Assessing model accuracy

David Orme

October 23 2016

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

MODIS Confusion Matrix

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 can have reasonable accuracy!

	Pred. Needle	**Pred. Broad**	**Sum**
Obs. Needle	754	748	1502
Obs. Broad	2520	2400	4920
Sum	3274	3148	6422

$$A = \frac{754 + 2400}{6422} = 49.1\%$$

Accuracy

And so can stupid ones: 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

Prevalence is simple the proportion of the observed positive outcomes:

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

Accuracy

And note that accuracy is affected by prevalence

	Pred. Needle	Pred. Broad	Sum
Obs. Needle	0	35	35
Obs. Broad	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. Needle	**Pred. Broad**
Obs. Needle	True Positive	False Negative
Obs. Broad	False Positive	True Negative

Prediction outcomes

Other much less intuitive names do get used:

	Pred. Needle	**Pred. Broad**
Obs. Needle	True Positive	Type II Error
Obs. Broad	Type I Error	True Negative

Rates of outcomes

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

	Pred. Needle	**Pred. Broad**
Obs. Needle	True Positive Rate	False Negative Rate
Obs. Broad	False Positive Rate	True Negative Rate

Rates of outcomes

Calculate those values:

	Pred. Needle	**Pred. Broad**	**Sum**
Obs. Needle	$\frac{1460}{1502} = 97.2\%$	$\frac{42}{1502} = 2.8\%$	1502
Obs. Broad	$\frac{31}{4920} = 0.6\%$	$\frac{4889}{4920} = 99.4\%$	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. Needle	Pred. Broad	Sum
Obs. Needle	1460	42	1502
Obs. Broad	2010	2910	4920
Sum	3470	2952	6422

Outcome rates for the new model above

	Pred. Needle	Pred. Broad
Obs. Needle	97.2%	2.8%
Obs. Broad	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:

$$\kappa = \frac{A - P_{exp}}{1 - P_{exp}}$$

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 expected agreement by chance (P_{exp})

	Pred. Needle	Pred. Broad	p
Obs. Needle	0.054	0.180	0.234
Obs. Broad	0.178	0.588	0.766
p	0.232	0.768	1.000

$$\kappa = \frac{0.989 - (0.054 + 0.588)}{1 - (0.054 + 0.588)} = 0.993$$

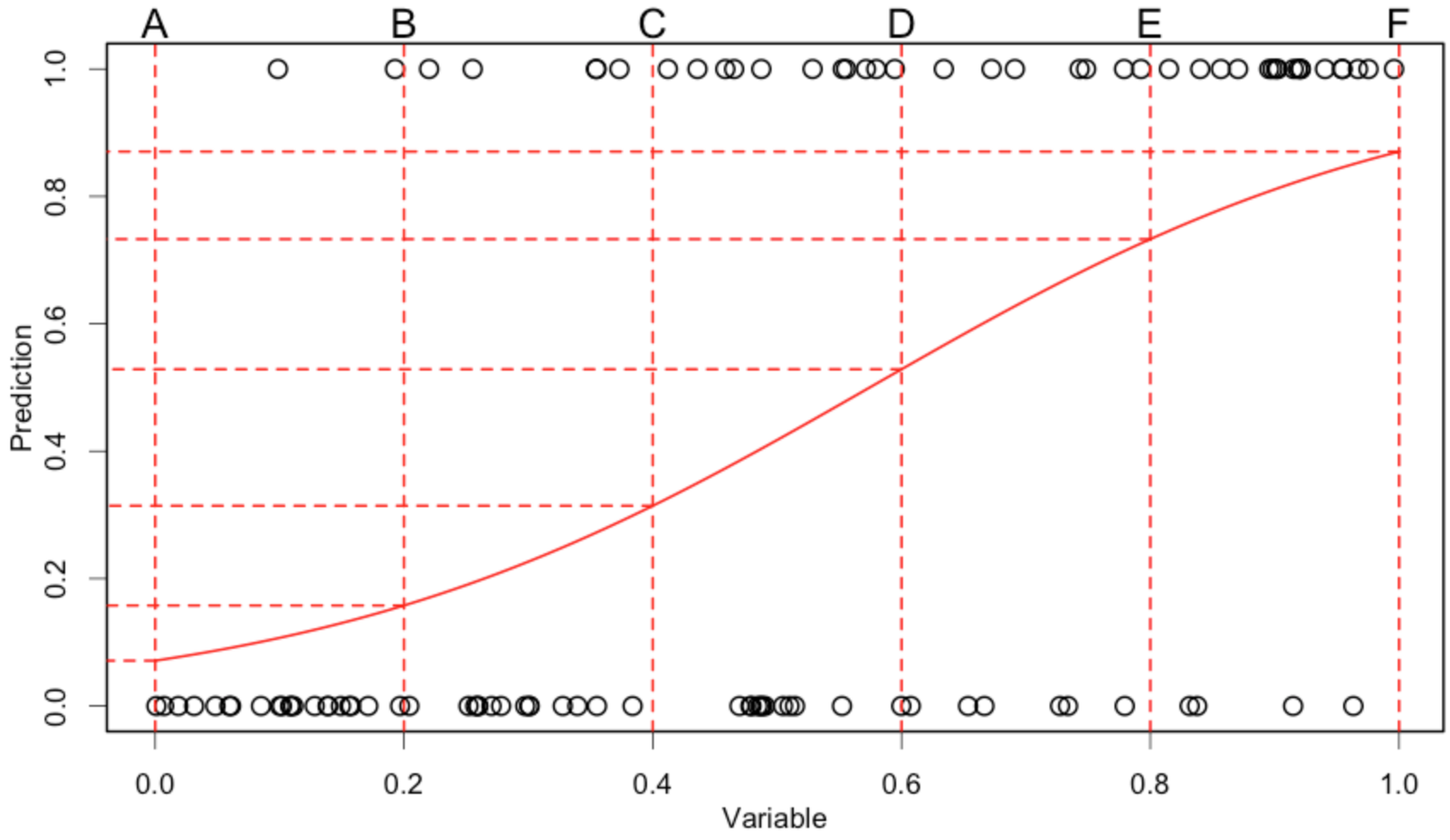
True Skill Statistic

An alternative measure is TSS:

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

- Sensitivity is in $[0, 1]$
- Specificity is also in $[0, 1]$
- TSS is between 1 (perfect) and -1 (perfectly imperfect!)
- TSS is unaffected by prevalence.

Threshold model



Threshold model

A	0	1
0	43	0
1	57	0

C	0	1
0	36	7
1	22	35

E	0	1
0	18	25
1	4	53

B	0	1
0	41	2
1	36	21

D	0	1
0	25	18
1	10	47

F	0	1
0	0	43
1	0	57

Threshold model

A	value
Sens	1
Spec	0
TSS	0

B	value
Sens	0.953
Spec	0.368
TSS	0.322

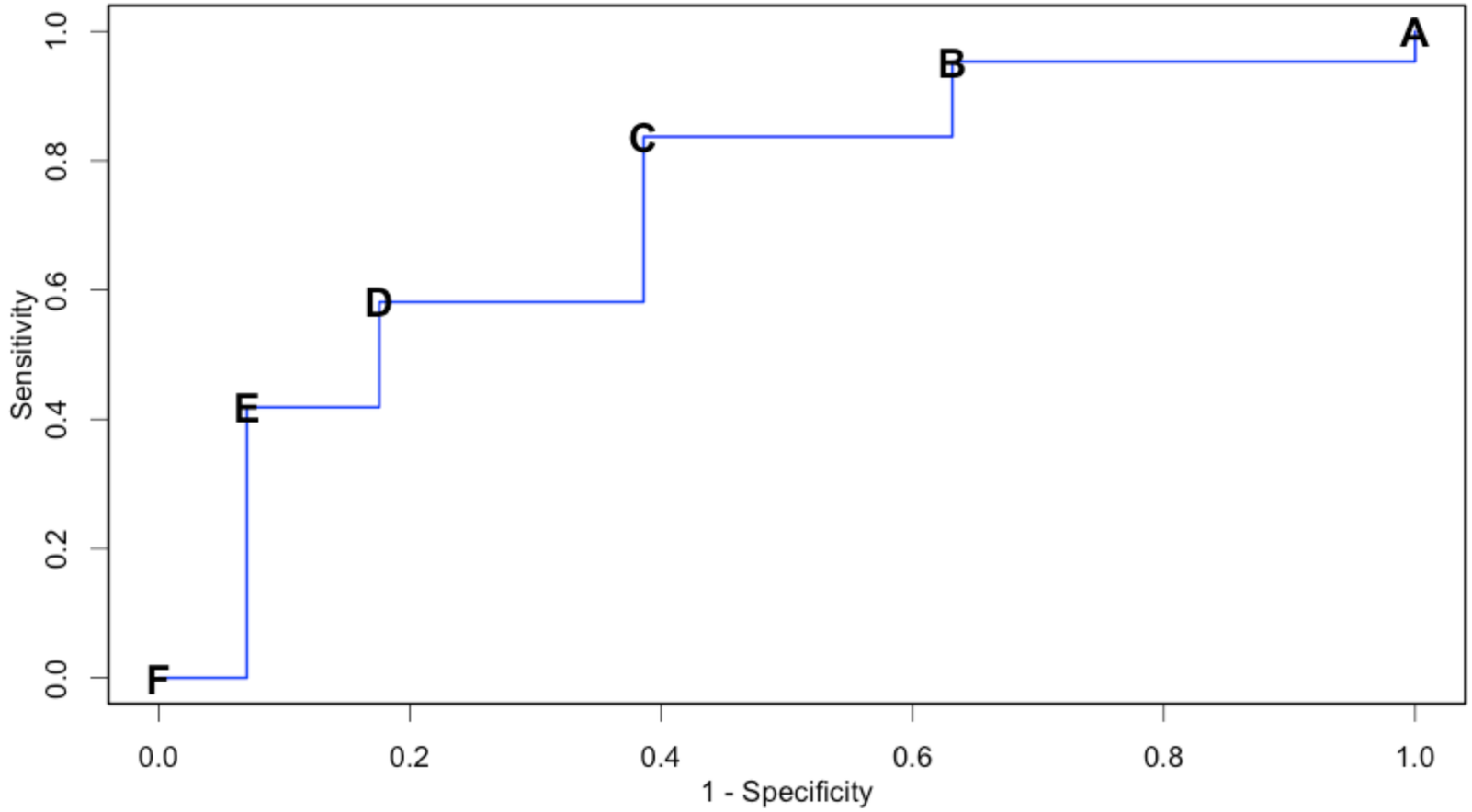
C	value
Sens	0.837
Spec	0.614
TSS	0.451

D	value
Sens	0.581
Spec	0.825
TSS	0.406

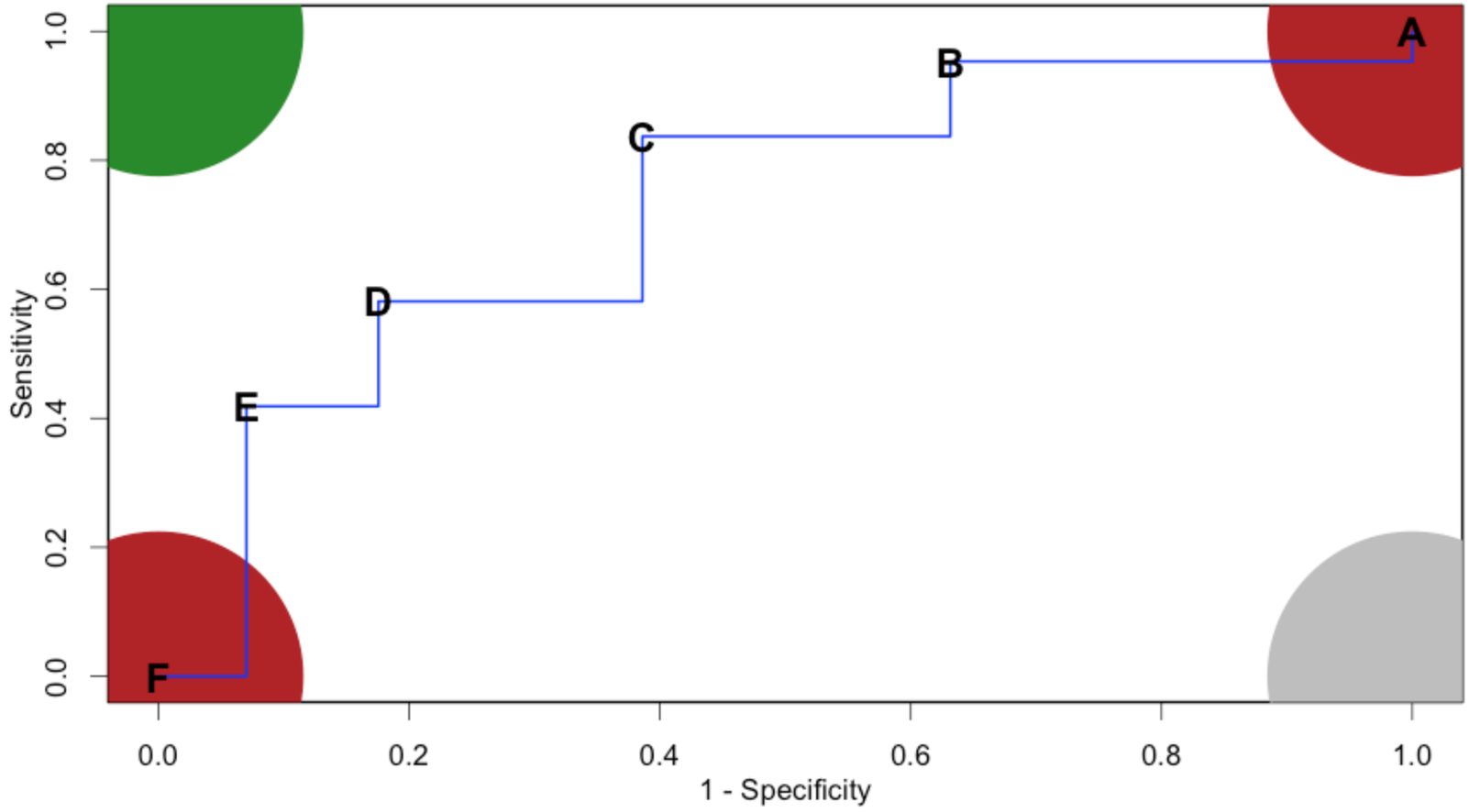
E	value
Sens	0.419
Spec	0.930
TSS	0.348

F	value
Sens	0
Spec	1
TSS	0

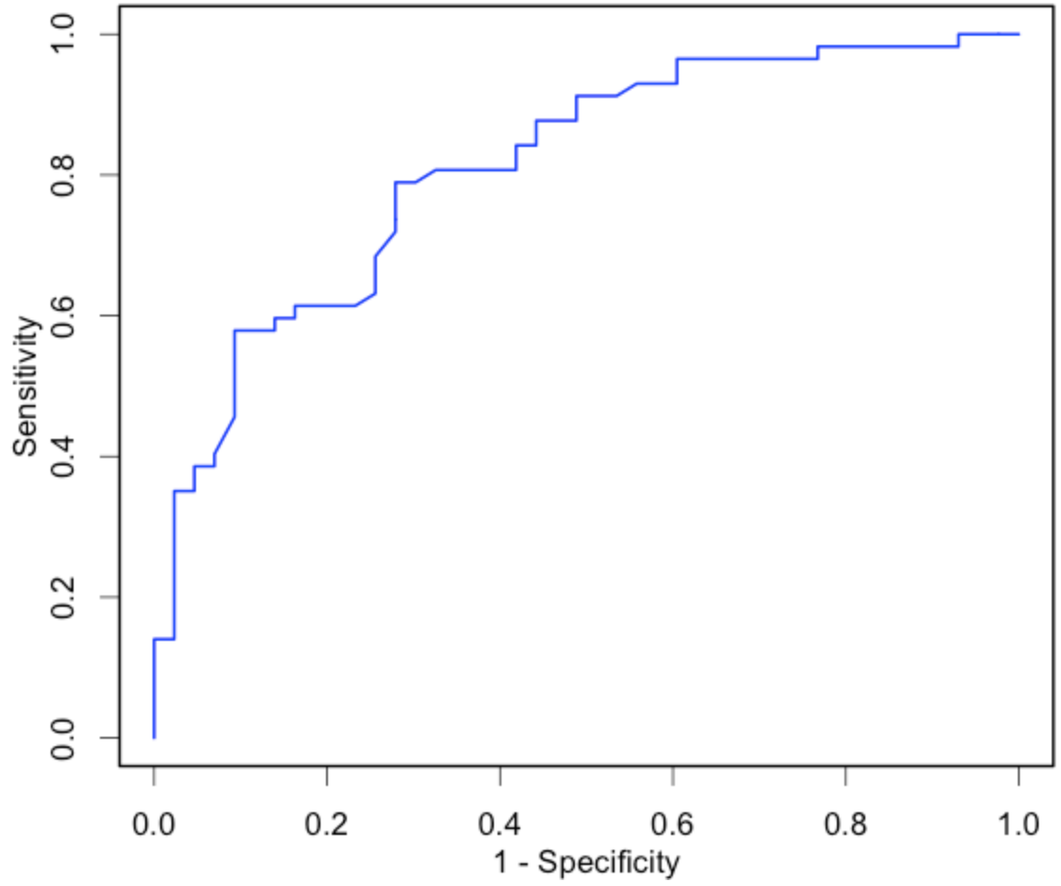
ROC Curve


A
B
C
D
E
F
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

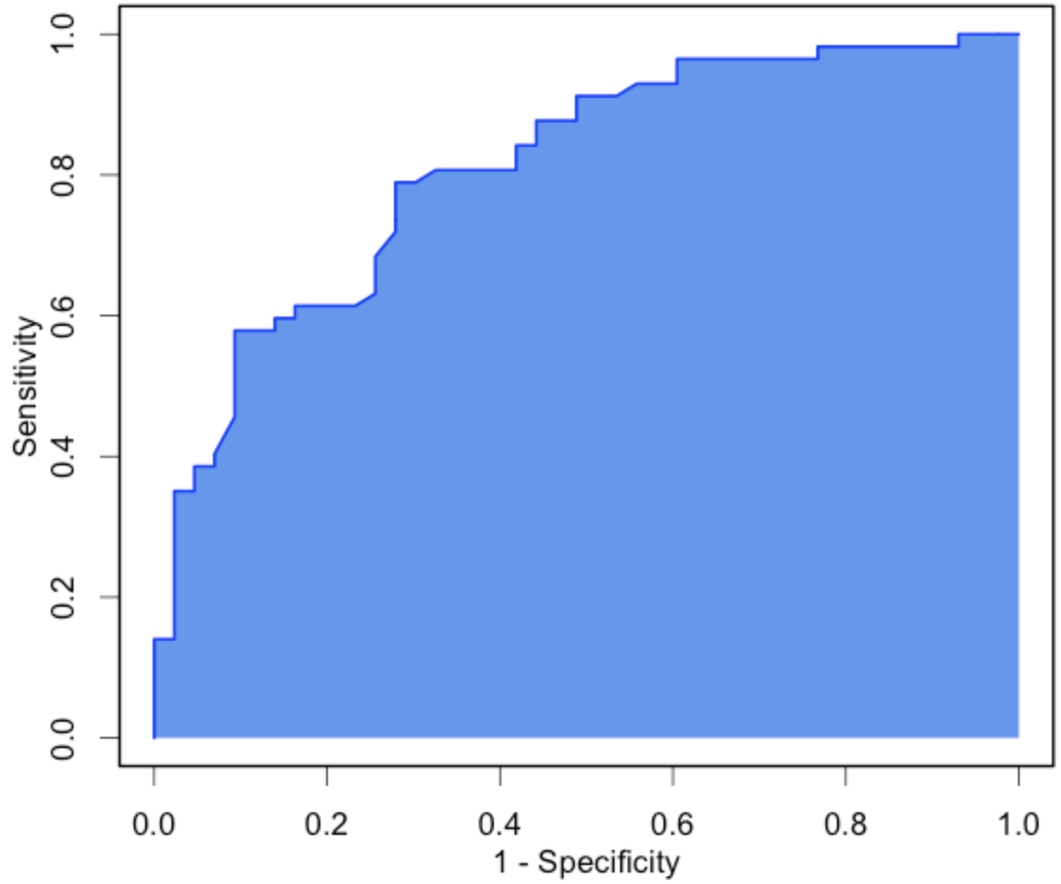
ROC Curve


A
B
C
D
E
F
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

ROC Curve



Area under ROC curve (AUC)

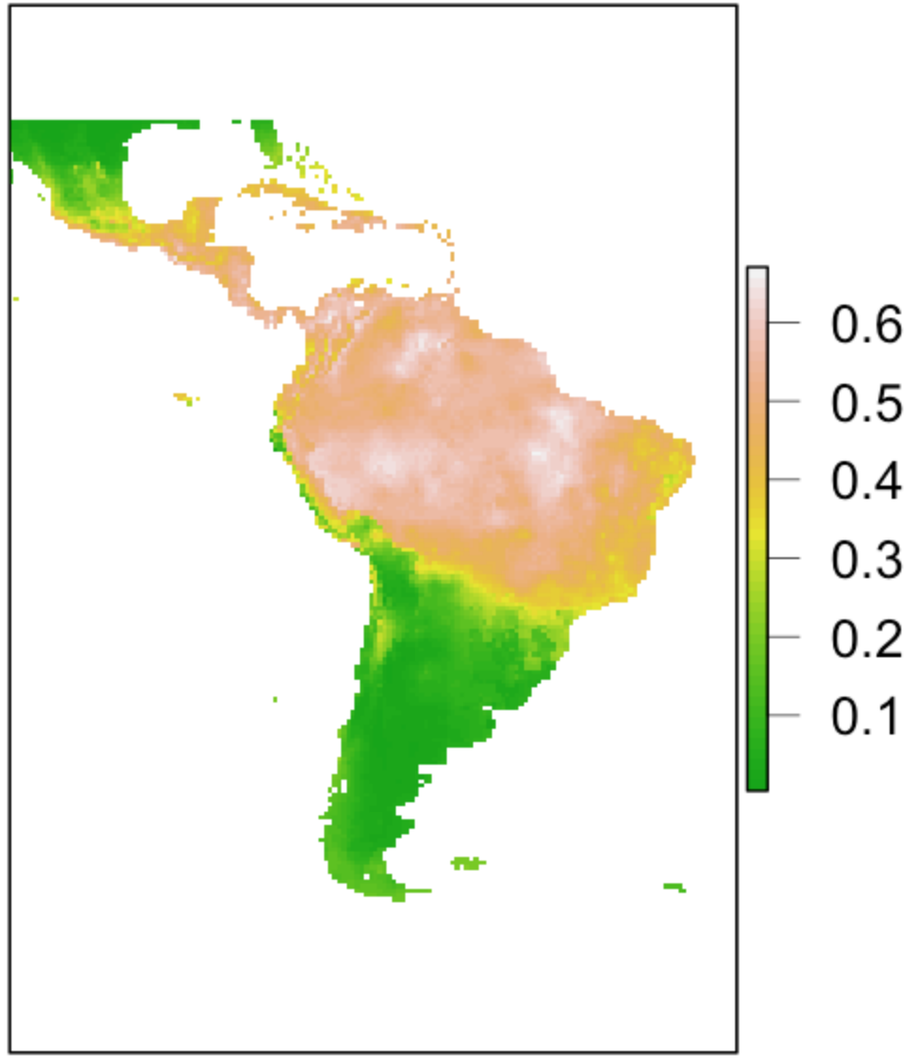


Area under ROC curve (AUC)

- AUC varies between 0 and 1.
- It is a threshold independent measure of model performance
- A model with an AUC of 0.5 is doing no better than random.

Species Distribution Models

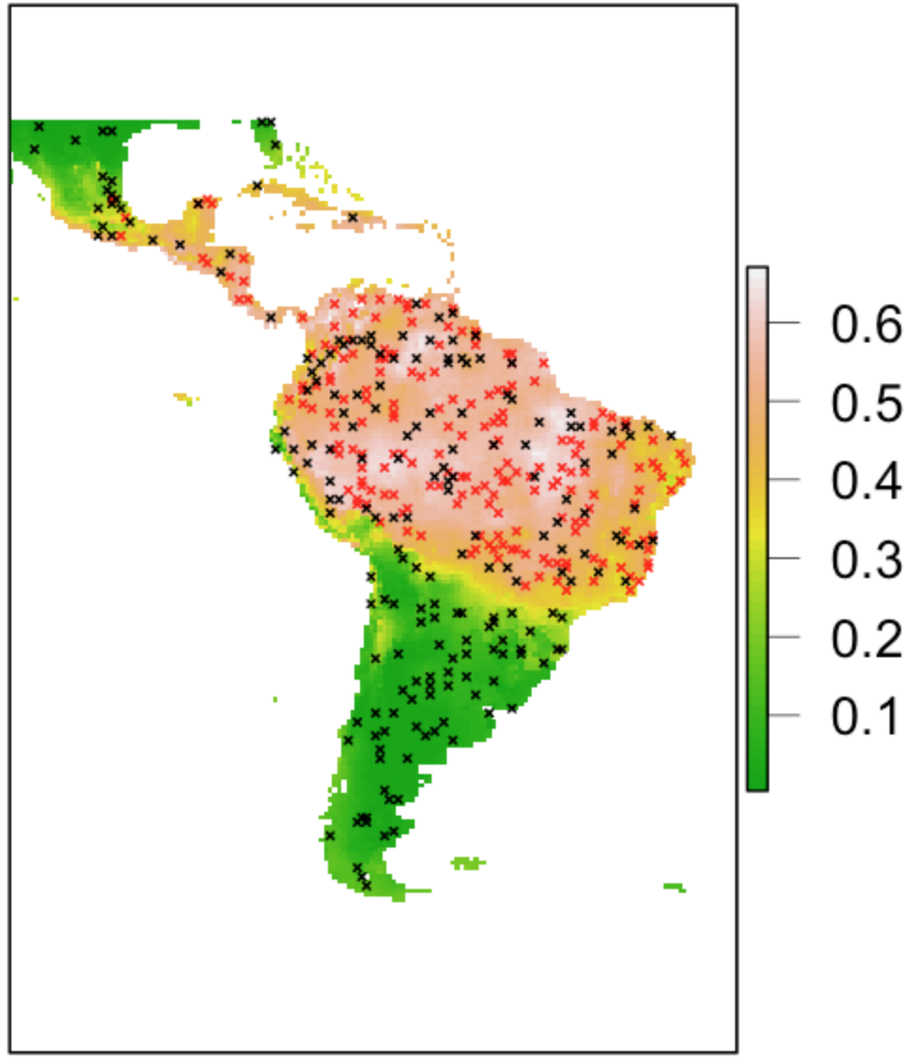
Kinkajou (*Potos flavus*)



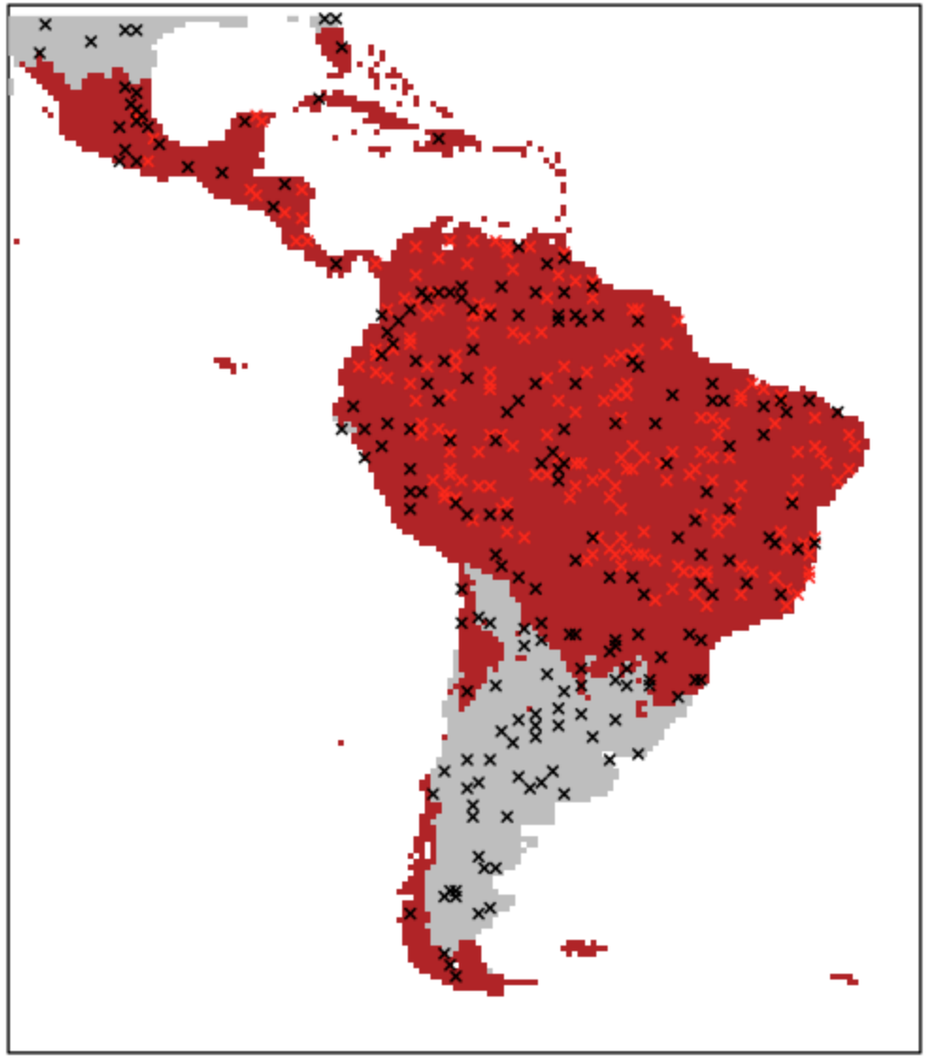
Species Distribution Models

Kinkajou (*Potos flavus*)

- Observed (red)
- Background (black)



Species Distribution Models

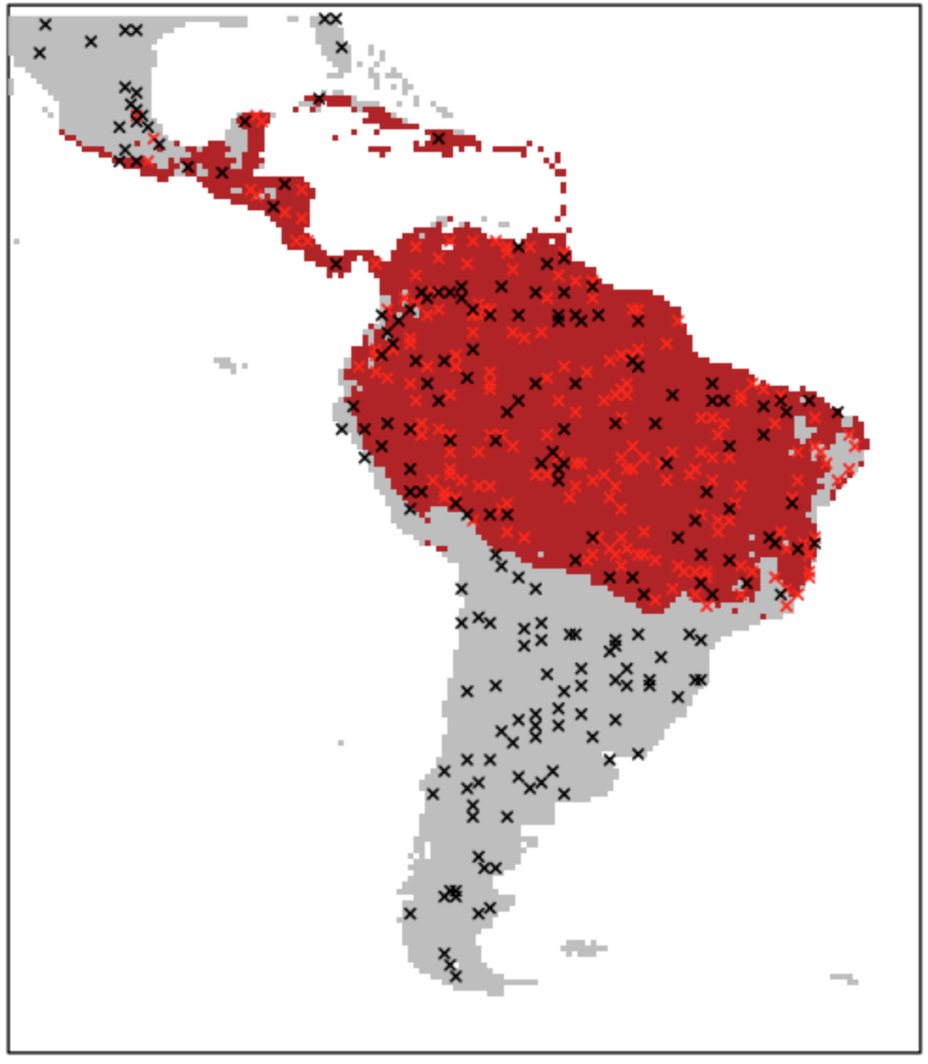


Threshold = 0.1

	Present	Absent
Obs	200	0
Back	146	54

	value
Sens	1.00
Spec	0.27
TSS	0.27

Species Distribution Models

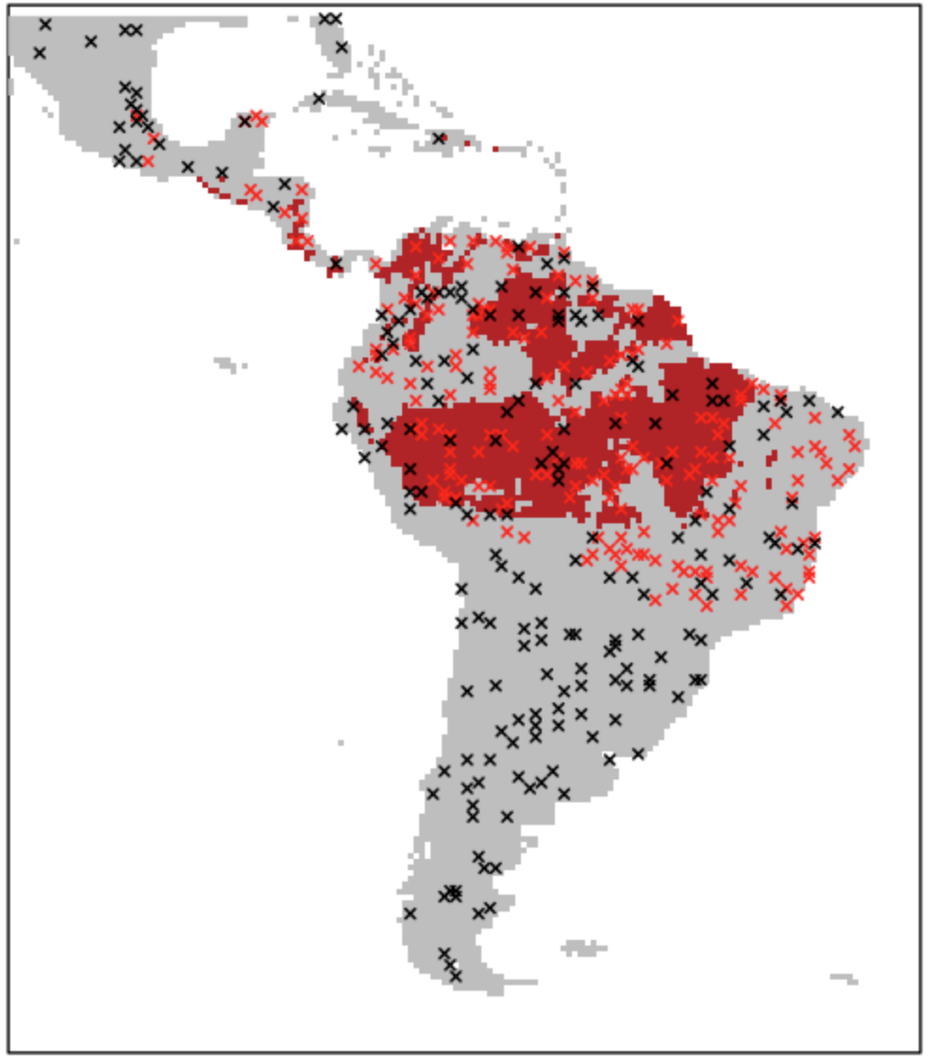


Threshold = 0.4

	Present	Absent
Obs	184	16
Back	103	97

	value
Sens	0.920
Spec	0.485
TSS	0.405

Species Distribution Models



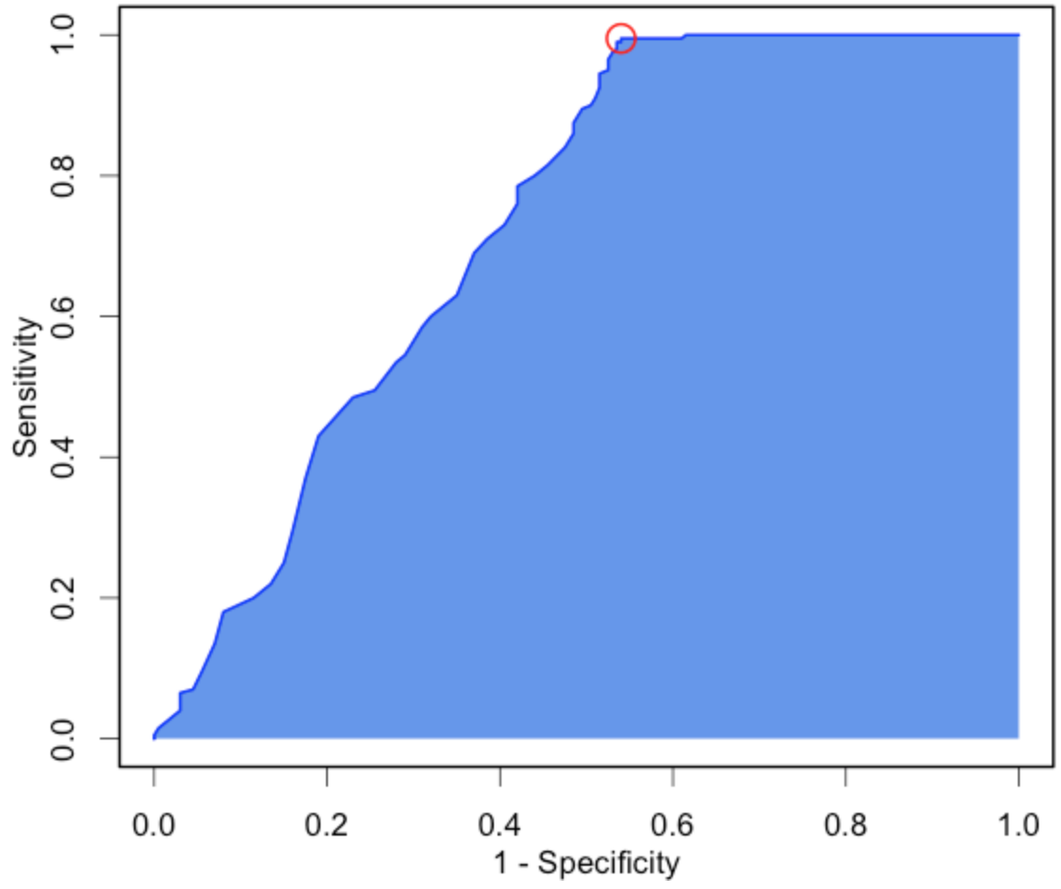
Threshold = 0.55

	Present	Absent
Obs	76	124
Back	36	164

	value
Sens	0.38
Spec	0.82
TSS	0.20

AUC for the Kinkajou

Maximum sensitivity + specificity shown in red.



Species Distribution Models

Threshold = 0.331