

Biodiversity Informatics Facility

Species Distribution Modeling

Guide [Online Resources](#) [Training](#) [Search](#)

Developing a species' distribution model
Species Distribution Modeling > Guide to species distribution modeling > Developing a species' distribution model

Modeling algorithms
This online guide has been adapted from a Network for Conservation Practitioners (NCP) synthesis document developed by Richard G. Pearson, titled *Species' distribution modeling for conservation educators and practitioners*. The document, in its entirety, is also [available for download on the pdf \(677kb format\)](#).

http://biodiversityinformatics.amnh.org/index.php?section_id=105&content_id=345


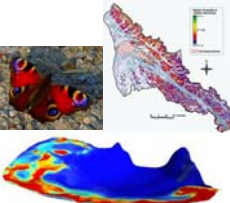


Table 3. Some published methods for species' distribution modeling.

Method(s) ¹	Model/software name ²	Species data type ³	Key Reference/URL
Gower Metric	DOMAIN ⁴	presence-only	Carroll et al. 1992 http://www.cbr.umd.edu
Ecological Niche Factor Analysis (ENFA)	ESOMAPPER ⁵	presence and background	Hirzel et al. 2002 http://www.univ-lorraine.fr/~esomapper/
Maximum Entropy	MAXENT ⁶	presence and background	Elith et al. 2006 http://www.scripps.edu/~schapero/maxent/
Genetic algorithm (GA)	GAPP ⁷	pseudo-absence ⁸	Shabazi and Peters 1999 http://www.itsmacor.org/Reiki/sgaps/
Artificial Neural Network (ANN)	SPECIES	presence and absence (or pseudo-absence)	Pearson et al. 2002
Regression: generalized linear model (GLM), generalized additive model (GAM), boosted regression trees (BRT), multivariate adaptive regression splines (MARS)	Implemented in R ⁹	presence and absence (or pseudo-absence)	Elith et al. 2002 Elith et al. 2006 Elith et al. 2008 Elith et al. 2011
Multiple methods	ESOMOD	presence and absence (or pseudo-absence)	Thuiller 2003
Multiple methods	OpenModeller	depends on method implemented	http://www.modeller.sourceforge.net/

MAXENT Species Distribution Modelling

Αδρή Περιγραφή, software, και ένα case study



MAXENT

Πλεονεκτήματα με μια ματιά

- Presence-only
- Σχετικά πρόσφατη μέθοδος
- Υψηλή απόδοση ακρίβειας προβλέψεων σχετικά με παλαιότερους αλγόριθμους
- Ακόμη και με μικρό αριθμό δειγμάτων

MAXimum ENTropy

- Από τα δεδομένα παρουσίας:

Αποδίδει κατανομή πιθανότητας παρουσίας (*probability distribution*),

ουσιαστικά ένα δείκτη καταλληλότητας βιοτόπου (*Habitat Suitability Index*)

MAXimum ENTropy

- Από τα δεδομένα παρουσίας:

Βασική αρχή:

Η εκτίμηση πρέπει να συμβαδίζει με ότι είναι γνωστό από τα δεδομένα.

Παραδοχές που δεν υποστηρίζονται από τα δεδομένα αποφεύγονται.

MAXimum ENTropy

- Από τα δεδομένα παρουσίας:

Βασική αρχή:

Έτσι υπολογίζεται η κατανομή πιθανότητας που μεγιστοποιεί την εντροπία, "a measure of dispersedness", δηλ. την πιο «απλωμένη» κατανομή (πλησιέστερη στην ομοιόμορφη)

με βάση τους περιορισμούς από τα δεδομένα παρουσίας και το εύρος των περιβαλλοντικών μεταβλητών στην περιοχή έρευνας

Μέγιστη εντροπία

- Η εντροπία της πληροφορίας (information entropy) ή απλώς εντροπία, είναι ένα μέτρο της ποσότητας πληροφορίας που περιέχεται σ' ένα μήνυμα.
- Όπως δηλαδή κατά την κλασσική έννοια η εντροπία αποτελεί ένα μέτρο του πλήθους των πιθανών μικροκαταστάσεων ενός συστήματος,
- η εντροπία κατά τη θεωρία της πληροφορίας (information theory) αποτελεί ένα μέτρο του πλήθους των πιθανών 'μεταφράσεων' που περιέχει ένα μήνυμα

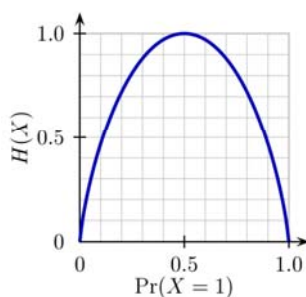
Μέγιστη εντροπία

- Ας θεωρήσουμε την περίπτωση στριψίματος ενός νομίσματος.
- Τα πιθανά αποτελέσματα είναι δύο.
– Κορώνα ή γράμματα.
- Σε κάθε ρίψη στην περίπτωση ενός κανονικού νομίσματος η πιθανότητα να έρθει κορώνα είναι ίση με την πιθανότητα να έρθει γράμματα και ίση με 1/2 (50% δηλαδή) .

Μέγιστη εντροπία

- Κάθε ρίψη αντιστοιχεί και σε ένα bit πληροφορίας.
-
- Σε αυτήν την περίπτωση η εντροπία είναι μέγιστη, αφού και η αβεβαιότητα είναι μέγιστη, και περιέχεται η μέγιστη δυνατή ποσότητα πληροφορίας στο μήνυμα (ρίψη του νομίσματος).
- Αν το νόμισμα ήταν κίβδηλο, τότε θα έφερνε διαρκώς π.χ. κορώνα, η πιθανότητα θα ήταν 1 (100%), η αβεβαιότητα ανύπαρκτη και η εντροπία μηδενική.

Μέγιστη εντροπία



Μέγιστη εντροπία

- Βλέπουμε δηλαδή ένα ενδιαφέρον χαρακτηριστικό της εντροπίας:
- Είναι μέγιστη όταν οι πιθανότητες είναι ίσες και τότε το μήνυμα εμπεριέχει τη μέγιστη δυνατή ποσότητα πληροφορίας (ή το μέγιστο δυνατό ρυθμό μετάδοσής της).

ΕΝΤΡΟΠΙΑ

Το εύρος των δυνατοτήτων της πιθανής μεταβολής

«**Η μέτρηση θα πρέπει να είναι μέγιστη**, αν όλα τα αποτελέσματα είναι εξίσου πιθανό (η αβεβαιότητα είναι μεγαλύτερη όταν όλα τα πιθανά γεγονότα είναι ισοπίθανες)» Μέγιστη Εντροπία καταχώρηση της Wikipedia (επιστήμη των πληροφοριών)

ΕΝΤΡΟΠΙΑ

Η **εντροπία** είναι η έννοια μέσω της οποίας μετράται η **αταξία**, της οποίας η μέγιστη τιμή αντικατοπτρίζει την πλήρη **αποδιοργάνωση** (ομογενοποίηση των πάντων) και ισοδυναμεί με την παύση της **ζωής** ή αλλιώς της **εξέλιξης**.



Άμμος ασχημάτιστη (αγνοήστε τη γραφή), μπορεί να πάρει άπειρα σχήματα



ΕΝΤΡΟΠΙΑ

Άμμος σε γλυπτό. Μπορεί μόνο να επανέλθει σε ασχημάτιστη άμμο



ΕΝΤΡΟΠΙΑ

MAXENT

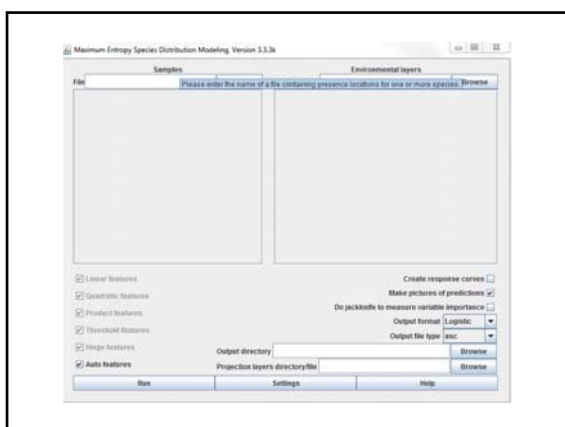
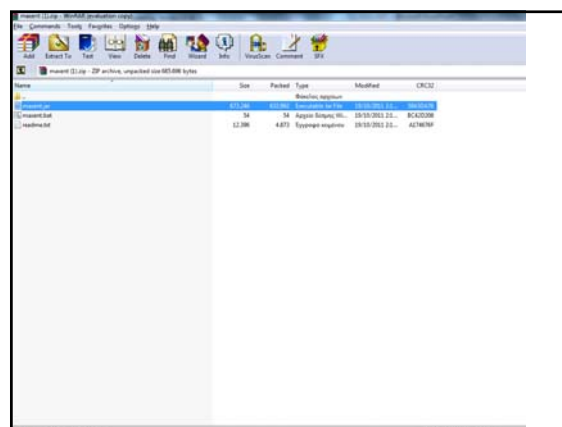
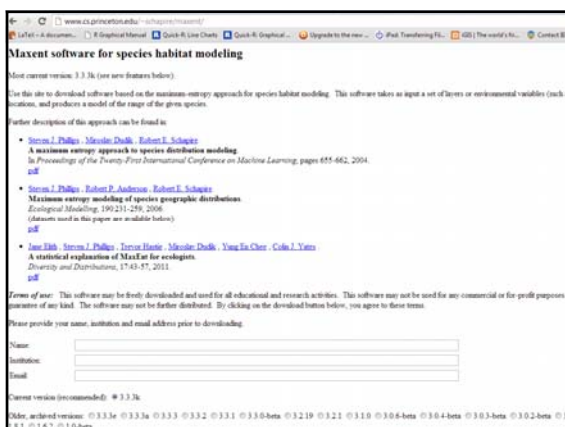
- Μαθηματικά.....

Εφαρμογή Java

- <http://www.cs.princeton.edu/~schapire/maxent/>

Υποβολή φόρμας και download εφαρμογών και tutorial file και data σε συμπιεσμένο φάκελο

Η εφαρμογή ανοίγει από το αρχείο maxent.jar



directory Κατανομής (Samples)

- Το αρχείο κατανομής, απλό *.csv με x,y, και species name.
- π.χ.

Species	X	Y
Αρπία θηλαστίς	630596	562336
Αρπία θηλαστίς	624798	492368
Αρπία θηλαστίς	590598	559817

directory Κατανομής (Samples)

- Το αρχείο κατανομής, απλό *.csv με x,y, και species name.
- π.χ.

```

species,longitude,latitude
bradypus_variegatus,-65.4,-10.3833
bradypus_variegatus,-65.3833,-10.3833
bradypus_variegatus,-65.1333,-16.8
bradypus_variegatus,-63.6667,-17.45
bradypus_variegatus,-63.85,-17.4
    
```

Directory Μεταβλητών Environmental Layers

- Μεταβλητές Περιβάλλοντος - Αρχεία ASCII GRID (π.χ. από μετατροπή οποιοδήποτε raster αρχείου στο GIS), συνεχής και κατηγορικές μεταβλητές
- ASCII files (ESRI or DIVA-GIS formats) ομαδοποιημένα σε ένα φάκελο. Δεν χρειάζεται μάσκα...

✓ Με ίδιο Μέγεθος Κελιού (cell size) και Έκταση (extent), δηλ. να «πατάνε» ακριβώς το ένα στο άλλο

Features

- Το MAXENT μοντελοποιεί τη σχέση των μεταβλητών με την παρουσία των ειδών σαν features (σύνολο μετατροπών των περιβαλλοντικών μεταβλητών),
- επιτρέποντας την εφαρμογή πολύπλοκων σχέσεων (π.χ. Interactions, Regression Splines)

Χαρακτηριστικά

- Το μέγεθος του δείγματος ορίζει τον τύπο των features που χρησιμοποιούνται στα μοντέλα όταν χρησιμοποιούνται τα default settings για την αποφυγή υπερ-παραμετροποιημένων μοντέλων

(linear, quadratic, threshold, hinge, product interaction)

Μικρό δείγμα → Μεγαλύτερο δείγμα

Classes of features:

1. Linear* → variable itself
2. Quadratic → square of variable
3. Product → product of two variables
4. Threshold → binary transformation (0, 1) of a continuous variable using a threshold
5. Hinge → like a linear feature, but constant below a threshold

* Categorical data: Binary feature → variable itself



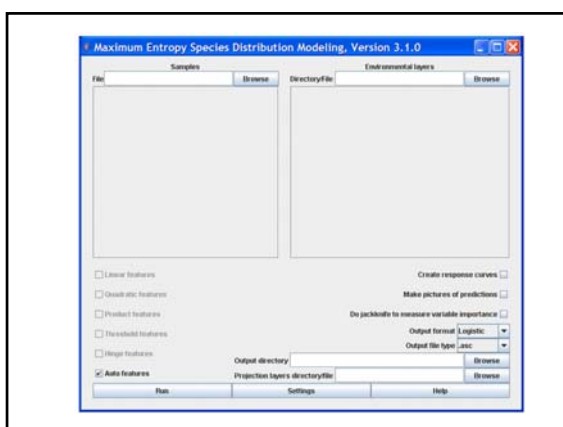
Ποια χαρακτηριστικά χρησιμοποιούνται για?

• Για να περιοριστεί η πιθανότητα κατανομής της μέγιστης εντροπίας (πιο απλωμένα), η οποία καθορίζει την πιθανότητα κατανομής των ειδών (πρόβλεψη εξόδου)

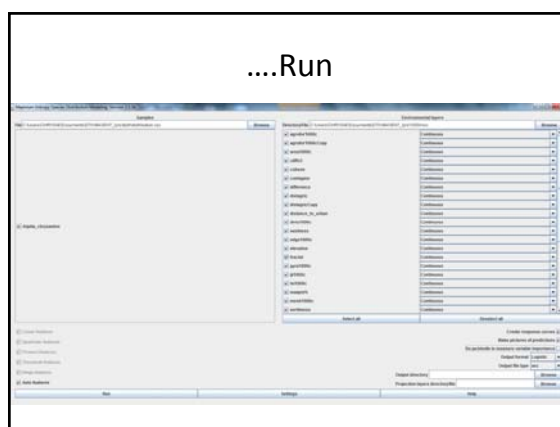
Constraints:

Linear* → mean
 Quadratic → variance
 Product → covariance
 Threshold → προσαρμογή σε μια αυθαίρετη απόκριση
 Hinge → like linear (but constant below a threshold)

* Categorical data: Binary feature → proportion

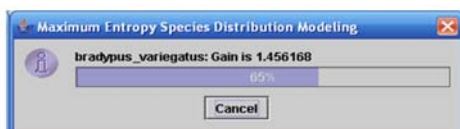


....Run



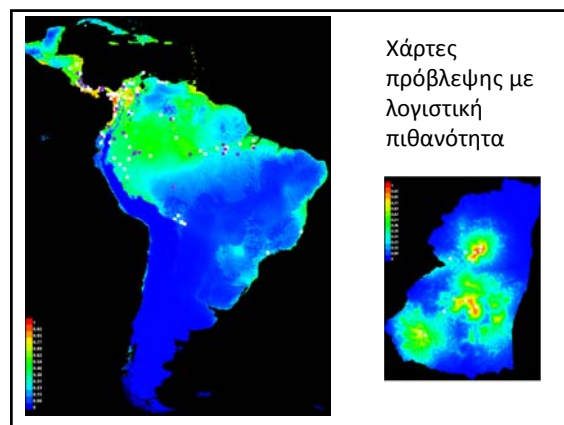
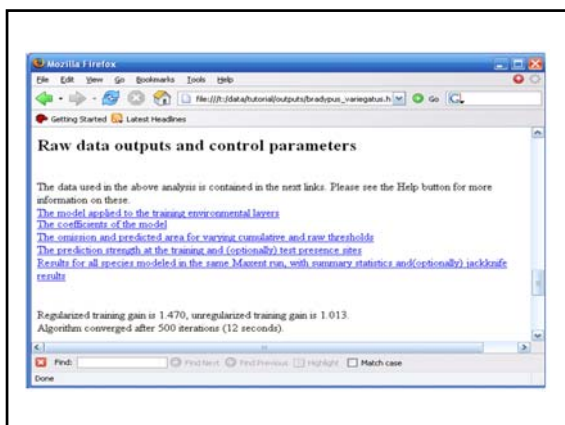
Model Fit

- Μια προσέγγιση με βάση το training gain, αντίστοιχο με την απόκλιση (deviance) των Γενικών Γραμμικών Μοντέλων.
- Μεγαλύτερο training gain, καλύτερο μοντέλο

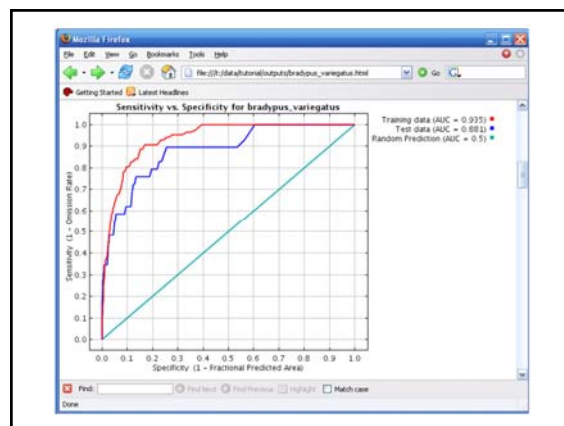
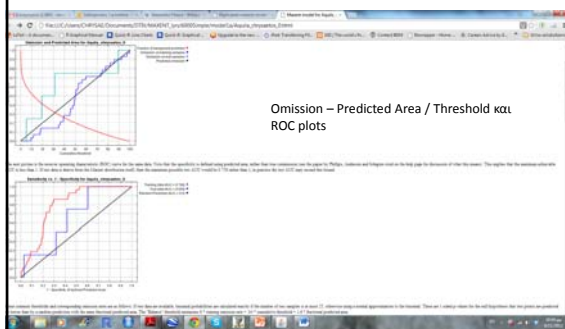


Output Directory

- Αποθηκεύει σε μορφή html την περιλήψη του μοντέλου, γραφήματα ως εικόνες και αρχεία csv με αποτελέσματα



Περίληψη μοντέλων




Ποια είναι η πιο σημαντική παράμετρος;

- Μια φυσική εφαρμογή των μοντέλων κατανομής ειδών είναι να απαντηθεί το ερώτημα, ποια από τις μεταβλητές έχουν μεγαλύτερη σημασία για τα είδη που διαμορφώνεται;
- Υπάρχουν περισσότεροι από ένας τρόποι για να απαντηθεί το ερώτημα αυτό?

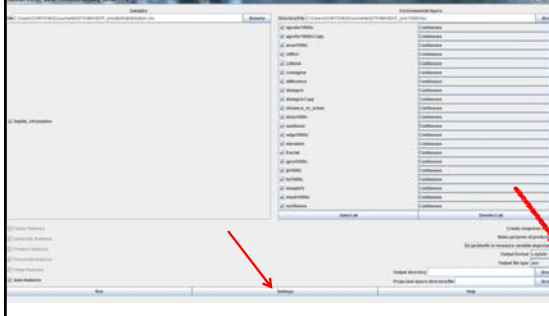
Συμβολή μεταβλητών στο μοντέλο (με βάση τη συμμετοχή της μεταβλητής στο συνολικό training gain).

Variable	Percent contribution	Permutation importance
distwater	40.2	57.9
distunpaved	35.8	18.9
distrocks	8.2	0
distags	7.3	9.9
disturban	4.2	7.2
distpaved	2.9	1.9
open500	0.8	1.2
sidi	0.5	2.9
distopen	0	0
distunimpv	0	0
mile4500	0	0
broad500	0	0

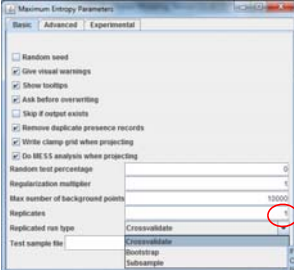
Καμπύλες απόκρισης (Response curves), marginal και ανά μεταβλητή χωριστά



επιπλέον δυνατότητες, επαναλήψεις και variable jackknife



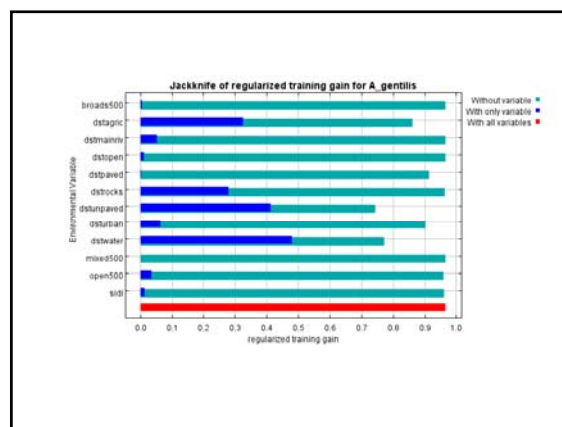
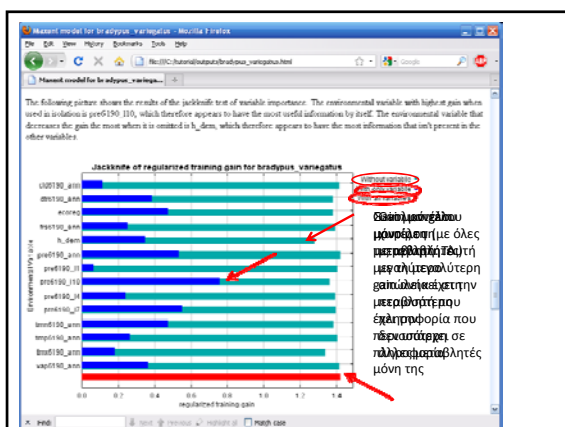
Από το μενού settings, δημιουργία επαναλήψεων, διαχωρισμός του δείγματος σε σύνολα training και test με πολλούς τρόπους



Jackknife

Χρήσιμο για τη διερεύνηση των πιο σημαντικών μεταβλητών, τρέχει το μοντέλο $2n + 1$ φορές με όλες τις μεταβλητές, με καθεμία ξεχωριστά και αφαιρώντας από μία μεταβλητή

(όπου n ο αριθμός περιβαλλοντικών μεταβλητών)



Case Study:

- Κατανομή του Χρυσαιτού (*Aquila chrysaetos*, L.) στη Θράκη...

Το είδος

Μεγάλο αρπακτικό, 3,5-6 kg, φτερά μέχρι 2,27 m

Ζει πολλά χρόνια, 0-2 μικρά ανά έτος

Διατηρεί μόνιμες επικράτειες ~100 km²

Το είδος

Φ
βι
...
N
σι
μι

©TealIn / nocturnalsoldier.org

κα
/α

Στη Θράκη τρέφεται κυρίως με χελώνες

Κατανομή και Ανάλυση

- Δείγμα 35 επικρατειών στην Οροσειρά της Ροδόπης
- Μέγεθος Κελιού 100X100 m

Μεταβλητές Περιβάλλοντος I

- Χάρτες Κάλυψης Γης:

Γκρουπ κατηγοριών του CORINE LAND COVER 2000

Δάση, Σκληροφυλλική Βλάστηση (μακκί),
Εντατικές Καλλιέργειες, Αγροδασικές Περιοχές,
Ανοίγματα, Μεταβατική Βλάστηση
(Λιβάδια→Δάσος), Δομημένο Τοπίο, Υγρότοποι

Μεταβλητές Περιβάλλοντος II

- Τοπογραφία

Υψόμετρο, Δείκτες Έντασης Ανάγλυφου, Έκθεση
Υψομετρική διαφορά, Πυκνότητα
Κορυφογραμμών

Μεταβλητές Περιβάλλοντος III

- Ανθρωπογενείς επιδράσεις, ψηφιοποιημένες από google earth, κτηματολόγιο

Πυκνότητες κύριων δρόμων, αποστάσεις από
κύριους δρόμους – οικισμούς - λατομεία

Μεταβλητές Περιβάλλοντος IV

- Δείκτες τοπίου βασισμένοι στο CLC, εξαγωγή δεικτών στο FRAGSTATS 4.0

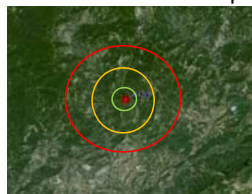
Contagion, Gyrate, Shape, Fractal, Shannon D.I.
κ.α.

3 διαφορετικές κλίμακες

- 1000 m (Επίπεδο Χώρου φωλεοποίησης)
- 3000 m (Επίπεδο κέντρου επικράτειας, σημαντικό για central place foragers όπως ο χρυσαετός)
- 6000 m επίπεδο χωροκράτειας

ASCII rasters

- Υπολογισμός τιμών (% κάλυψης, μ.ο., πυκνότητες, απόλυτες τιμές) σε moving window από το κέντρο του κελιού



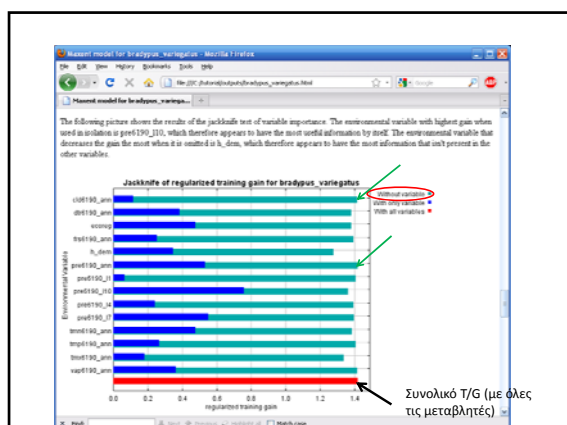
- 1000 m
- 3000 m
- 6000 m

2 στρατηγικές επιλογής μοντέλων

- Μία προσέγγιση πιο formal, βασισμένη σε διαστήματα εμπιστοσύνης του μέσου training gain (Jackknife method)
- Κρατώντας από κάθε κατηγορία μεταβλητών (κάλυψη, τοπογραφία, τοπίο, ανθρωπογενής) τη σημαντικότερη (Μεταβλητή ανά κατηγορία)

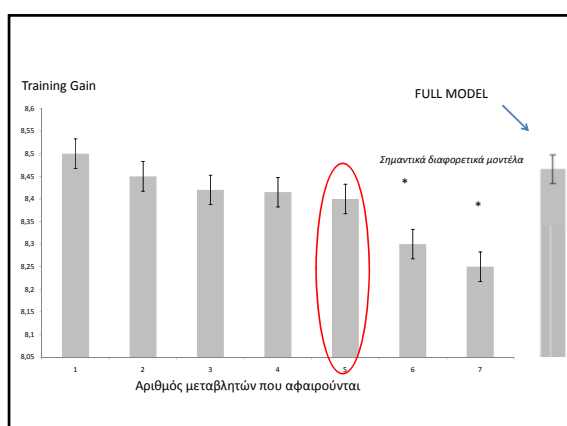
Μοντέλα Jackknife

- Ξεκινούν με το σύνολο των μεταβλητών (Πλήρες μοντέλο)
- Κάθε μοντέλο τρέχει σε 10 crossvalidation training datasets του 90 % των δειγμάτων.
- Η μεταβλητή με τη μικρότερη απώλεια του training gain (λιγότερο χρήσιμη με βάση τη λειτουργία jackknife) αφαιρείται διαδοχικά και το μοντέλο τρέχει ξανά ώσπου να μείνει μόνο μία μεταβλητή.



Μοντέλα Jackknife

Για κάθε μοντέλο υπολογίζεται το μέσο training gain και διαστήματα εμπιστοσύνης. Τελικά επιλέγεται με βάση της αρχής του Όκκαμ (parsimony) το μοντέλο με training gain που δεν διαφέρει σημαντικά από το πλήρες, με τις λιγότερες μεταβλητές

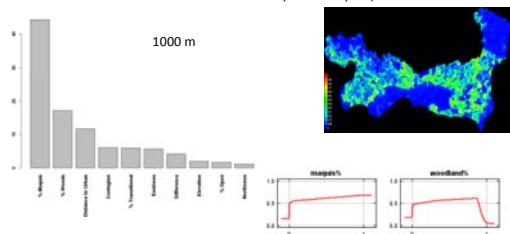


Μεταβλητή ανά Κατηγορία

- Σε κάθε κατηγορία κρατούμε τη μεταβλητή με το μεγαλύτερο average training gain
- Το τελικό μοντέλο συνδυάζει τις πιο σημαντικές μεταβλητές ανά κατηγορία

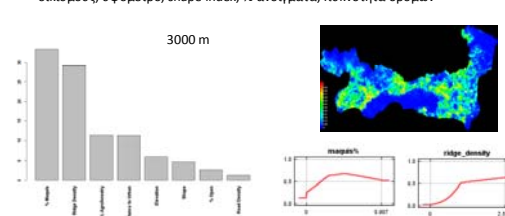
Αποτελέσματα - Jackknife μοντέλα

Contagion, % ανοίγματα, % μεταβατική βλάστηση, % μακκί, % έλκασ: έκθεση, επίκαιμιακή διαφορά, απόσταση από οικισμούς



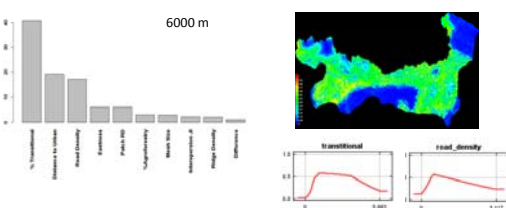
Αποτελέσματα - Jackknife μοντέλα

% μακκί, πυκνότητα κορυφογραμμών, αγροδασικές περιοχές, απόσταση από οικισμούς, υψόμετρο, share index, % ανοίγματα, πυκνότητα δρόμων



Αποτελέσματα - Jackknife μοντέλα

% μεταβατική βλάστηση, απόσταση από οικισμούς, πυκνότητα δρόμων, μέση Έκθεση (A-Δ), Πυκνότητα Patch Richness, % Αγροδασικά, Mesh Size, Interspersion – Juxtaposition Index, Πυκνότητα Κορυφογραμμών, μέση Υψομετρική Διαφορά



Ξεκινήστε εδώ

- <http://www.cs.princeton.edu/~schapire/maxent/>
- http://biodiversityinformatics.amnh.org/index.php?section_id=105&content_id=345

Επεκτάσεις

- Εξαγωγή αποτελεσμάτων και περαιτέρω αναλύσεις στο R (<http://www.r-project.org/>)
- Εφαρμογή ENMTools, με AIC / BIC συγκρίσεις μοντέλων, εφαρμογή διερεύνησης πολυσυσχετίσης κ.α. (<http://enmtools.blogspot.gr/>)

Phillips, S. J., Anderson, R. P. and Schapire, R. E. 2006: Maximum Entropy modelling of species geographic distributions. *Ecological Modelling*. 190: 231-259

Phillips, S. J. and Dudik, M. 2008: Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*. 31:161-175

Elith, J., Phillips, S. J., Hastie, T., Dudik, M., Chee, Y. E. and Yates, C., J. 2011: A statistical explanation of MaxEnt for Ecologists. *Diversity and Distributions*. 17: 43-57

Baldwin, R. A. 2009: Use of Maximum Entropy modelling in wildlife research. *Entropy* 11:854-866

Yost, A., C., Petersen, S. L., Gregg, M., and Miller R. 2008: Predictive Modelling and Mapping sage grouse (*Centrocercus urophasianus*) nesting habitat using Maximum Entropy and a long-term dataset from Southern Oregon. *Ecological Informatics*. 3: 375-386