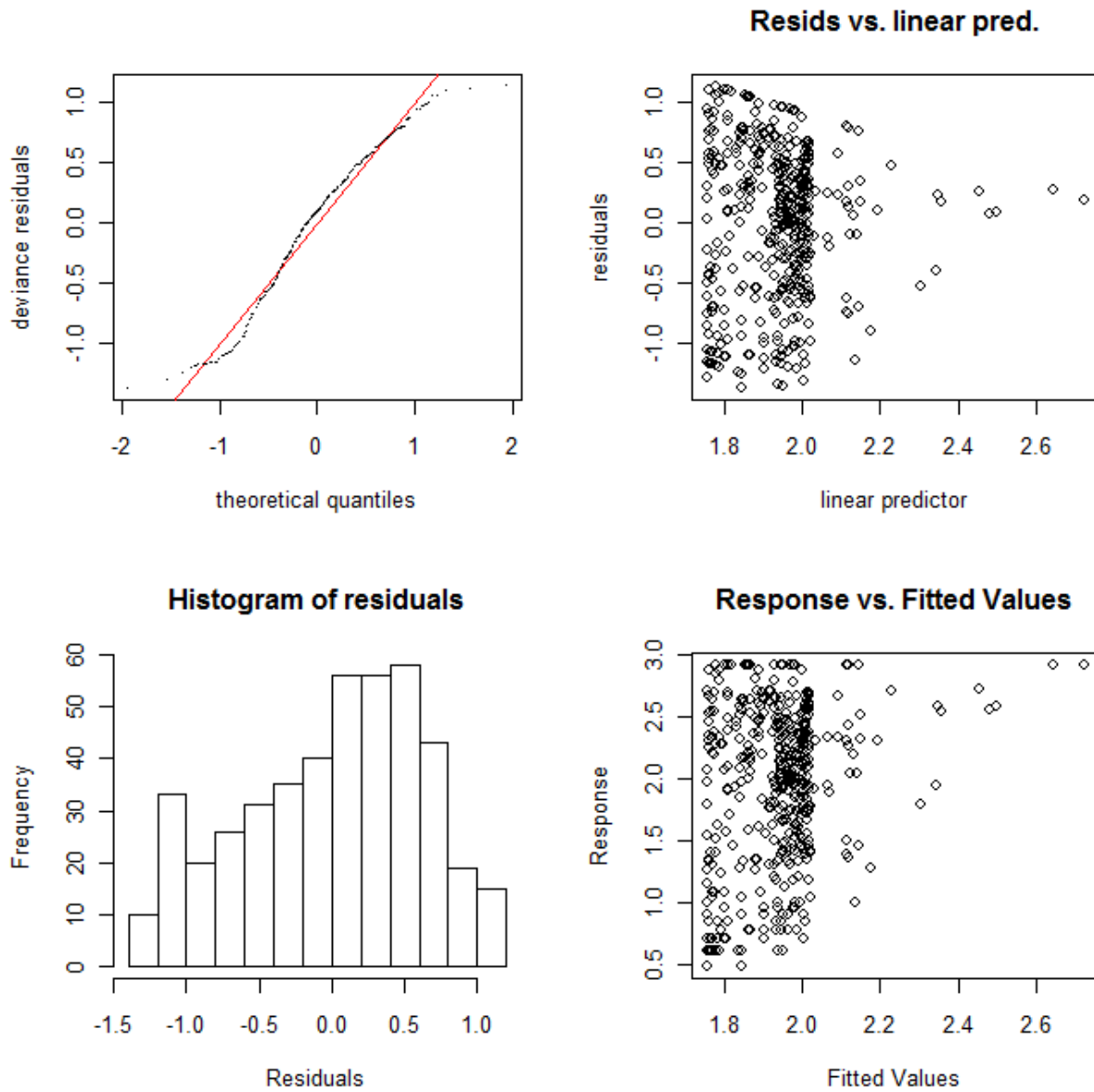


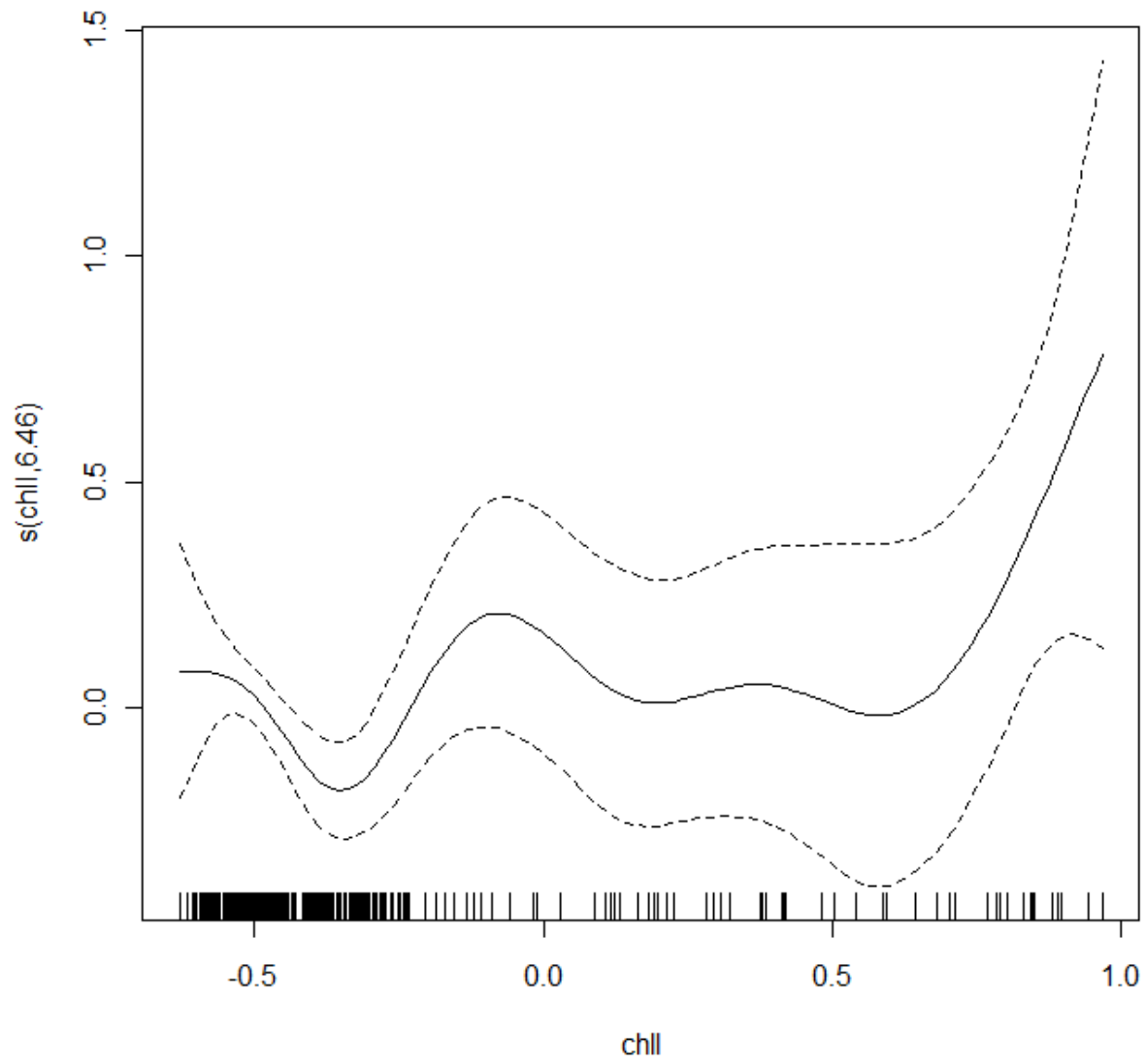
SST: Deviance explained = 2.33%, p-value: 0.0868, edf=3.303



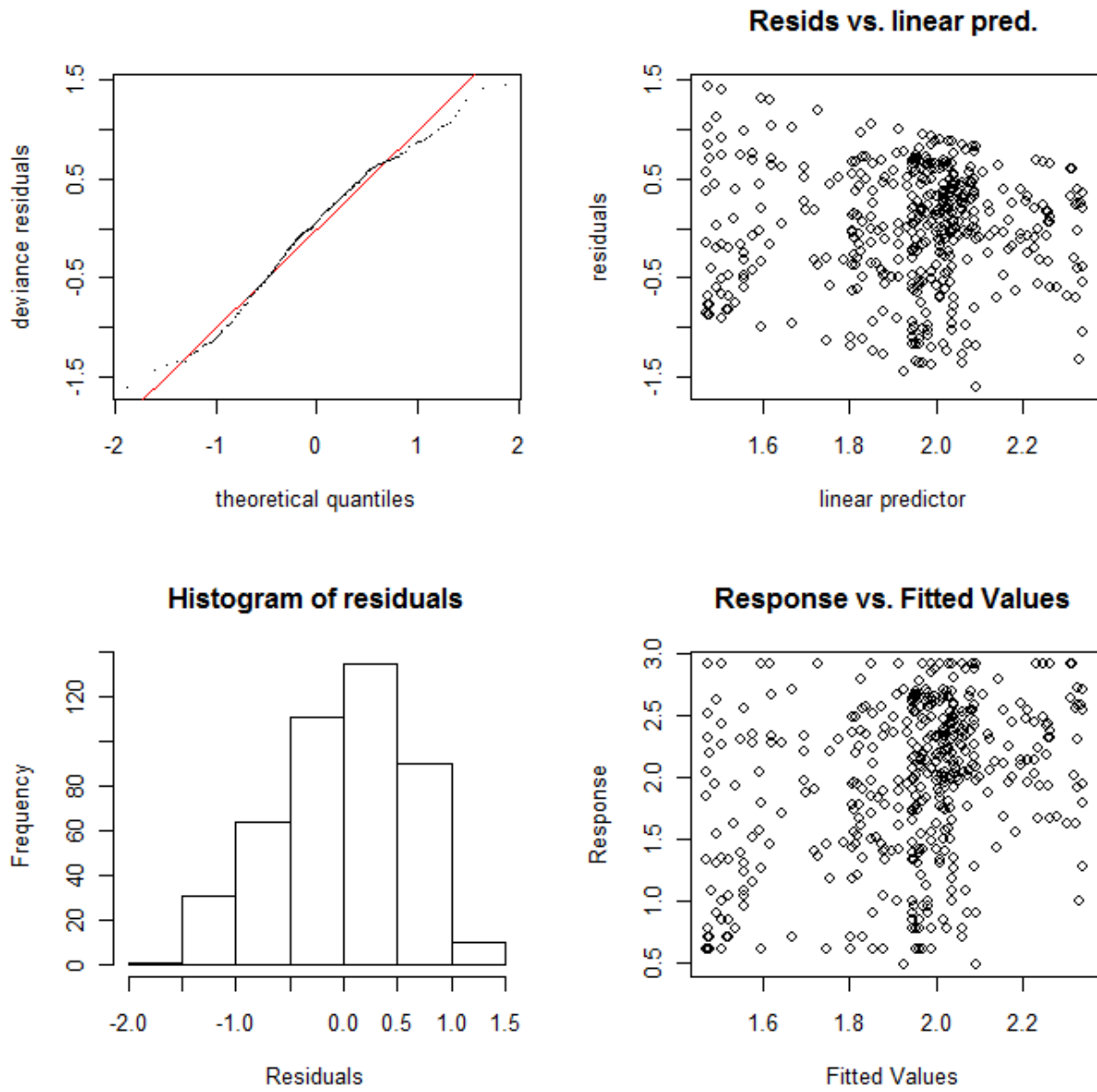


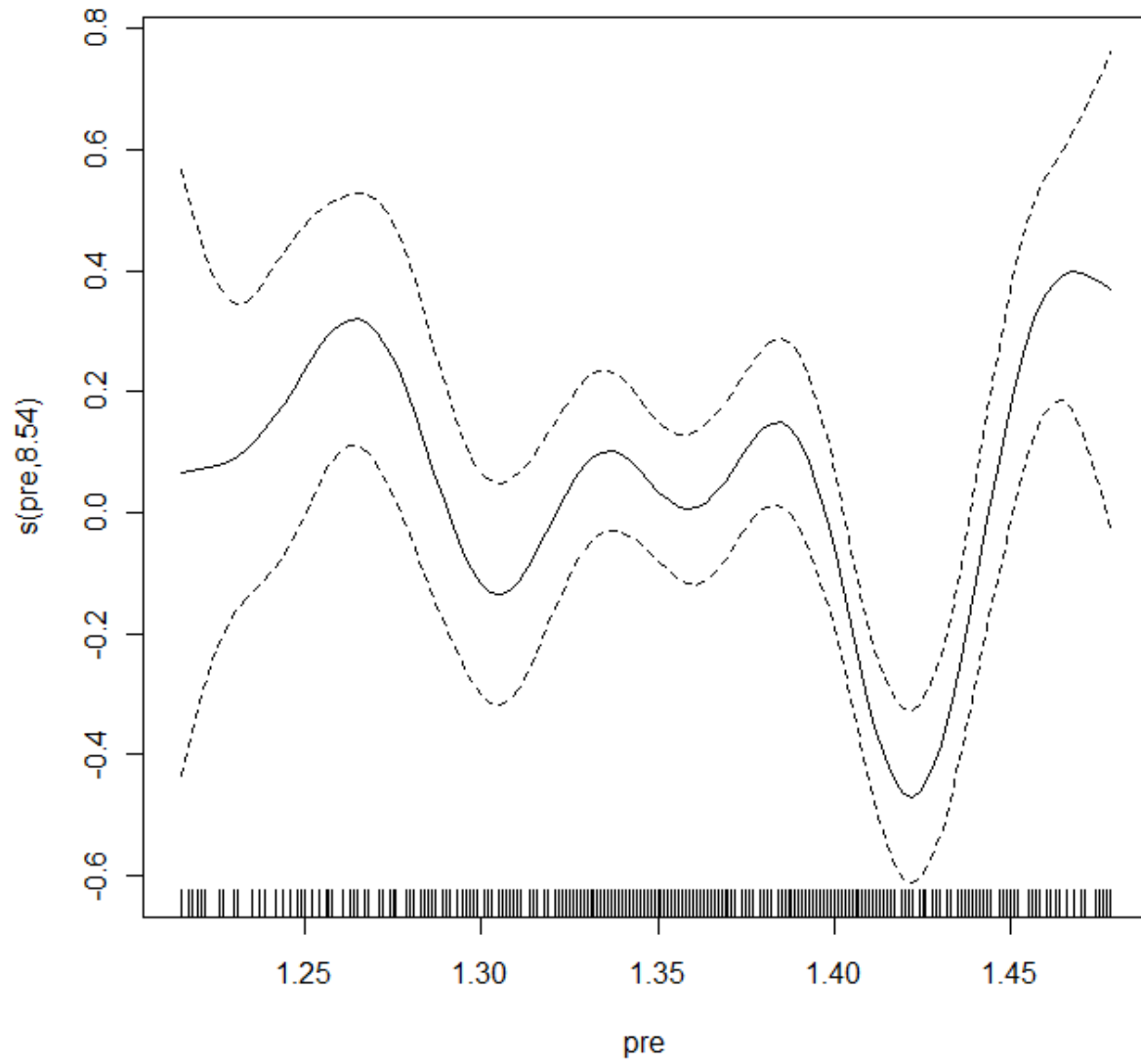
CHL: Deviance explained = 4.9%, p-value: 0.0104, edf=6.456



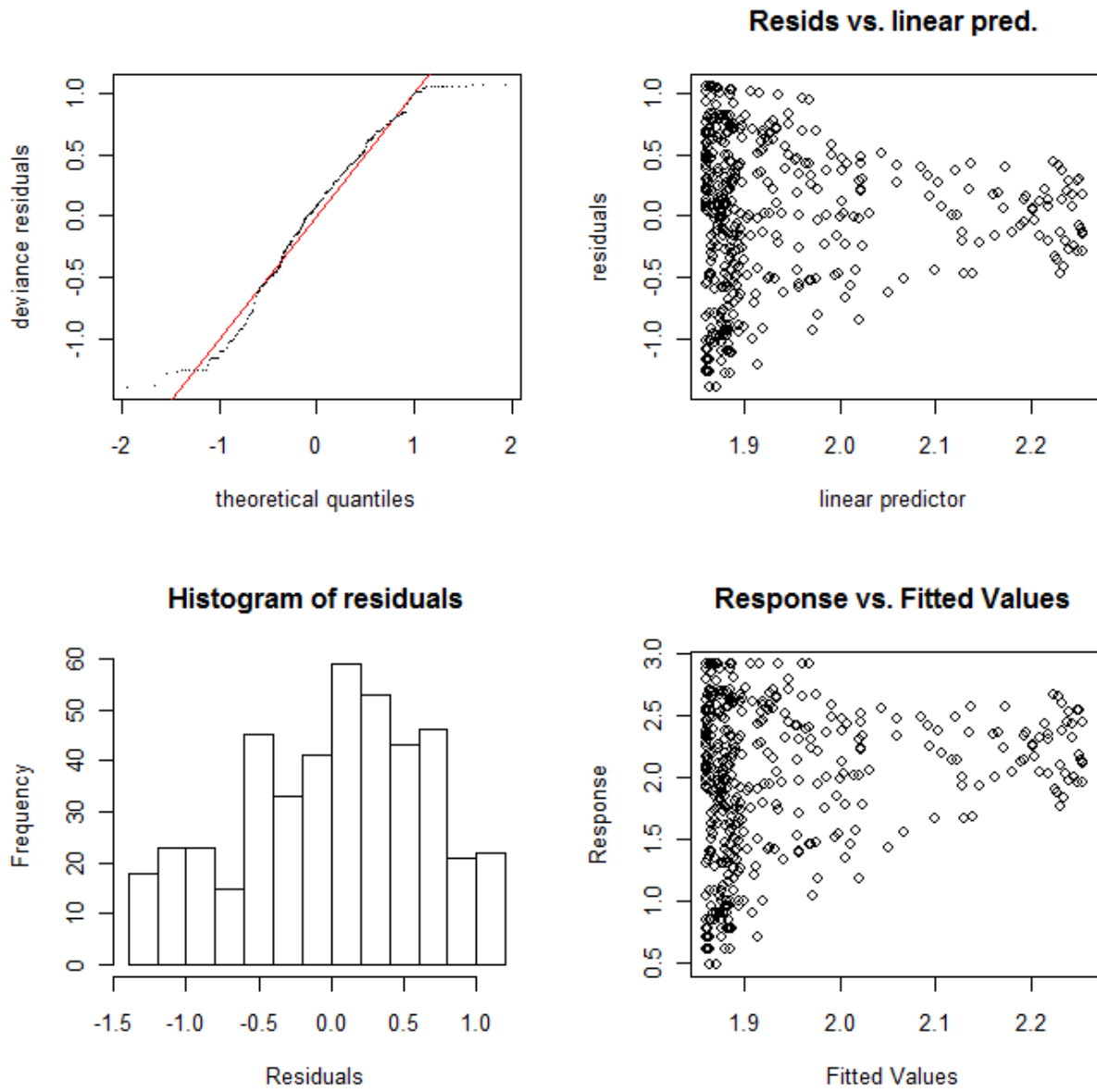


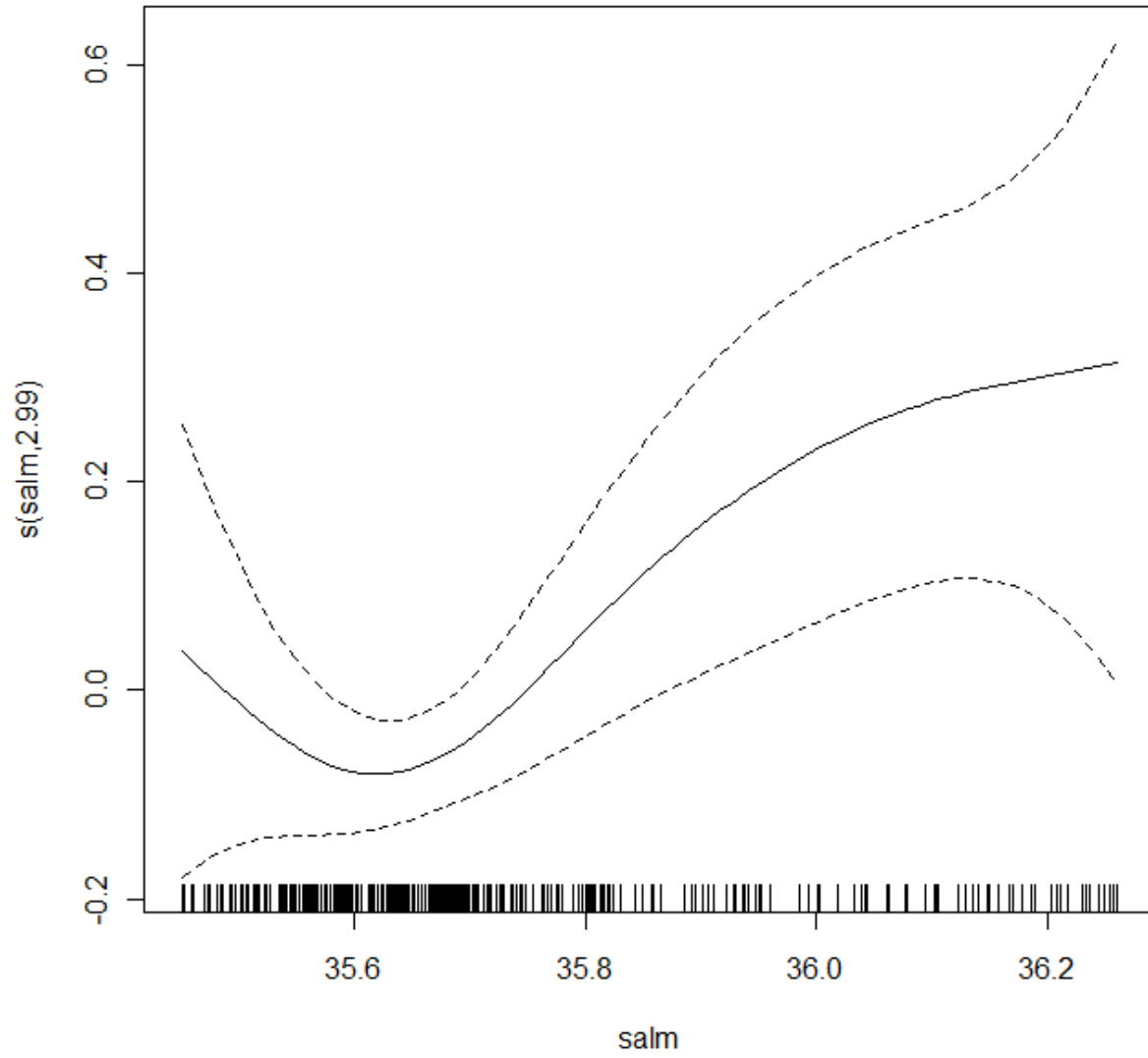
PRE: Deviance explained = 11.6%, p-value: 2.40e-08, edf=8.535



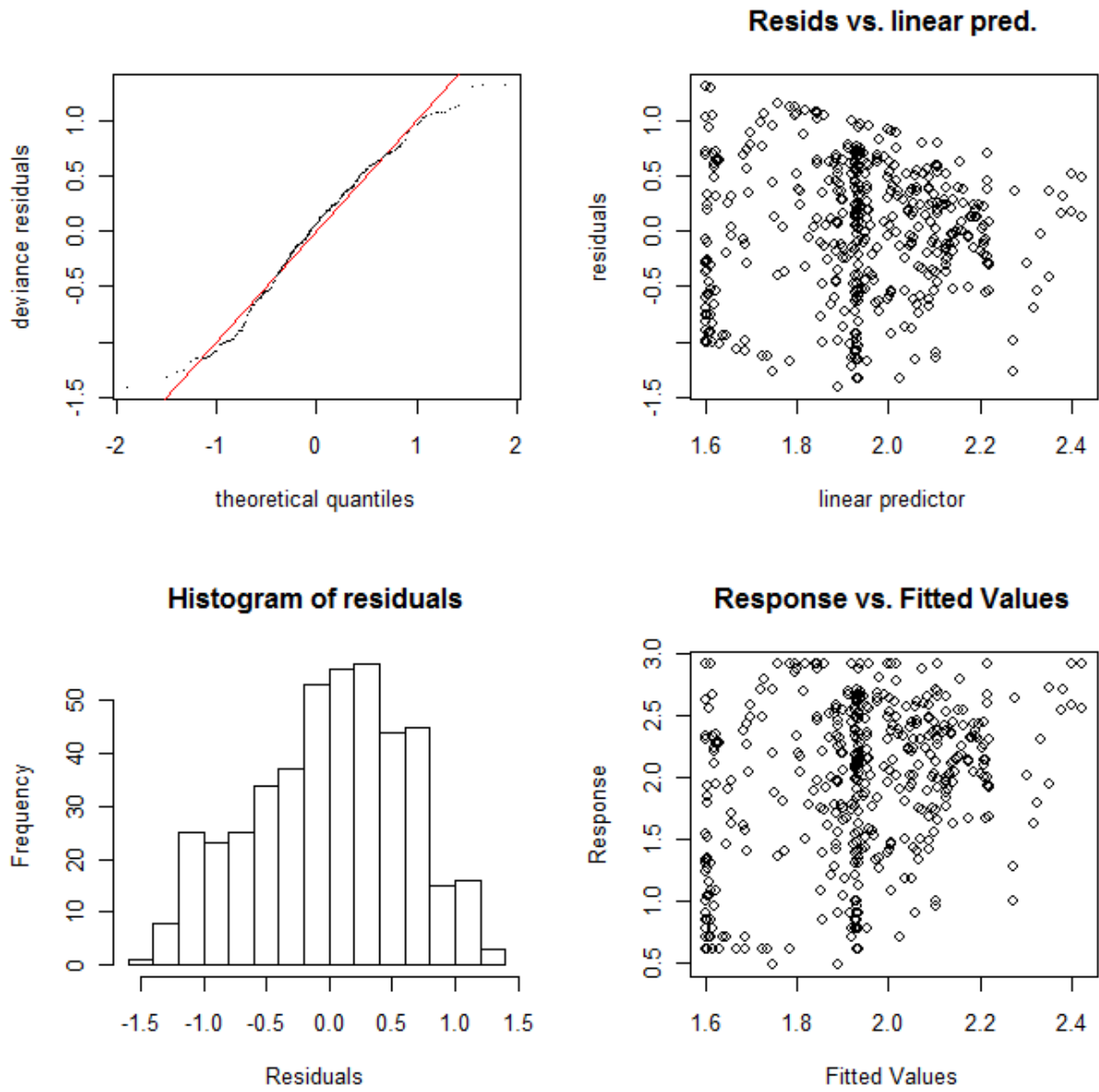


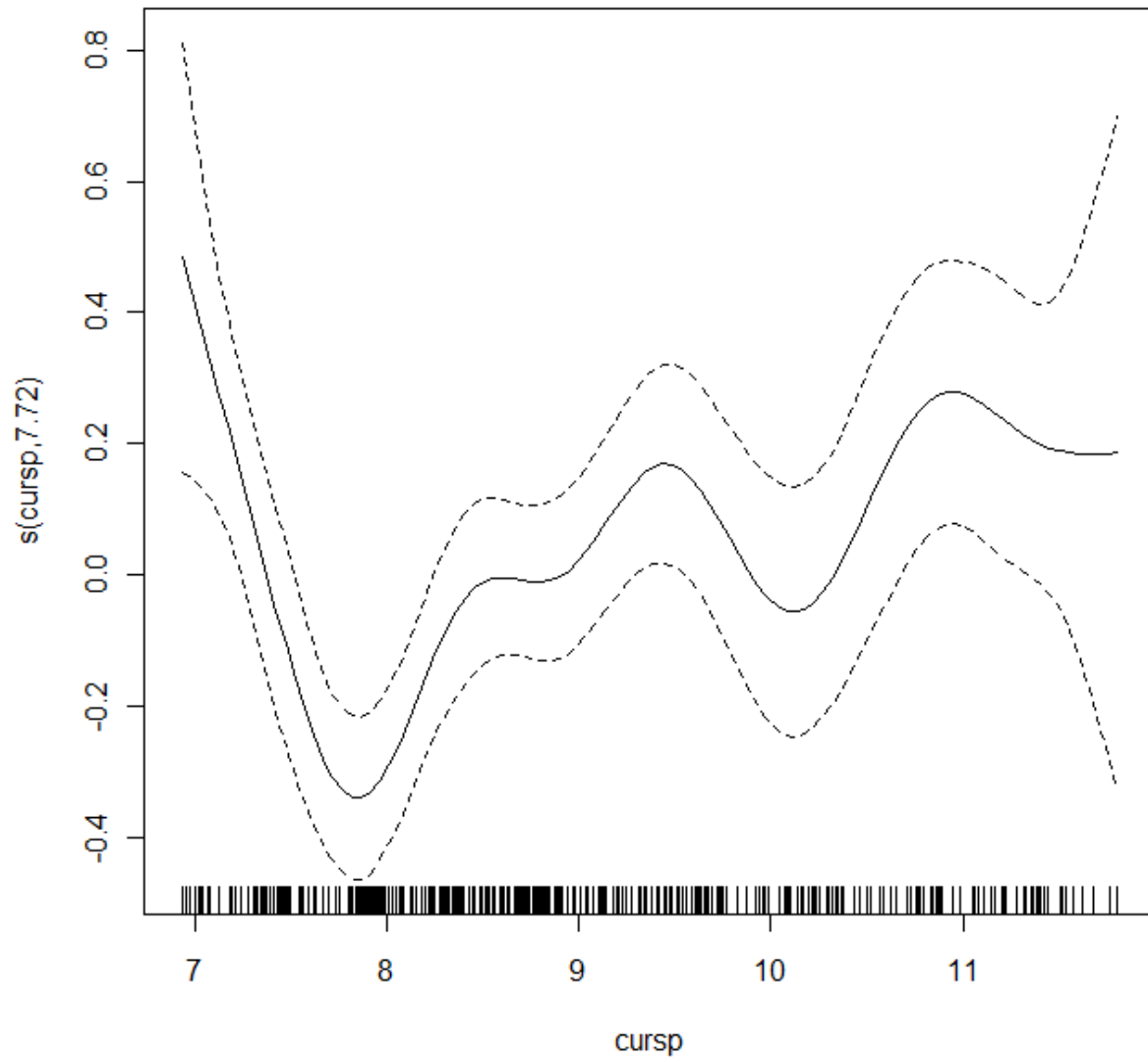
SAL: Deviance explained = 3.73%, p-value: 0.00287, edf=2.988



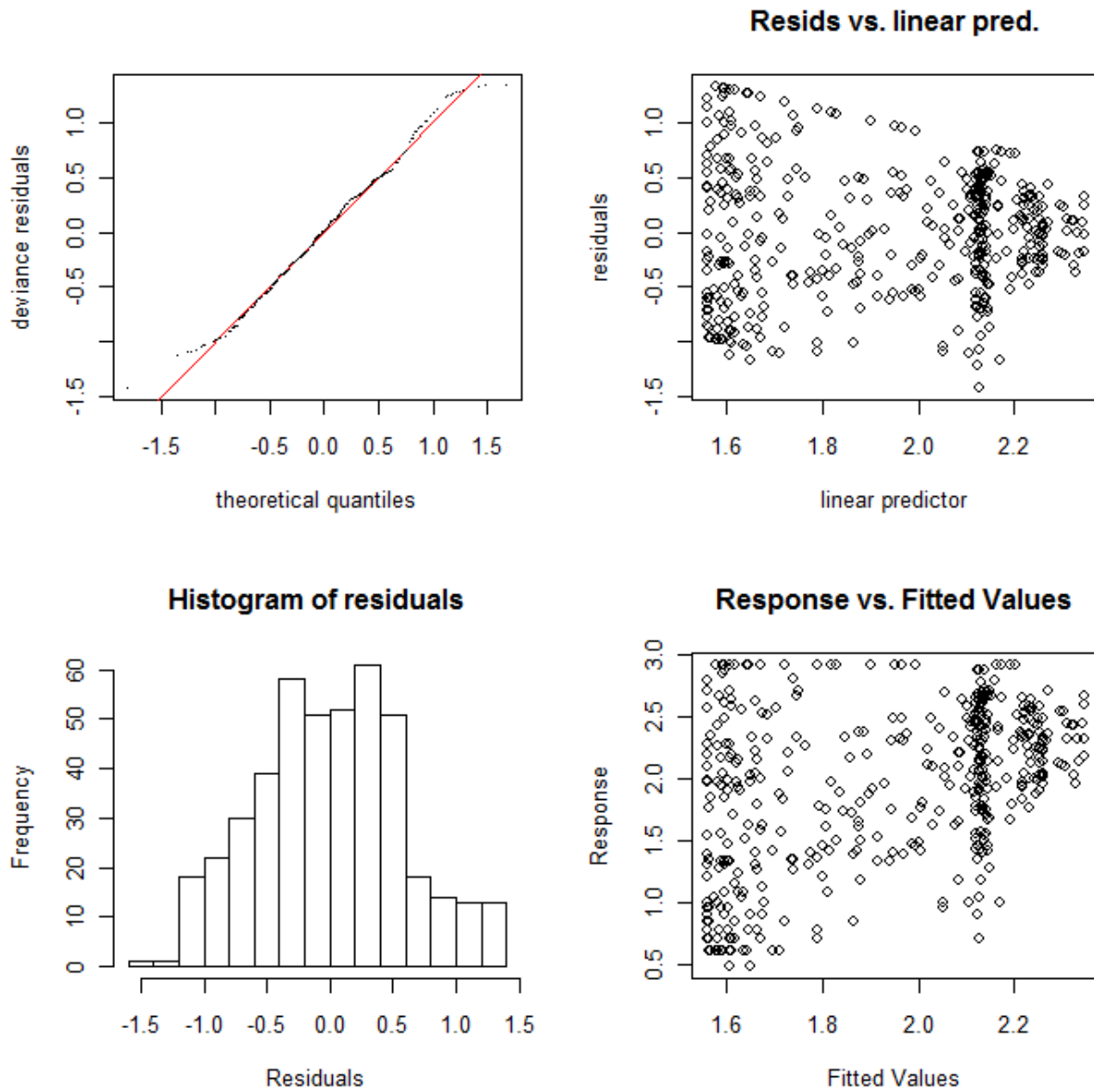


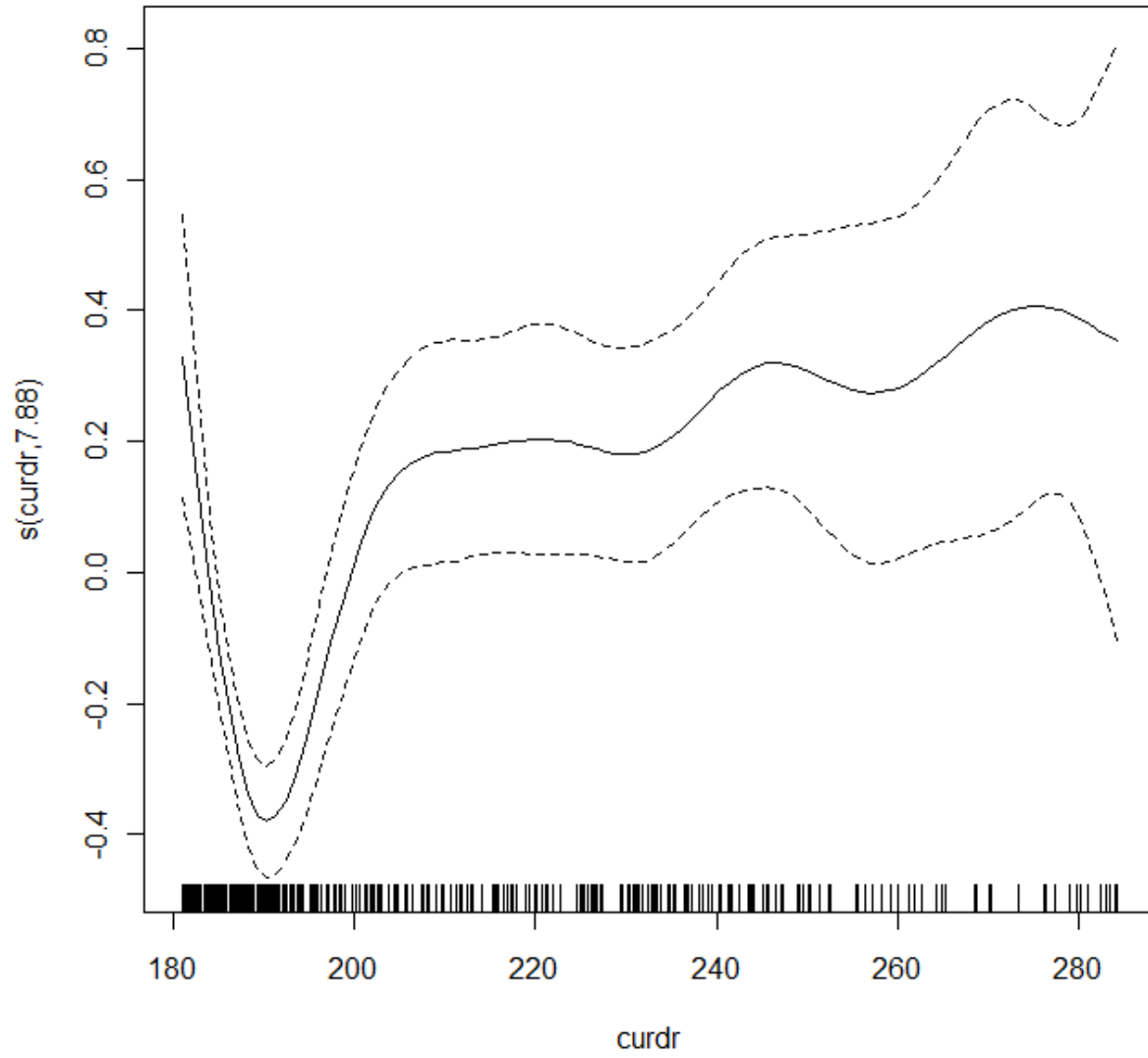
CURSP: Deviance explained = 9.69%, p-value: 2.23e-06, edf=7.719



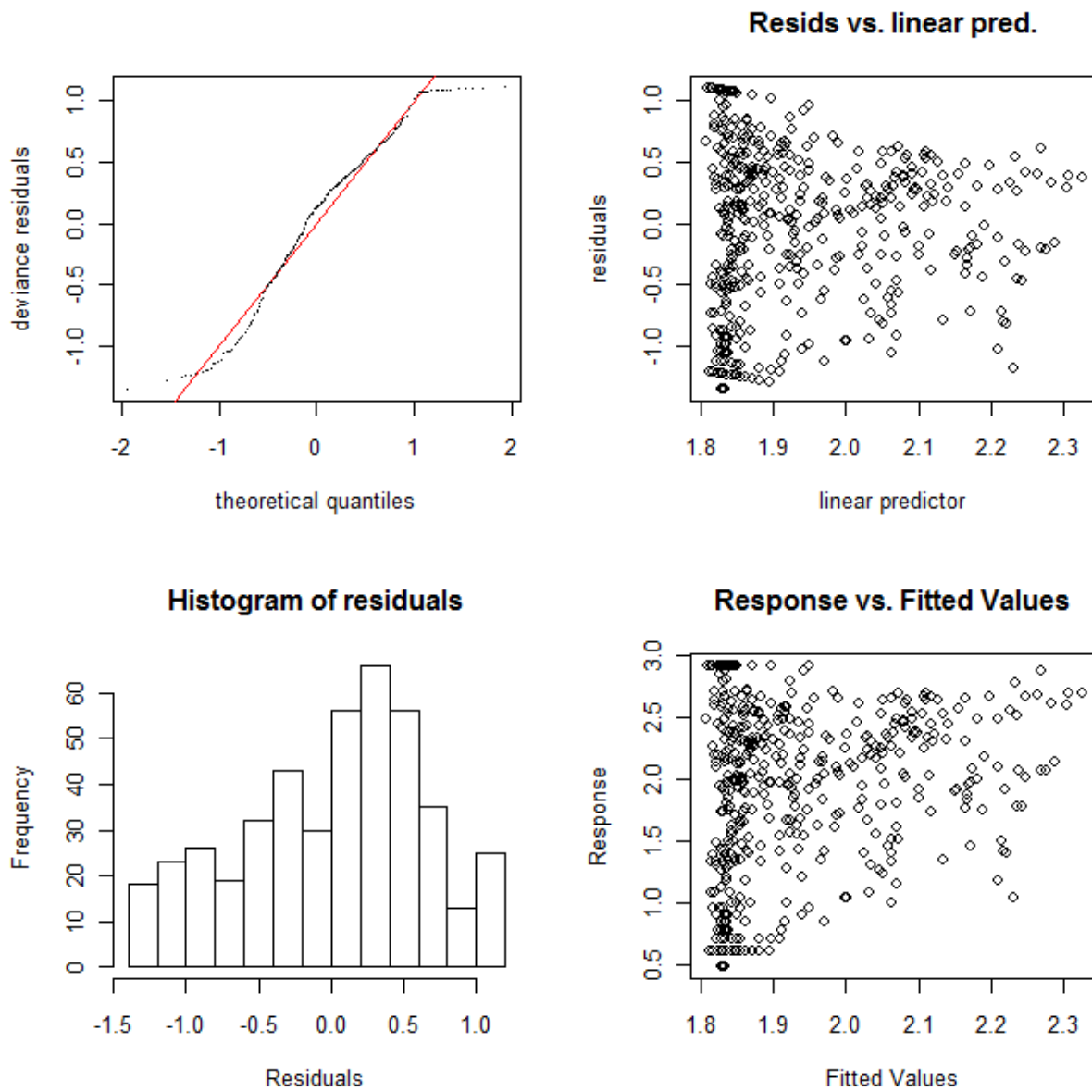


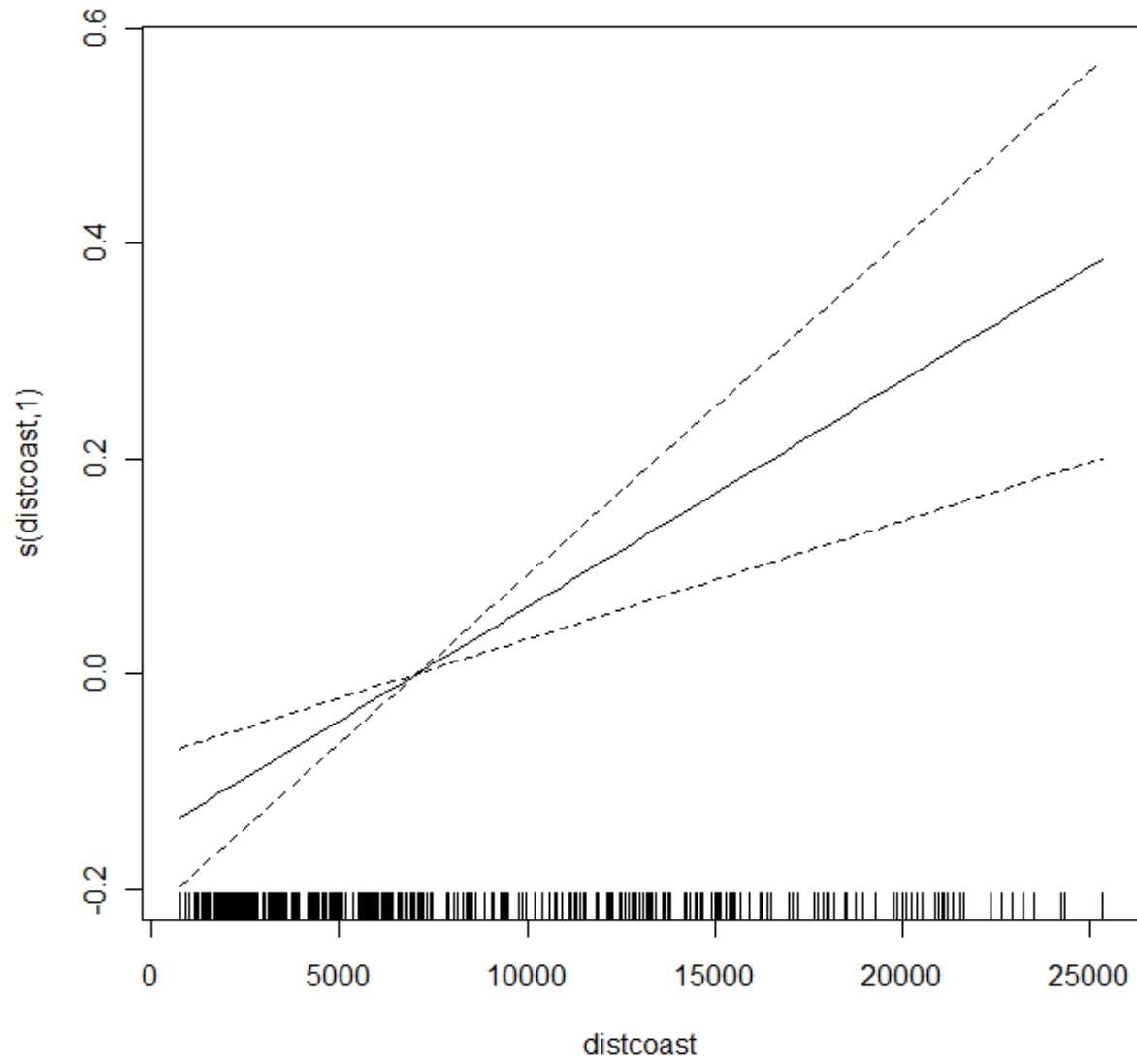
CURDR: Deviance explained = 17.8%, p-value: 6.77e-14, edf=7.884



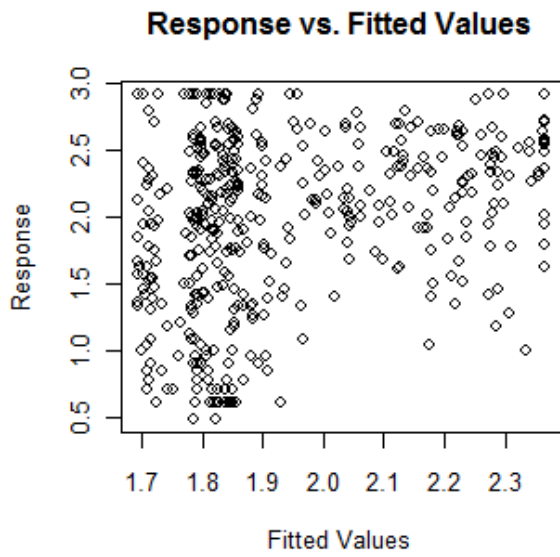
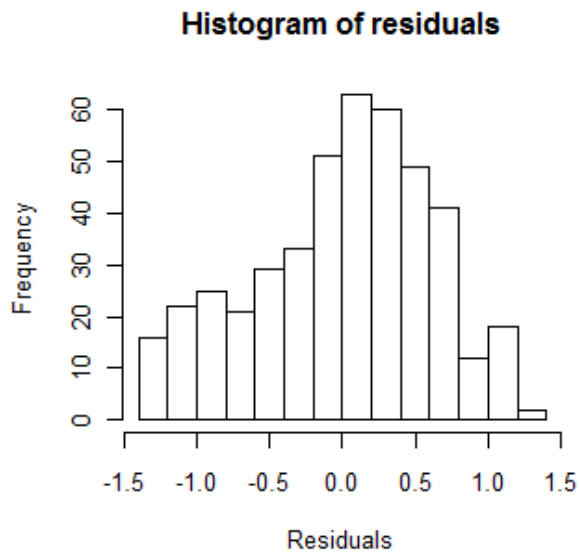
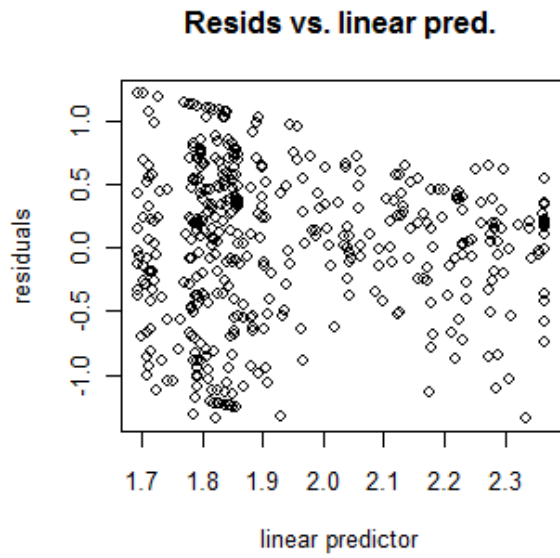
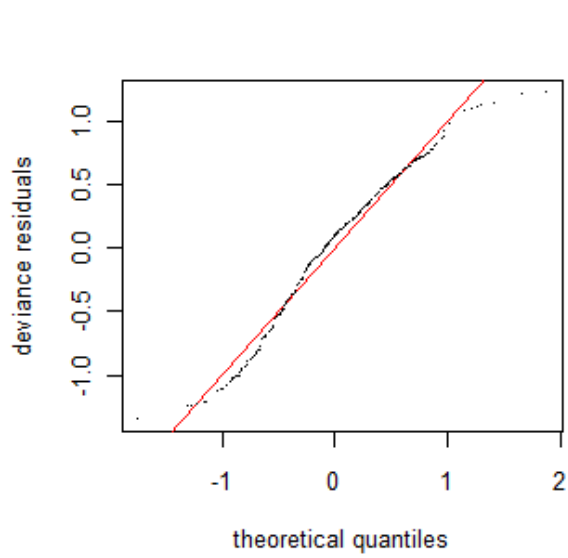


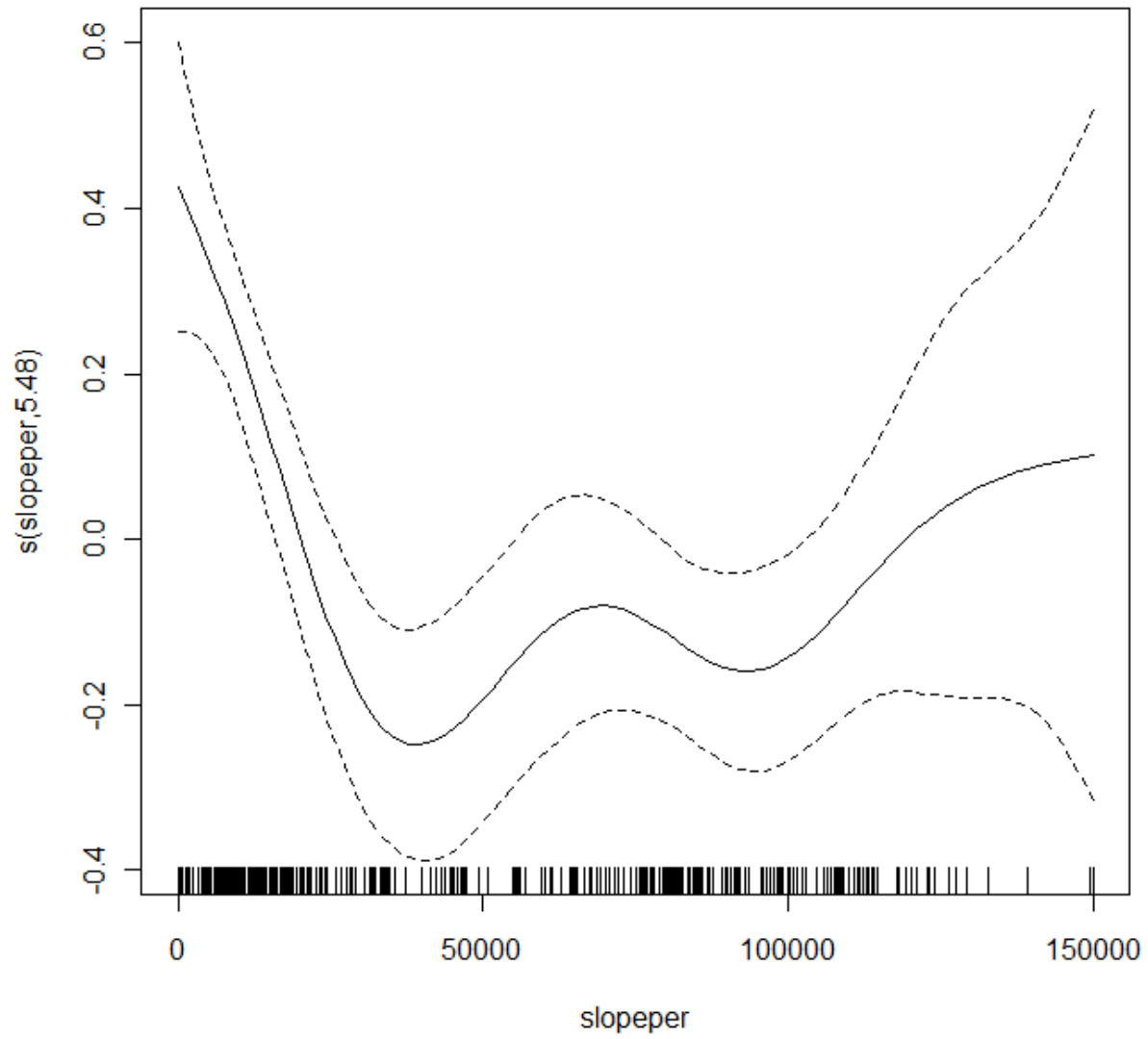
DISTCOAST: Deviance explained = 3.79%, p-value: 3.81e-05, edf=1



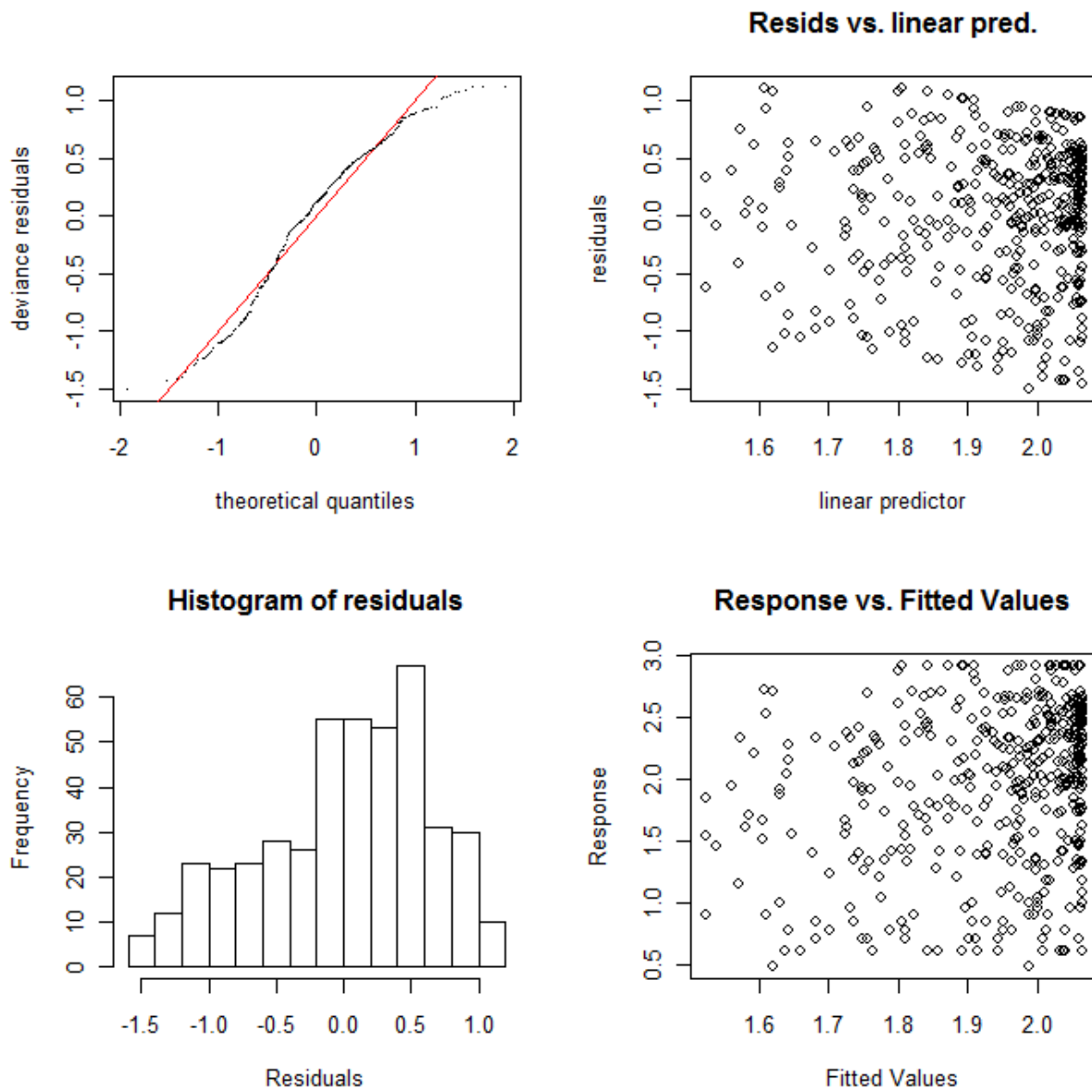


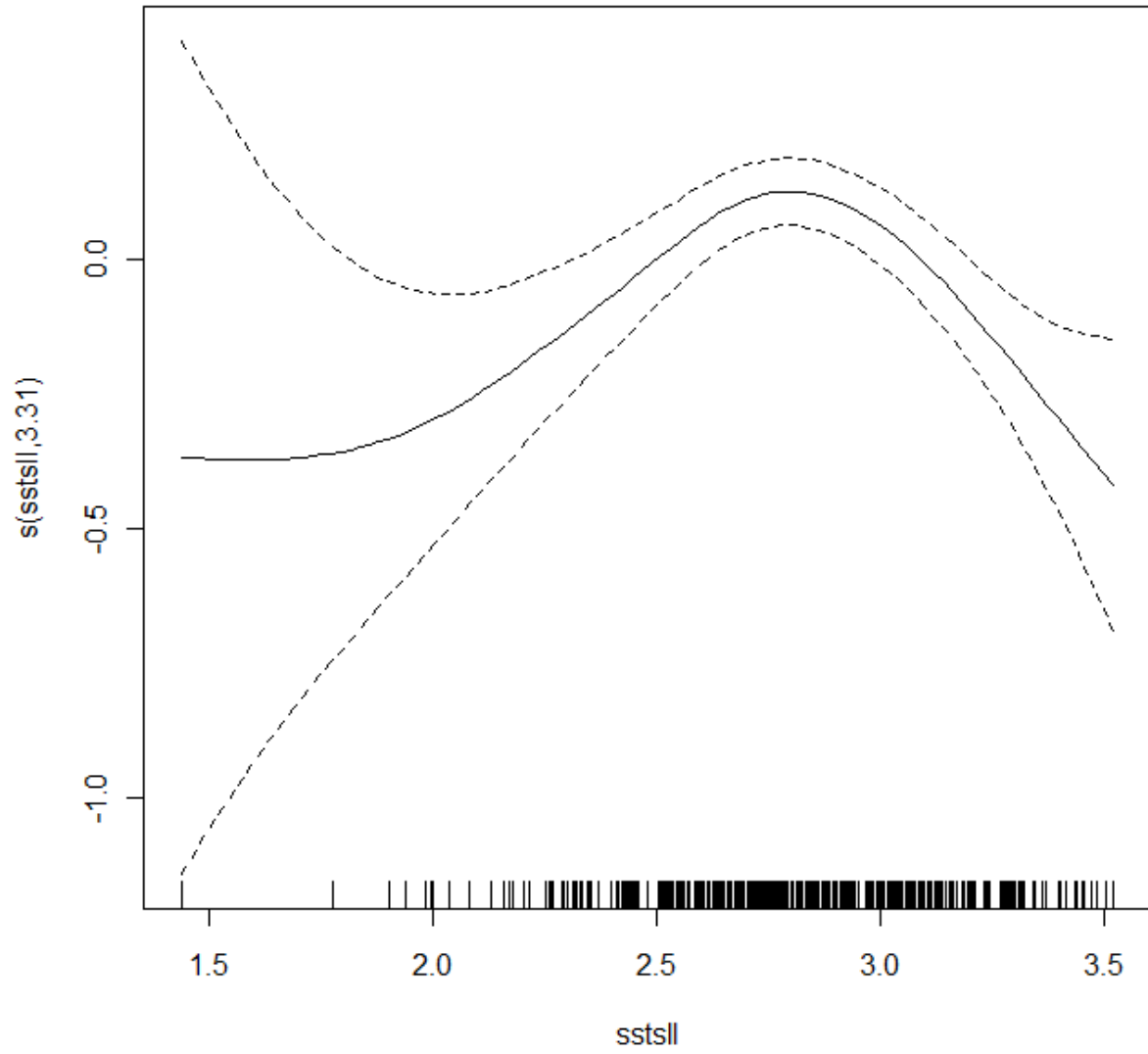
SLOPEPER: Deviance explained = 10.4%, p-value: 5.89e-08, edf=5.482



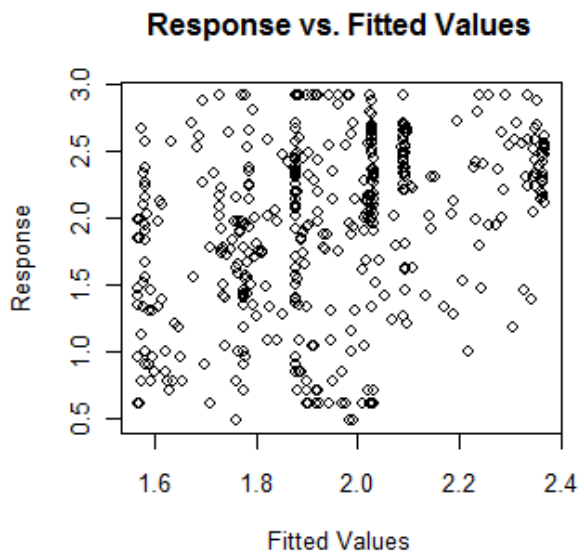
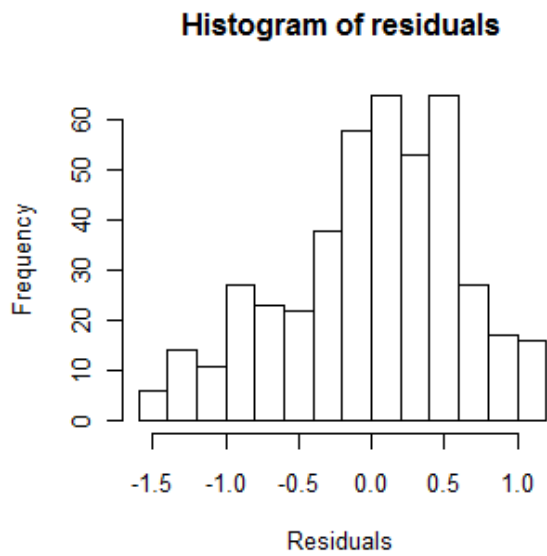
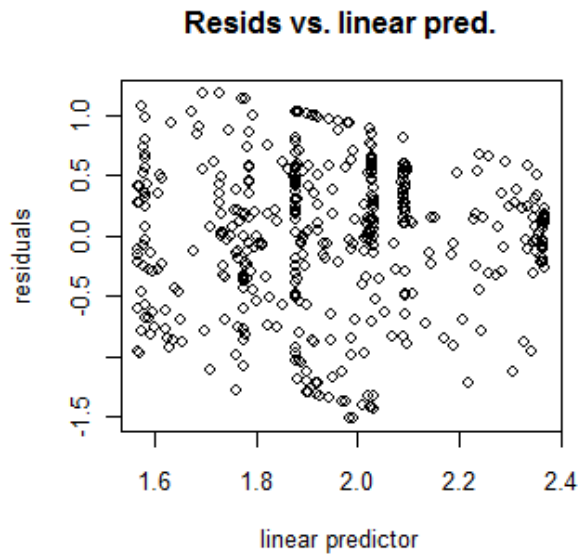
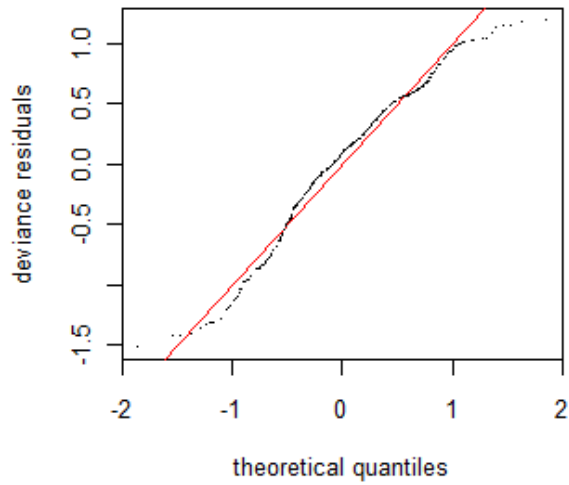


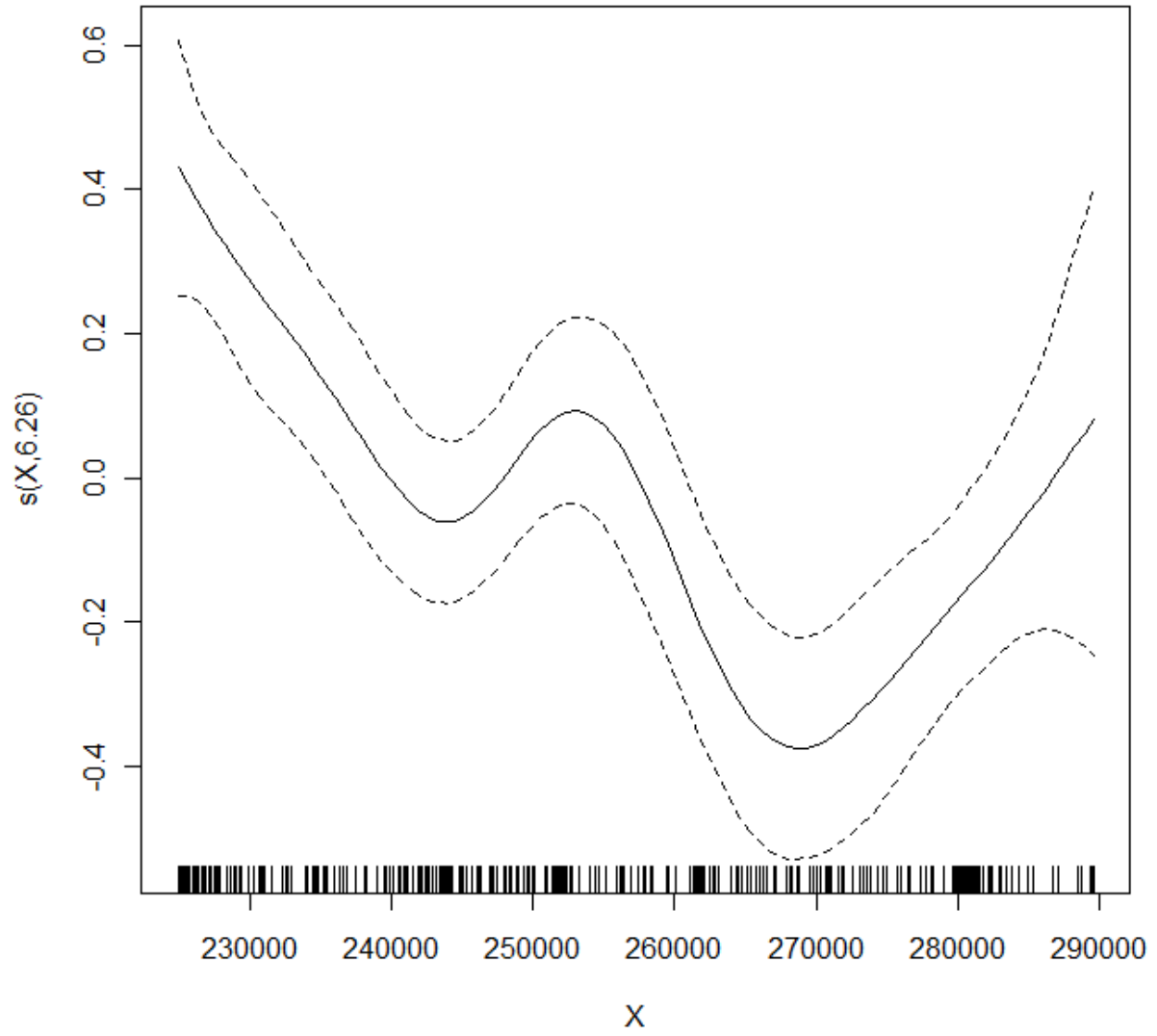
SSTSL: Deviance explained = 5.74%, p-value: 0.000189, edf=3.308



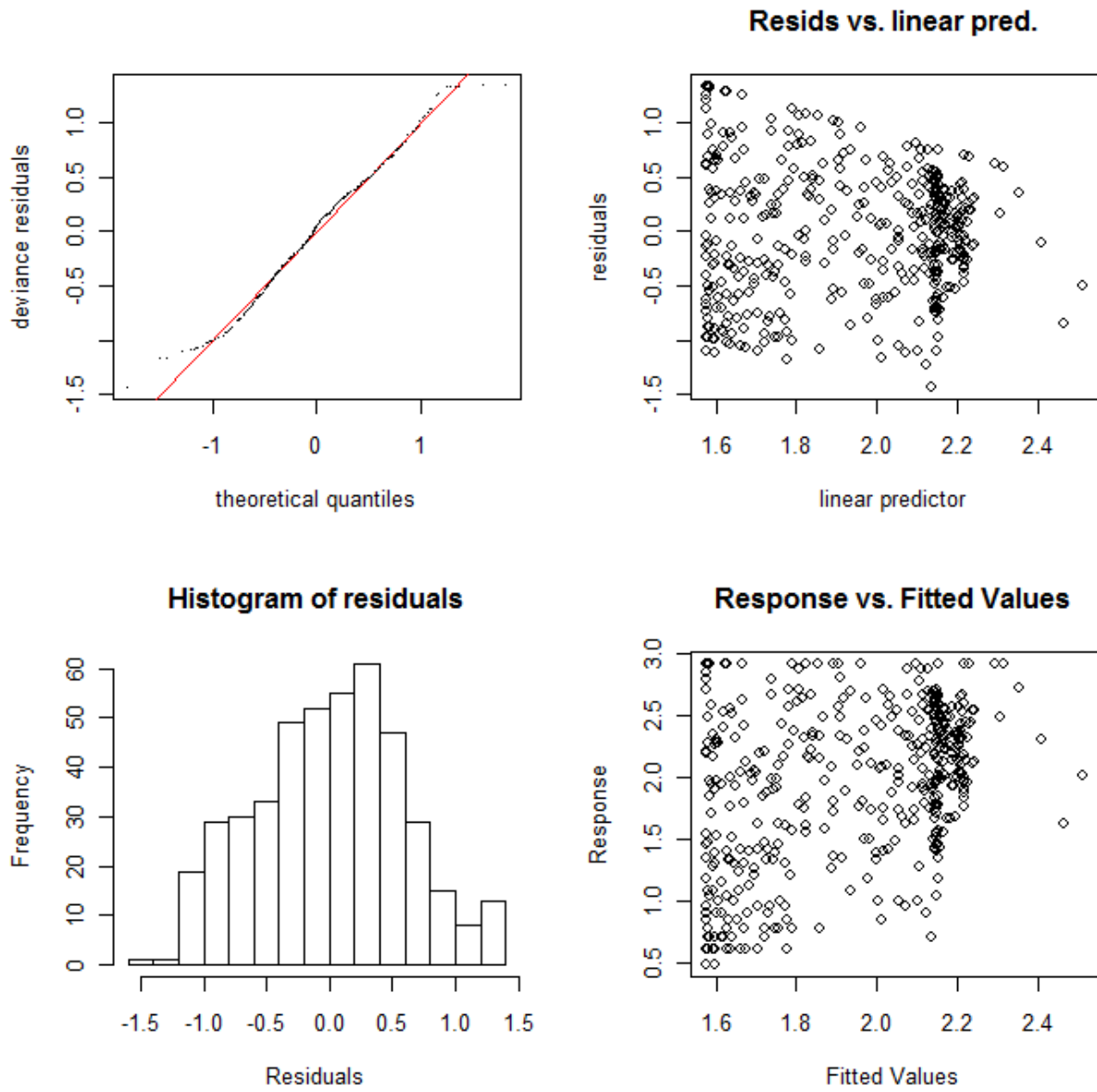


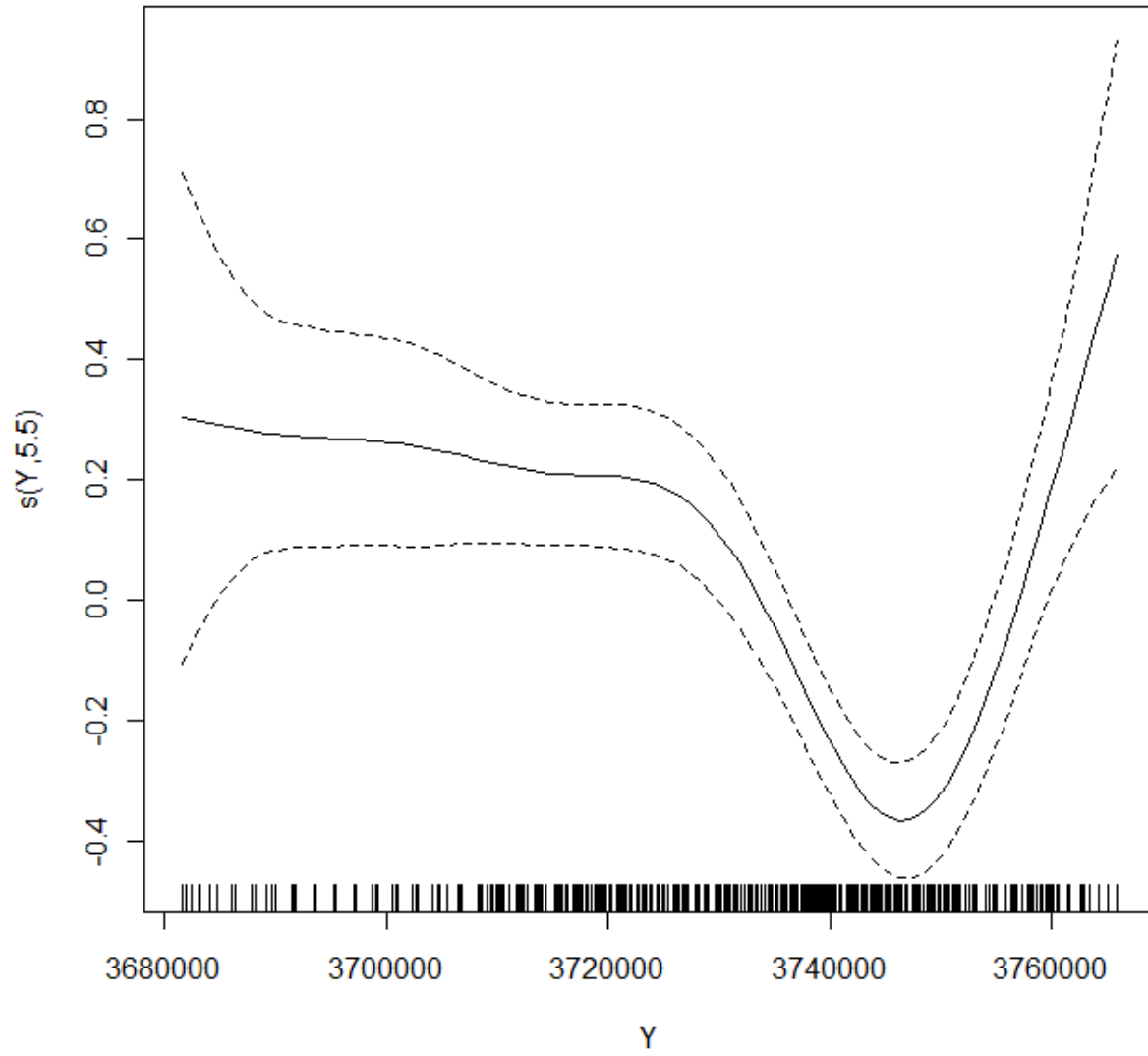
X: Deviance explained = 12.6%, p-value: 1.72e-09, edf=6.261



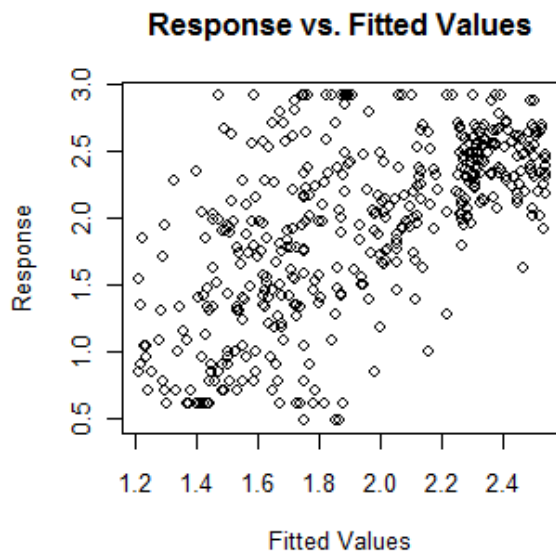
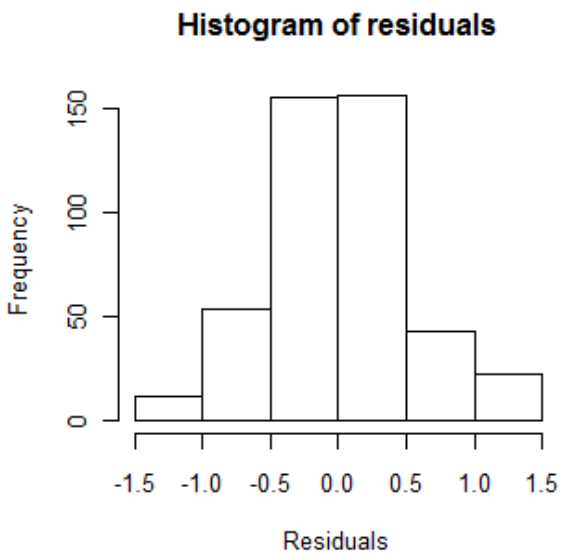
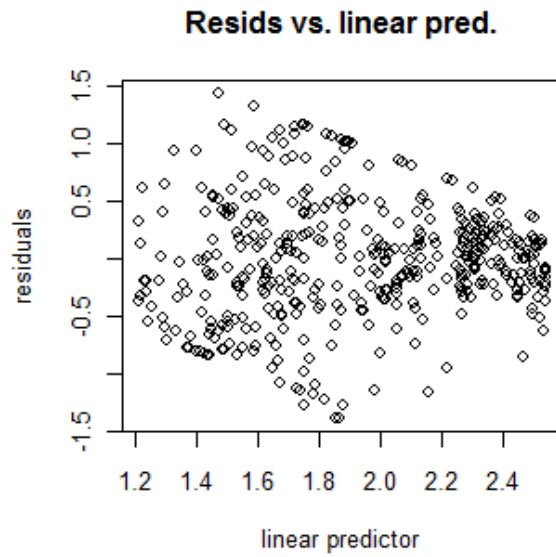
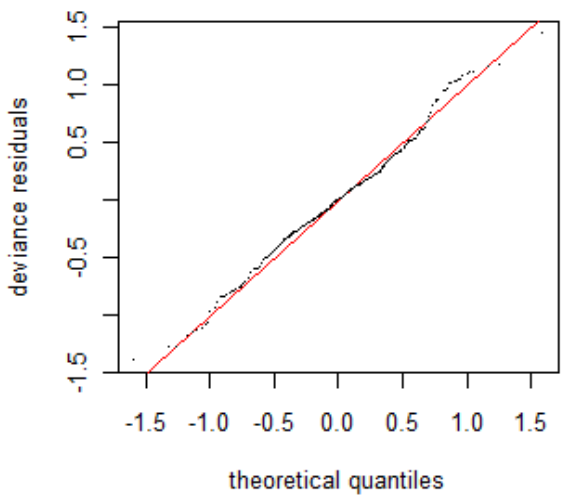


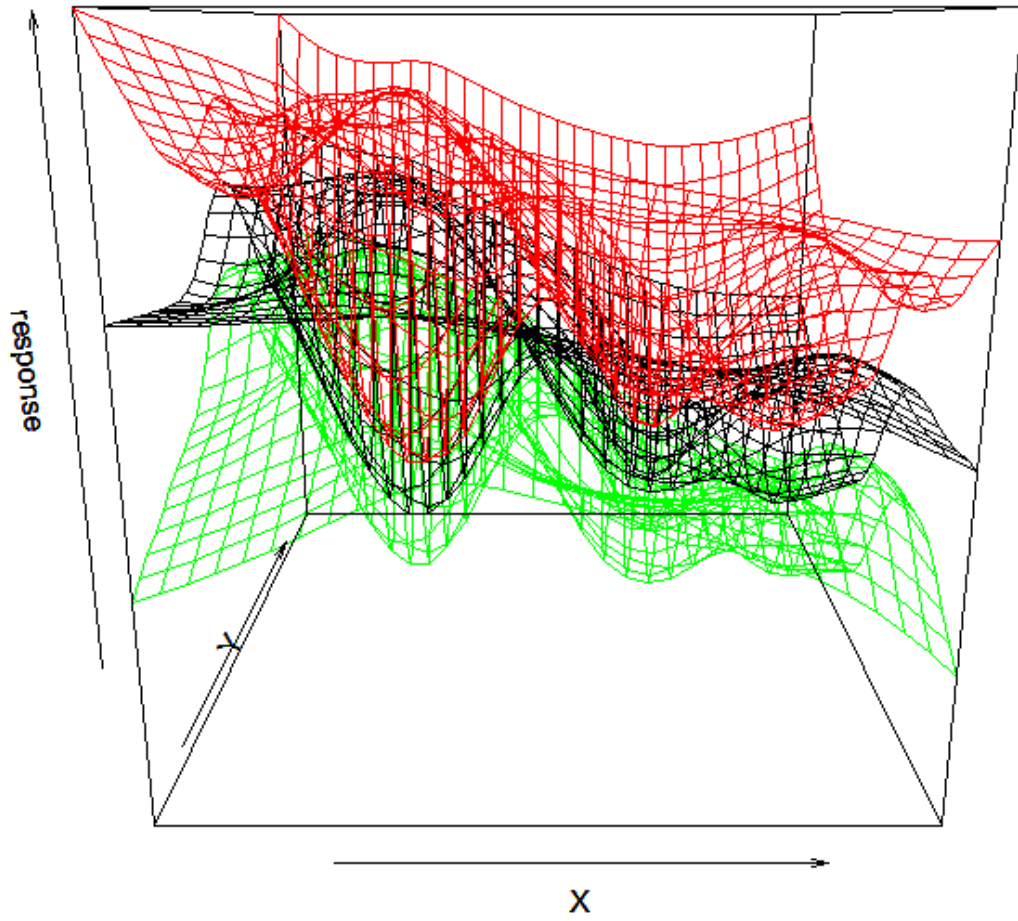
Y: Deviance explained = 16.2%, p-value: 2.96e-13, edf=5.503



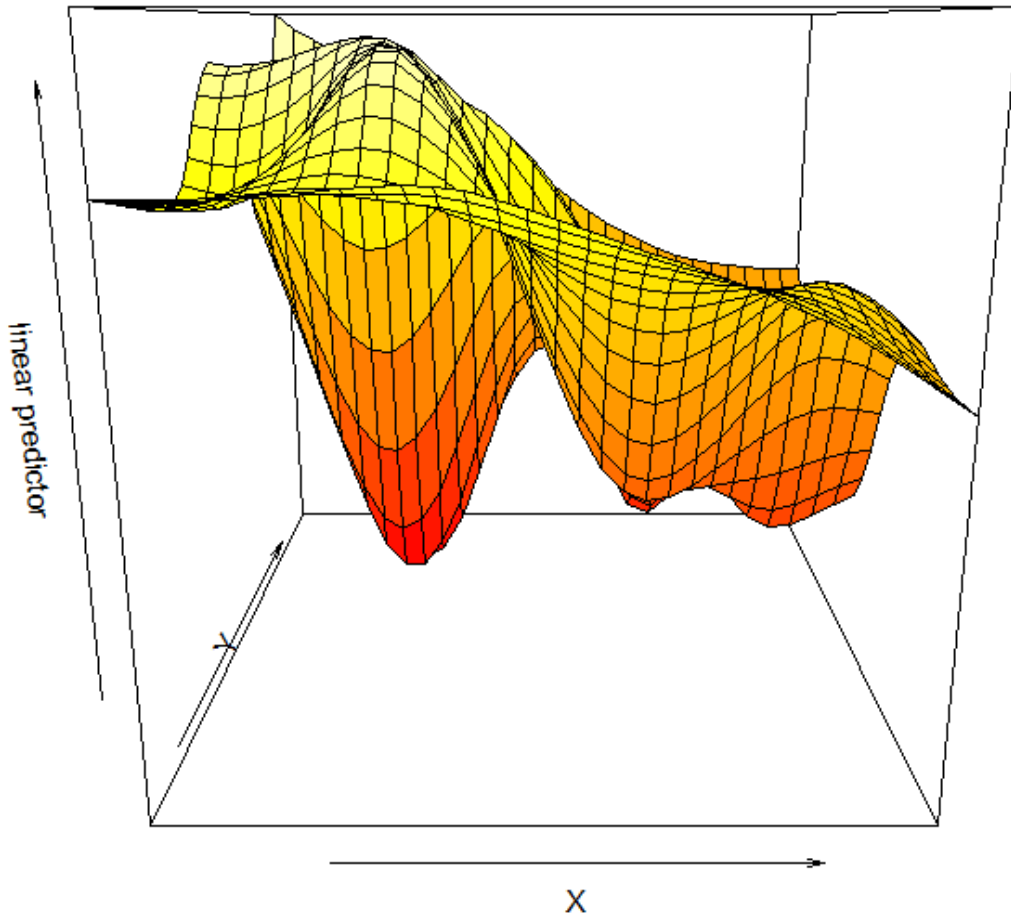


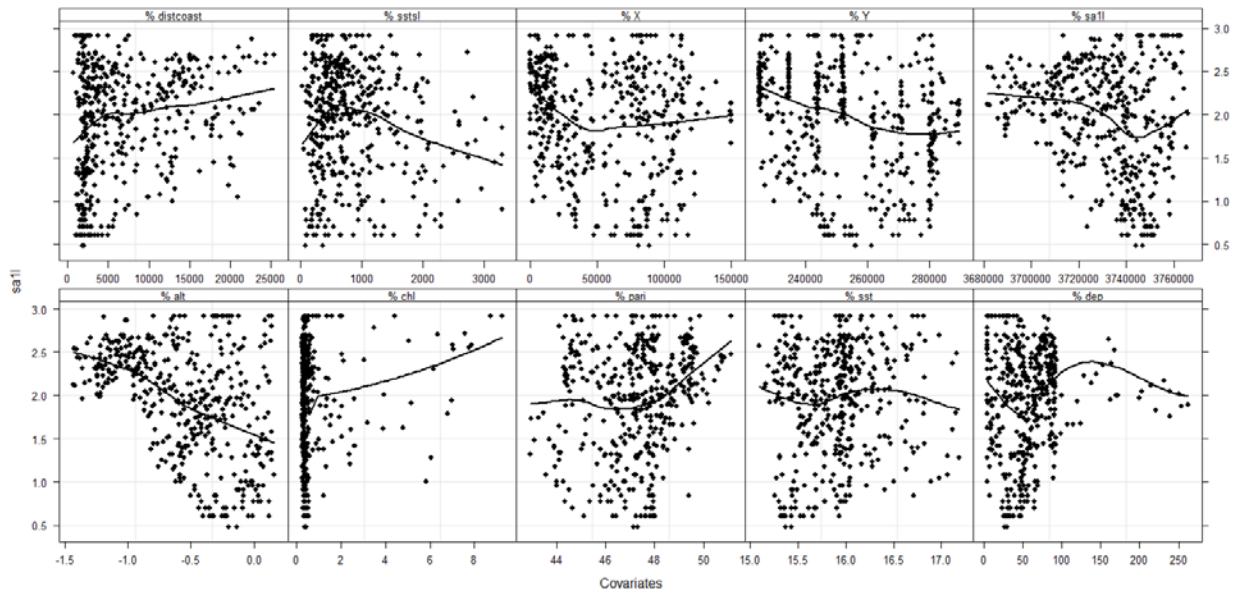
X,Y: Deviance explained = 38.3%, p-value: 2e-16, edf=22.98





red/green are +/- 2 s.e.





Cleveland dotplots και εξομαλυντές των επεξηγηματικών μεταβλητών με την εξαρτημένη μεταβλητή για την επισκόπηση της κατανομής των τιμών των πρώτων και της σχέσης τους με την δεύτερη

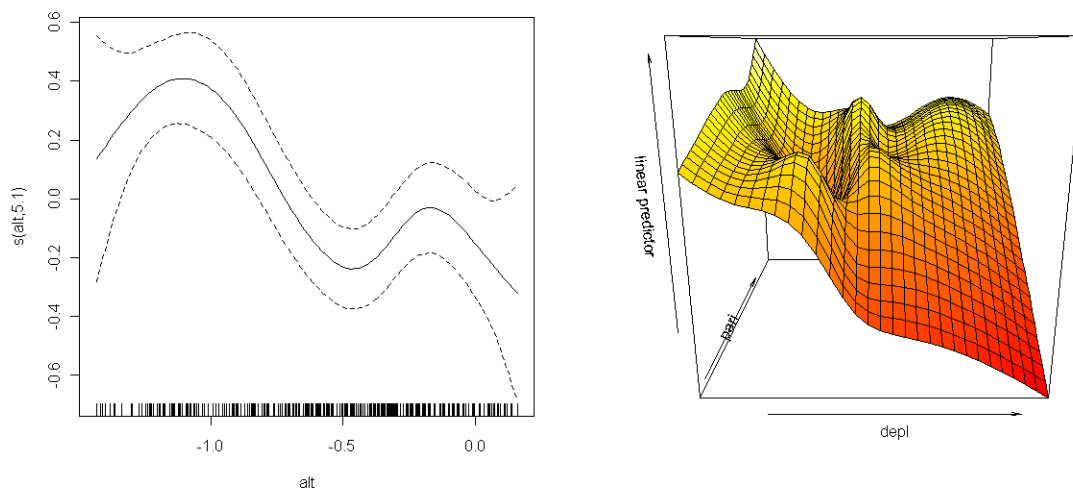
9. Παράρτημα III

Σχέσεις ανεξάρτητων μεταβλητών με την εξαρτημένη μεταβλητή, όπως εξάγονται από κάποιες προσεγγίσεις μοντελοποίησης της κατανομής ατόμων ειδών με δεδομένα παρουσίας-απουσίας.

1. Μοντέλα αφθονίας

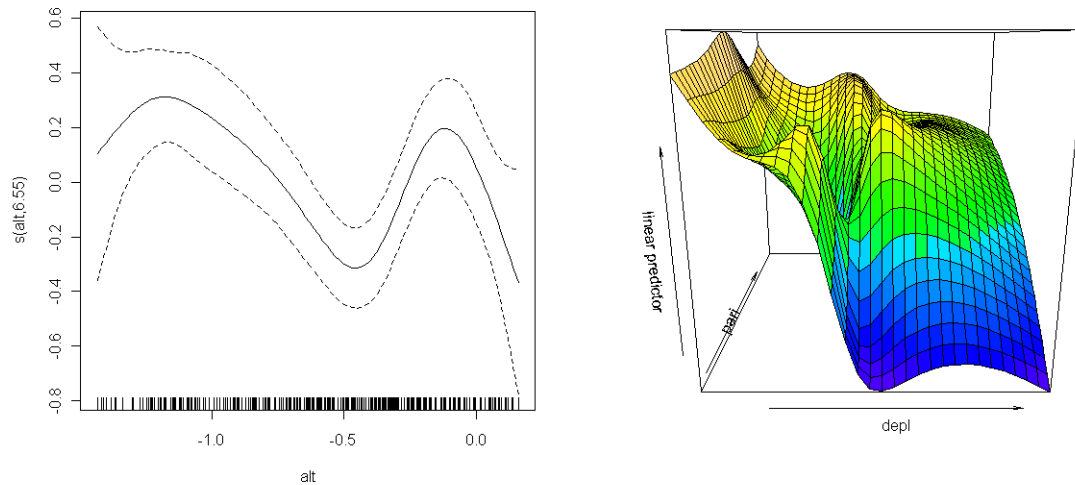
Από τα μοντέλα αφθονίας που χρησιμοποιήθηκαν στην παρούσα εργασία μόνο τα GAMs δίνουν αυτοματοποιημένα γραφήματα και αποτελέσματα σχέσεων μεταξύ των ανεξάρτητων και της εξαρτημένης μεταβλητής. Τα ASNN απαιτούν πρόσθετες εφαρμογές ώστε να εξαχθεί η παραπάνω πληροφορία, όπως περιγράφονται στους Elith et al. (2005). Οι μέθοδοι Kriging βασίζονται στην αυτοσυσχέτιση των δεδομένων και δεν χρησιμοποιούν περιβαλλοντικές επεξηγηματικές μεταβλητές, οπότε στα αποτελέσματα τους φαίνεται η σχέση αυτοσυσχέτισης μεταξύ ζευγών τιμών του αρχικού δείγματος.

1.1 GAM1



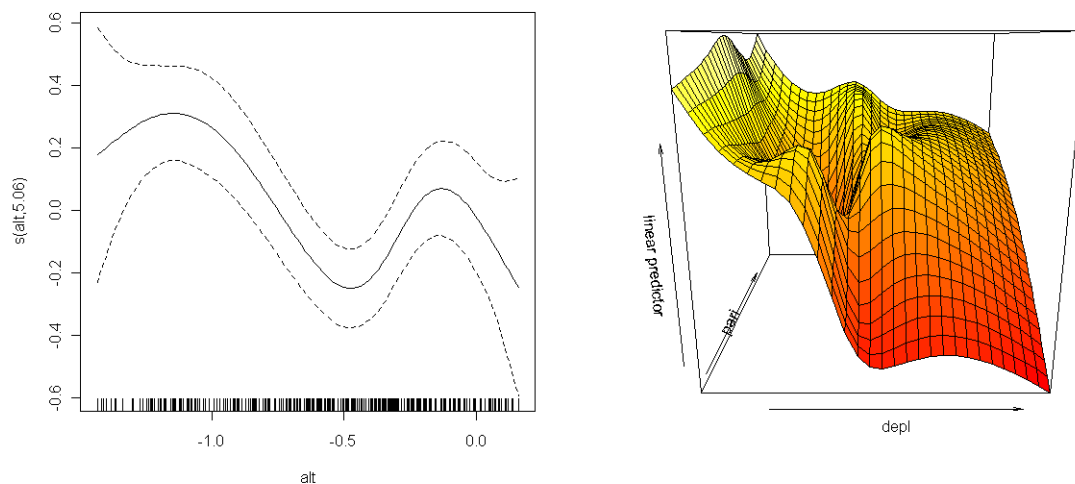
Τόσο η ALT όσο και ο παράγοντας αλληλεπίδρασης PAR:DEP είναι στατιστικά σημαντικοί ($p\text{-value} \ll 0.05$) στο μοντέλο GAM1. Οι υπόλοιπες μεταβλητές εξαιρέθηκαν από τη διαδικασία επιλογής μεταβλητών και το ίδιο ισχύει για όλα τα μοντέλα παλινδρόμησης που ακολουθούν. Οπότε σύμφωνα με το GAM1 η ALT και ο παράγοντας αλληλεπίδρασης PAR:DEP είναι οι στατιστικά σημαντικές μεταβλητές που ερμηνεύουν μεγάλο μέρος της διακύμανσης της εξαρτημένης μεταβλητής.

1.2 GAM2



Τόσο η ALT όσο και ο παράγοντας αλληλεπίδρασης PAR:DEP είναι στατιστικά σημαντικοί ($p\text{-value} \ll 0.05$) στο μοντέλο GAM2. Οπότε σύμφωνα με το GAM2 η ALT, ο παράγοντας αλληλεπίδρασης PAR:DEP και ο κατηγορικός παράγοντας DDND είναι οι στατιστικά σημαντικές μεταβλητές που ερμηνεύουν μεγάλο μέρος της διακύμανσης της εξαρτημένης μεταβλητής. Η προσθήκη του κατηγορικού παράγοντα αυξάνει την εξηγούμενη διακύμανση συγκριτικά με το GAM1.

1.3 GAMM



Όπως και στις παραπάνω περιπτώσεις έτσι και εδώ η ALT και ο παράγοντας αλληλεπίδρασης PAR:DEP είναι στατιστικά σημαντικοί ($p\text{-value} \ll 0.05$). Είναι οπτικά διακριτές οι ομοιότητες μεταξύ του GAM2 και GAMM. Οπότε σύμφωνα με το GAMM η ALT, ο παράγοντας αλληλεπίδρασης PAR:DEP, ο κατηγορικός παράγοντας DDND και ένας παράγοντας που μοντελοποιεί τις χωρικές συσχετίσεις των τιμών της εξαρτημένης μεταβλητής

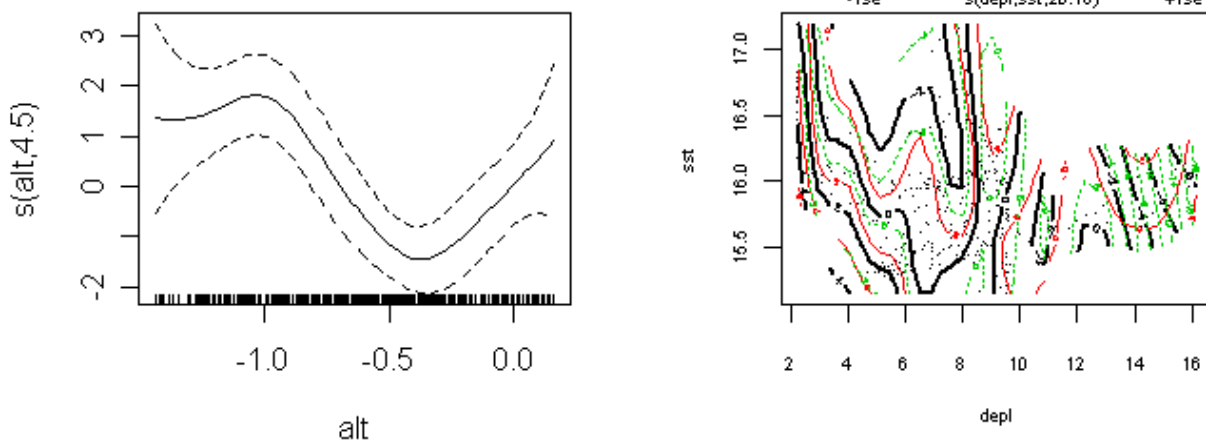
είναι οι στατιστικά σημαντικές μεταβλητές που ερμηνεύουν μεγάλο μέρος της διακύμανσης της εξαρτημένης μεταβλητής. Παρά την προσθήκη του παράγοντα αυτοσυσχέτισης δεν φαίνεται να βελτιώνεται η απόδοση του GAMM περισσότερο από το GAM2 παρά το γεγονός ότι αυξάνεται η πολυπλοκότητα του σημαντικά.

2. Μοντέλα παρουσίας-απουσίας

Από τα μοντέλα παρουσίας-απουσίας που αναπτύχθηκαν μόνο ορισμένα έχουν την ικανότητα να αναπαριστούν τις σχέσεις εξαρτημένης-ανεξάρτητων μεταβλητών. Αυτά είναι τα GAMs, GAMMs, BRTs, MARS, MAXENT.

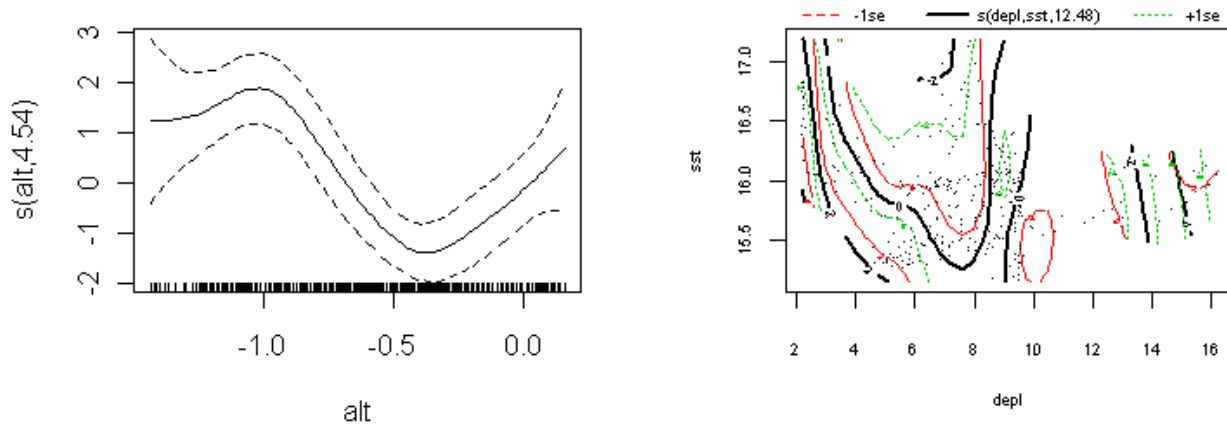
2.1 GAMs

Τόσο η ALT όσο και ο παράγοντας αλληλεπίδρασης SST:DEP είναι στατιστικά σημαντικοί ($p\text{-value} \ll 0.05$) στο μοντέλο GAM, στο οποίο επίσης συμμετέχει και η κατηγορική μεταβλητή DDND. Στα γραφήματα που ακολουθούν φαίνεται η επίδραση των επεξηγηματικών μεταβλητών στο μοντέλο. Οι υπόλοιπες μεταβλητές εξαιρέθηκαν από τη διαδικασία επιλογής μεταβλητών και το ίδιο ισχύει για όλα τα μοντέλα παλινδρόμησης που ακολουθούν. Οπότε σύμφωνα με το GAM η ALT και ο παράγοντας αλληλεπίδρασης SST:DEP είναι οι στατιστικά σημαντικές μεταβλητές που ερμηνεύουν προσθετικά μεγάλο μέρος της διακύμανσης της εξαρτημένης μεταβλητής.



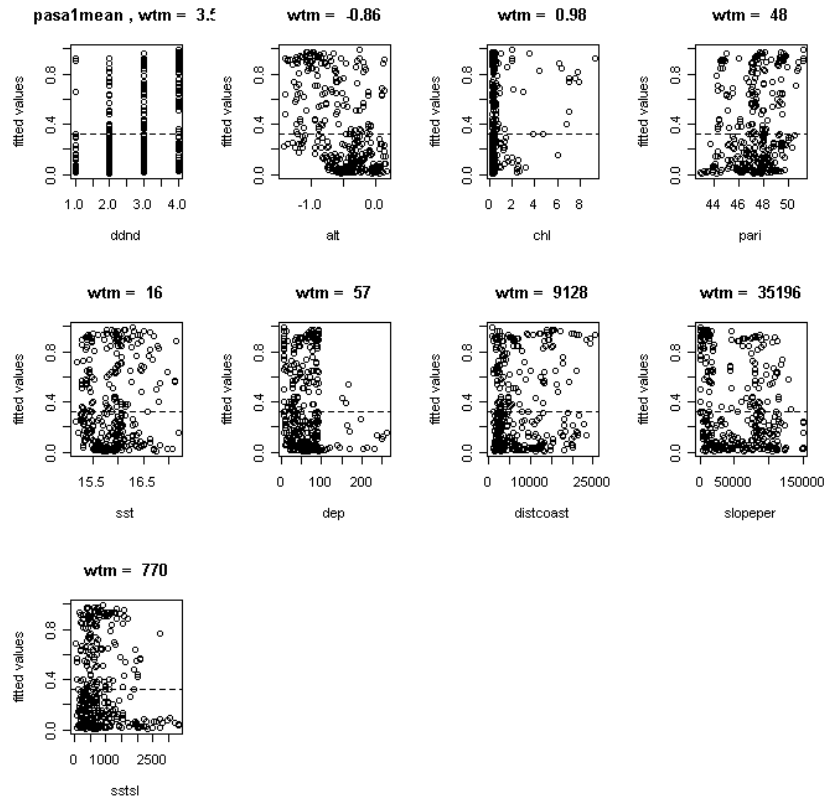
2.2 GAMMs

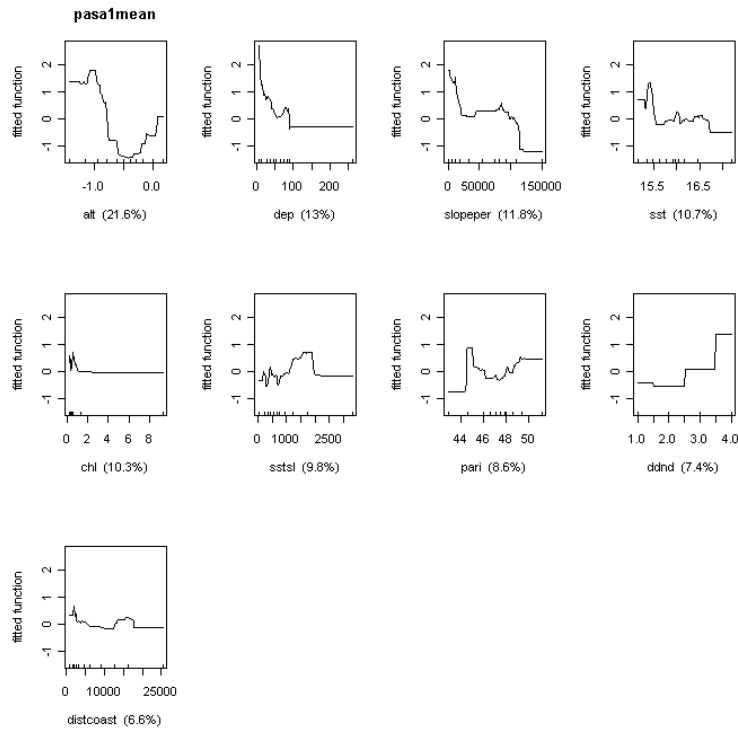
Αντίστοιχη είναι η εικόνα και στο GAMM, όπου παρατηρούνται μικροδιαφορές σε σχέση με το GAM. Άλλωστε το δεύτερο είναι ένθετο του πρώτου. Ακολουθούν τα γραφήματα που περιγράφουν τις επιδράσεις των ανεξάρτητων μεταβλητών στην διακύμανση των τιμών της εξαρτημένης μεταβλητής. Εδώ πρέπει να σημειωθεί ότι το GAMM εμπεριέχει επιπλέον του GAM ένα παράγοντα χωρικής συσχέτισης των αρχικών δεδομένων, ο οποίος είναι επίσης στατιστικά σημαντικός στην εφαρμογή του μοντέλου.



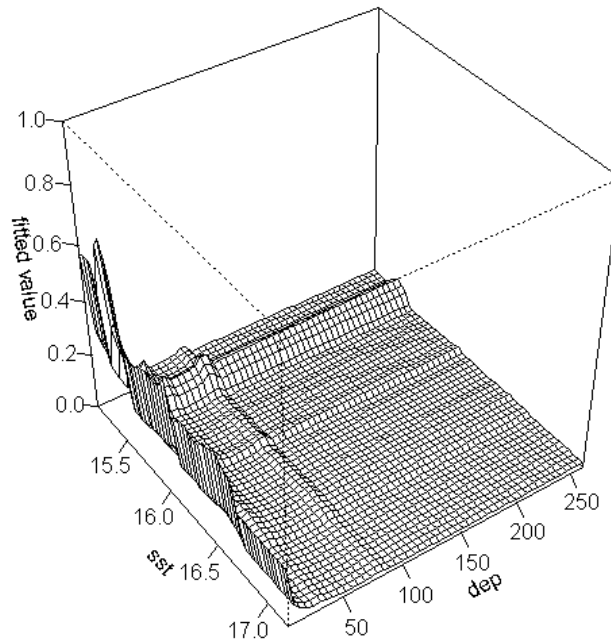
2.3 BRTs

Παρακάτω απεικονίζονται οι μεταβλητές που μετείχαν στο BRT μοντέλο. Τα γραφήματα μερικής εξάρτησης δείχνουν τη διακύμανση της εκάστοτε μεταβλητής και έχουν κεντραριστεί ώστε να έχουν μηδενικό μέσο γύρω από την κατανομή των δεδομένων. Στο BRT τελικά αξιοποιούνται περισσότερες μεταβλητές από τα μοντέλα παλινδρόμησης, με διαφορετική βαρύτητα η καθεμία, όπως φαίνεται και παρακάτω.





Ακολουθεί ένα γράφημα μερικής εξάρτησης, τριών διαστάσεων, που απεικονίζει την ισχυρότερη αλληλεπίδραση (SST:DEP) μεταξύ των μεταβλητών του μοντέλου. Αυτή είναι σε ταύτιση με την αλληλεπίδραση που έχουν αναδείξει τα μοντέλα GAM και GAMM, ως την πλέον σημαντική. Όλες οι μεταβλητές, πλην των δύο που απεικονίζονται διατηρούνται σταθερές στο μέσο τους, για την απόδοση του γραφήματος.



Ο παρακάτω πίνακας παρουσιάζει τη σχετική συνεισφορά των ανεξάρτητων μεταβλητών που χρησιμοποιήθηκαν στην ανάπτυξη του BRT μοντέλου. Υπάρχει σχετική αντιστοιχία με τα αποτελέσματα των μοντέλων παλινδρόμησης, αν και στην προσέγγιση των BRT η διαδικασία επιλογής μεταβλητών δεν απέκλισε καμία από τις μεταβλητές που δοκιμάστηκαν. Τα BRTs άλλωστε διαχειρίζονται καλύτερα από τα GAMs αλληλεπιδράσεις μεταβλητών και συγγραμικότητες, οπότε δεν υπόκεινται σε τόσο έντονους περιορισμούς ως προς τη χρήση μεταβλητών, όπως τα GAMs.

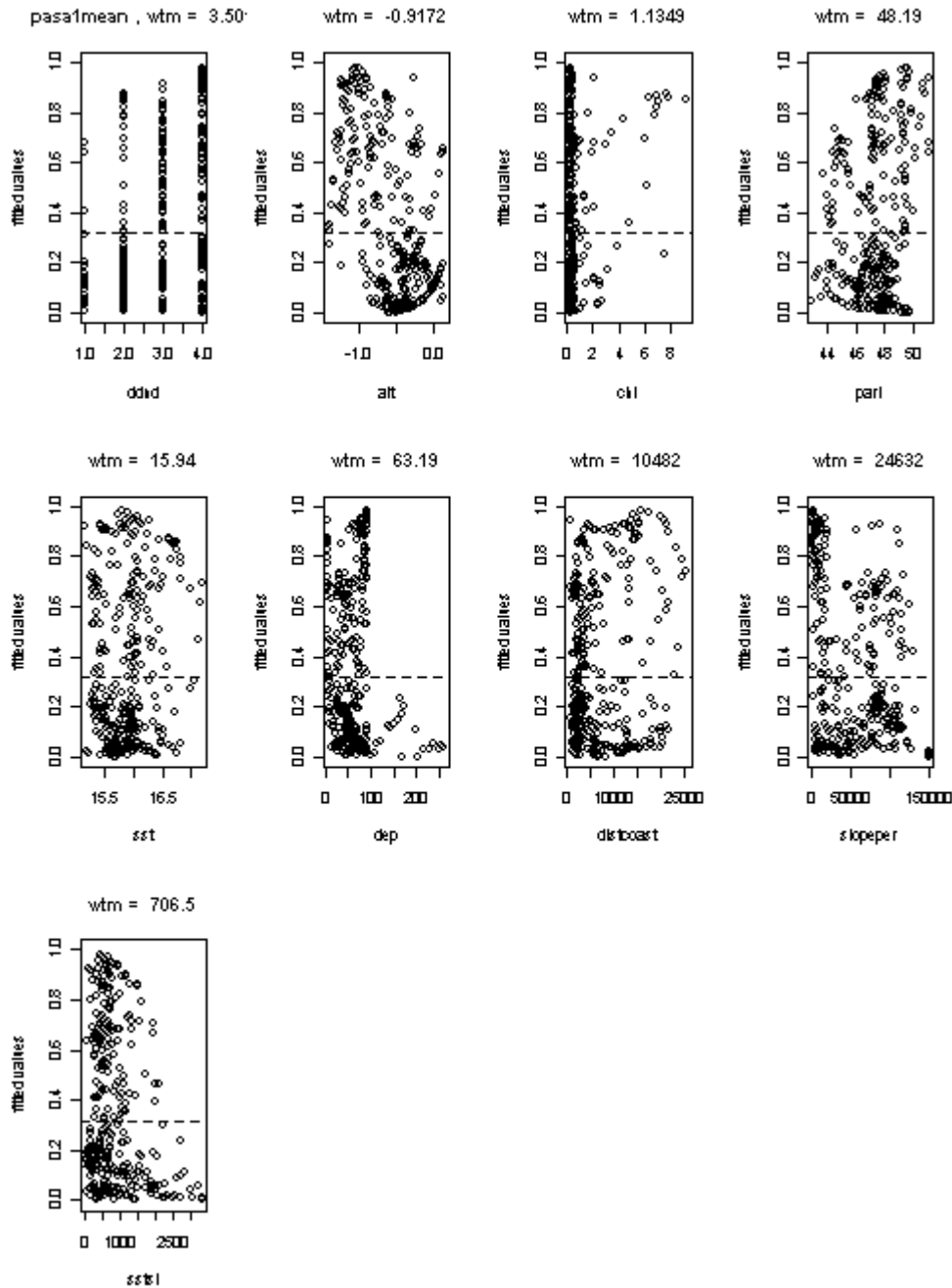
Μεταβλητή	Αναλογική συνεισφορά στο μοντέλο
ALT	21.617720
DEP	12.996731
Slopeper	11.844079
SST	10.744043
CHL	10.342609
SSTsl	9.816546
PAR	8.622069
DDND	7.449729
DistCoast	6.566474

2.4 MARS

Ο παρακάτω πίνακας παρουσιάζει τη σχετική συνεισφορά των ανεξάρτητων μεταβλητών που χρησιμοποιήθηκαν στην ανάπτυξη του MARS μοντέλου. Υπάρχει σχετική αντιστοιχία με τα αποτελέσματα των μοντέλων παλινδρόμησης και BRTs, αν και στην προσέγγιση των MARS η αρχική διαδικασία επιλογής μεταβλητών δεν απέκλισε καμία από τις μεταβλητές που δοκιμάστηκαν. Ωστόσο, σε μια εκ των μεταβλητών (DistCoast) αποδίδεται μηδενική βαρύτητα. Είναι χαρακτηριστικό στον πίνακα ότι υπάρχουν μικρές διαφοροποιήσεις ως προς τη βαρύτητα των μεταβλητών γεγονός που μπορεί να αποδοθεί στον διαφορετικό τρόπο που προσεγγίζει η εκάστοτε τεχνική τα δεδομένα και τις αλληλεπιδράσεις τους ώστε να αυξήσει την εφαρμογή της στην κατανομή των δεδομένων που καλείται να προσομοιώσει.

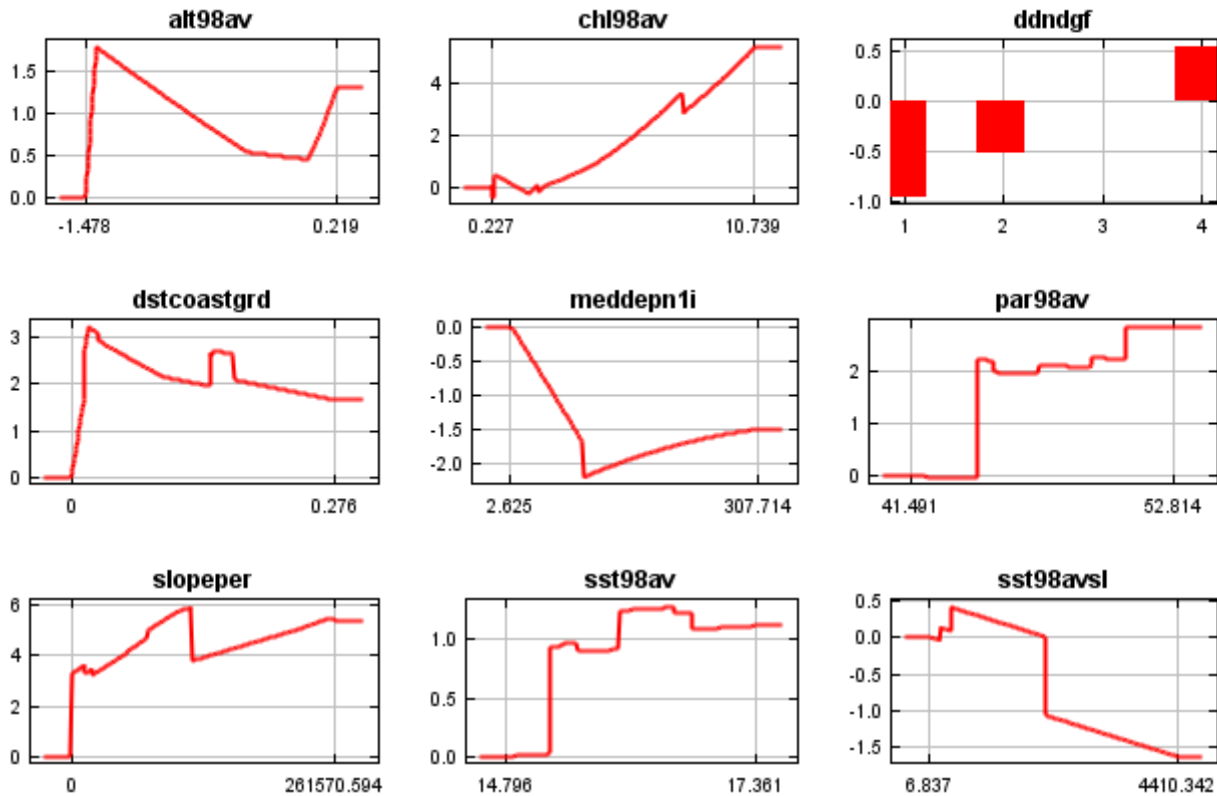
	Μεταβλητή	delta dev	Βαθ. ελευθερίας	p-value
1	ALT	83.98	3	4.30255e-18
2	DDND	37.28	1	1.023166e-09
3	SST	16.759	2	0.0002295612
4	DEP	13.776	2	0.001019904
5	Slopeper	13.209	1	0.0002786319
6	CHL	8.916	2	0.01158444
7	PAR	7.836	1	0.00512057
8	SSTsl	4.688	1	0.03038179
9	DistCoast	0	0	0

Τα γραφήματα που ακολουθούν δείχνουν την κατανομή των σφαλμάτων (άξονας Y) της κάθε μεταβλητής ως προς τις τιμές της (άξονας X).

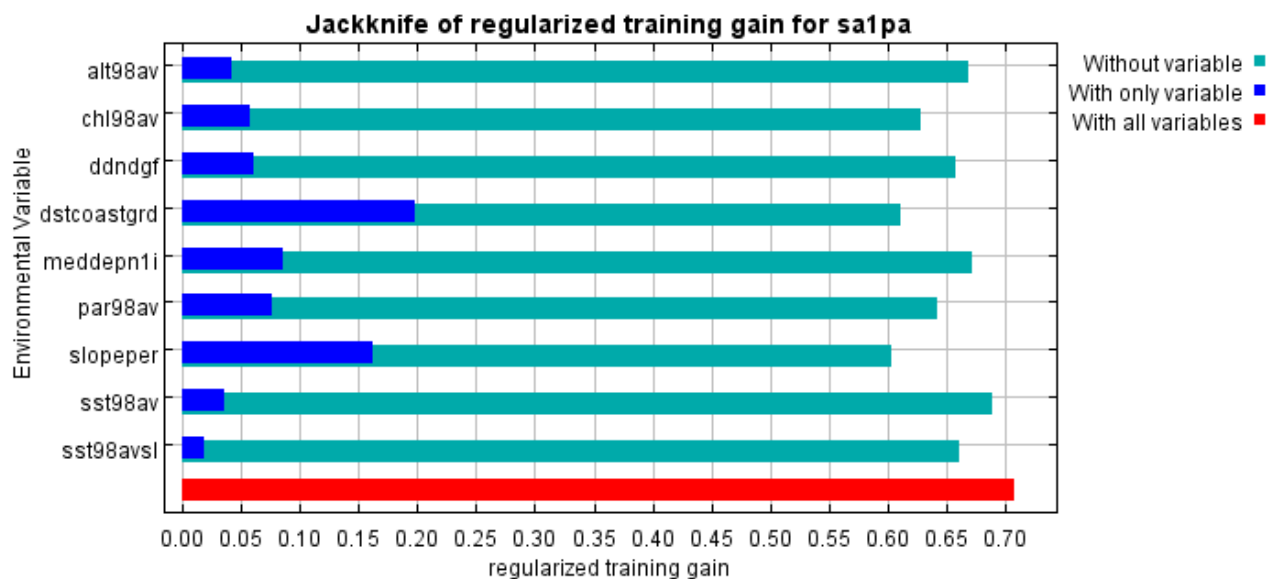


2.5 MAXENT

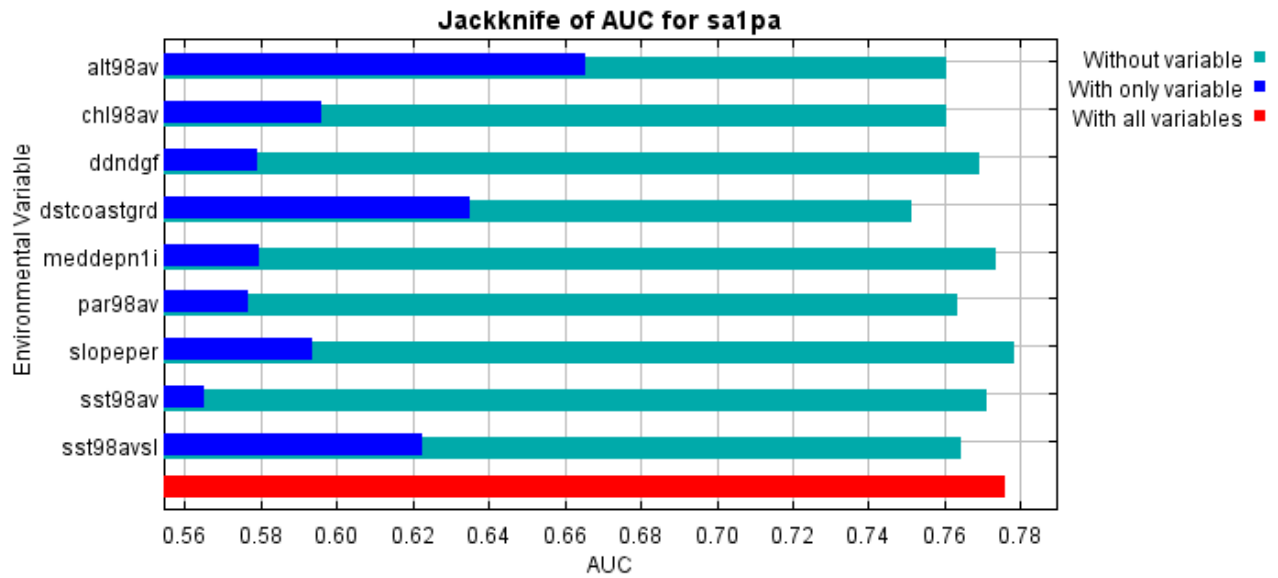
Οι παρακάτω καμπύλες δείχνουν πως κάθε μεταβλητή επηρεάζει την πρόβλεψη του MAXENT. Πρέπει να σημειωθεί ότι οι καμπύλες αναφέρονται στην οριακή επίδραση της μεταβολής μιας αποκλειστικά μεταβλητής τη φορά, ενώ οι άλλες διατηρούνται σταθερές στη μέση τιμή τους. Από την άλλη το μοντέλο δύναται να αξιοποιεί την παράλληλη μεταβολή όλων των μεταβλητών.



Η παρακάτω εικόνα δείχνει τα αποτελέσματα ενός τεστ jackknife για την σημαντικότητα των μεταβλητών στο μοντέλο. Η μεταβλητή με τη μεγαλύτερη συνεισφορά στο μοντέλο όταν χρησιμοποιείται μόνη της είναι η DistCoast, οπότε θεωρητικά αποτελεί τη μεταβλητή με την πλέον χρήσιμη πληροφορία. Από την άλλη η συμμετοχή της μεταβλητής Slopeper δείχνει ότι η παρουσία της εμπεριέχει το μεγαλύτερο ποσό πληροφορίας που δεν είναι παρόν σε άλλη μεταβλητή.



Τέλος παρουσιάζεται ένα ανάλογο jackknife τεστ που χρησιμοποιεί το AUC ως κριτήριο συμμετοχής των μεταβλητών στο μοντέλο.



Σε αυτή την εικόνα είναι φανερό ότι υπάρχουν κάποιες ομοιότητες με τις μεταβλητές που επιλέγουν άλλες τεχνικές μοντελοποίησης (π.χ. ALT ως τη σημαντικότερη), όμως υπάρχουν αρκετές διαφορές ως προς την ιεράρχηση των υπολοίπων.

10. Παράρτημα IV

Στο παράρτημα αυτό παρατίθενται οι δημοσιεύσεις και ανακοινώσεις που προήλθαν από τη διατριβή:

- Palialexis A, Georgakarakos S, Karakassis I, Lika K, Valavanis VD (2011). Fish distribution predictions from different points of view: Comparing Associative Neural Networks, Geostatistics and Regression Models. *Hydrobiologia* **670**, 165–188.
- Palialexis A, Georgakarakos S, Karakassis I, Lika K, Valavanis VD (2011). Prediction of marine species distribution from presence/absence acoustic data: Comparing the fitting efficiency and the predictive capacity of resource selection functions. *Hydrobiologia* **670**, 241–266.
- Palialexis A, Georgakarakos S, Lika K, Valavanis VD (2009). Comparing novel approaches used for prediction of species distribution from presence/absence acoustic data. Proceedings of the **Second International Conference on Environmental Management, Engineering, Planning and Economics (CEMEPE 09)**. June 21-26, 2009, Mykonos, Greece. 379-85.
- Palialexis A, Georgakarakos S, Lika K, Valavanis VD (2009). Use of GIS, remote sensing and regression models for the identification and forecast of small pelagic fish distribution. Proceedings of the **Second International Conference on Environmental Management, Engineering, Planning and Economics (CEMEPE 09)**. June 21-26, 2009, Mykonos, Greece. 1743-49.
- Valavanis VD, GJ Pierce, AF Zuur, A Palialexis, A Saveliev, I Katara, J Wang (2008). Modelling of Essential Fish Habitat based on Remote Sensing, Spatial Analysis and GIS. *Hydrobiologia* **612(1)**, 5-20