Potential Habitat Modelling for Snow Leopard using MaxEnt with emphasis on model analysis and validation

by

Anto R. Dominic

Research Colloquium (Apr 2010),

Faculty of Forest and Environment

HNE, Eberswalde

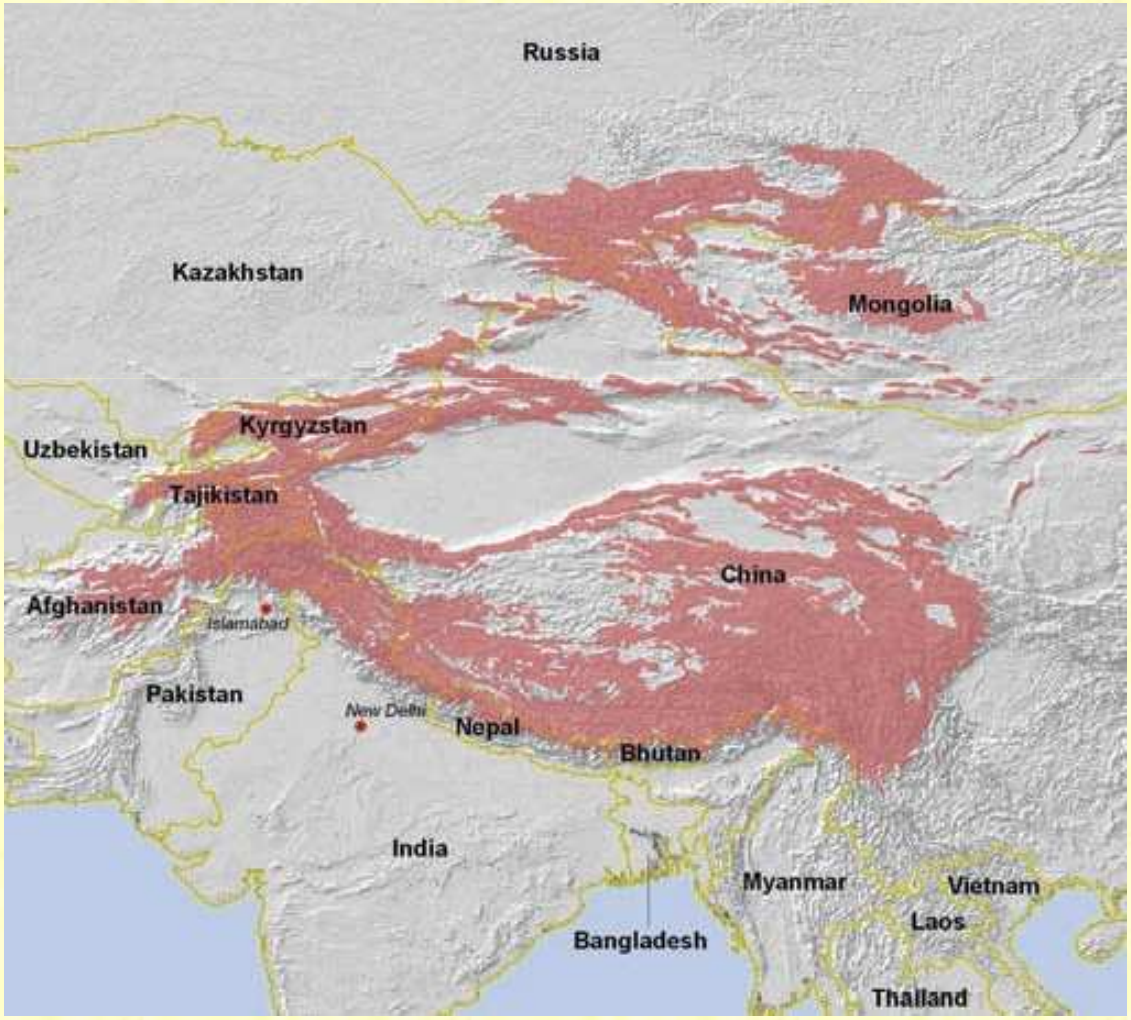
Fachhochschule Eberswalde

Contents

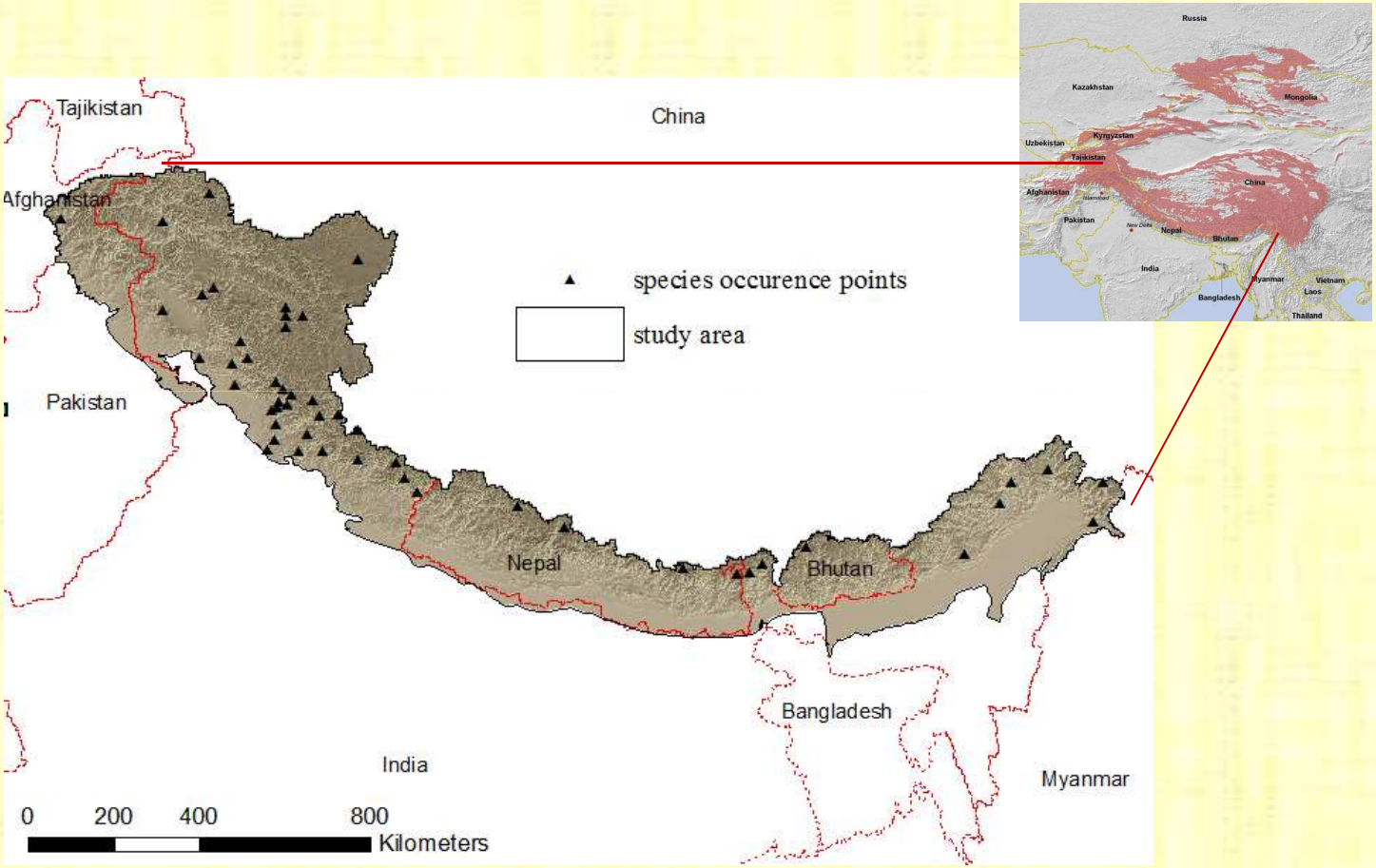
Introduction
the snow leopard (*Uncia uncia*)

- Top predator of the higher Central Asian mountain ranges
- Habitat spread over an area of 3 million sq.kms
- the population is scattered in a few islands spread over this area
- Inhabit steep, rugged areas and are well adapted to the extreme cold conditions.
- Diet: mostly native ungulates and domestic livestock
- Population : 4000 – 7500 approx.
- Status: endangered (IUCN, Red Book)

Introduction - study area



Introduction - study area



Introduction - MaxEnt

- Based on the maximum entropy principle (Jaynes 1957)
- The algorithm aims to extract maximum information possible from the available data while making least assumptions of what is not given
- Been used successfully in different fields especially in natural language processing
- MaxEnt - a sequential update algorithm iteratively gathering information (gain)
- Useful for presence only species habitat modelling
- Can be used to project habitat suitability spatially and temporally
- Is many times faster than most available habitat suitability models available
- Model evaluation uses pseudo absence points

Methodology

- Topographic features (elevation, slope, aspect, land cover, roads, rivers and water bodies), meteorological features (19 bioclimatic variables), anthropogenic features (population density)
- And snow leopard observation locations
- Run model with all features – select the relevant ones
- Run model in cross validation mode
- Assess the effect of the features and validate the model

Model analysis and validation techniques

- Model analysis
  - Two sets of response curves (effect of the feature on habitat suitability)
  - Jackknife tests (effect of each feature on the gain of the model)
- Model validation
  - ROC plots (using pseudo absences)
  - Boyce index (k-fold cross validation and Spearman rank correlation)

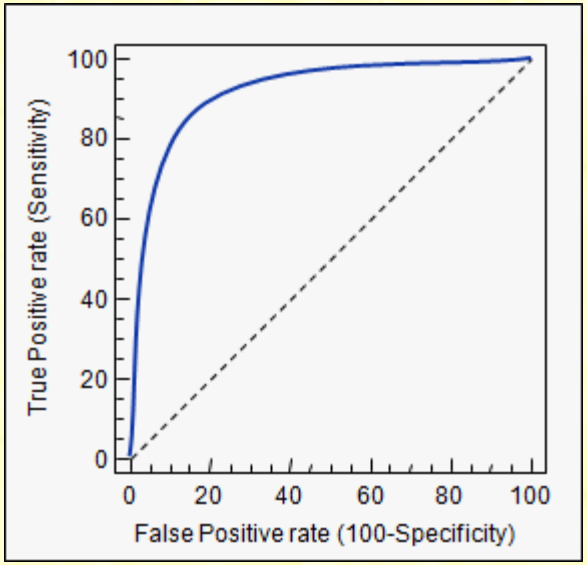
ROC plot

Model's ability to differentiate presence from absence is measured

Sensitivity = *a/(a+c)* or *TPR*

Specificity = *d/(b+d)* or *100-FPR*

	actual		
test	+ve	-ve	
+ve	*TP (a)*	*FP (b)*	*a+b*
-ve	*FN (c)*	*TN (d)*	*c+d*
	a+c	*b+d*	*N*



http://www.medcalc.be/manual/roc.php

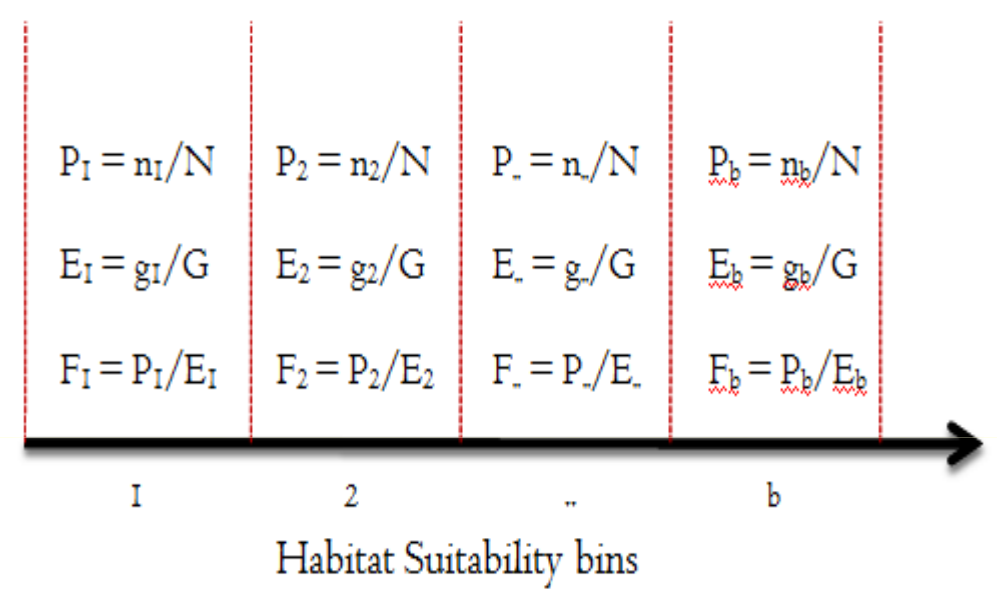
Perfect prediction: AUC = 1

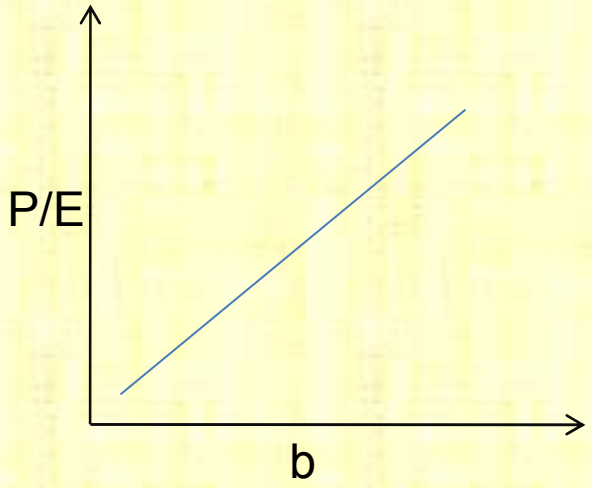
Random prediction: AUC = 0.5

Boyce index

- Has been so far used only to validate Resource Selection Function (RSF) habitat models.
- Model's ability to predict different levels of suitability is measured (predicted to expected ratio)
- Perform Spearman rank correlation between the binned suitability ranks and the predicted to expected ratio

Boyce index





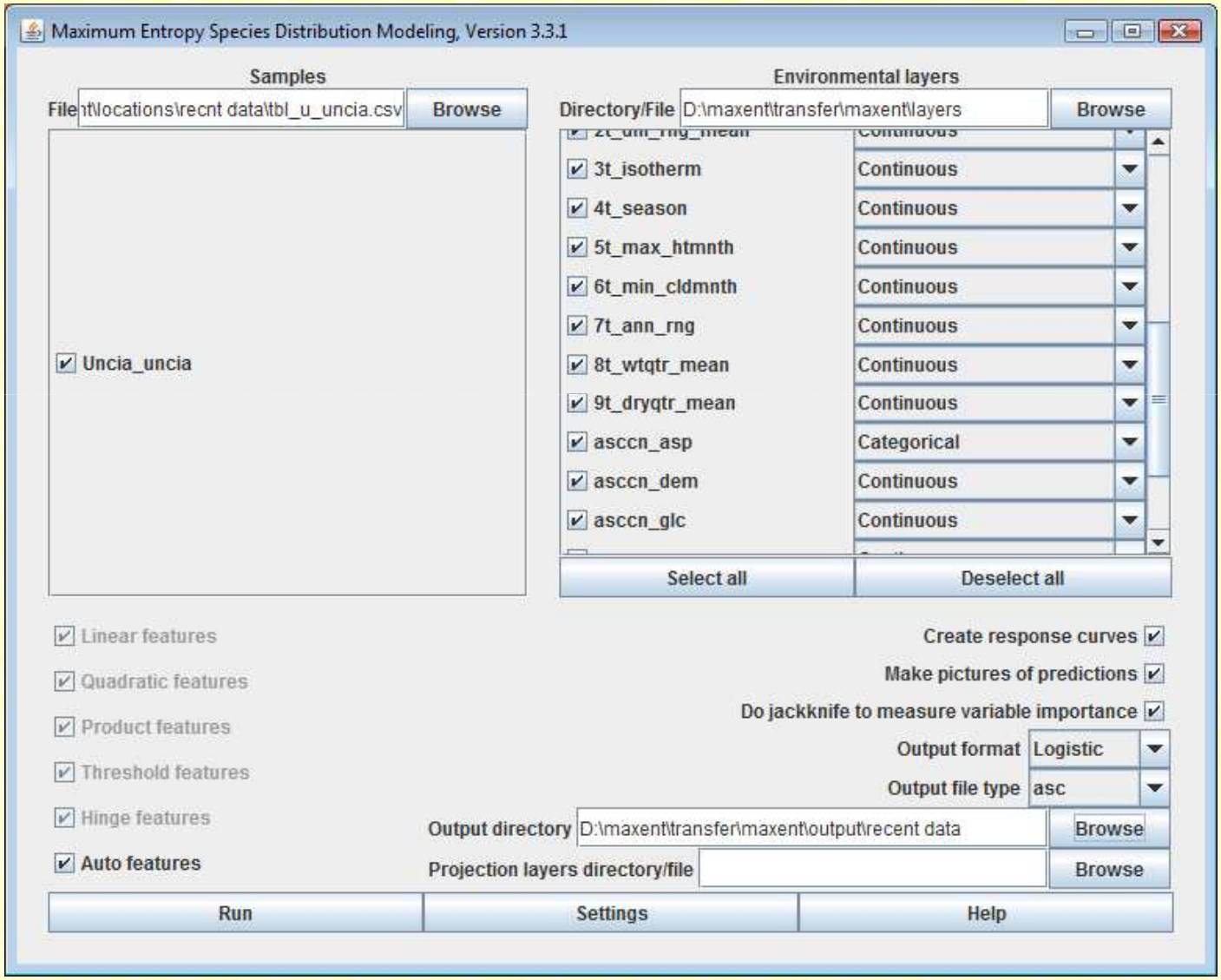
N – total number of sample points
n – number of sample points in each bin
F – freq of points to grid cells for each bin
P – predicted frequency

G – total number of grid cells
g – number of grid cells in each bin
b – number of bins
E – expected frequency

Data preparation

- From DEM, slope and aspect were derived
- Distances to roads, rivers and water bodies
- These and the meteorological and population rasters were clipped and resampled to 5x5 km grids sharing the same origin
- All rasters converted to ASCII grids (with identical headers) for running in MaxEnt
- The sample points were extracted with only latitude and longitude as fields and converted to *.csv format

Modelling – MaxEnt interface



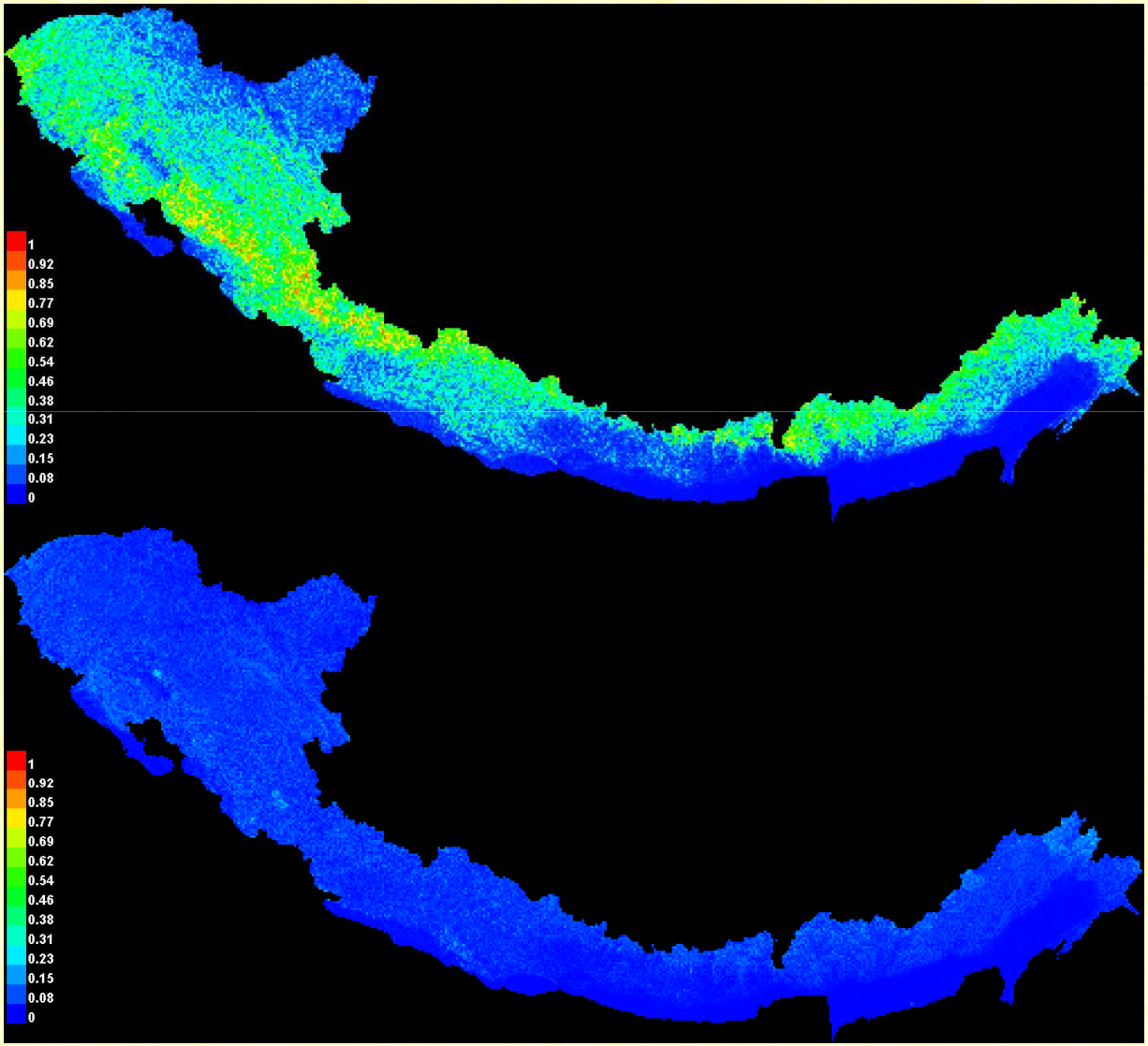
Modelling – batch run

selected features: >5% contribution + annual mean temp and land cover types

predictor variables	percentage contributions in 10 model runs										mean	SD
aspect	18.1	17.1	17.3	17	16.4	17.6	18	17.5	17.4	17.4	17.38	0.49
elevation	15.8	12.2	14.4	15.7	16.1	17.6	15.7	17.7	16.4	17.1	15.87	1.63
19p_coldqtr	12.8	12	17.4	13	15.6	14.8	17.3	13.6	14.9	13.4	14.48	1.86
4t_seasonlity	10	11.7	10.2	13.4	12.5	11.9	10.1	13.3	11.1	11.8	11.6	1.25
slope	7.9	10.7	9.8	8.6	5.5	9.9	5.9	6.8	9.1	8.6	8.28	1.74
pop	6.9	8.7	7.5	6.9	6.1	4.9	7.9	6.9	8	6.8	7.06	1.06
3t_isotherm	5.5	4.3	4.8	3.2	3.9	4.9	4.4	4.6	4.6	3.6	4.38	0.67
land_cover	4.5	3.9	3.9	4.9	4.3	3.2	5.1	4.7	4	4	4.25	0.57
8t_mean_wetqtr	5.1	0.8	4.6	4.5	5.3	0.6	5.4	5.5	3.9	1	3.67	2.04
roads_dist	3.8	3.5	3.4	3.2	3.5	3.2	3.1	3.8	3.7	3.7	3.49	0.26

Modelling

5-fold cross validation runs with 20% set aside for testing in each run



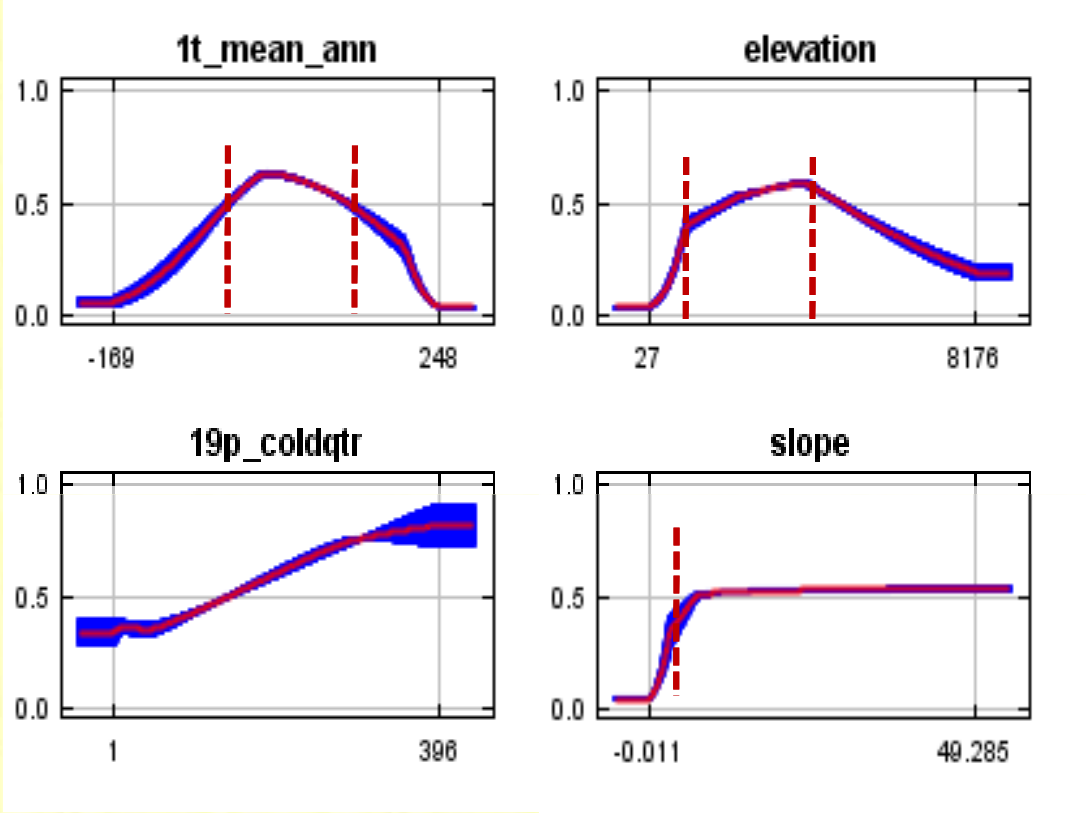
Logistic suitability

range: 0 - 1

Mean habitat suitability

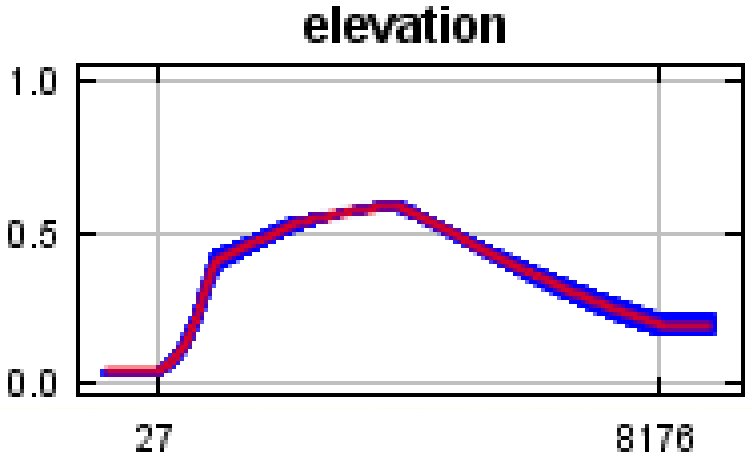
Standard deviation of the habitat suitability

Model analysis – response curves



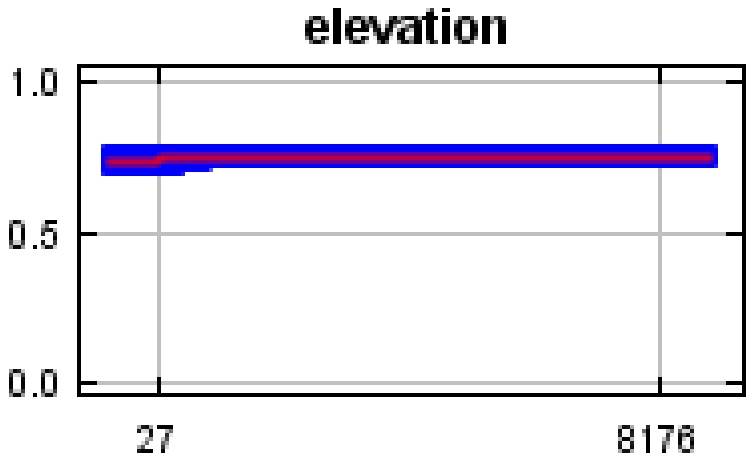
With Mean annual temperature, elevation and slope, the changes in suitability values are in accordance with what we know of snow leopard habiatt preferences.
Though precipitation (here that of the cold quarter) does affect the cat itself, it could reflect prey habitat preferences.

Model analysis – response curves

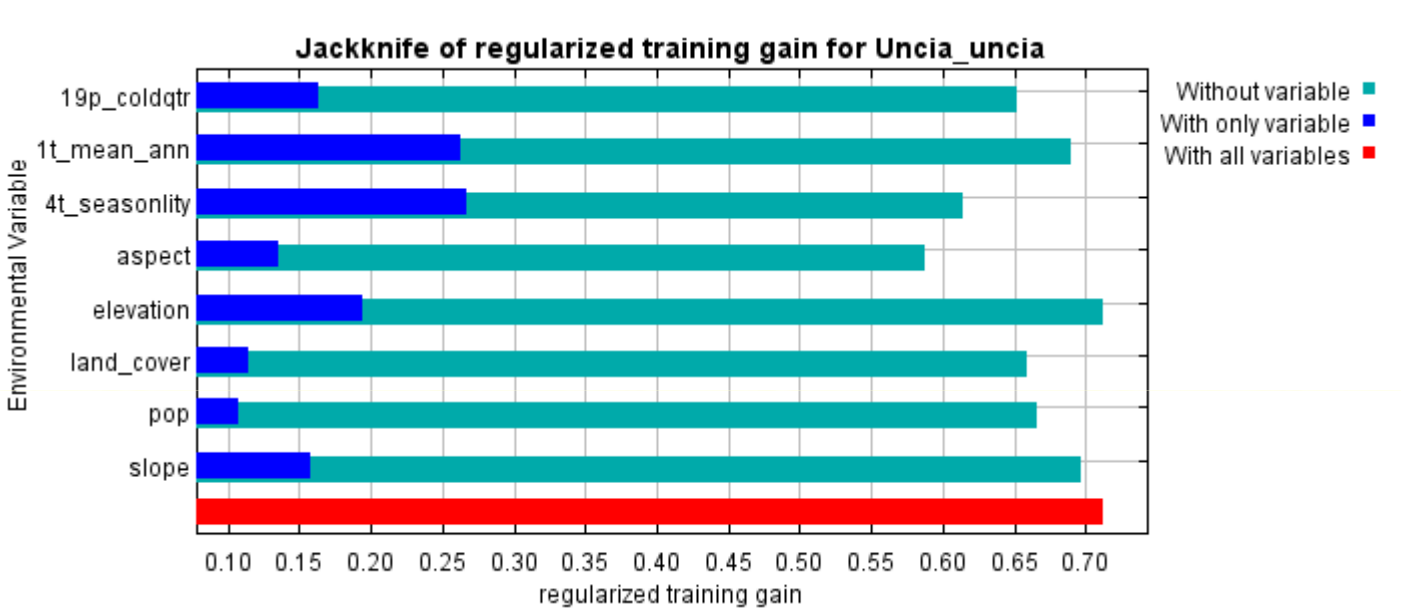


Logistic suitability with only elevation, reflects fairly well the preferred elevation for snow leopard habitat (~1000m – 4000m)

Logistic suitability for elevation when all the others are held constant, shows an even high suitability (0.7) , due to correlations with slope, population and may be also temperature.



Model analysis – jackknife tests of features



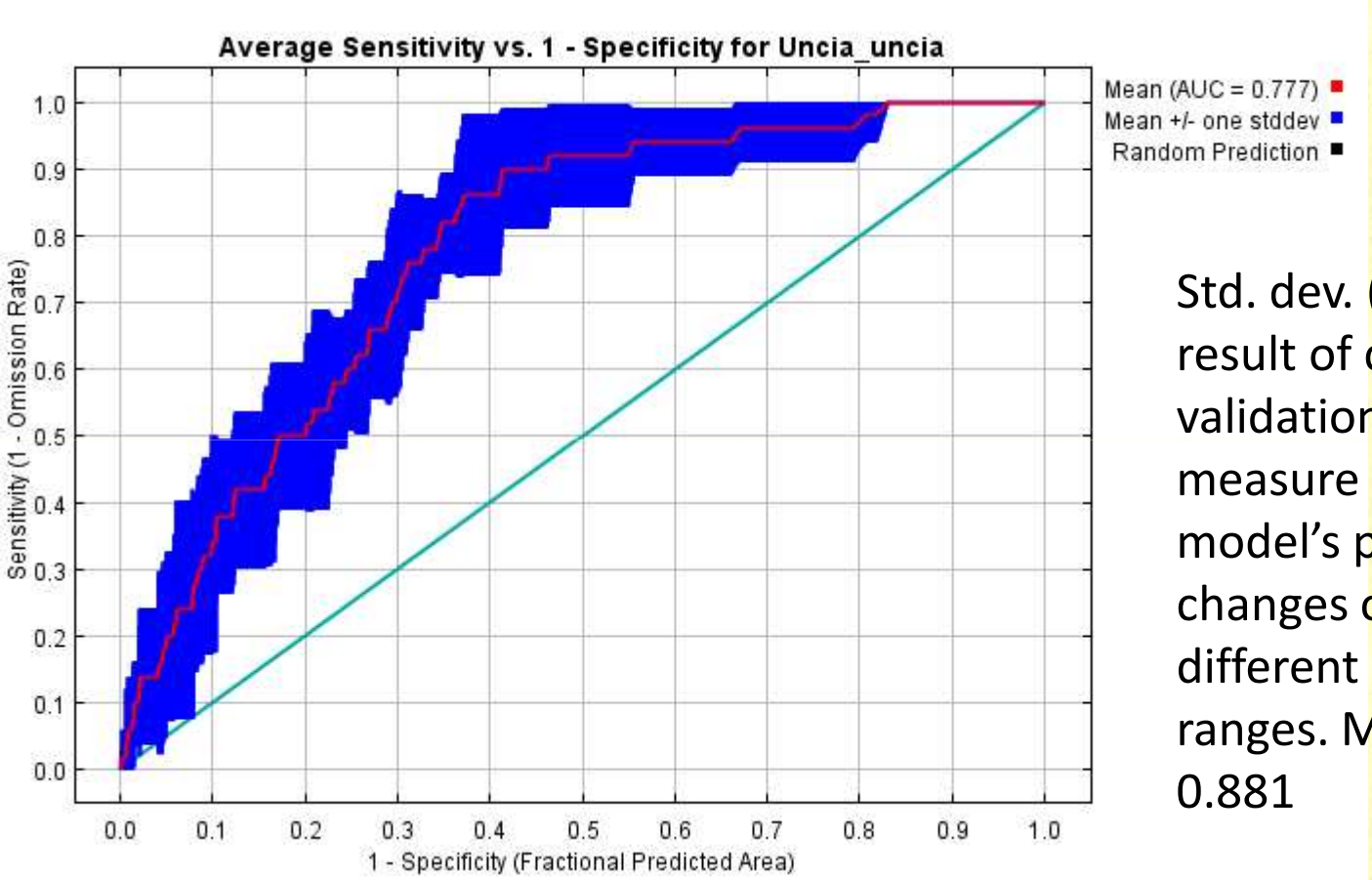
Red bar: total gain using all the features
Light blue bars: model gain without the corresponding feature
Dark blue bars: model gain with only the corresponding feature

This graph is useful to weigh the amount of information contained in the variables and decide on whether a particular variable is important for model building or not.

Model validation – ROC plot

- ROC plot built using random background points (pseudo-absences) instead of absences.
- Specificity is defined using predicted area rather than absence data.
- Default no. of random points: 10000, of which some fall in the suitable habitat (as predicted by MaxEnt)
- If the species prevalence is 50%, max AUC is (1-prevalence/2) or 0.75
- Prevalence is what MaxEnt is trying to estimate and so cannot be known with certainty.
- ROC plots using pseudo absences result in slightly over fitted models and less predictive power.

Model validation – ROC plot



Std. dev. (blue) as a result of cross-validation is a measure of how the model's predictability changes over the different suitability ranges. Max AUC: 0.881

Model validation – Boyce index

- 6 bins created from habitat suitability range
- The resulting maps imported to ArcGIS
  - Predicted frequency calculated by intersecting sample points used with the distribution map
  - Expected frequency calculated using Raster Calculator

classes	P_1/E_1	P_2/E_2	P_3/E_3	P_4/E_4	P_5/E_5	mean	SD
1	0.053	0.05	0.058	0.057	0.059	**0.056**	**0.004**
2	0.625	0.619	0.406	0.499	0.476	**0.525**	**0.095**
3	1.283	0.867	1.146	1.21	1.089	**1.119**	**0.158**
4	2.948	4.091	3.336	2.788	3.8	**3.393**	**0.552**
5	8.391	9.232	7.806	9.192	6.929	**8.31**	**0.974**
6	12.676	15.701	6.978	6.224	12.227	**10.761**	**4.035**
rho	**1**	**1**	**0.943**	**0.943**	**1**	**1**	
p	**0.003**	**0.003**	**0.017**	**0.017**	**0.003**	**0.003**	

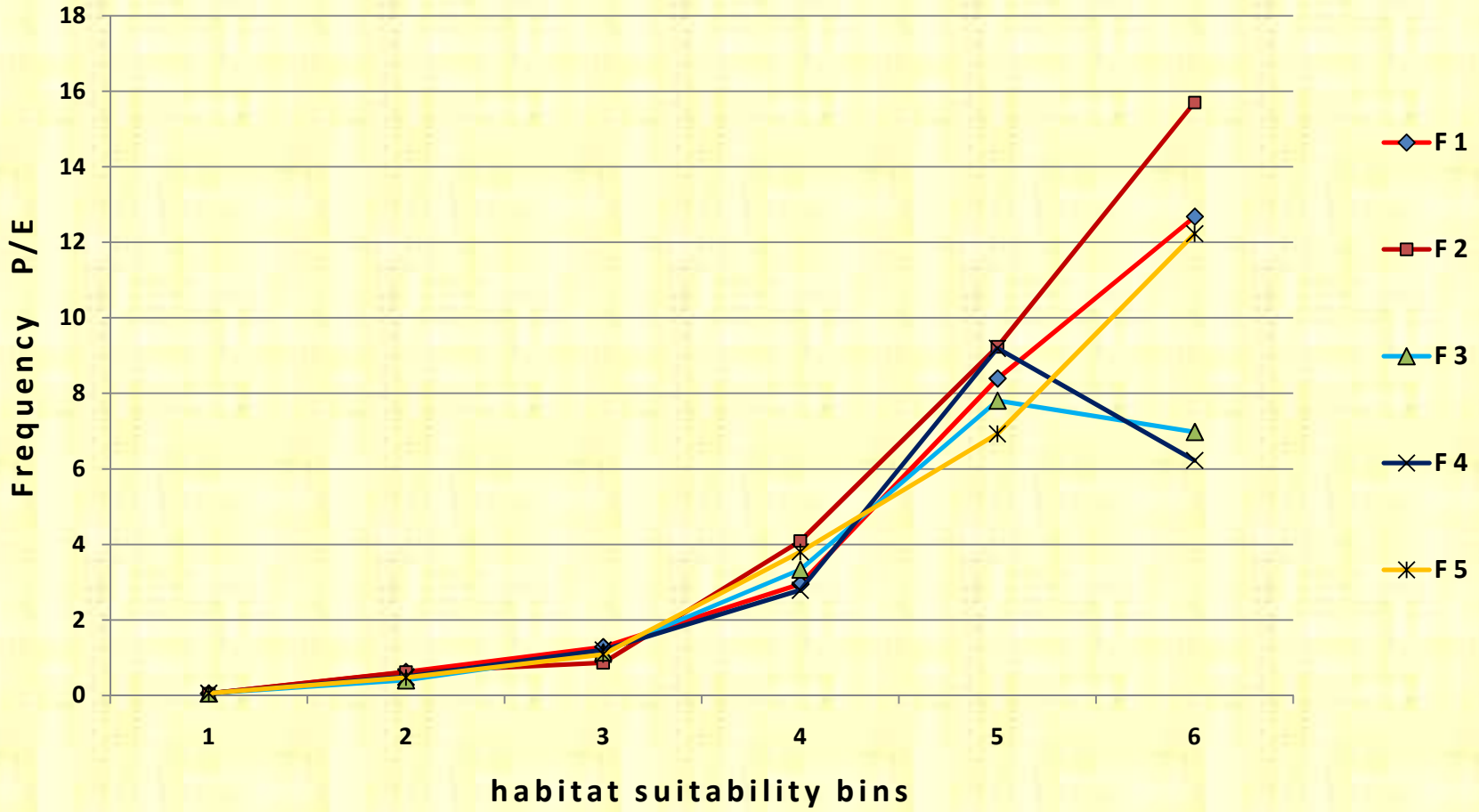
Spearman's rho: $1 - \frac{6 \sum d^2}{n(n^2 - 1)}$

n – number of bins

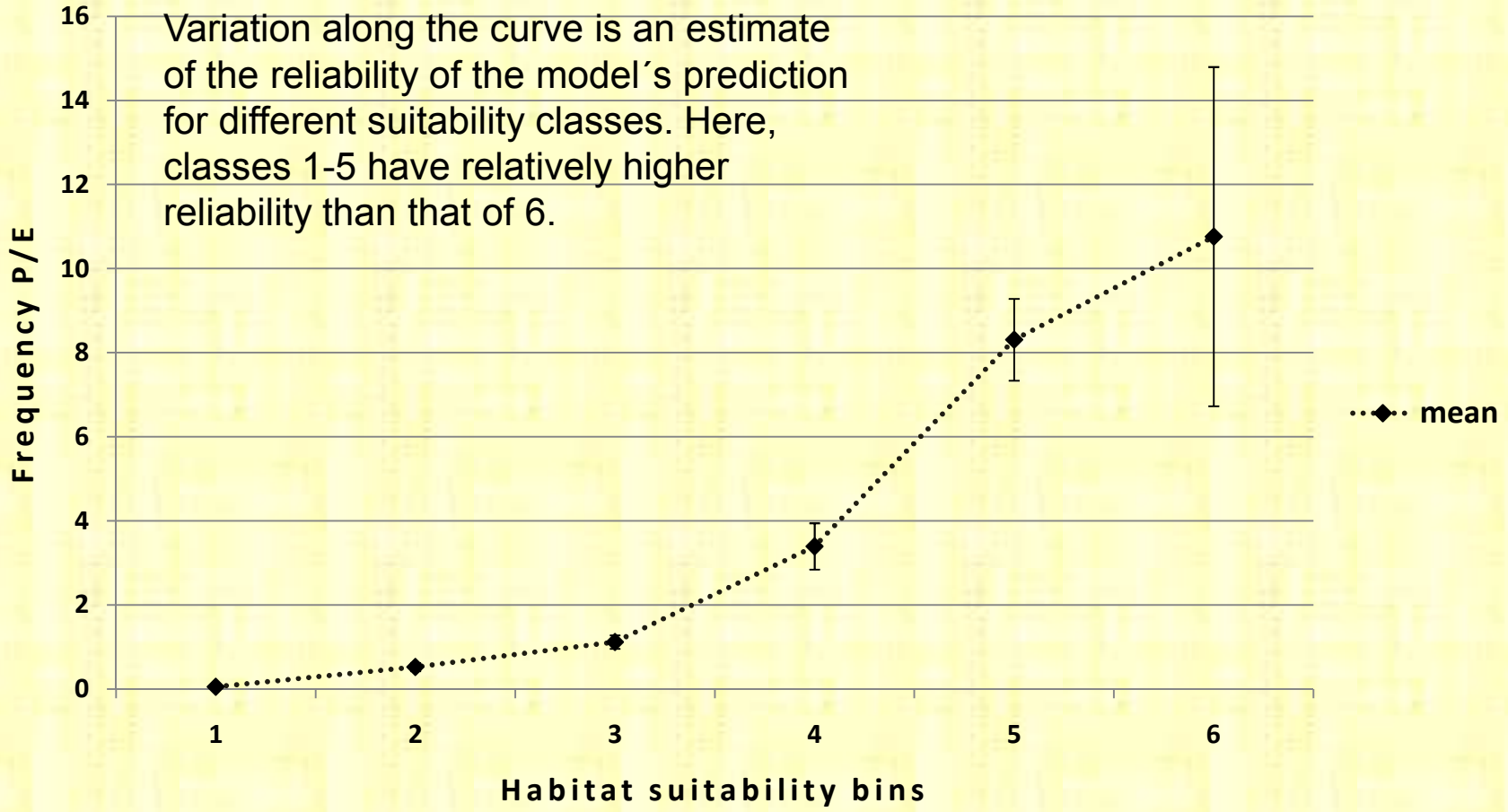
d – class-wise difference between ranks

Mean significance is < 0.01 and <0.05 in all cases

Model validation – Boyce index



Model validation – Boyce index



Conclusion

- While MaxEnt is being applied more widely, validation methods are under continuous development
- Boyce index provides a suitable alternative way of model validation.
- Habitat models help to delineate areas of suitable habitat for management plans
- Realised niche can be estimated from potential habitat by removing areas of unsuitable habitat
- Identify areas that require more focused study and survey

Acknowledgements

- Thanks to
  - KORA and staff, Switzerland
  - Dr. F. Zimmerman, my supervisor
  - MaxEnt Google Groups http://groups.google.com/group/Maxent
- Data sources
  - Global Mammal Assessment (GMA) databank, CatSG, Switzerland
  - Gridded population of the world from http://sedac.ciesin.columbia.edu/gpw
  - Bioclimatic variables from http://www.worldclim.org/download
  - DEM data from http://srtm.csi.cgiar.org/
  - Global land cover data from http://ionia1.esrin.esa.int/index.asp
  - Roads, rivers and water bodies from www.diva-gis.org

Acknowledgements (contd.)

- Software used
  - ArcInfo 9.3.1, ESRI, USA
  - R 2.9.0, R Development Core Team, Austria
  - MS Office, Microsoft, USA
  - MaxEnt 3.3.1, AT&T Labs and American Museum of Natural History, USA
- References (works of ...)
  - Phillips, S. J., Anderson, R. P., Schapire, R. E, and Dudík (for MaxEnt)
  - Boyce, M. S, Guisan, A, Hirzel, A. H (for ROC plots and Boyce index)
  - Jackson, R. (for snow leopard)
  - For a complete reference list contact author