ASSESSING MODEL ACCURACY

DAVID ORME

OVERVIEW

- The confusion matrix
- Measures of model accuracy
- Thresholds for continuous predictions
- Application to Species Distribution Models

MODIS LAND COVER CLASSIFICATION

Site Class	Class Name	Classification Outcome															
		1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	Total
1	Evergreen Needleleaf	1460	42	18	11	266	7	9	17	23	10	15	21	2	0	0	1901
2	Evergreen Broadleaf	31	4889	0	14	14	11	18	79	23	17	4	38	10	0	1	5149
3	Deciduous Needleleaf	87	0	104	25	118	0	0	4	0	0	0	10	0	0	0	348
4	Deciduous Broadleaf	22	56	16	384	278	0	3	11	1	3	0	47	82	0	0	903
5	Mixed Forest	405	63	94	148	1355	3	1	27	7	8	40	41	17	0	0	2209
6	Closed Shrubland	34	35	2	12	5	140	124	29	15	30	2	158	19	0	8	613
7	Open Shrubland	10	12	3	9	1	41	1002	33	45	203	0	210	6	0	213	1788
8	Woody Savanna	62	133	0	16	110	11	104	577	141	71	0	221	22	0	3	1471
9	Savanna	10	53	1	0	21	18	48	93	440	43	1	252	79	0	16	1075
10	Grasslands	2	16	0	2	20	4	179	6	101	632	0	249	13	0	363	1587
11	Pmnt Wtlnd	63	24	0	5	28	23	1	2	36	2	89	1	7	0	0	281
12	Cropland	6	75	2	7	16	8	61	42	132	133	2	5168	183	0	18	5853
14	Cropland/Natural Vegn	2	133	0	48	28	2	8	16	66	8	1	320	832	0	7	1471
15	Snow+ice	1	0	0	0	0	1	2	0	0	0	5	1	0	1297	5	1312
16	Barren	0	2	1	0	0	1	162	4	5	126	3	56	5	14	3537	3916
	Total	2195	5533	241	681	2260	270	1722	940	1035	1286	162	6793	1277	1311	4171	29877

Accuracy = 21906 / 29877 = 73.3%

A SIMPLER CONFUSION MATRIX

Zoom in on just two of those categories:

Site Class	Class Name	1	2
1	Evergreen Needleleaf	1460	42
2	Evergreen Broadleaf	31	4889

Model predicts: Is this evergreen forest needleleaf or broadleaf

ACCURACY

Easy to calculate **accuracy**:

	Pred. Needle	Pred. Broad	Sum
Obs. Needle	1460	42	1502
Obs. Broad	31	4889	4920
Sum	1491	4931	6422

$$A = \frac{1460 + 4889}{1460 + 4889 + 42 +31} = 98.9\%$$

ACCURACY

But **random** models have ~50% accuracy!

	Pred. Needle	Pred. Broad	Sum
Obs. Needle	737	765	1502
Obs. Broad	2520	2400	4920
Sum	3257	3165	6422

$$A = \frac{737 + 2400}{6422} = 48.8\%$$

ACCURACY

Bad models: **everything is a broadleaf**

	Pred. Needle	Pred. Broad	Sum
Obs. Needle	0	1502	1502
Obs. Broad	0	4920	4920
Sum	0	6422	6422

$$A = \frac{0 + 4920}{6422} = 76.6\%$$

PREVALENCE

Proportion of the observed positive outcomes

	Pred. Pos	Pred. Neg	Sum
Obs. Pos	1460	42	1502
Obs. Neg	31	4889	4920
Sum	1491	4931	6422

$$\mbox{Prevalence} = \frac{1502 }{6422} = 0.234$$

ACCURACY

And **accuracy is affected by prevalence**

	Pred. Pos	**Pred. Neg**	**Sum**
Obs. Pos	0	35	35
Obs. Neg	0	6407	6407
Sum	0	6442	6442

$$A = \frac{0 + 6407}{6422} = 99.5\%$$

PREDICTION OUTCOMES

Giving some simple names to the four outcomes:

	Pred. Pos	**Pred. Neg**
Obs. Pos	True Positive	False Negative
Obs. Neg	False Positive	True Negative

PREDICTION OUTCOMES

Other more confusing names do get used:

	Pred. Pos	**Pred. Neg**
Obs. Pos	True Positive	Type II Error
Obs. Neg	Type I Error	True Negative

RATES OF OUTCOMES

Divide the four outcomes by the **observed** positive and negative counts to give **rates**:

	Pred. Pos	**Pred. Neg**
Obs. Pos	True Positive Rate	False Negative Rate
Obs. Neg	False Positive Rate	True Negative Rate

RATES OF OUTCOMES

Calculate those values:

	Pred. Pos	Pred. Neg	Sum
Obs. Pos	$\frac{1460}{1502}=97.2\%$	$\frac{42}{1502}=2.8\%$	1502
Obs. Neg	$\frac{31}{4920}=0.6\%$	$\frac{4889}{4920}=99.4\%$	4920

SENSITIVITY AND SPECIFICITY

Sensitivity

- Another name for the True Positive Rate
- The proportion of correctly predicted positive observations

Specificity

- Another name for the True Negative Rate
- The proportion of correctly predicted negative observations

SENSITIVITY AND SPECIFICITY

	Pred. Pos	Pred. Neg	Sum
Obs. Pos	1460	42	1502
Obs. Neg	2010	2910	4920
Sum	3470	2952	6422

	Pred. Pos	Pred. Neg
Obs. Pos	97.2%	2.8%
Obs. Neg	40.9%	59.1%

COHEN'S KAPPA

Cohen's kappa (κ) is a measure of agreement that rescales accuracy (A) to account for chance agreement (P_e):

$$\kappa = \frac{A - P_e}{1- P_e}$$

It can take values from $-\infty$ to 1, where 1 is perfect prediction and anything below zero is worse than chance.

COHEN'S KAPPA

Multiply proportions of observed and predicted to get probability of each outcome

	Pred. Pos	**Pred. Neg**	**Sum**
Obs. Pos	**1460**	42	1502
Obs. Neg	31	4889	4920
Sum	1491	4931	6422

$$P_{YY} = \frac{1491}{6422} \times \frac{1502}{6422} = 0.054$$

COHEN'S KAPPA

	Pred. Pos	Pred. Neg	p
Obs. Pos	0.054	0.180	0.234
Obs. Neg	0.178	0.588	0.766
p	0.232	0.768	1.000

$$ P_e = P_{YY} + P_{NN} = 0.054 + 0.588 = 0.642 $$

$$ \kappa = \frac{0.989 - 0.642}{1- 0.642} = 0.969 $$

TRUE SKILL STATISTIC

Journal of Applied Ecology
BRITISH ECOLOGICAL SOCIETY

Free Access

Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS)

OMRI ALLOUCHE, ASAF TSOAR, RONEN KADMON

First published: 12 September 2006 | https://doi.org/10.1111/j.1365-2664.2006.01214.x

TRUE SKILL STATISTIC

An alternative measure is TSS:

$$ \mbox{TSS} = \mbox{Sensitivity} + \mbox{Specificity} - 1 $$

$$ \mbox{TSS} = [0, 1] + [0, 1] - 1 $$

- TSS = 1 (perfect)
- TSS = 0 (random)
- TSS = -1 (always wrong)
- Unaffected by prevalence.

WAIT, NO. NOT TSS!

Nature Conservation

Home Articles About

About Pensoft Books Journ

Research Article

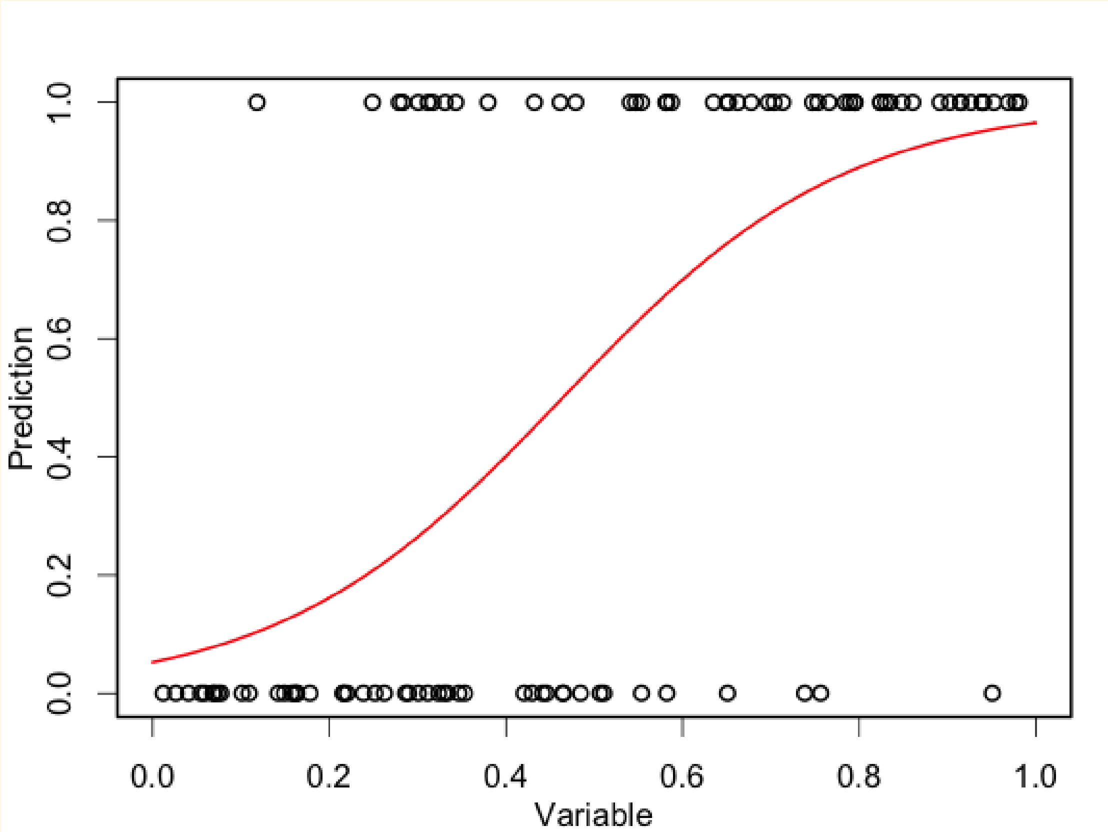
Nature Conservation 35: 97-116
https://doi.org/10.3897/natureconservation.35.33918 (20 Jun 2019)

Two alternative evaluation metrics to replace the true skill statistic in the assessment of species distribution models

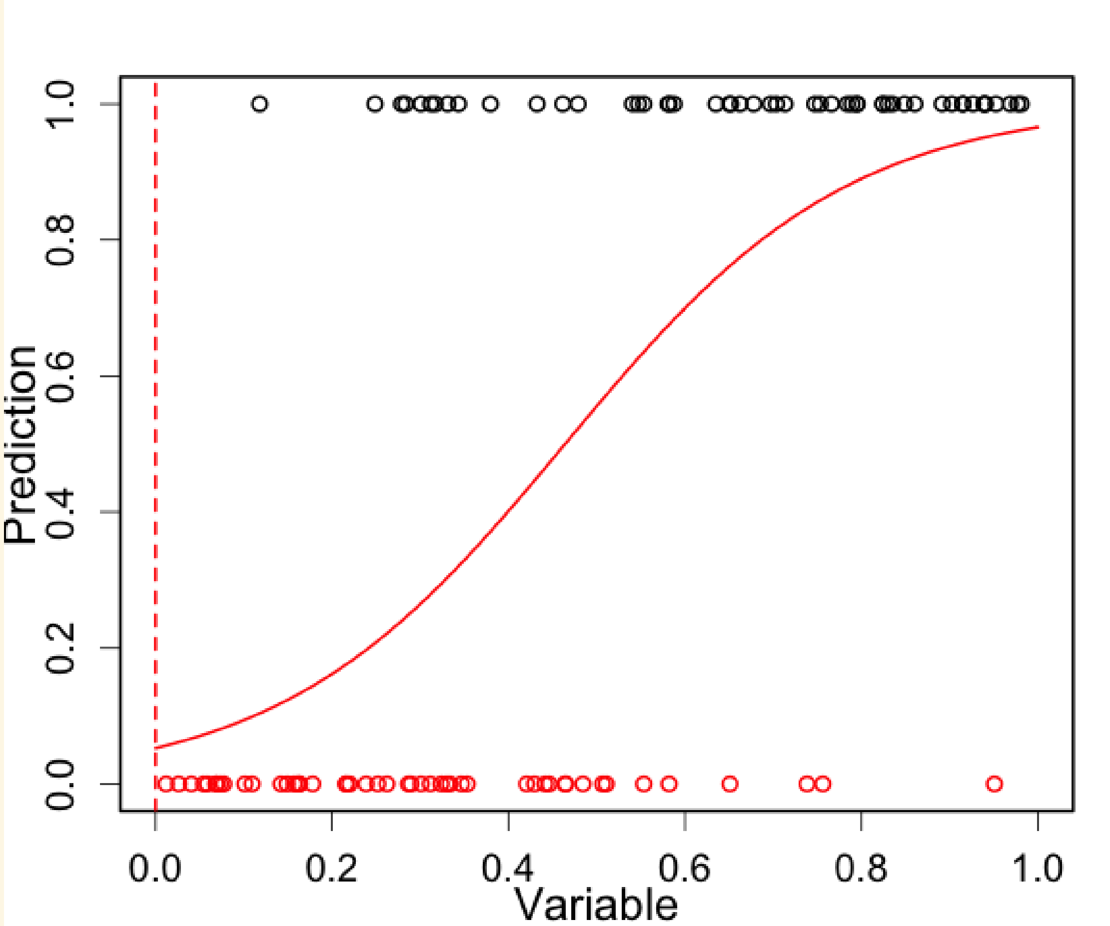
Rainer Ferdinand Wunderlich, Yu-Pin Lin, Johnathen Anthony, Joy R. Petway

PROBABILISTIC CLASSIFICATION

A model predicting the probability of success / presence



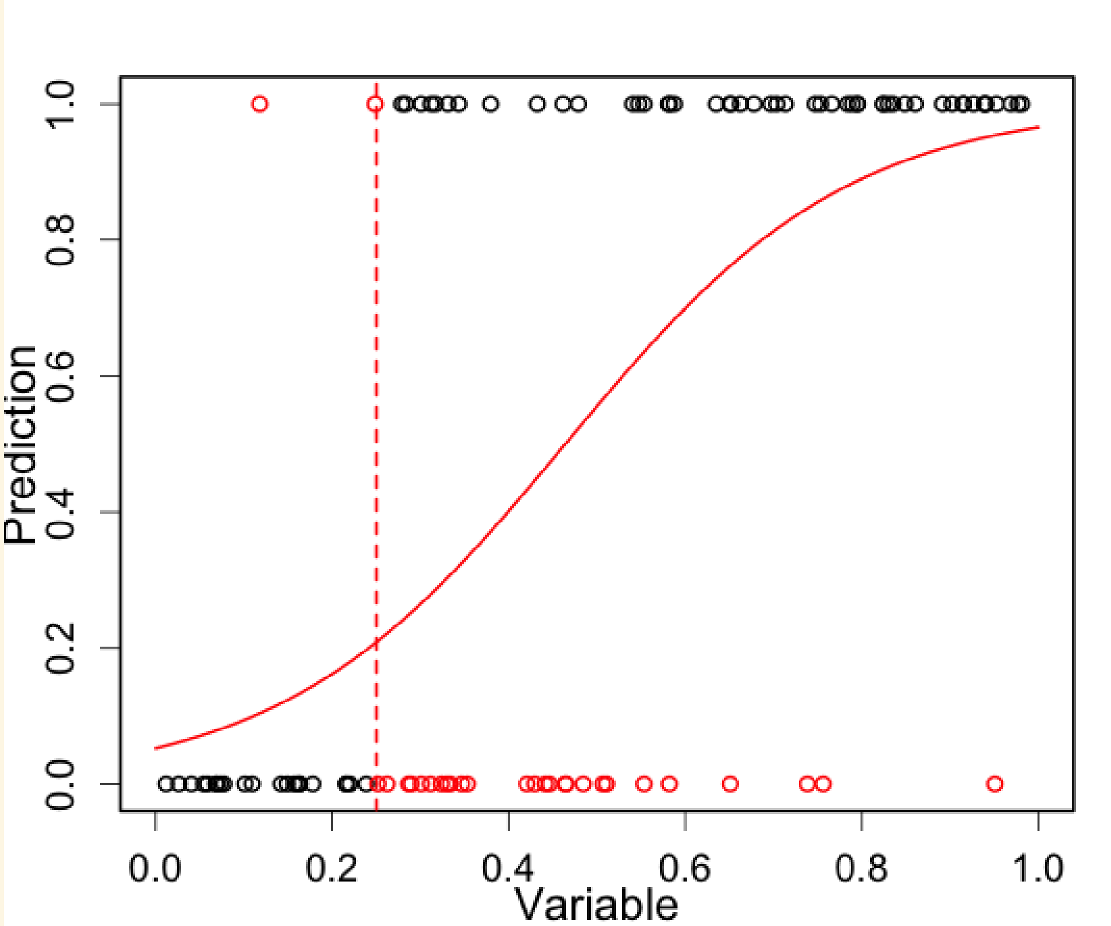
THRESHOLD MODEL



	0	1
1	0	52
0	0	48

	value
Sens	1
Spec	0
TSS	0

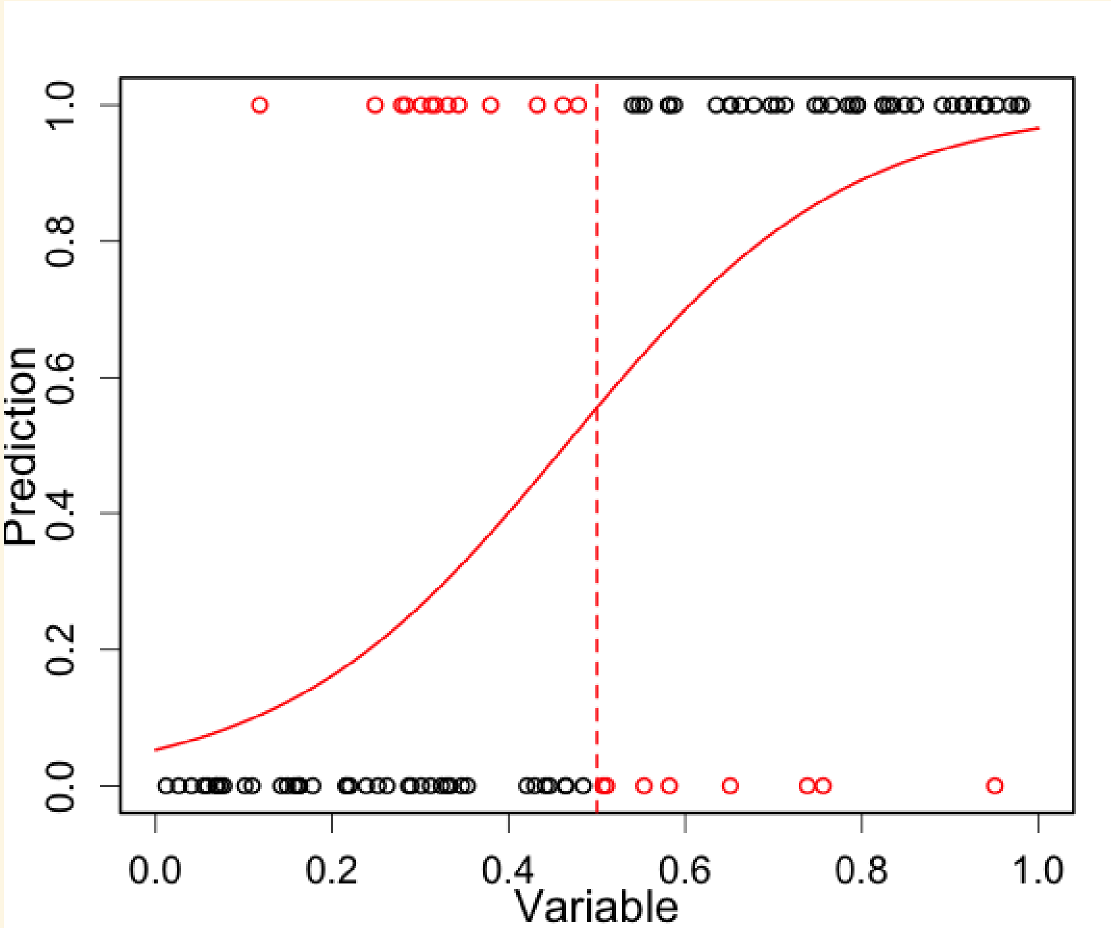
THRESHOLD MODEL



	0	1
1	2	50
0	22	26

	value
Sens	0.962
Spec	0.458
TSS	0.420

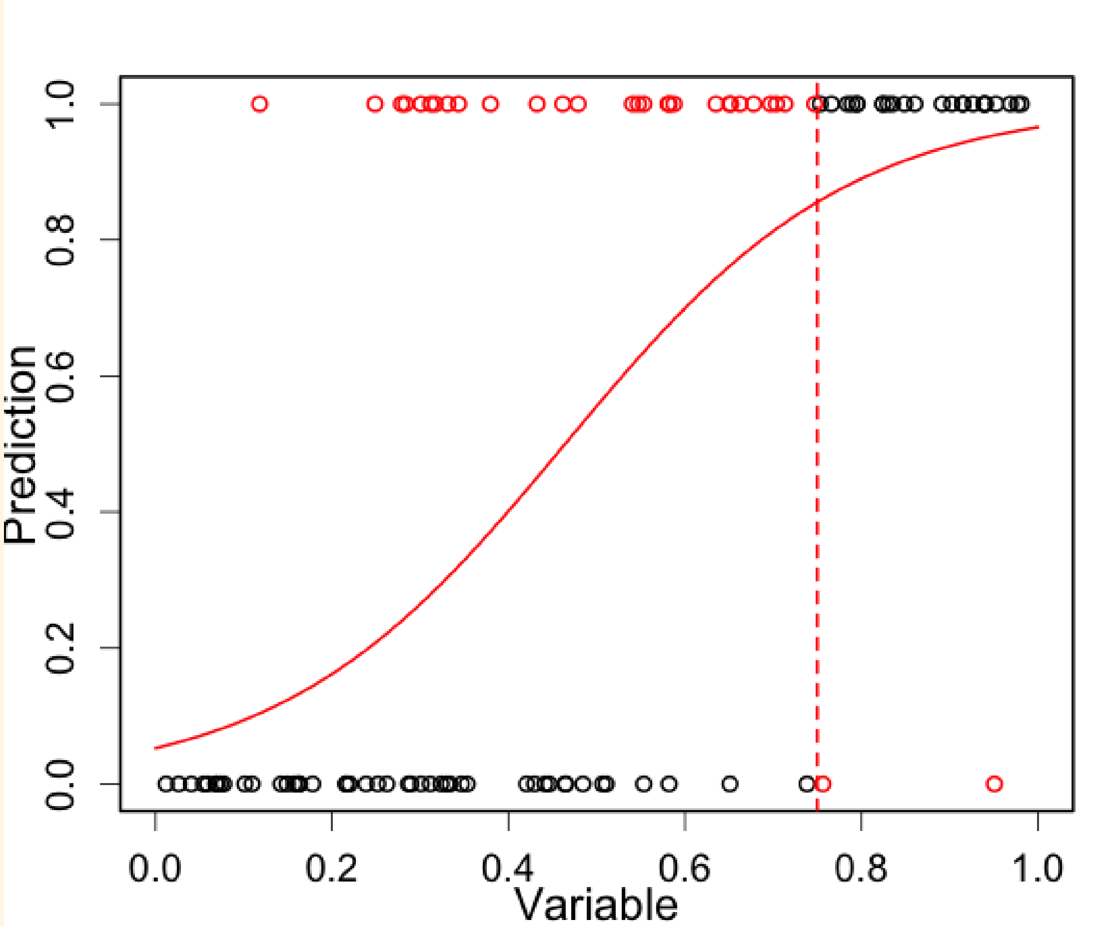
THRESHOLD MODEL



	0	1
1	13	39
0	40	8

	value
Sens	0.750
Spec	0.833
TSS	0.583

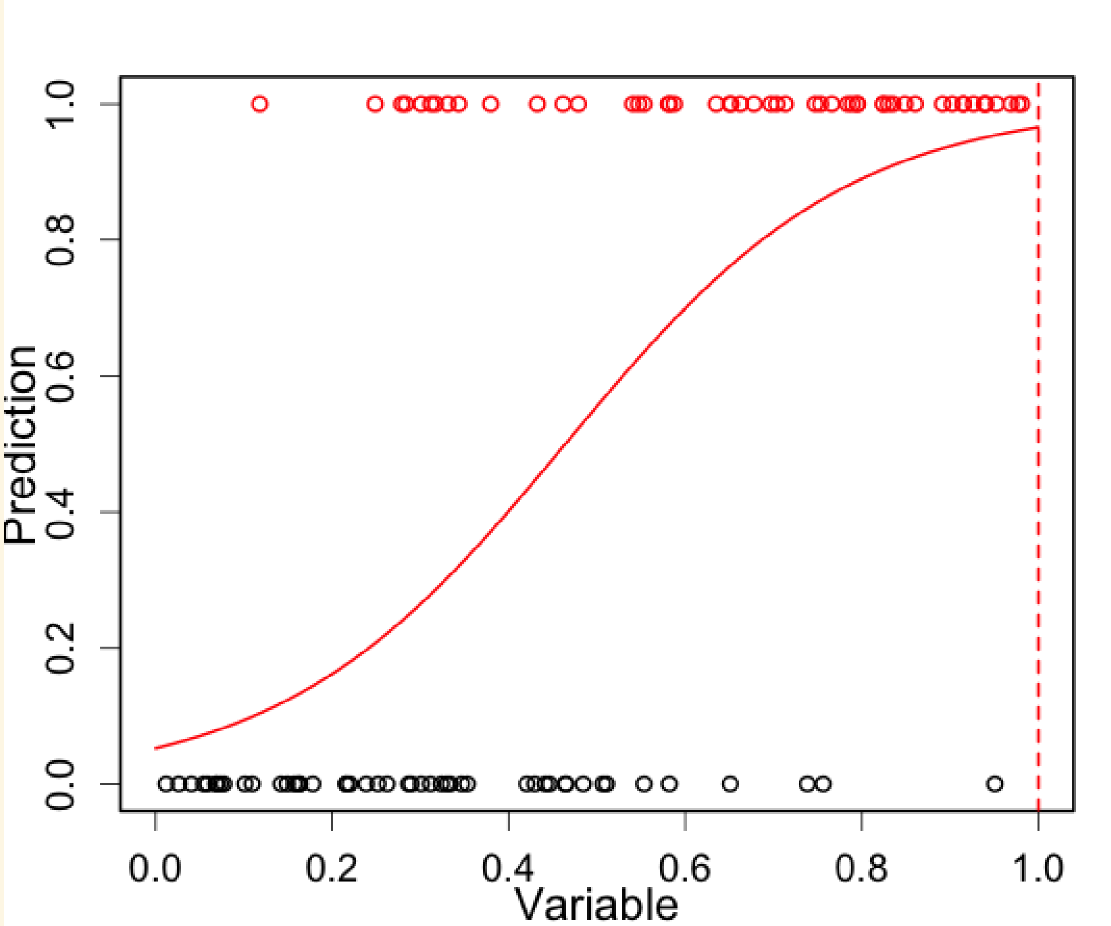
THRESHOLD MODEL



	0	1
1	28	24
0	46	2

	value
Sens	0.462
Spec	0.958
TSS	0.420

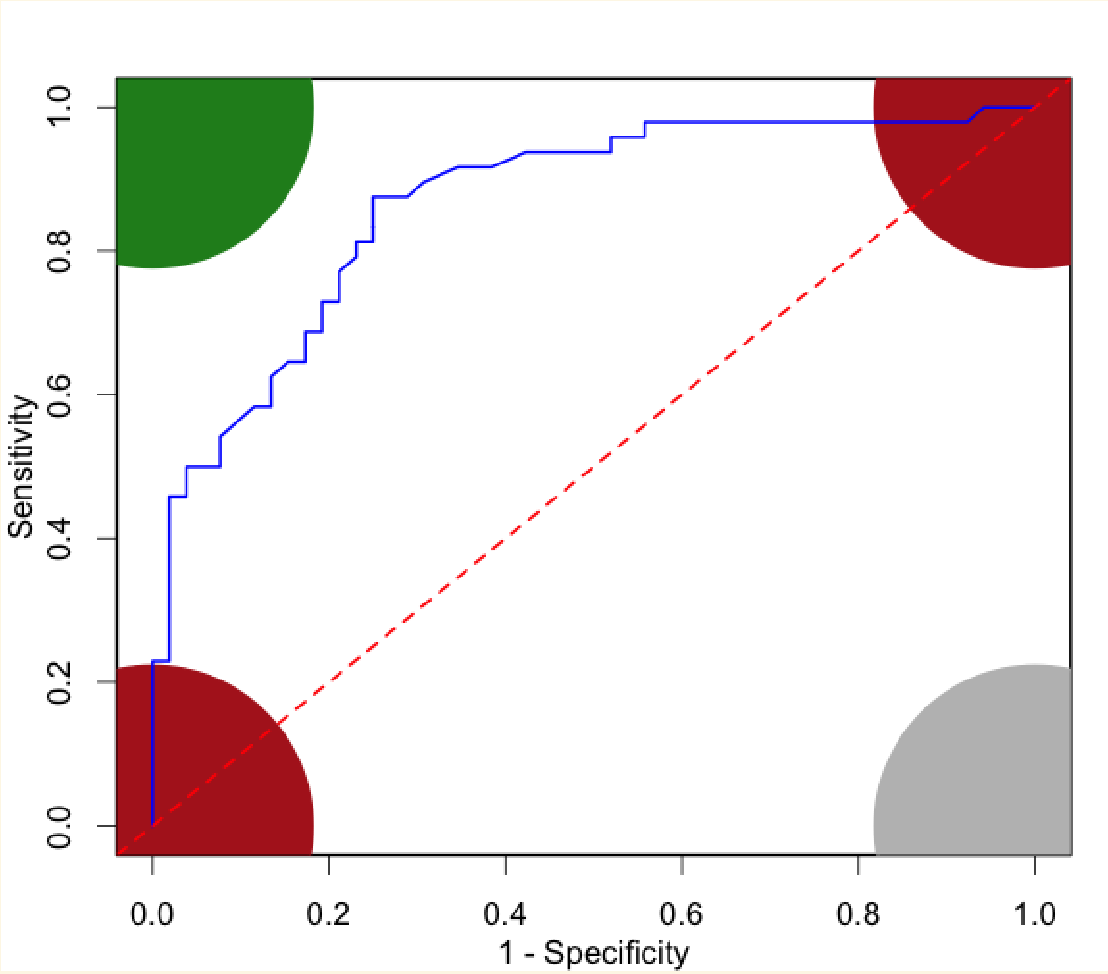
THRESHOLD MODEL



	0	1
1	52	0
0	48	0

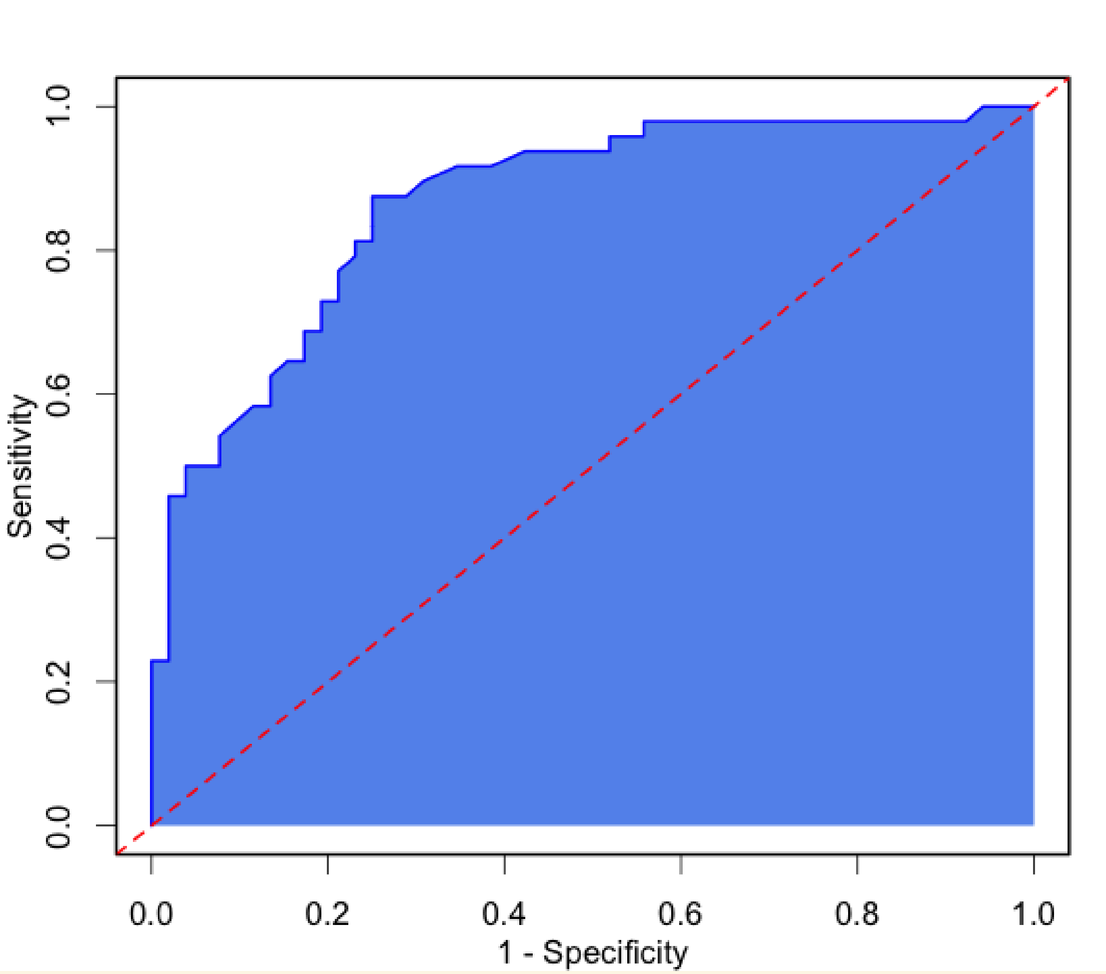
	value
Sens	0
Spec	1
TSS	0

ROC CURVE



- Receiver operating characteristic (ROC)
- A random model gives the red line

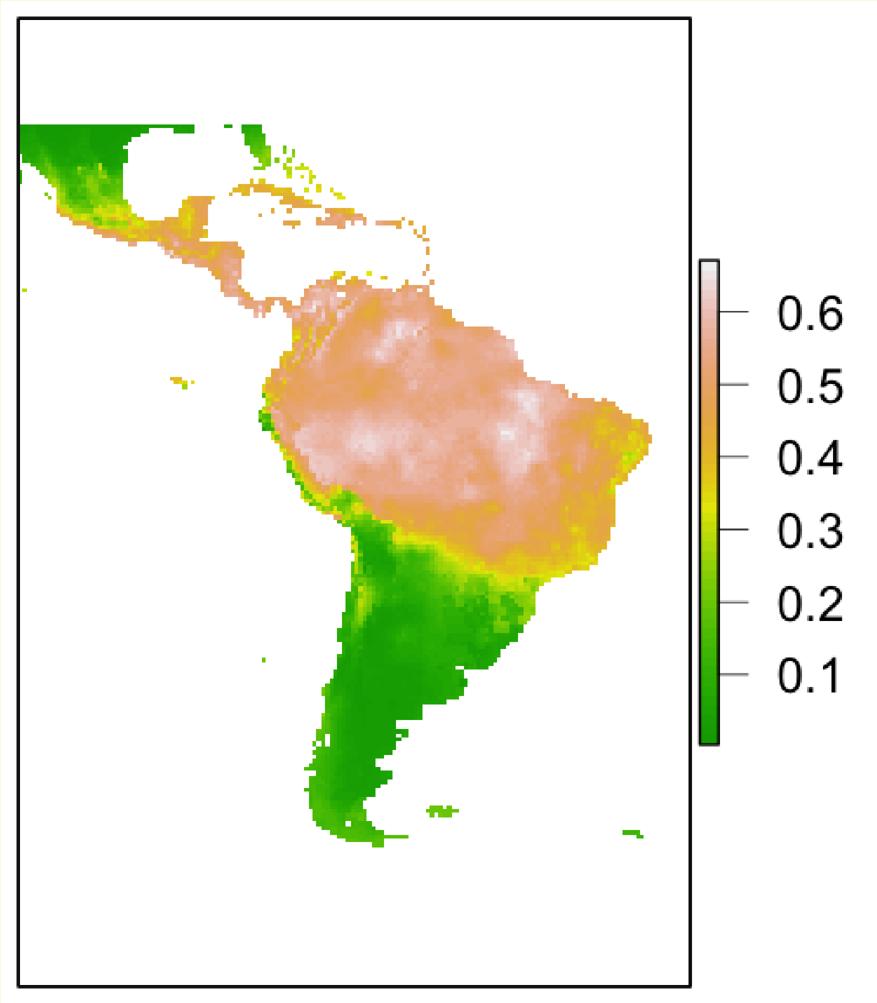
AREA UNDER ROC CURVE (AUC)



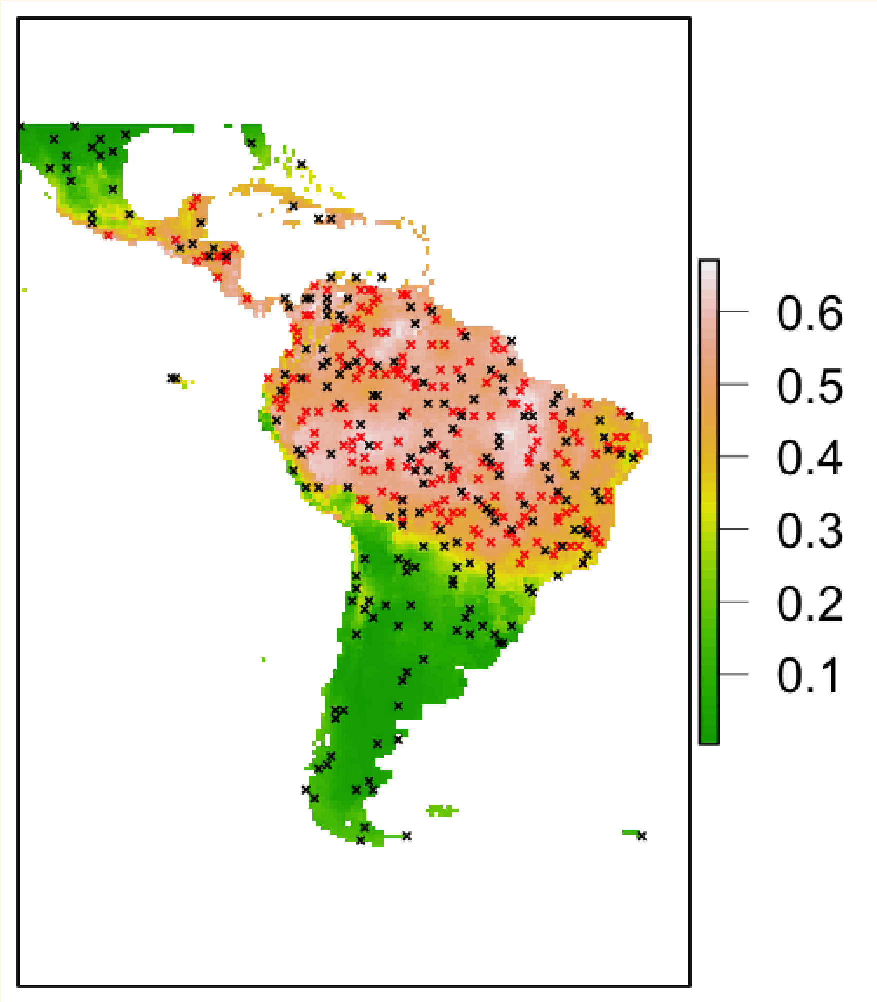
- AUC varies between 0 and 1.
- AUC = 0.5 is random.
- Overall model performance

SPECIES DISTRIBUTION MODELS

Kinkajou (*Potos flavus*)

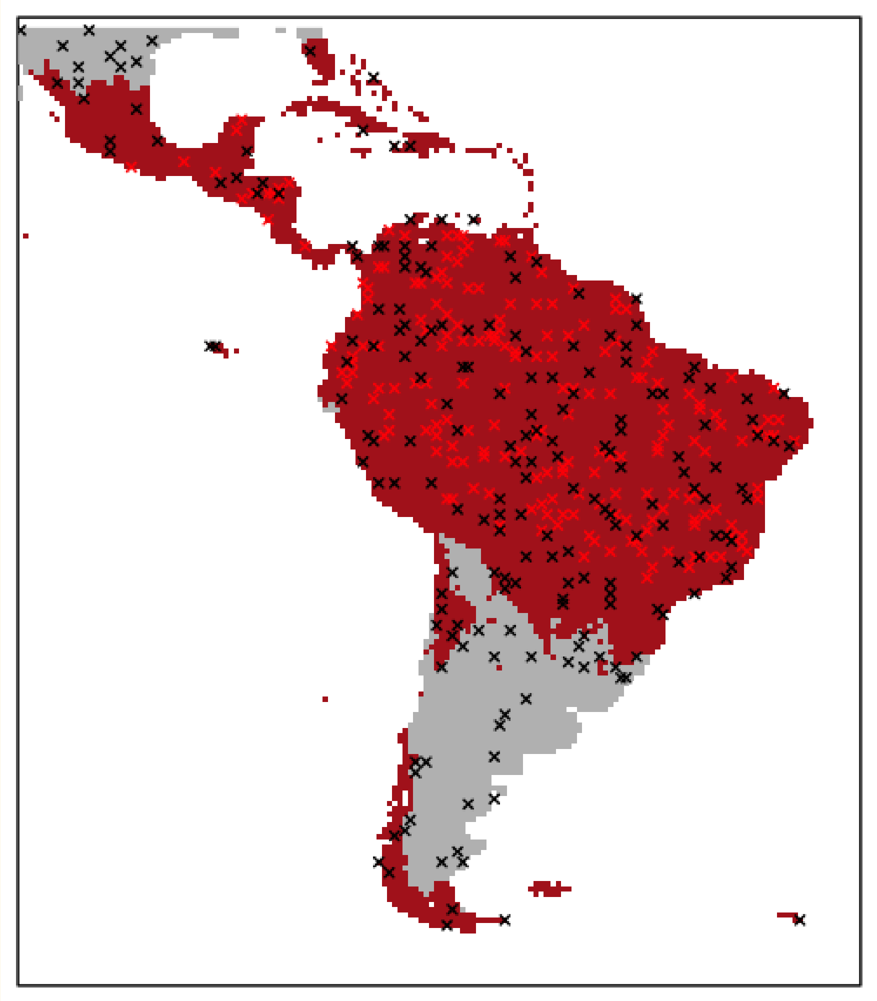


SPECIES DISTRIBUTION MODELS



- Observed (red)
- Background (black)

SPECIES DISTRIBUTION MODELS

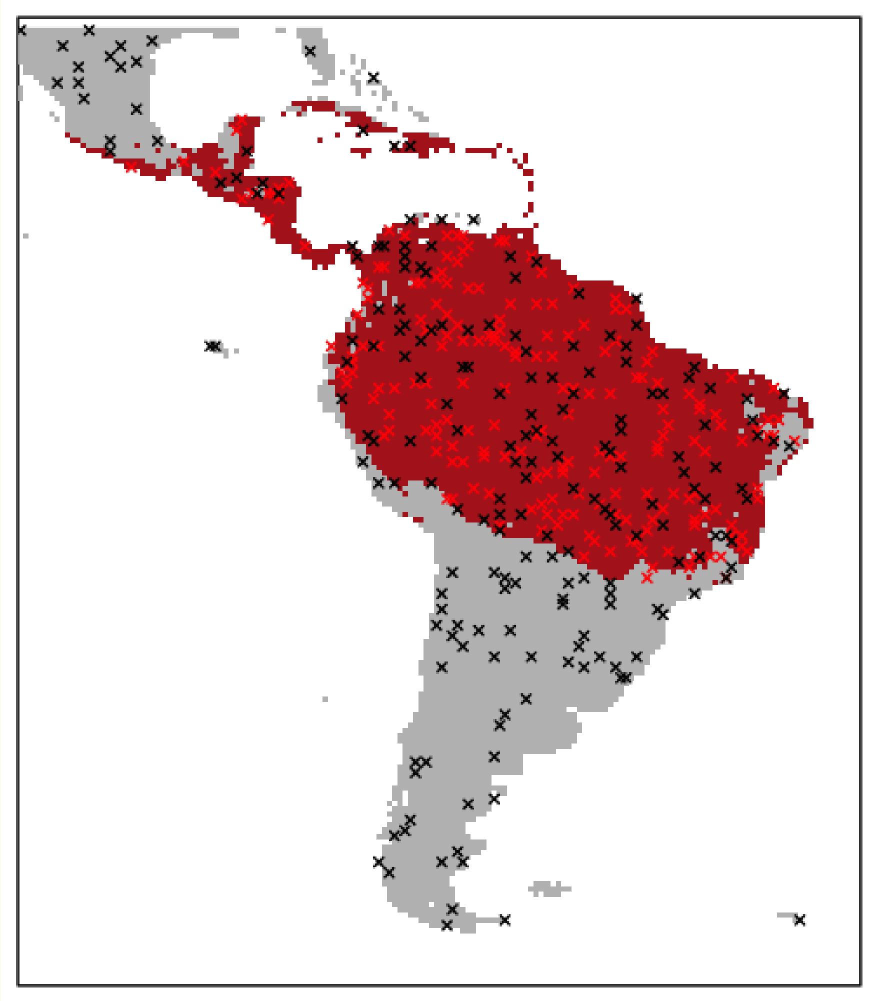


Threshold = 0.1

	Present	Absent
Obs	200	0
Back	162	38

	value
Sens	0.19
Spec	1.00
TSS	0.19

SPECIES DISTRIBUTION MODELS

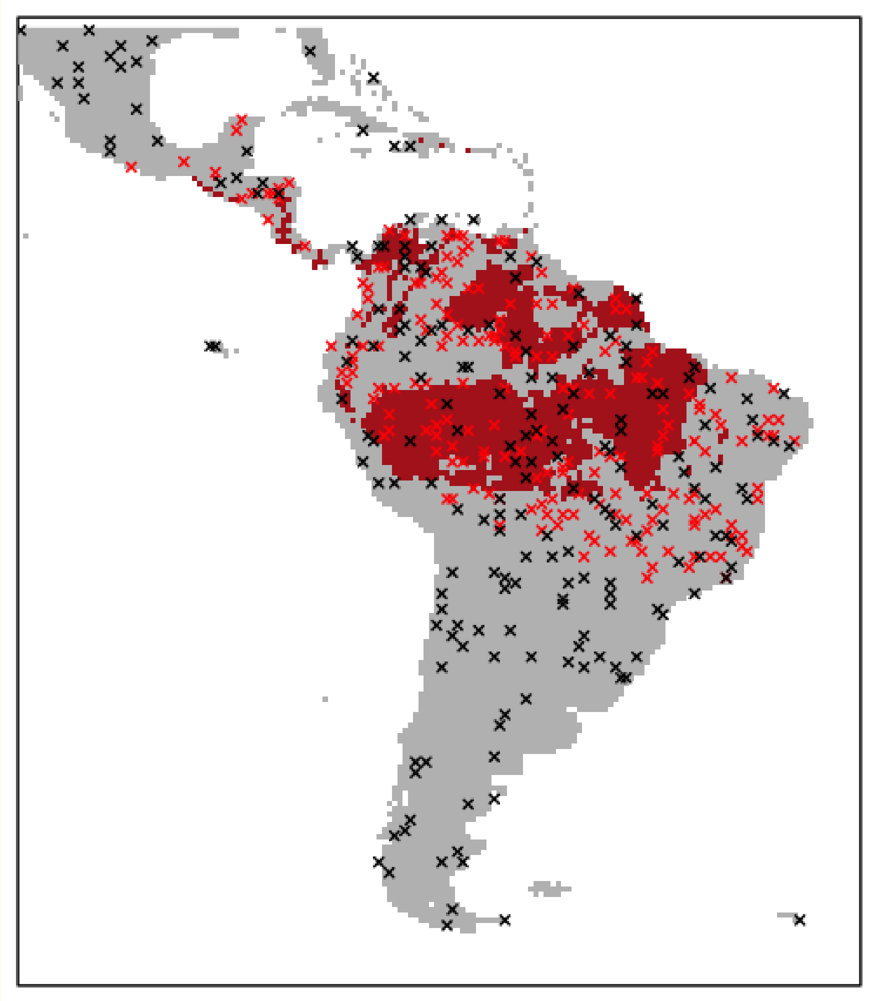


Threshold = 0.4

	Present	Absent
Obs	187	13
Back	111	89

	value
Sens	0.445
Spec	0.935
TSS	0.380

SPECIES DISTRIBUTION MODELS



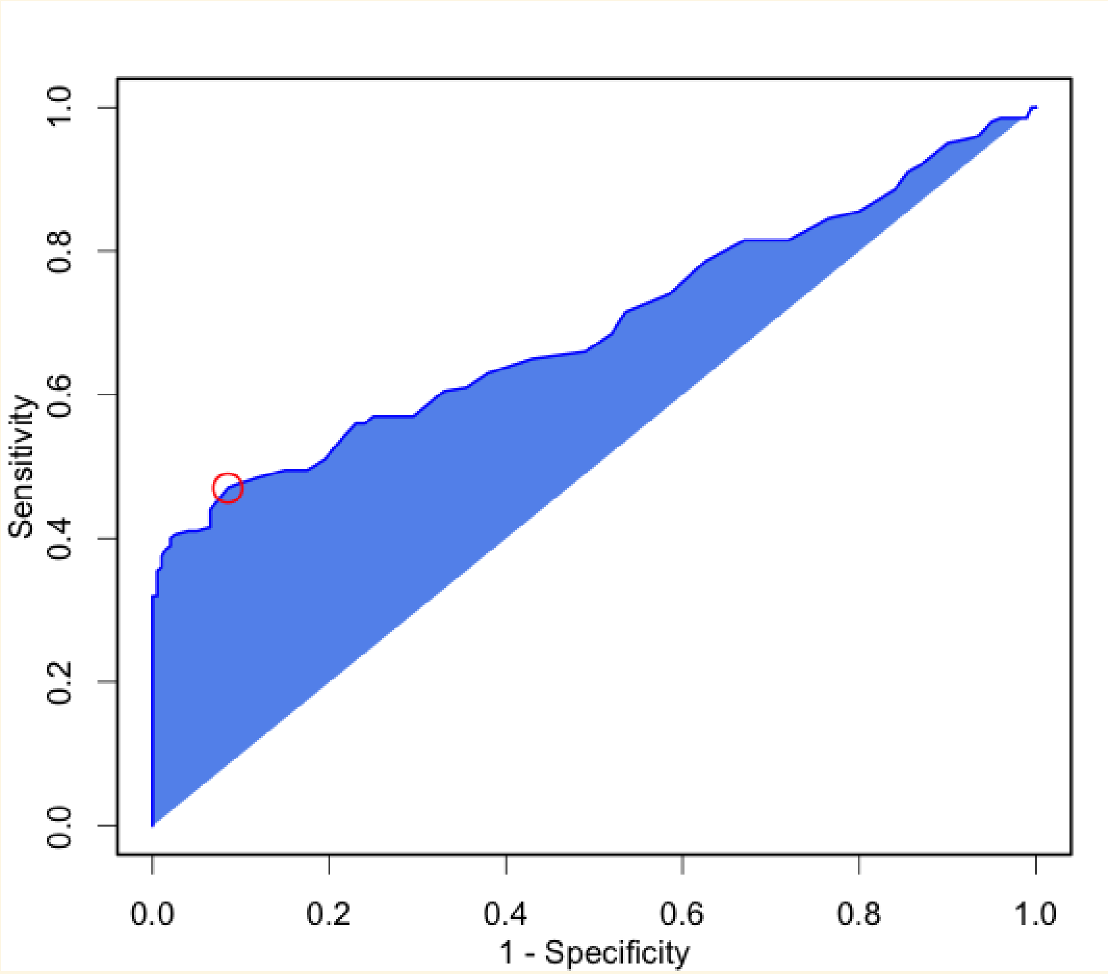
Threshold = 0.55

	Present	Absent
Obs	67	133
Back	37	163

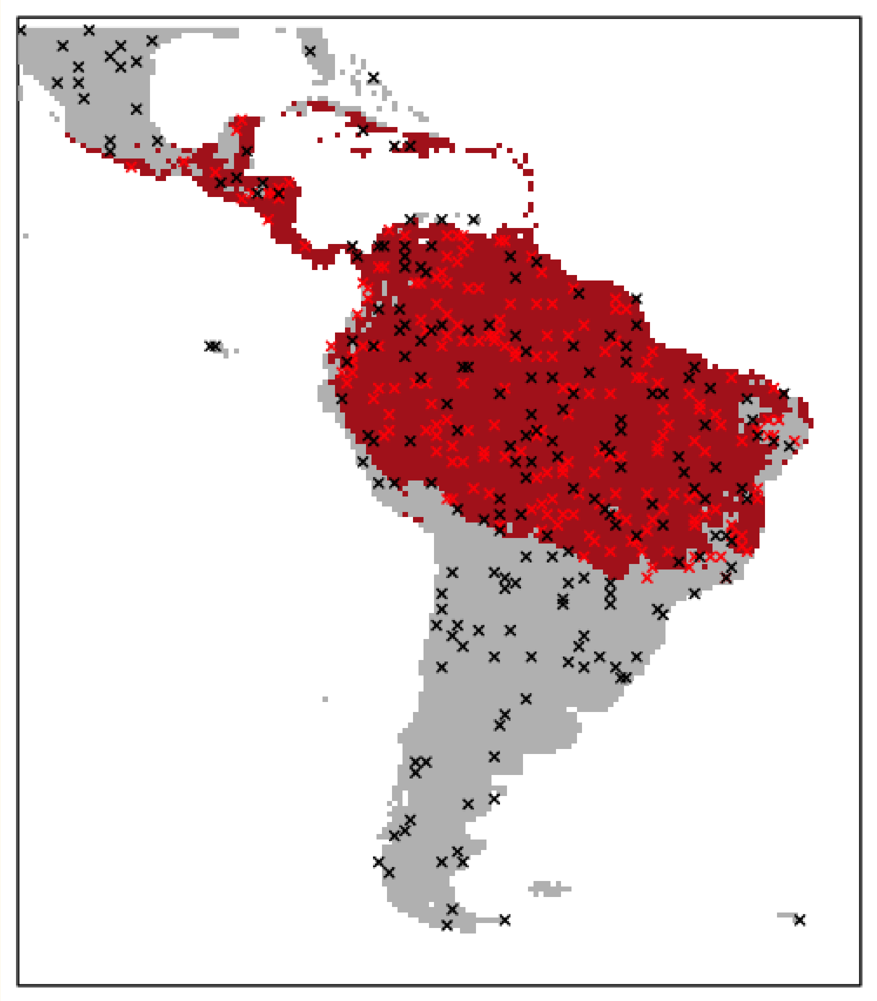
	value
Sens	0.815
Spec	0.335
TSS	0.150

AUC FOR THE KINKAJOU

Maximum sensitivity + specificity shown in red.



SPECIES DISTRIBUTION MODELS



Threshold = 0.411

	Present	Absent
Obs	183	17
Back	106	94

	value
Sens	0.470
Spec	0.915
TSS	0.385

THRESHOLD CHOICES

Method	Definition
Fixed value	An arbitrary fixed value (e.g. probability = 0.5)
Lowest predicted value	The lowest predicted value corresponding with an observed occurrence record
Sensitivity-specificity equality	The threshold at which sensitivity and specificity are equal
Sensitivity-specificity sum maximization	The sum of sensitivity and specificity is maximized
Maximize Kappa	The threshold at which Cohen's Kappa statistic is maximized
Equal prevalence	Propn of presences relative to the number of sites is equal in prediction and calibration data