

Modelagem de nicho ecológico e distribuição de espécies



Marília Melo Favalesso

Laboratório de Investigações Biológicas (LINBIO)



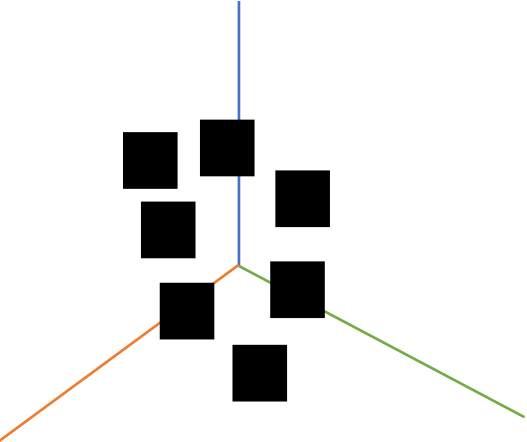
Agradecimentos pelo CRTL + C e CRTL + V

- **Thadeu Sobral de Souza** (LEEC – UNESP – Rio Claro, SP)
thadeusobral@gmail.com
- **Maurício Humberto Vancine** (LEEC – UNESP – Rio Claro, SP)
mauricio.vancine@gmail.com

Onde paramos:

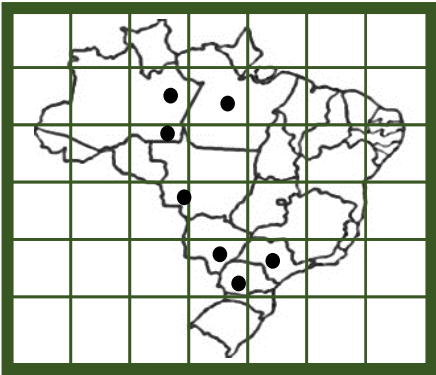
VARIÁVEIS ECOLÓGICAS

Modelagem matemática
Regressão logística, distância, bioclim,
MAXENT, etc

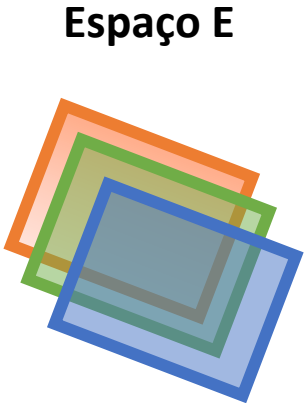


**Projeta o modelo no
espaço geográfico**

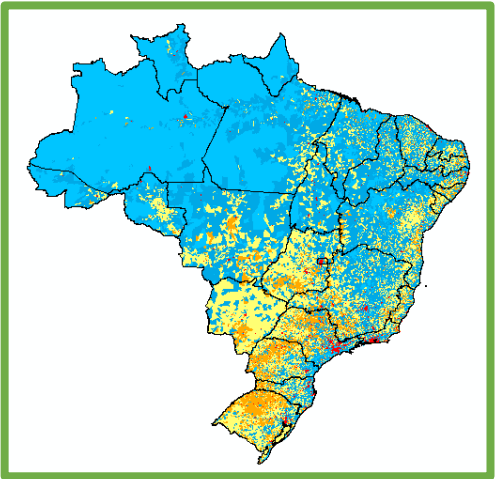
ESPAÇO GEOGRÁFICO



Espaço M



Espaço E



Mas e a modelagem matemática?

“É a arte de transformar problemas da realidade em problemas matemáticos e resolvê-los interpretando suas soluções na linguagem do mundo real”.

Basicamente criamos uma **função matemática** que representa (ou chega perto de representar) uma realidade.

Representação do pé-grande



Pé-grande real



“(...) Os modelos não representam TODAS as características do mundo real, mas representam aquelas essências para entender um problema específico, dados os parâmetros utilizados/disponíveis e os pressupostos assumidos (...)”

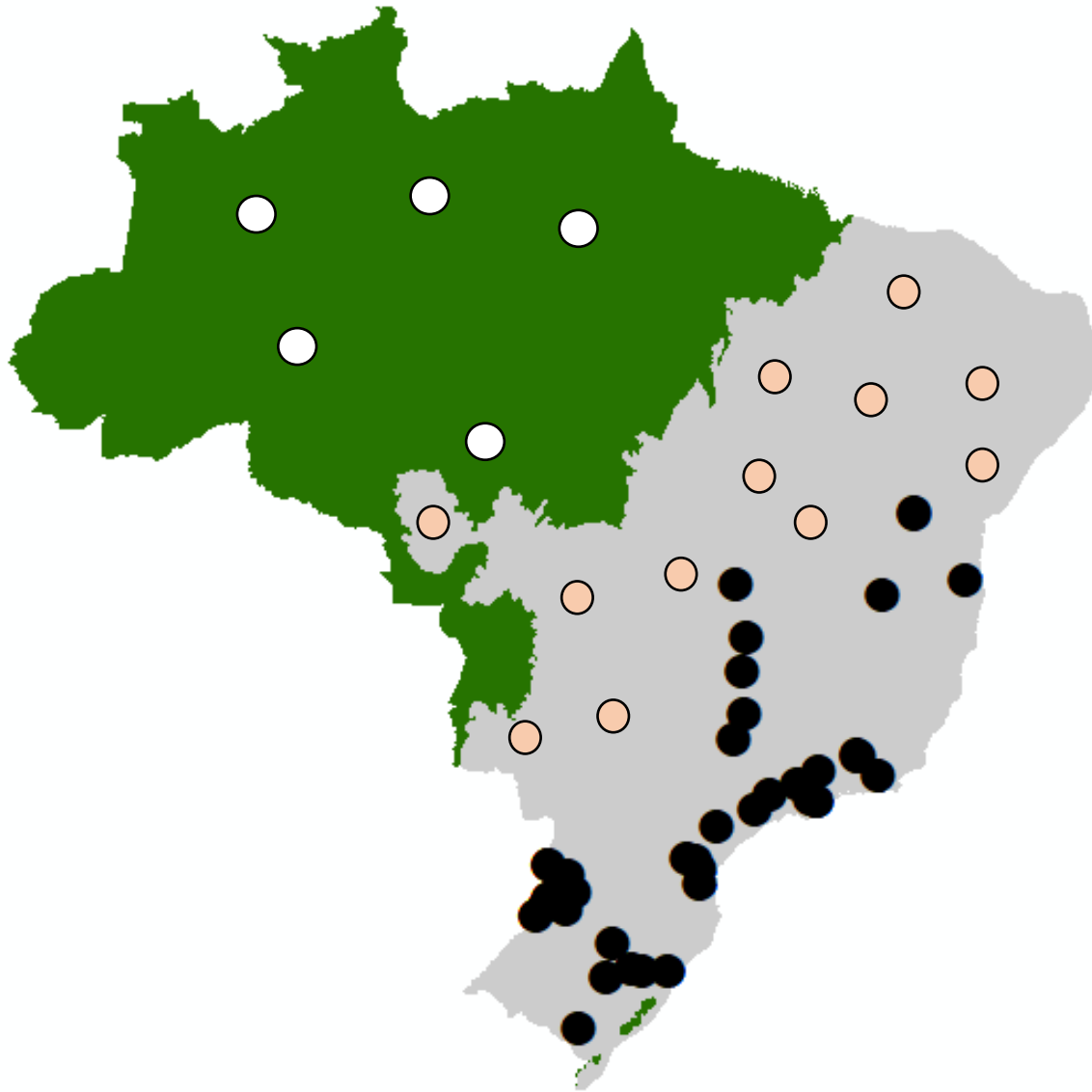
Modelos a partir de algoritmos...

“**Algoritmo** é uma sequência finita de instruções bem definidas e não ambíguas, cada uma das quais devendo ser executadas mecânica ou eletronicamente em um intervalo de tempo finito e com uma quantidade de esforço finita (...)”



“(...) Um algoritmo não representa, necessariamente, um programa de computador, e sim os passos necessários para realizar uma tarefa (...)”

Outros conceitos importantes



- **Presença:**
Locais onde a espécie foi observada ou registrada.
- **Ausência:**
Locais onde foi confirmada a não ocorrência da espécie.
- **Pseudo-ausência:**
Pontos gerados artificialmente para representar locais onde a espécie provavelmente não ocorre.
- **Background:**
Conjunto de pontos representando a disponibilidade das condições ambientais no espaço de estudo.

Para a modelagem de nicho e distribuição de espécies...

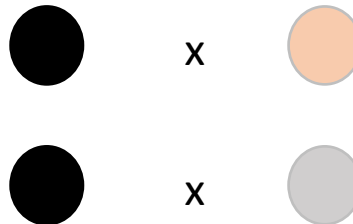
Apenas presença de espécies

Bioclim
Distância euclidiana
Distância de Mahalanobis
Distância de Gower (domain)
ENFA (análise ecológica de nicho fatoria)



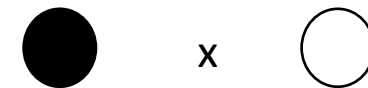
Presença vs. Pseudo-ausência

GARP (genetic algoritmo for rule-se production)
MAXENT (máxima entropia)
SVM (support vector machine)



Presença vs. Ausência

GLMz (generalized linear model)
GAM (generalized additive model)
FDS (flexible discriminant analysis)
MARS (multivariate adaptive reg. Splines)



Para a modelagem de nicho e distribuição de espécies...

Apenas presença de espécies

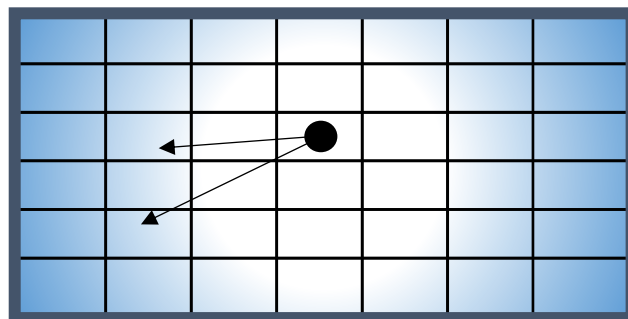
Presença vs. Pseudo-ausência

Presença vs. Ausência

Bioclim
Distância euclidiana
Distância de Mahalanobis
Distância de Gower (domain)
ENFA (análise ecológica de nicho fatoria)

GARP (genetic algoritmo for rule-
se production)
MAXENT (máxima entropia)
SVM (support vector machine)

GLMz (generalized linear model)
GAM (generalized additive model)
FDS (flexible discriminant analysis)
MARS (multivariate adaptive reg.
Splines)



Espaço E

- Ambiente é um gradiente
- Estima a distância, no espaço ambiental, entre cada região da área de estudo (célula) e o centroide do ponto de ocorrência.
- Geralmente a predição é contínua entre 0 e ∞

Para a modelagem de nicho e distribuição de espécies...

Apenas presença de espécies

Bioclim
Distância euclidiana
Distância de Mahalanobis
Distância de Gower (domain)
ENFA (análise ecológica de nicho fatoria)

Presença vs. Pseudo-ausência

GARP (genetic algoritmo for rule-se production)
MAXENT (máxima entropia)
SVM (support vector machine)

Presença vs. Ausência

GLMz (generalized linear model)
GAM (generalized additive model)
FDS (flexible discriminant analysis)
MARS (multivariate adaptive reg. Splines)

- Presença conhecida
- Pseudo-ausência



“black box” ou cofres

Para a modelagem de nicho e distribuição de espécies...

Apenas presença de espécies

Bioclim
Distância euclidiana
Distância de Mahalanobis
Distância de Gower (domain)
ENFA (análise ecológica de nicho fatoria)

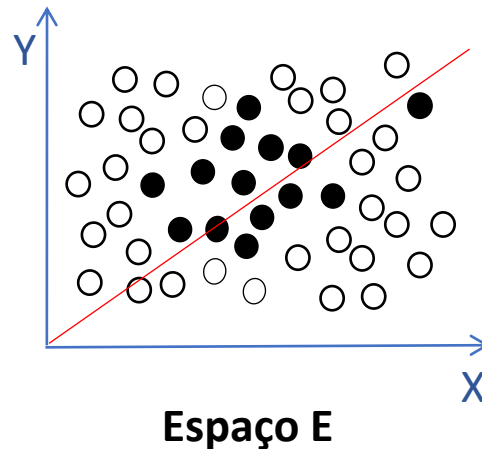
Presença vs. Pseudo-ausência

GARP (genetic algoritmo for rule-se production)
MAXENT (máxima entropia)
SVM (support vector machine)

Presença vs. Ausência

GLMz (generalized linear model)
GAM (generalized additive model)
FDS (flexible discriminant analysis)
MARS (multivariate adaptive reg. Splines)

● Presença
○ Ausência



Mas qual tipo de modelagem é melhor?

Adendo 1



Ausência ou pseudo-ausência?



- Certeza sobre os dados!
- Esforço amostral

Mas qual tipo de modelagem é melhor?

Adendo 2

Pé-grande



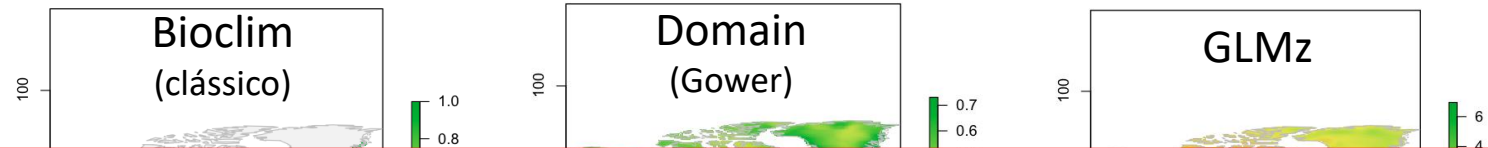
Chewbacca



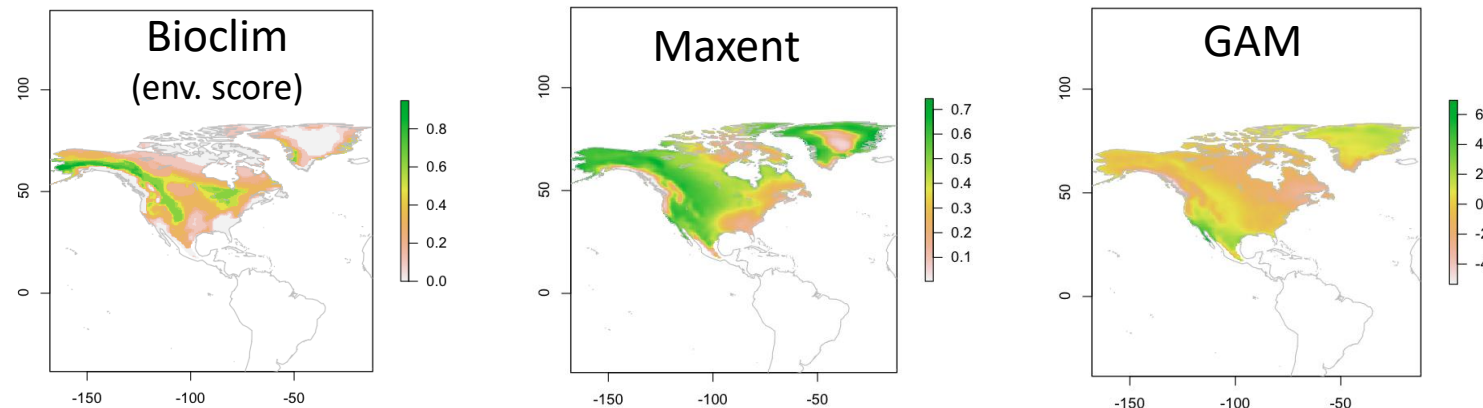
A espécie está corretamente identificada?

Mas qual tipo de modelagem é melhor?

Adendo 3



Utilizar métricas estatísticas para validação/avaliação do modelo!



Mas qual tipo de modelagem é melhor?

Validação vs. avaliação



Modelo de somente presença!

Mas qual tipo de modelagem é melhor?

Validação vs. avaliação

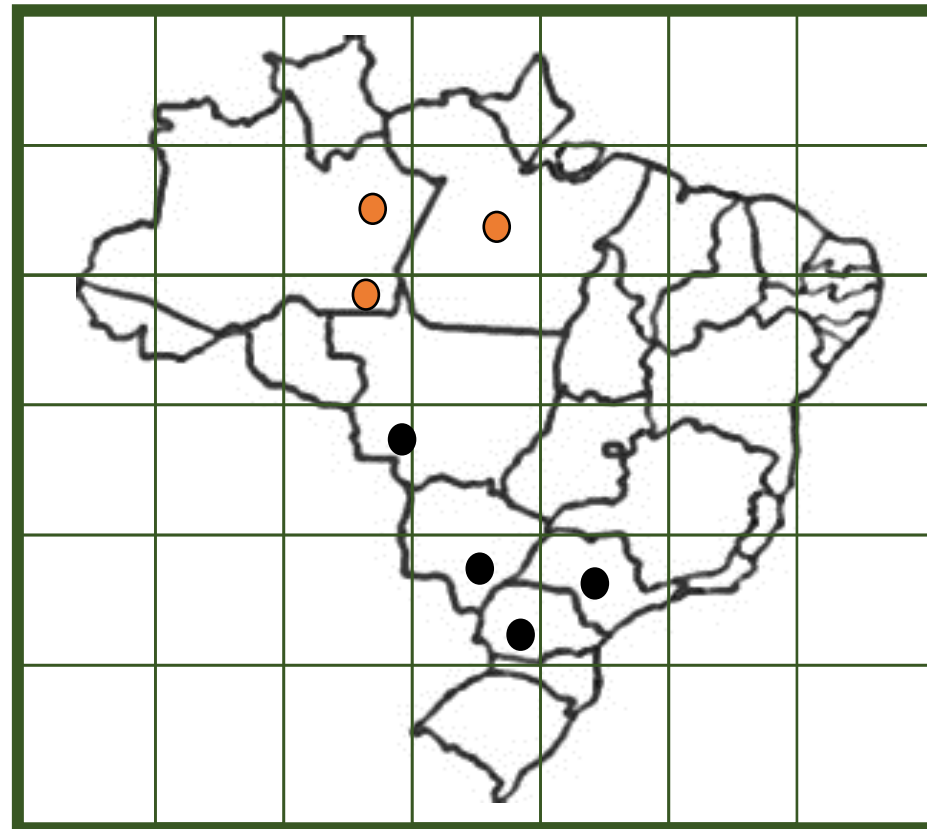


Modelo de somente presença!

Mas qual tipo de modelagem é melhor?

Validação vs. avaliação

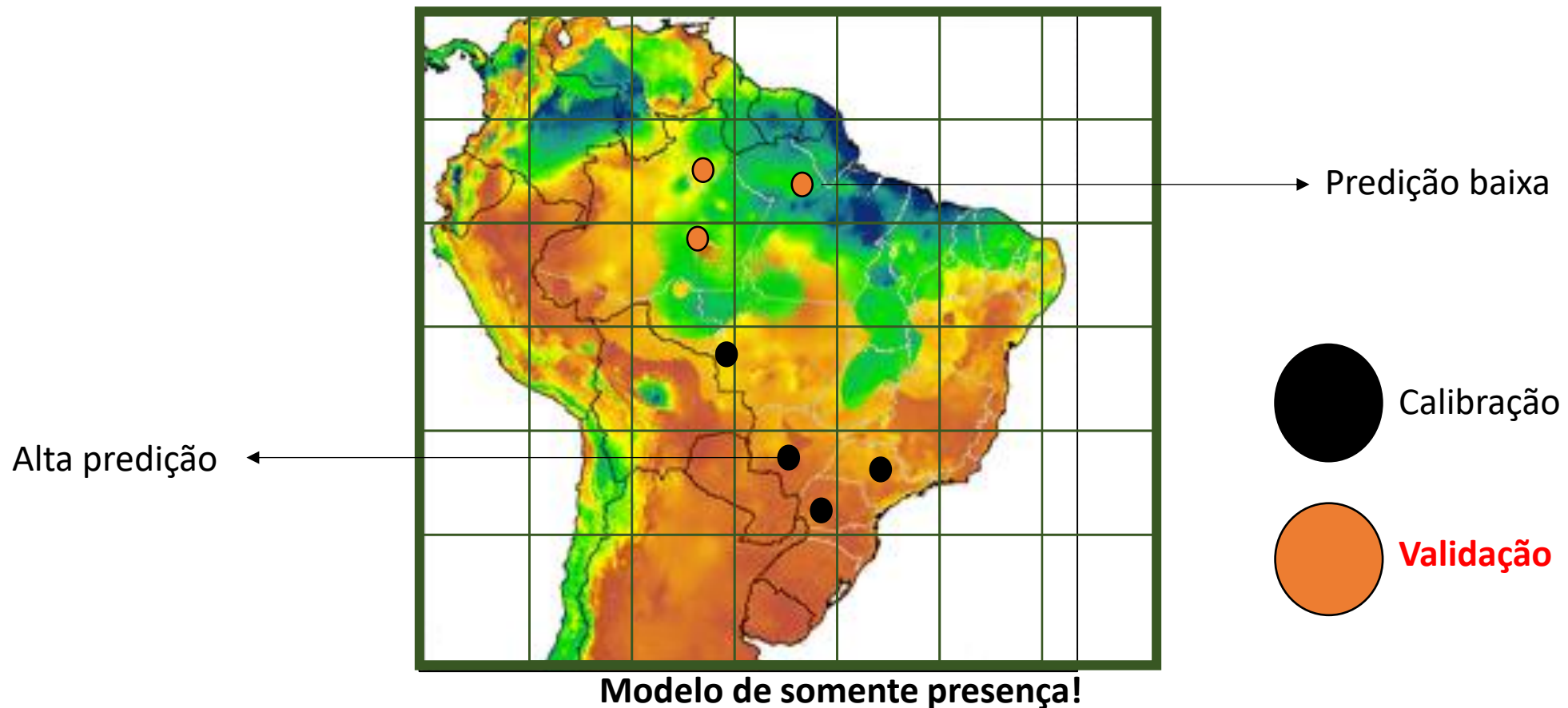
?



Modelo de somente presença!

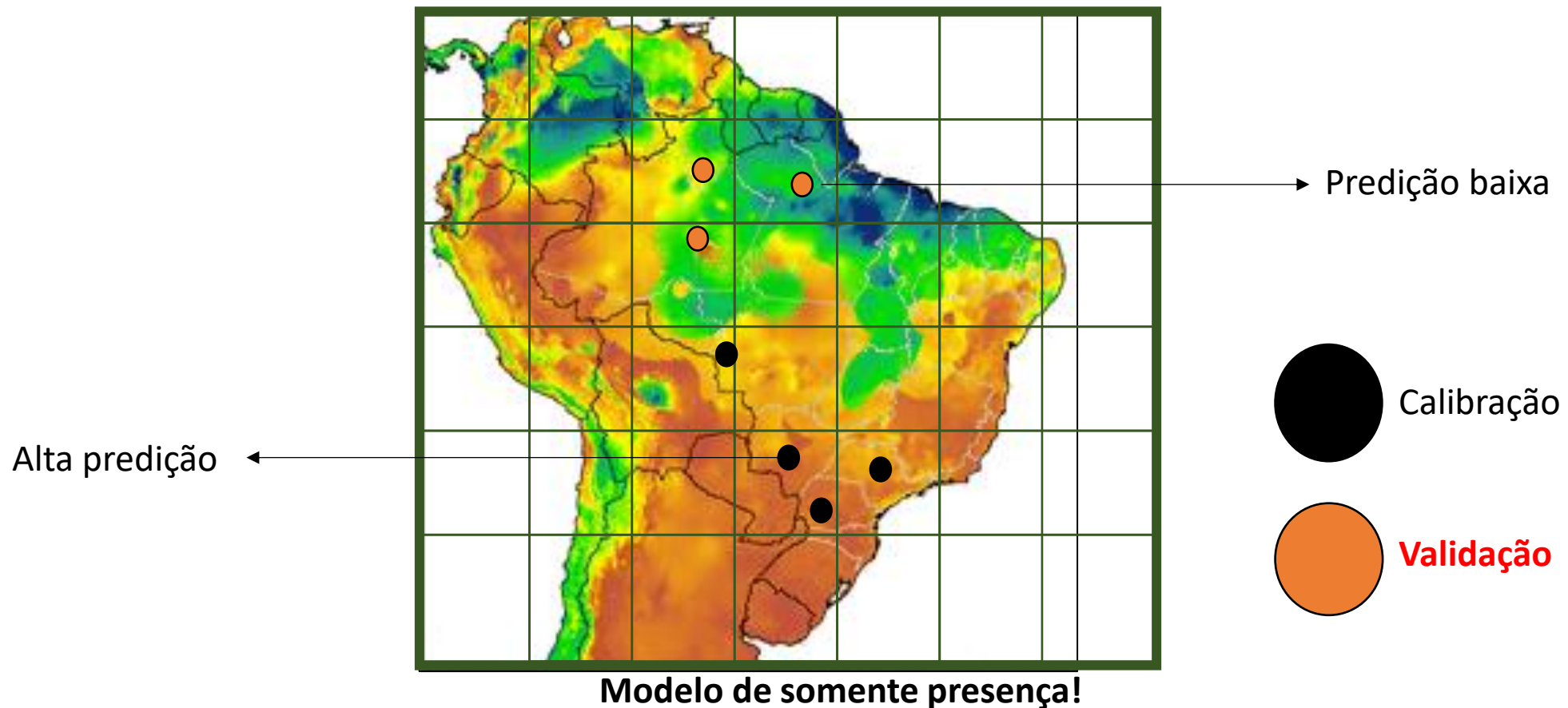
Mas qual tipo de modelagem é melhor?

Validação vs. avaliação



Mas qual tipo de modelagem é melhor?

Validação vs. avaliação



Mas qual tipo de modelagem é melhor?

Validação vs. avaliação

Estipula um “limiar” (threshold)

Sensibilidade (% de acertos do modelo)

$$4/7 = 0.5714 * 100\% = 57,14\%$$

Omissão (% de erros do modelo)

$$3/7 = 0.4286 * 100\% = 42,86\%$$

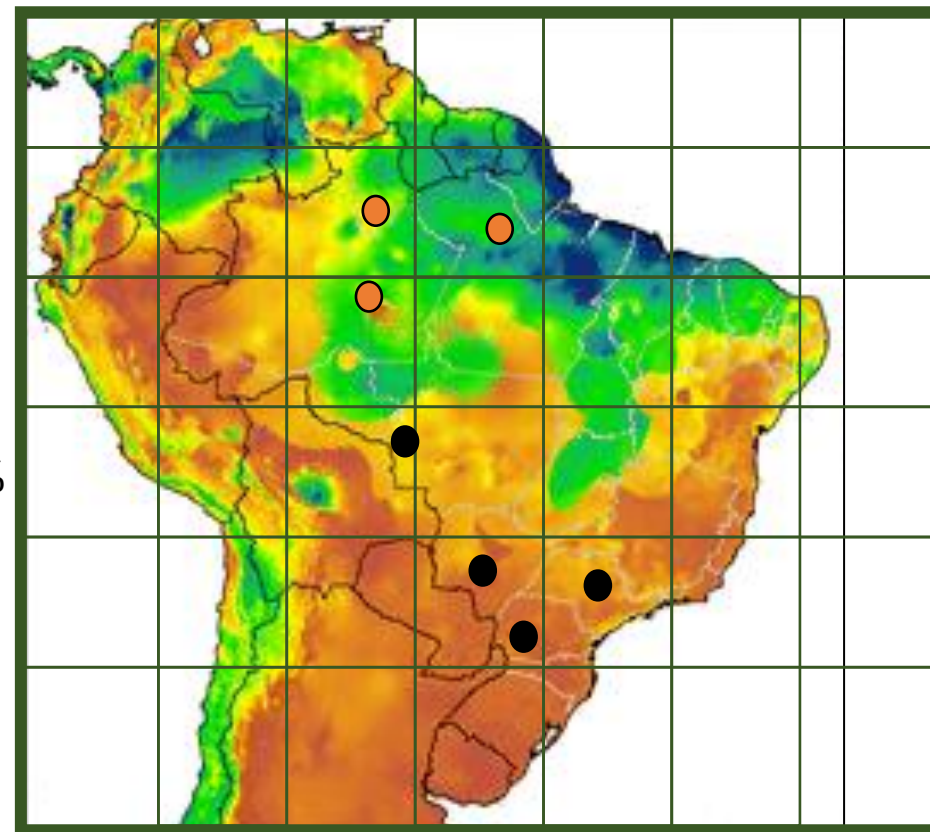
Sensibilidade + omissão = 100%



Calibração



Validação

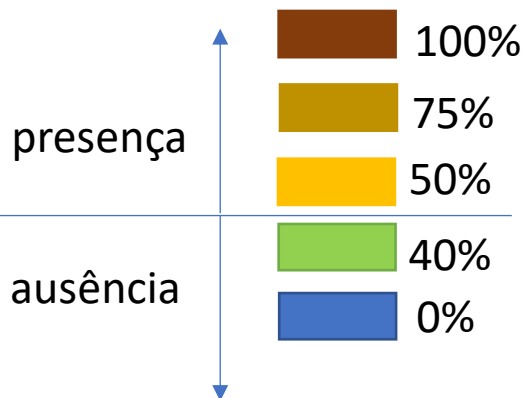


Modelo de somente presença!

H0: Modelo é igual a um modelo nulo



P-valor ~ Distribuição binomial



Presença mínima da espécie

Mas qual tipo de modelagem é melhor?

Validação vs. avaliação



Não temos tempo... Mas existem outras métricas de validação do modelo – tudo vai depender da natureza do seu modelo!

Curva ROC/índice AUC **Independente de limiar!**

Índice TSS

Índice KAPA

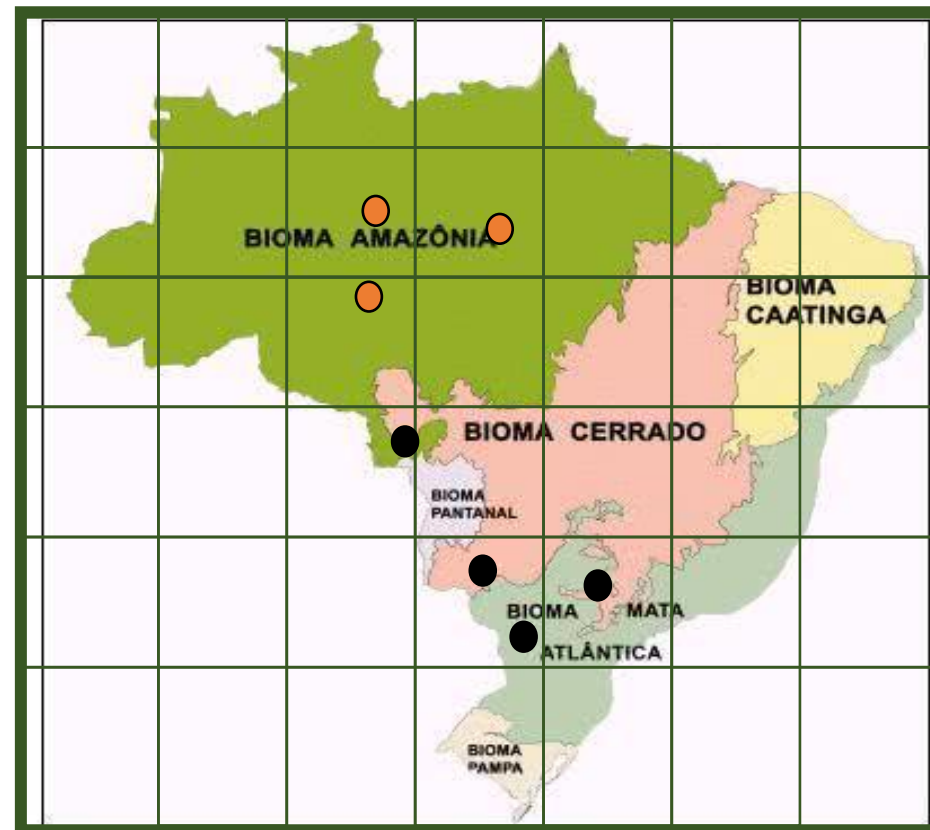
Outros tipos de teste estatísticos

Outros..

Mas qual tipo de modelagem é melhor?

Validação vs. avaliação

?



Calibração



Validação



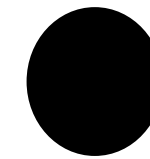
n

Modelo de somente presença!

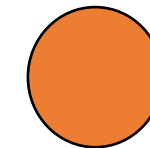
Mas qual tipo de modelagem é melhor?

Validação vs. **avaliação**

?



Calibração



Avaliação!!

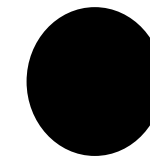
**Mesmas métricas, mas
agora elas avaliam e não
validam!**

Modelo de somente presença!

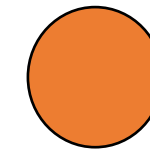
Mas qual tipo de modelagem é melhor?

Validação vs. **avaliação**

?



Calibração



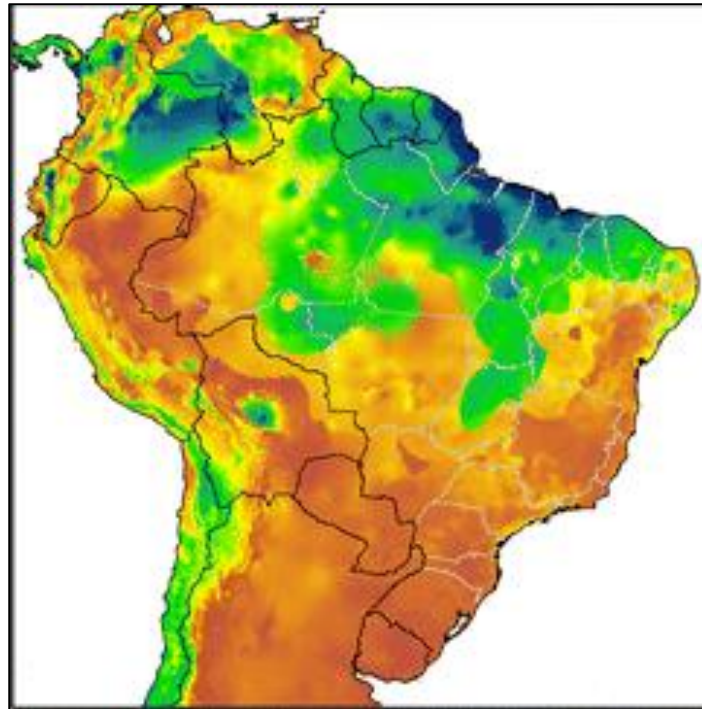
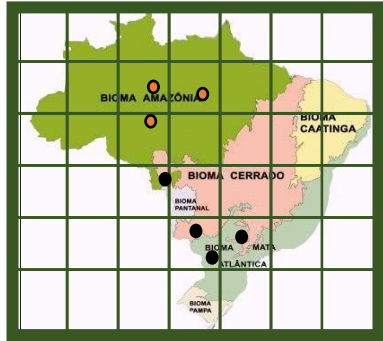
Avaliação!!

**Mesmas métricas, mas
agora elas avaliam e não
validam!**

Modelo de somente presença!

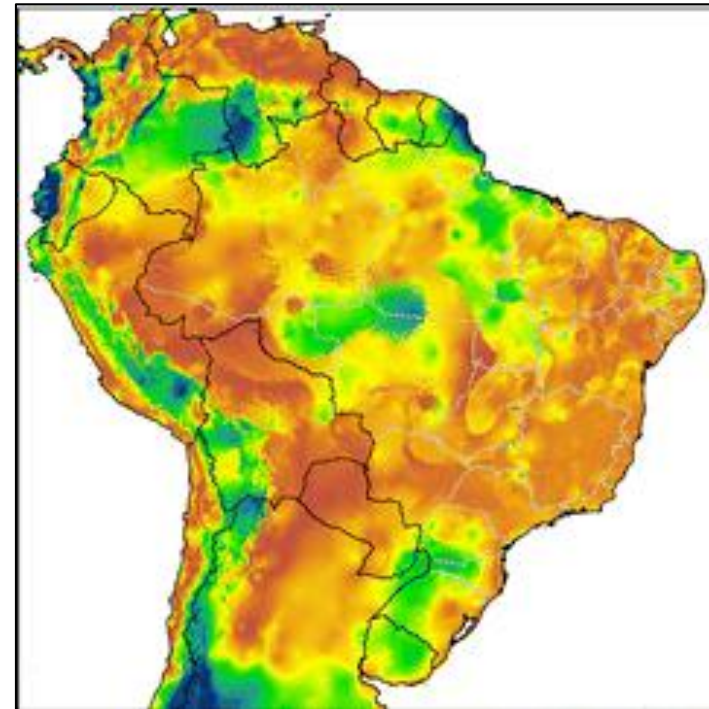
Mas qual tipo de modelagem é melhor?

Validação vs. **avaliação**



Réplica 1

$$5/7 = 0.7143 * 100 = 71,43\%$$



Réplica 2

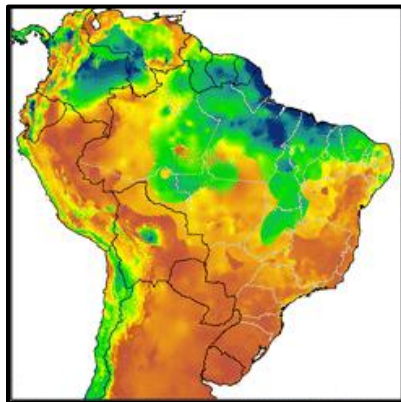
$$7/7 = 1 * 100\% = 100\%$$

Sensibilidade

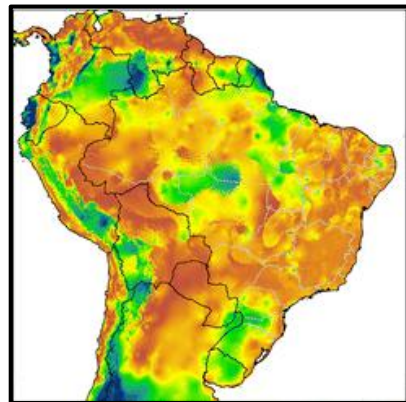


Mas qual tipo de modelagem é melhor?

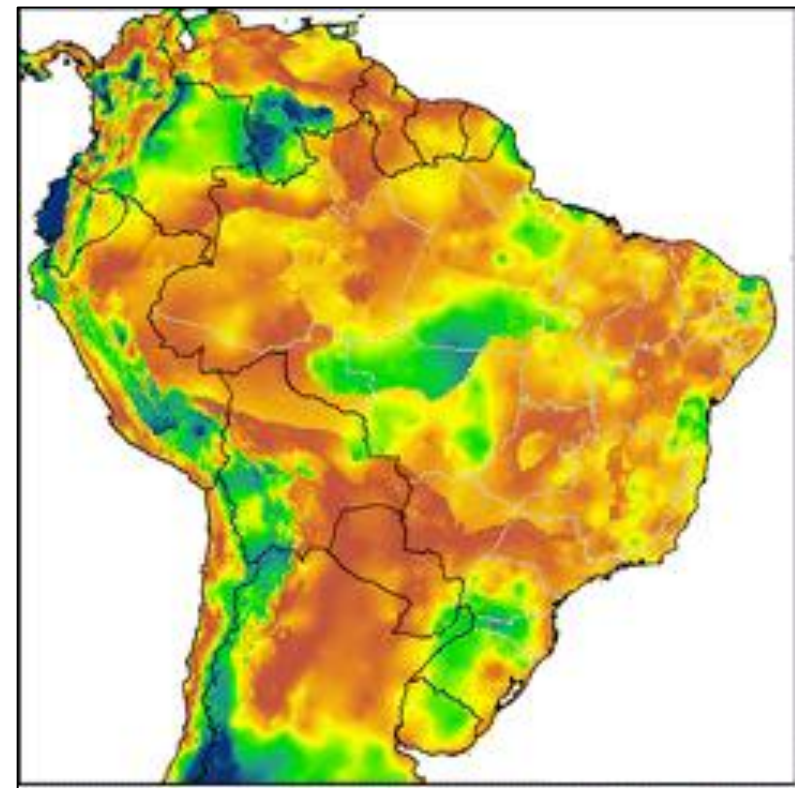
Validação vs. **avaliação**



+



=
Ensemble



2

$$(71,43\% + 100\%) / 2 = 85,71\%$$

Sensibilidade

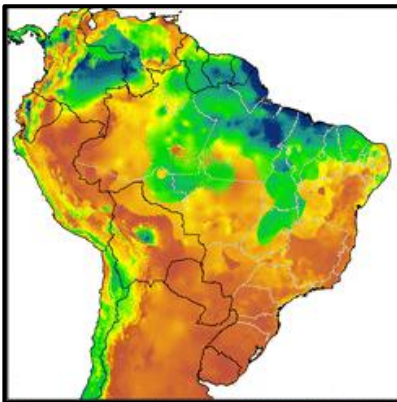
Mapa final!

Mas qual tipo de modelagem é melhor?

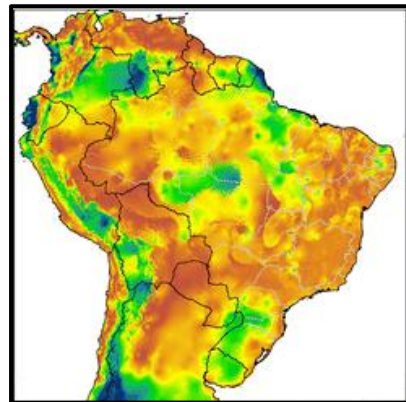
Dois caminhos

Seleciona o “melhor modelo a partir das métricas”

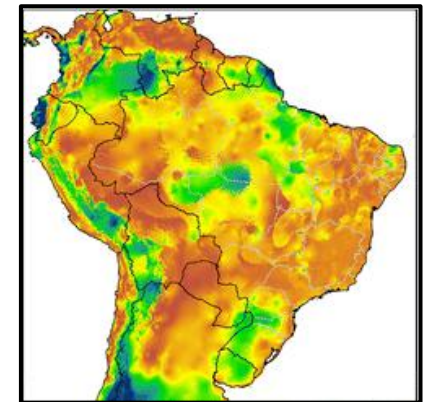
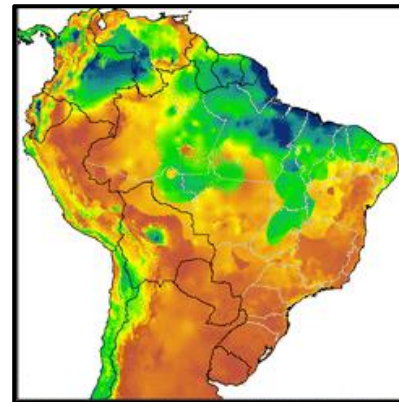
Faz o ensemble e média das métricas para comprovar a adequabilidade do modelo



ou



+





Obrigado