



TUTORIAL DE MODELOS DE DISTRIBUIÇÃO DE ESPÉCIES: GUIA PRÁTICO USANDO O MAXENT E O ARCGIS 10

Jeronymo Dalapicolla¹

¹E-mail: jdalapicolla@usp.br

APRESENTAÇÃO

Esse tutorial foi feito em novembro de 2015, durante a disciplina *LCF5883 – Modelagem e Distribuição de Espécies para a Conservação da Biodiversidade* na Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ/USP) em Piracicaba-SP. Essa disciplina foi lecionada pelos professores: Dra. Katia M. P. M. de Barros Ferraz (ESALQ/USP), Dr. Milton Cezar Ribeiro (UNESP/Rio Claro) e a doutoranda Flávia Pinto (UNESP/Rio Claro). Além disso esse material conta com dicas de Alan Braz (Universidade Federal do Rio de Janeiro), Ana Carolina Loss (Universidade Federal do Espírito Santo), João Paulo Hoope (Universidade Federal do Espírito Santo) e algumas imagens de Flávia Pinto (UNESP/Rio Claro) e de artigos clássicos da área.

As técnicas de modelagem são novas e muito dinâmicas, assim as metodologias usadas aqui podem vir a ser consideradas antiquadas no futuro. O intuito desse material é ajudar as pessoas que estão começando nessa área. Há um tutorial com uma explicação teórica sobre a modelagem de distribuição de espécies e os principais problemas na hora de produzir os modelos. Esse tutorial será referido aqui como a parte 1 do tutorial. É aconselhado ler esse tutorial teórico antes de iniciar este guia prático. Todos os passos foram feitos no Windows 7 e com o ArcGIS 10.1, não há versão do ArcGIS para IOS.

Esse material será atualizado periodicamente por mim, dentro do possível, acrescentando técnicas novas, corrigindo erros de digitação e problemas metodológicos que surgirem, por isso o *feedback* é tão importante. Qualquer dúvida, esclarecimento ou sugestão podem me escrever. A intenção no futuro é trocar o ArcGIS pelo QGIS e construir os modelos dentro do ambiente R. Espero que esse material seja útil a vocês como foi para mim.

Jeronymo Dalapicolla.



SUMÁRIO

I) Programas utilizados no tutorial.....	4
II) Dicas para melhorar o desempenho do computador	4
1) Mudar a região do computador	4
2) Padronizar os nomes de pastas e arquivos	4
3) Local para salvar os projetos.....	5
4) Projeção dos modelos	6
III) Pré-Análise.....	8
1) Aquisição e Qualidade dos Dados Bióticos	8
a) Eliminar pontos da mesma localidade	10
b) Eliminar outliers	10
c) Eliminar problemas taxonômicos.....	10
d) Eliminar pontos históricos	10
e) Eliminar pontos de sede de município.....	11
2) Aquisição de Dados Abióticos	15
3) Criar uma máscara da área de estudo/background	18
a) Criar uma máscara a partir de um shapefile	18
b) Criar área de amortecimento para os limites da máscara.....	20
c) Desenhar à mão a própria máscara	22
4) Padronizar os dados abióticos.....	26
a) Cortar variáveis ambientais.....	26
b) Reamostrar a resolução de uma camada	27
c) Converter as camadas ambientais.....	28
d) Cortar, reamostrar e converter várias variáveis ambientais ao mesmo tempo	29
e) Alinhamento das variáveis ambientais	32
5) Eliminar dados abióticos redundantes	34
a) Calcular a correlação entre as variáveis ambientais	34
b) Escolher as variáveis ambientais	36
6) Eliminar dados bióticos redundantes.....	39
a) Rarefação Simples	39
b) Rarefação por Heterogeneidade Ambiental	40
7) Análise do Espaço Ambiental	44
a) Seleção dos pontos de background por Bias Grid ou Bias File	44
b) Estimar o Espaço Ambiental.....	48
IV) Modelagem	50
1) O programa MaxEnt.....	50



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



a) Checklist de arquivos necessários.....	50
b) Parâmetros do MaxEnt.....	50
2) Erros mais comuns ao executar o MaxEnt.....	54
a) Tamanho do pixel (<i>Cell size</i>) fora do padrão	54
b) Número de linhas e colunas ou alinhamento diferentes.....	54
c) Pontos fora das camadas ambientais	54
d) Caminhos longo demais.....	54
e) Não foi possível achar uma variável ambiental quando está projetando	54
V) Pós-Análise	55
1) Verificar a qualidade do modelo estatisticamente.....	55
2) Cortar o modelo pelo threshold	58
3) Validação externa	60
4) Categorização da área adequada	61
5) Calcular o tamanho da área adequada de um modelo.....	63
6) Cálculo da matriz de confusão e de outros índices estatísticos	66
7) Gráficos de probabilidade de ocorrência para as variáveis ambientais.....	71
VI) Referências.....	73
VII) Apêndices.....	74



I) Programas utilizados no tutorial

Os programas que serão usados durante esse tutorial são os seguintes:

ArcGIS 10.1 ou superior: com as toolboxes *Analysis Tools*, *Spatial Analyst Tools*, *Conversion Tools*, *Data Management Tools*, *SDMToolsbox* (<http://sdmtoolbox.org/>; Brown 2014) e *LEEC_SDM_Toolbox* (propriedade intelectual do Dr. Milton Cezar Ribeiro (UNESP/Rio Claro), é necessário autorização prévia para usar essa toolbox). Alternativas ao LEEC_SDM serão apresentadas quando necessário.

MaxEnt 3.3.3k

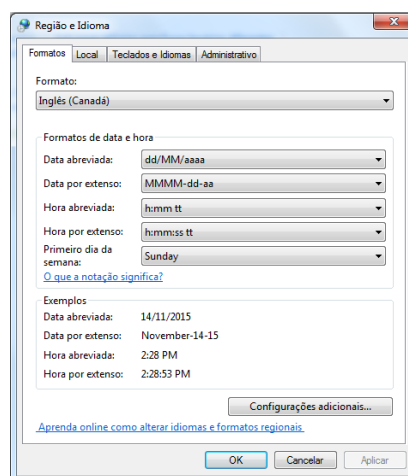
Java: Atualizado para a última versão disponível.

RStudio ou R: com os pacotes “raster”, “rgdal” e “vegan”.

II) Dicas para melhorar o desempenho do computador

1) Mudar a região do computador: No português o separador das casas decimais é a vírgula, o separador de milhar é ponto e o separador de colunas nos arquivos .csv é o ponto-e-vírgula. No inglês são respectivamente o ponto, a vírgula e a vírgula. Os arquivos lidos e os *output* dos programas são gerados respeitando os separadores em inglês. Para evitar a necessidade de substituir ponto por vírgula e substituir ponto-e-vírgula por vírgula em vários arquivos é recomendado mudar a região do computador. Esse processo é melhor do que apenas mudar o separador de decimal do Excel (que não corrige a leituras dos arquivos .csv). Para mudar a região do seu computador utilizando Windows 7:

INICIAR >> PAINEL DE CONTROLE >> Em **exibir na forma de categoria**, procure a categoria **RELÓGIO, IDIOMA E REGIÃO >> REGIÃO E IDIOMA >>** Na aba **FORMATOS**, escolha qualquer formato em inglês na caixa de diálogo **FORMATO**, no exemplo abaixo foi escolhido **INGLÊS (CANADÁ) >> OK**.



2) Padronizar os nomes de pastas e arquivos: Sempre que for colocar nome em pastas e arquivos **NUNCA** utilize espaços e caracteres não reconhecidos pelo inglês como acentos gráficos, vírgulas, pontos ou cedilha. Alguns comandos podem dar erros por causa desses caracteres.



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



Lembre-se que os arquivos em formato *GRID* (formato gerado e lido pelo ArcGIS) terão no máximo 13 caracteres, não gaste esse espaço colocando sublinhados/*underlines*. Sempre padronize as siglas, por exemplo, *Nectomys* = **nec**; presente = **pres**; último glacial = **lgm**; mapa binário = **bi**. Crie um arquivo *.txt* de metadados para lembrar o que significa cada sigla.

3) Local para salvar os projetos: Sempre escolha o caminho mais curto para salvar e executar as análises no ArcGIS e no MaxEnt, na maioria dos casos os erros nas análises são porque os arquivos são salvos em subpastas do “*Meus Documentos*” e o programa não consegue lê-los. Uma dica: salve tudo na raiz do computador, no **C:** ou próximo a ele. Eu tenho duas pastas para modelagem no C:, uma chamada **LAYERS** com as camadas ambientais brutas, sem modificações e com os *shapefiles* (arquivos para a criação de mapas) e as pastas dentro dela estão organizadas por região de corte. A outra pasta no C: é para cada espécie que eu for modelar, com seis subpastas **LAYERS** (para as camadas ambientais em formato *.asc* e já prontas para modelar; **OUTPUT** (onde o MaxEnt vai salvar os resultados); **SAMPLES** (onde fica o arquivo *.csv* com os pontos de ocorrência pronto para ser usado no Maxent); **ESCOLHA_CAMADAS** (onde salvo os resultados dos testes de correlação e PCA para a escolha das camadas ambientais); **RAREFACAO_PONTOS** (onde guardo os resultados das análises de redução do viés amostral) e **MASCARA** (onde tem a forma do corte das camadas ambientais, um *shapefile* que representa a minha área de interesse). Isso é um exemplo, você pode criar a melhor maneira de organizar seus arquivos, mas não caia na besteira de colocar todos esses arquivos na mesma pasta, pois depois nunca se lembrará o que cada um representa. Lembre-se que nos arquivos *GRID*, o nome só pode conter 13 caracteres.

Exemplo:

C:

LAYERS

BRASIL

Uso_do_solo.tif

Hidrografia_ANA.tif

NEOTROPICAL

Politico_paises.shp

MUNDIAL

WORLDCLIM_PRES_30S

WORLDCLIM_LGM_2_5_MIN_MIROC

C:

NECTOMYS_SQUAMIPES

MASCARA_CORTE_CAMADAS

LAYERS

PRESENTE

bio_1.asc

LGM_MIROC



OUTPUT

PRESENTE

LGM_MIROC

RAREFACAO_PONTOS

SIMPLES_10KM

HETEROG_10KM

ESCOLHA_CAMADAS

SAMPLES

`samples_nec_maxent.csv`

4) Projeção dos modelos: O ArcGIS permite vários tipos de projeção (*datum*) para os mapas e camadas ambientais. Se arquivos com diferentes projeções (*datum*) forem usados na mesma análise, provavelmente a análise não funcionará. Uma das soluções é mudar a projeção de cada arquivo, só que é cansativo e tem um problema: se perde a projeção original do arquivo. Uma opção é predefinir uma projeção para o projeto inicial antes de carregar qualquer arquivo, isso não muda o dado original, apenas converte momentaneamente os mapas e camadas que forem incluídos no projeto para a projeção predefinida para que as análises funcionem. Para isso siga os passos:

Abra o ArcGIS 10.1 >> aba **TABLE OF CONTENTS** >> clique com o botão direito do mouse em **LAYERS** >> **PROPERTIES** >> aba **COORDINATE SYSTEM** >> Escolha uma projeção de interesse e dê **OK**.

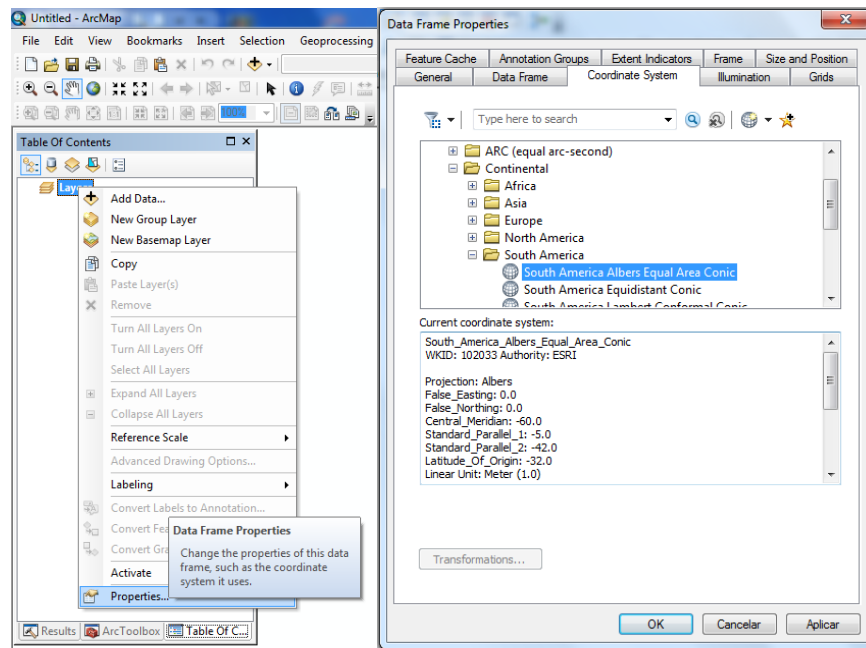
As projeções dentro de **GEOGRAPHIC COORDINATE SYSTEMS (GCS)** dão as projeções em coordenadas geográficas e as projeções dentro de **PROJECTED COORDENATE SYSTEMS (PCS)** dão as projeções em distâncias (metros, km, jardas).

As projeções mais usadas em GCS são: **GEOGRAPHIC COORDINATE SYSTEMS >> WORLD >> WGS1984** ou **GEOGRAPHIC COORDINATE SYSTEMS >> WORLD >> SOUTH AMERICA >> SAD 1969**. A projeção de PCS recomendada é a Albers: **PROJECTED COORDENATE SYSTEMS >> CONTINENTAL >> SOUTH AMERICA ALBERS EQUAL AREA CONIC**.



LaMaB: Laboratório de Mastozoologia e Biogeografia

Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



Dalapiccola, J. 2016. Tutorial de modelos de distribuição de espécies: guia prático usando o MaxEnt e o ArcGIS 10. Laboratório de Mastozoologia e Biogeografia, Universidade Federal do Espírito Santo, Vitória. Disponível em: <http://blog.ufes.br/lamab/tutoriais>



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



III) Pré-Análise

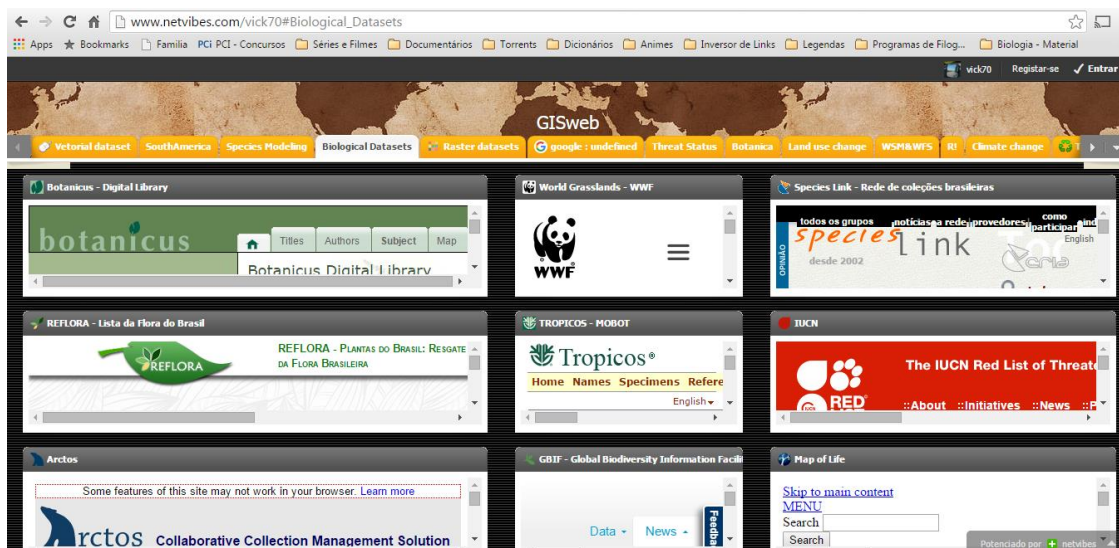
1) Aquisição e Qualidade dos Dados Bióticos

Os pontos podem ser coletados em coleções científicas, museus, literatura científica especializada, bancos de dados *online* ou podem ser cedidos por pesquisadores em campo. Os principais bancos de dados online são:

- **GBIF** (Global Biodiversity Information Facility, www.gbif.org);
- **SpeciesLink** (www.splink.cria.org.br);
- **VertNet** (VertNet: Distributed Databases With Backbone, <http://vertnet.org/>);
- **Arctos** (Collaborative Collection Management Solution, <http://arctos.database.museum/>).

Há um site montado e mantido pela pesquisadora Flávia Pinto do Laboratório de Ecologia Espacial e Conservação da UNESP/Rio Claro que reúne todas as informações disponíveis sobre modelagem em um único lugar. Lá se encontra vários arquivos interessantes como uma lista dos locais na internet onde você coletar dados bióticos:

- www.netvibes.com/vick70



É necessário organizar os pontos encontrados em planilhas. Algumas colunas são indispensáveis. Lembrando que o separador decimal da planilha deverá ser o ponto (Veja como mudar isso em **DICAS PARA MELHORAR O DESEMPENHO DO COMPUTADOR** no início do tutorial). Segue o exemplo de uma planilha usada para dados de *Nectomys squamipes*:

UFES_MAM							
A	B	C	D	E	F	G	H
1	Longitude	Latitude	Ano de Coleta	Fonte Primária	Autores da Fonte Primária	Fonte Secundária	Autores da Fonte Secundária
7	-38.90000	-15.85000	2001	UFMG	Moura	2096	Fazenda Taquara
8	-39.03333	-13.51667	2000	UFMG	Moura	2095	Fazenda Subauma
9	-39.49583	-15.41917	2008	MBML-MAMIFEROS	E.H.Barros; J.Dalla & A.Durão	2876	Gasoduto Cacimbas- Catu- BR
13	-39.66667	-13.85000	2001	UFMG	Moura	2097	Fazenda Alto São Roque

Dalapicolla, J. 2016. Tutorial de modelos de distribuição de espécies: guia prático usando o MaxEnt e o ArcGIS 10. Laboratório de Mastozoologia e Biogeografia, Universidade Federal do Espírito Santo, Vitória. Disponível em: <http://blog.ufes.br/lamab/tutoriais>



LaMaB: Laboratório de Mastozoologia e Biogeografia

Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



	H	I	J	K	L	M	N	O	P	Q	R
	Localidade	Município	Estado	País	Tipo de Coordenada	Fonte da Coordenada	Observações				
7	Fazenda Taquara	Belmonte	Bahia	Brasil	Sede do Município	Geoloc					
8	Fazenda Subauma	Cairu	Bahia	Brasil	Localidade Específica	Autores					
9	Gasoduto Cacimbas - Catu - BR	Camacan	Bahia	Brasil	Sede do Município	Autores					

A planilha tem 14 colunas:

- **LONGITUDE:** Em graus decimais. É interessante que a longitude esteja primeiro que a latitude porque o ArcGIS considera a primeira coluna sempre como a longitude;
- **LATITUDE:** Em graus decimais;
- **ANO DE COLETA:** Ano que o espécime foi coletado. Cuidado, não é o ano do trabalho publicado, é o ano da coleta, a maioria dos trabalhos não traz essa informação. É importante para separar dados históricos e atuais;
- **FONTE PRIMÁRIA:** Onde você adquiriu a informação, museu, artigo, tese. É a sua fonte primária;
- **AUTORES DA FONTE PRIMÁRIA:** Coletor, autores dos artigos, etc;
- **FONTE SECUNDÁRIA:** Se sua fonte primária cita outro artigo para a informação da localidade, ou o número de tombo do museu;
- **AUTORES DA FONTE SECUNDÁRIA:** Coletor, autores dos artigos da fonte secundária;
- **LOCALIDADE:** localidade informada pela fonte primária ou secundária;
- **MUNICÍPIO:** Município ou equivalente;
- **ESTADO:** estado ou equivalente;
- **PAÍS:** país.
- **TIPO DE COORDENADA:** Se a coordenada é de uma localidade específica ou uma sede de município;
- **FONTE DA COORDENADA:** Se a coordenada foi informada pelos autores das fontes ou você que pesquisou em outros *gazetteers*;
- **OBSERVAÇÕES:** Qualquer informação que ache relevante e que não foi contemplada nas colunas anteriores.

É comum encontrarmos localidades específicas no nível de Unidade de Conservação, Fazenda Santa Maria ou 4km SE de tal cidade, mas sem a coordenada. É possível encontrar a coordenada utilizando *gazetteers* como:

- <http://www.fallingrain.com/world/>
- <http://www.heavens-above.com/countries.asp#B>
- http://www.getty.edu/research/conducting_research/vocabularies/tgn/index.html
- <http://www.tageo.com/index.htm>
- http://earth-info.nga.mil/gns/html/cntry_files.html#B



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



As coordenadas podem ser encontradas, também, através do:

- **Google Earth:** <http://earth.google.com/>
- **Google Maps:** <https://www.google.com.br/maps>
- **GeoLoc:** <http://splink.cria.org.br/geoloc>

Também é comum encontrarmos as coordenadas em UTM ou em graus, minutos e segundos. É necessário convertê-las em graus decimais por meio desses sites, em especial a do *speciesLink*:

- <http://splink.cria.org.br/conversor?criaLANG=pt>
- <http://www.cellspark.com/UTM.html>
- <http://home.hiwaay.net/~taylorc/toolbox/geography/geoutm.html>
- http://gis.wvdep.org/convert/llutm_conus.php
- <http://www.fcc.gov/mb/audio/bickel/DDDMSS-decimal.html>

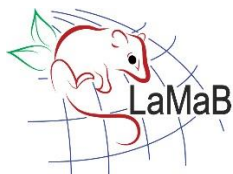
TODOS os pontos de ocorrência coletados devem ser conferidos um a um. É recomendado o uso o *GoogleMaps* (<https://www.google.com.br/maps>), *GoogleEarth* (<http://earth.google.com/>) ou a ferramenta *Info XY* do *speciesLink* (<http://splink.cria.org.br/infoxy?criaLANG=pt>) para verificar se as coordenadas batem com as localidades indicadas. **ATENÇÃO!** No *GoogleMaps* e *GoogleEarth* a latitude vem primeiro que a longitude na hora de pesquisar uma localidade. Lembre-se de inverter as posições das mesmas. Após conferir todas as coordenadas é hora de eliminar os pontos redundantes:

a) *Eliminar pontos da mesma localidade:* Classifique as células em valores decrescentes de latitude ou longitude e veja quais pontos tem as mesmas coordenadas e fique com apenas um deles. Esse processo pode ser automatizado no ArcGIS por meio do *SDMToolbox* (ver adiante);

b) *Eliminar outliers:* Projete seus pontos em uma mapa no ArcGIS ou QGIS e elimine da planilha os pontos que caírem no mar ou muito fora da distribuição conhecida da espécie-alvo, por exemplo, a espécie que tenho interesse de modelar é da Mata Atlântica e tem um ponto no Cerrado ou Caatinga, mesmo que você tenha certeza que é da sua espécie, esse ponto pode indicar uma migração ou um movimento do indivíduo e pode não representar a distribuição da espécie-alvo. Melhor seria descartar o ponto. A não ser que a espécie ocorra também no Cerrado.

c) *Eliminar problemas taxonômicos:* Grupos com taxonomia instável como pequenos mamíferos, anuros, invertebrados e alguns grupos de plantas podem estar com a nomenclatura antiga nos bancos de dados *online* e museus. Tente fazer a busca pelos seus sinônimos. Se uma das populações de uma espécie foi elevada ao nível específico e os dois nomes são válidos, cuidado! Desconfie de pontos indicados por trabalhos que não tenham amostras genéticas ou que não tenham analisado os espécimes-testemunho.

d) *Eliminar pontos históricos:* É comum encontrarmos pontos de coletas bem mais antigos (do século XIX ou início do XX). Os dados das camadas ambientais do WorldClim foram construídas com dados de 1950 a 2000, então registros de coletas abaixo de 1950 não são recomendados. E com pequenos mamíferos sempre é bom utilizar pontos mais recentes que a última revisão taxonômica.



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



No caso de *Nectomys squamipes*, a última revisão é de Bonvicino 1994, dada que coincide com a difusão de novas tecnologias como GPS. Meu recorte temporal dos pontos de ocorrência foi de 1994 a 2013. Se você for acrescentar camadas ambientais que representam mudanças nas paisagens (uso do solo, cobertura vegetal), então seu recorte temporal não pode ser muito grande, porque as paisagens mudam muito, e um ponto de 1950 com certeza não vai representar a paisagem de 2015. Não há regras para definir o recorte temporal do modelo, mas é correto indicar no trabalho qual recorte está usando e porquê.

e) *Eliminar pontos de sede de município*: Vários registros de ocorrência serão da sede do município da localidade. Se mesmo utilizando os *gazetteers* não foi possível chegar à coordenada específica, retire esses pontos das análises.

Depois de todo esse trabalho veja a quantidade de pontos que restaram na planilha, se for entre 20 e 50 pontos você tem uma boa quantidade de registros. Se você tem menos de 14 pontos considere recolocar alguns pontos históricos e pontos de sede de municípios com pequena área de extensão (o erro da coordenada será menor) para aumentar o número de pontos pelo menos até 14 registros. Se mesmo assim não foi possível, não se preocupe o MAXENT também é bom trabalhando com poucos registros de ocorrência, até cerca de 8 registros, porém os parâmetros do MaxEnt mudam quando se tem poucos pontos (de 8 a 14 pontos, veja adiante). Se você tem menos que 8 pontos então o MaxEnt não é recomendado, é melhor modelar com o algoritmo da Distância Ambiental (*Environmental Distance* ou *ED*) no programa *OpenModeller* ou com *Random Forest*, mas isso não será discutido nesse tutorial.

Se você possuir bastantes pontos (40 ou mais) é interessante separar alguns pontos para o teste externo de validação (Ver a parte 1 desse tutorial). Uma dica: use os pontos que você não tem muito confiança para a validação externa, por exemplo, aqueles registros que as coordenadas não foram indicadas pelos autores ou coletores originais e que você as achou em *gazetteers* ou no *GeoLoc*. Por exemplo, eu tinha uma localidade, uma Unidade de Conservação (UC) no Espírito Santo, que estava georreferenciada pela sede do município. Contudo a UC é pequena e consegui a localização dela pelo *GeoLoc*, conferi no *GoogleMaps* e bate com a sede da UC. É recomendável que eu não utilize esse ponto para gerar o modelo, já que as coordenadas não são específicas. Contudo não descarto esse ponto, como estou fazendo a modelagem numa resolução de 5 km² (Veja adiante) o erro dessa localidade é pequeno e posso usá-la para a validação externa do modelo.

Enfim só use ótimos pontos para a modelagem, antigamente (menos de 10 anos) acreditava-se que muitos pontos eram necessários para a criação dos modelos. Hoje estudos mostram que poucos pontos com zero de erro no georreferenciamento são melhores que muitos pontos com um pouco de erro. A modelagem pode ser feita no MaxEnt com até 8 pontos. Não se preocupe se você ficar com 15 ou 20 pontos depois desses procedimentos. Contudo, esses poucos pontos devem representar toda a distribuição conhecida da espécie-alvo.



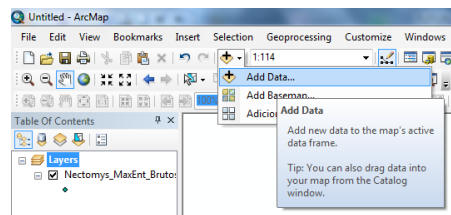
LaMaB: Laboratório de Mastozoologia e Biogeografia

Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo

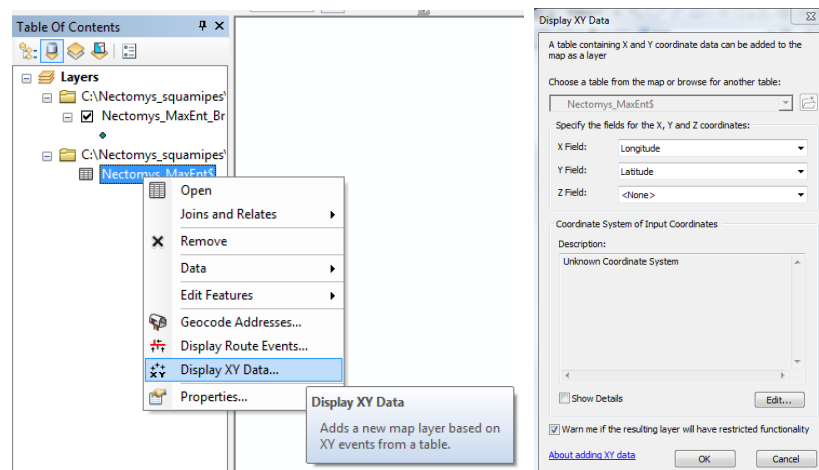


Neste ponto do tutorial você deve ter um banco de dados confiável e de qualidade, que possa ser usado no MaxEnt. Se você encontrou muitos registros da espécie-alvo então você também deve ter um outro banco de dados, com pontos de ocorrência **DIFERENTES DAQUELES QUE SERÃO USADOS NO MAXENT** e com pouco erro no georreferenciamento. Esse banco de dado será usado na Validação Externa. Contudo ainda há viés amostral e dados redundantes nestas planilhas. Para retirá-los leia os próximos tópicos. O último passo antes de trabalhar com as camadas ambientais é criar um *shapefile* dos seus pontos de ocorrência.

Abra o **ArcGIS** e importe a tabela de pontos final em **ADD DATA**.



Carregue a tabela no **TABLE OF CONTENTS**, e depois clique com o botão direito do *mouse* nela e escolha **DISPLAY XY DATA**.

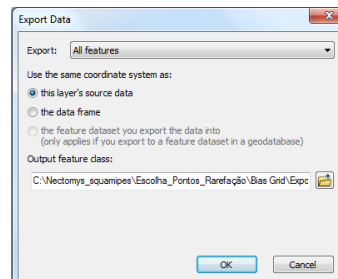


Escolha os campos de latitude em **Y FIELD** e longitude em **X FIELD** e clique em **EDIT** no campo de coordenada geográfica. Escolha a projeção em: **PROJECTED COORDINATE SYSTEMS >> CONTINENTAL >> SOUTH AMERICA ALBERS EQUAL AREA CONIC >> OK >> OK**.

Salve o *shapefile* com a projeção em Albers (use isso no nome do *shapefile*). Vá ao **TABLE OF CONTENTS >>** clique com o botão direito no *shapefile* de pontos em Albers e escolha **DATA >> EXPORT DATA**.



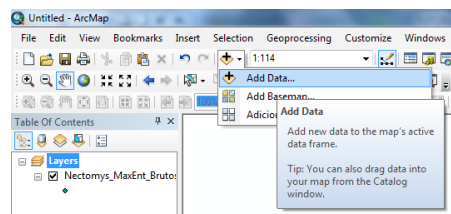
LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



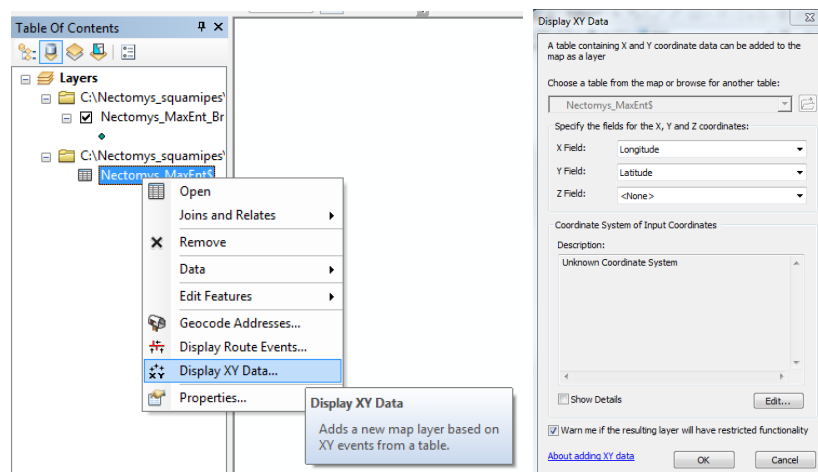
Em **EXPORT** deixe **ALL FEATURES**. Como **COORDINATE SYSTEM** pode deixar como **THIS LAYER'S SOURCE DATA**. Escolha o nome e onde salvar em **OUTPUT FEATURE CLASS** e depois clique **OK**.

Repita esse último passo de criar um *shapefile* dos pontos, mas ao invés de usar a projeção em Albers, crie um *shapefile* também em **WGS84**:

Abra o **ArcGIS** e importe a tabela de pontos final em **ADD DATA**.



Carregue a tabela no **TABLE OF CONTENTS**, e depois clique com o botão direito do *mouse* nela e escolha **DISPLAY XY DATA**.



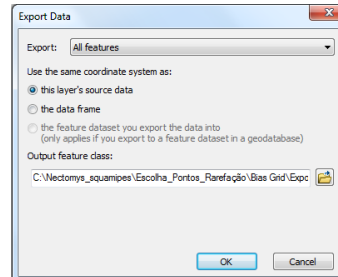
Escolha os campos de latitude em **Y FIELD** e longitude em **X FIELD** e clique em **EDIT** no campo de coordenada geográfica. Escolha a projeção em: **GEOGRAPHIC COORDINATE SYSTEMS >> WORLD >> WGS1984 >> OK >> OK**.



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



Salve o *shapefile* com a projeção em WGS84 (use isso no nome do *shapefile*). Vá ao **TABLE OF CONTENTS** >> clique com o botão direito no *shapefile* de pontos em WGS84 e escolha **DATA** >> **EXPORT DATA**.



Em **EXPORT** deixe **ALL FEATURES**. Como **COORDINATE SYSTEM** pode deixar como **THIS LAYER'S SOURCE DATA**. Escolha o nome e onde salvar em **OUTPUT FEATURE CLASS** e depois clique **OK**.

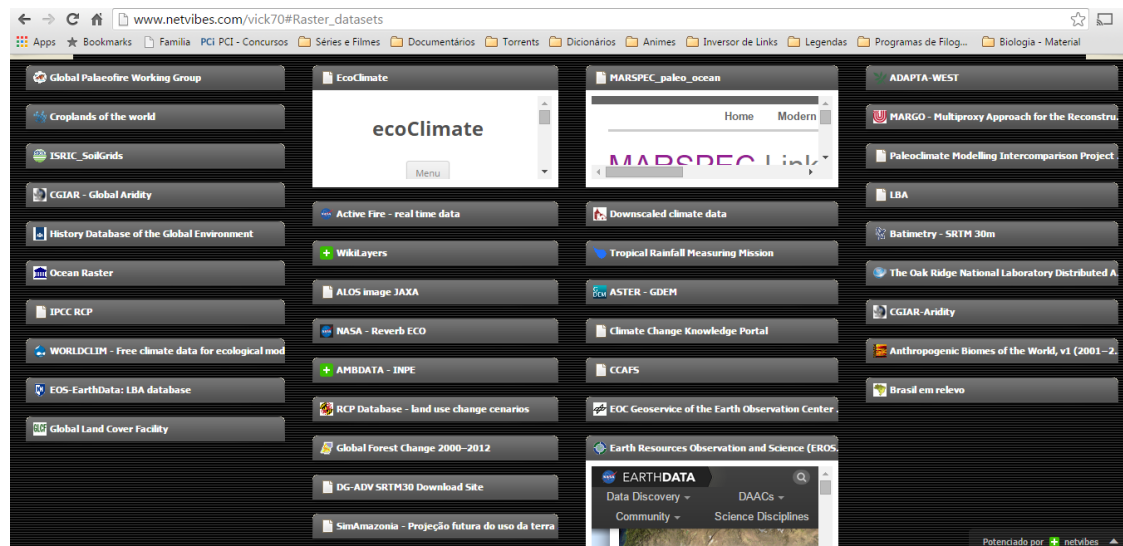
No fim desse processo você terá dois *shapefiles* de seus pontos, em projeções diferentes.



2) Aquisição de Dados Abióticos

É necessário o *download* de camadas ambientais de interesse em *sites* da *internet*. As camadas ambientais estão geralmente em formato *raster* (*GRID*, *TIFF* ou *IMG*). O site da pesquisadora Flávia Pinto possui vários links para *download* de vários tipos de camadas ambientais.

➤ http://www.netvibes.com/vick70#Raster_datasets



Há 19 variáveis climáticas mais utilizadas na modelagem e elas estão disponíveis no *site* do **WorldClim** (www.worldclim.org) de graça para *download*. O portal **ecoClimate** (<http://ecoclimate.org/downloads/>) em destaque na página acima disponibiliza as 19 variáveis climáticas (BioClim) para várias épocas, como a **MODERNA** (1950-1999), a **HISTÓRICA** (1900-1949) e a **PRÉ-INDUSTRIAL** (1760). Se há registro de pontos da espécie-alvo para uma dessas épocas é interessante então utilizar esses pontos históricos para a modelagem naquela época, ou para a validação externa. **IMPORTANTE! TODAS** as camadas devem possuir a mesma resolução, isto é, ter o mesmo *Cell size*, ou tamanho de pixel.

Há quatro tamanhos disponíveis para as camadas do *WorldClim* (10 min, 5 min, 2,5 min e 30 arc seg). Escolha uma resolução adequada para a sua pergunta. Se o problema envolve uma escala espacial pequena (bairro, cidade, estado) a melhor resolução (30 arc seg) é mais indicada. Se for um problema global, as camadas de 10 min são suficientes. Em nível de bioma e continente as camadas de 2,5 min já são suficientes (Gianini *et al.*, 2012).

Lembrando que a o último interglacial (LIG, há 120-140 mil anos) só está disponível a resolução de 30 arc seg. Como você tem que ter a mesma resolução de camadas, se for seu objetivo fazer paleomodelagem e projetar seu modelo para essa data, é necessário fazer o *download* dos dados do presente também em 30 arc seg ou então converter a camada do LIG para a resolução que você estiver usando (a opção mais fácil).

Dalapiccola, J. 2016. Tutorial de modelos de distribuição de espécies: guia prático usando o MaxEnt e o ArcGIS 10. Laboratório de Mastozoologia e Biogeografia, Universidade Federal do Espírito Santo, Vitória. Disponível em: <http://blog.ufes.br/lamab/tutoriais>

Lembrando também que a o último glacial máximo (LGM) 21 mil anos atrás não está disponível a resolução de 30 arc seg. Como você tem que ter a mesma resolução de camadas, se for seu objetivo fazer a transferibilidade do modelo para essa data, é necessário fazer o *download* dos dados do presente também em 2,5 min. Nesse caso você não pode converter a camada do LGM para uma resolução maior, só para uma menor. Para os dados do Holoceno Médio (6 mil anos atrás), há camadas para as mesmas resoluções que o presente. Dessa forma, se seu interesse é a paleomodelagem de todas as épocas, é melhor fazer os modelos com a resolução de 2,5 min e converter a do LIG para essa resolução. É a forma mais parcimoniosa e menos trabalhosa.

As variáveis ambientais do passado são construídas a partir de Modelos de Circulação Geral (*General Circulation Models – GCM*) que usam diferentes algoritmos para representar processos físicos que ocorrem na atmosfera, oceanos, criosfera e terras emersas. Como cada centro de pesquisa tem um algoritmo diferente, existem vários conjuntos de variáveis para a paleomodelagem com diferentes siglas e várias delas estão disponíveis no site do *WorldClim*. Os dados primários para a criação das camadas vêm de duas fontes: CMIP5 e PMIP3. Os GCM mais usados são CCSM4 e MIROC. Varela *et al.* (2015) avaliaram as diferenças e semelhanças entre esses conjuntos de dados. E viram que para uma análise completa é necessário fazer a modelagem com pelo menos quatro desses conjuntos, e se a modelagem for de multiespécies é necessário a exclusão de algumas variáveis ambientais.

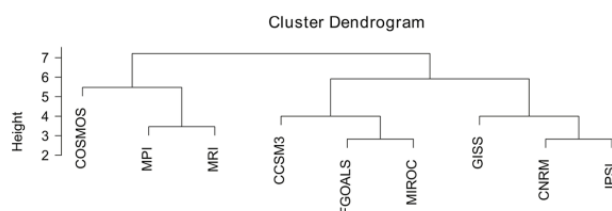


Fig 5. Hierarchical cluster grouping the nine GCMs by the correlation of their predictions for all 19 bioclimatic variables.

doi:10.1371/journal.pone.0129037.g005

O GCM MIROC e CCSM4 são bastantes similares, pois estão no mesmo grupo do dendrograma apresentado por Varela *et al.* (2015), então vocês tem que escolher apenas um deles. O ideal, segundo os autores, é escolher entre um dos 4 grupos de GCM:

GRUPO 1: COSMOS;

GRUPO 2: MPI ou MRI;

GRUPO 3: CCSM4, MIROC ou FGOALS;

GRUPO 4: GISS, CNRM ou IPSL.

Como esse é um procedimento novo, ainda não há consenso de que essa abordagem seja a melhor, mas é interessante perceber que dependendo do GCM escolhido (um dos 4 grupos) e da área modelada os resultados podem ser diferentes. A maioria dos trabalhos escolhem entre MIROC e CCSM4. O MPI também vem ganhando força para trabalhos na América do Sul. Varela et al. (2011)



LaMaB: Laboratório de Mastozoologia e Biogeografia

Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



traz mais informações sobre os cuidados ao se fazer modelagem para épocas passadas e de como usar dados de fósseis.

Table 1. Details of the nine general circulation models (GCMs) used in this comparative study.

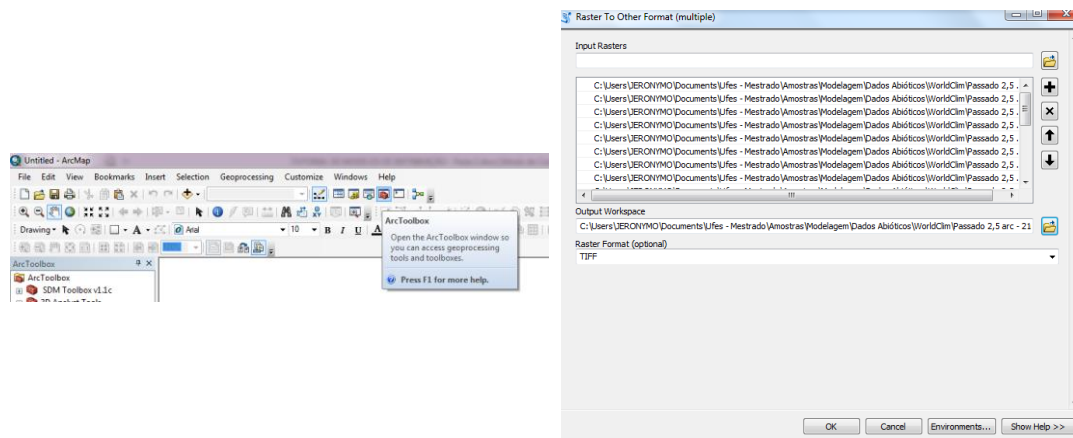
Model ID	Modeling Center	Resolution	# of years	Source	Release
CCSM4	National Center for Atmospheric Research, USA	0.9° × 1.25°	100	CMIP5/PMIP3	2012
CNRM-CM5	Centre National de Recherches Meteorologiques / Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique, France	1.4° × 1.4°	200	CMIP5/PMIP3	2012
COSMOS-ASO (FUB)	Freie Universität Berlin, Germany	3.75° × 3.7°	600	PMIP3	2012
GISS-E2-R	NASA Goddard Institute for Space Studies, USA	2.5° × 2.0°	100	CMIP5/PMIP3	2012
FGOALS-g2	National Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG). Institute of Atmospheric Physics (IAP), China	2.8° × 2.8°	100	CMIP5/PMIP3	2013
IPSL-CM5A-LR	Institut Pierre Simon Laplace, France	3.75° × 1.9°	200	CMIP5/PMIP3	2012
MIROC-ESM	Atmosphere and Ocean Research Institute (University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, Japan	2.8° × 2.8°	100	CMIP5/PMIP3	2012
MPI-ESM-P	Max Planck Institute for Meteorology, Germany	1.9° × 1.9°	100	CMIP5/PMIP3	2011
MRI-CGCM3	Meteorological Research Institute, Japan	1.1° × 1.1°	100	CMIP5/PMIP3	2012

All simulations were obtained from r11p1 ensemble member, except GISS (r11p151). The original resolutions of the maps, in decimal degrees (longitude° — latitude), are coarse (between 1° and 4°). Most models were run along a 100 years time-series after the spin-up period. Source: CMIP5, Coupled Model Intercomparison Project Phase 5 (<http://cmip-pcmdi.llnl.gov/cmip5/>) and PMIP3, Paleoclimate Modelling Intercomparison Project Phase 3 (<http://pmip3.lscce.jplisr.fr>).

doi:10.1371/journal.pone.0129037.t001

A melhor forma de trabalhar com as camadas ambientais no ArcGis é no formato *raster*, se você conseguiu outros formatos de camadas (.asc ou .bil) é recomendado transformá-las em *raster* (.GRID, .tiff ou .img). Para isso siga os passos abaixo:

Vá ao **ARCTOOLBOX >> CONVERSION TOOLS >> TO RASTER >> RASTER TO OTHER FORMAT (MULTIPLE)**



Input Rasters: Colocar as camadas que quer converter;

Output Workspace: Onde quer salvar as camadas convertidas;

Raster format (optional): em qual formato quer salvar.

Na ferramenta **TO RASTER** da **CONVERSION TOOLS** há várias outras formas de converter um *shapefile* ou outro vetor para *raster*. Só escolher a de seu interesse.



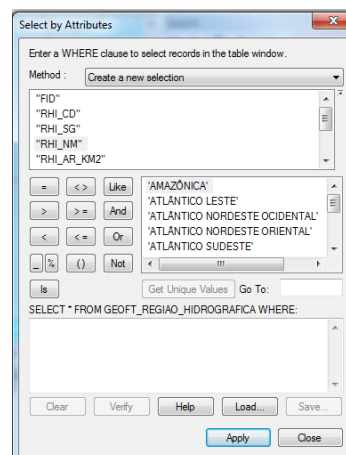
3) Criar uma máscara da área de estudo/background

Nessa etapa você já deve ter escolhido a área de onde serão retirados os pontos de *background* e a área onde será projetado o modelo, que podem ser as mesmas ou não (Veja a parte 1 desse tutorial). Nesse tópico será mostrado como criar máscaras para o corte das variáveis ambientais. As variáveis ambientais na maioria das vezes estão escala mundial ou continental, então precisamos cortá-las, em ambiente SIG, para restringi-las à área de interesse, de acordo com o objetivo da modelagem. Para cortá-las deveremos fazer uma máscara que terá o formato da nossa área de interesse. É aconselhado evitar as divisões políticas por país, estado ou quaisquer outras separações arbitrárias para a construção das máscaras. As distribuições das espécies não seguem essas divisões, por isso é preferível usar definições de bacias hidrográficas, biomas ou qualquer outra divisão “natural”.

a) *Criar uma máscara a partir de um shapefile*: Vou usar o exemplo do *Nectomys squamipes* um rato semiaquático que só ocorre em 5 bacias hidrográficas do Brasil. Se vou criar modelos para essa espécie minha área de estudo corresponderá às essas 5 bacias. Então preciso primeiramente de um *shapefile* das bacias hidrográficas do Brasil para então criar uma máscara só com as 5 bacias. Se meu *background* é a Mata Atlântica, vou procurar um *shapefile* de biomas, se for ecorregiões o *shapefile* será outro. Com a área com qual você quer modelar definida e com *shapefile* inicial que contenha essa área, você conseguirá criar a máscara.

Carregue o arquivo *shapefile* original, no caso o de bacia hidrográfica do Brasil, no ArcGIS.

No **TABLE OF CONTENTS** >> clique com o botão direito do *mouse* no *shapefile* de interesse e escolha **OPEN ATTRIBUTE TABLE** >> ir ao ícone  e escolha **SELECT BY ATTRIBUTE**, e a janela abaixo se abrirá:

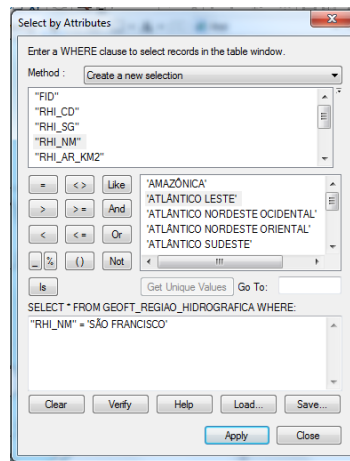


Method: Selecione **CREATE A NEW SELECTION**;

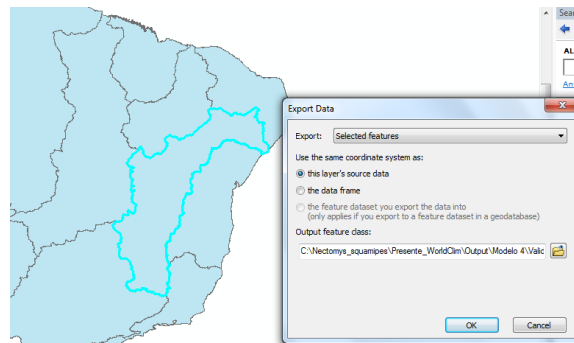
Na caixa de diálogo abaixo selecione o cabeçalho da coluna que contenha nome das bacias (ou do que você quer selecionar, no caso “RHI_NM”) e clique em **GET UNIQUE VALUES**. As variáveis que formam a coluna de interesse serão adicionadas. Crie uma fórmula explicando o que você quer selecionar: “**COLUNA TAL**” = “**NOME**” e clique em **APPLY**.



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



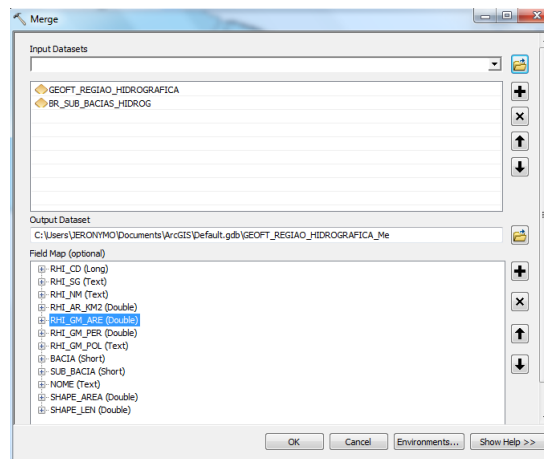
No mapa a bacia do São Francisco ficou selecionada, apresenta agora um contorno diferente. No **TABLE OF CONTENTS** >> clique com o botão direito do *mouse* no *shapefile* das bacias hidrográficas do Brasil escolha **DATA** >> **EXPORT DATA**.



Em **EXPORT** deixe **SELECTED FEATURES**. Como **COORDINATE SYSTEM** pode deixar como **THIS LAYER'S SOURCE DATA**. Escolha o nome e onde salvar em **OUTPUT FEATURE CLASS** e depois clique **OK**.

Faça isso para as outras bacias que você quer (ou outros componentes do mapa). Depois para uni-las em um único *shapefile* use a função **MERGE**.

No **ARCTOOLBOX** >> **DATA MANAGEMENT TOOLS** >> **GENERAL** >> **MERGE**



Dalapicolla, J. 2016. Tutorial de modelos de distribuição de espécies: guia prático usando o MaxEnt e o ArcGIS 10. Laboratório de Mastozoologia e Biogeografia, Universidade Federal do Espírito Santo, Vitória. Disponível em: <http://blog.ufes.br/lamab/tutoriais>



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



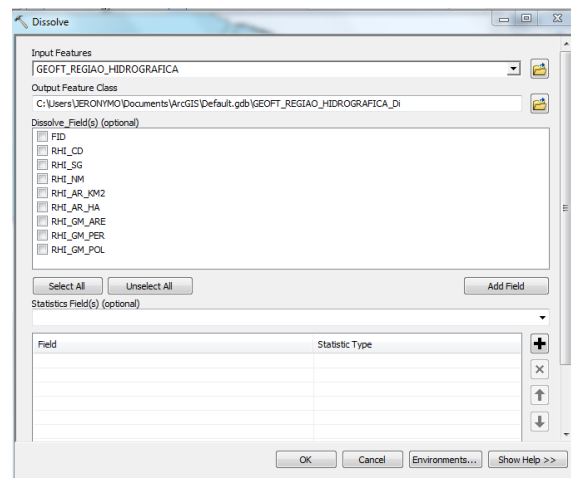
Input Datasets: Os *shapefiles* de cada bacia que você criou e que quer unir.

Output Dataset: Onde e com que nome vai salvar.

Field Map (optional): Que colunas o novo *shapefile* deve ter. **DELETE TODAS**, selecione uma por uma e no **X** do lado direito da caixa de diálogo apague uma por uma.

É necessário que a máscara seja limpa, sem coluna de dados, sem números, caso contrário pode ocorrer erros na hora de cortar as variáveis. Para isso é necessário usar a ferramenta **DISSOLVE**.

No **ARCTOOLBOX >> DATA MANAGEMENT TOOLS >> GENERALIZATION >> DISSOLVE**



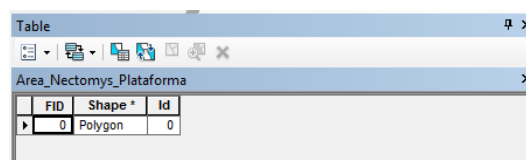
Input Features: Qual *shapefile* quer dissolver

Output Features class: Onde salvar e com que nome.

Dissolve Field(s) (optional): Desmarcar **TODOS** os campos, o campo marcado continuará no *shapefile* dissolvido. Clique em **OK**.

Pronto, sua máscara está feita. Para ter certeza que nenhum campo ficou nela, abra a tabela de atributos e veja se ela está vazia:

No **TABLE OF CONTENTS >>** clique com o botão direito do *mouse* no *shapefile* de interesse e escolha **OPEN ATTRIBUTE TABLE**. Se algo parecido com a imagem abaixo surgiu, então está tudo certo.



b) Criar área de amortecimento para os limites da máscara: É aconselhado criar uma área de amortecimento com a ferramenta **BUFFER** na sua máscara. Por exemplo, eu selecionei as 5 bacias hidrográficas como sendo a minha máscara, seguindo a delimitação de um *shapefile*. Contudo, posso possuir pontos de ocorrência em áreas próximas ao limite da máscara, inclusive do lado de fora da máscara por questão de poucos quilômetros. Seriam esses pontos *outliers* ou eles representam a



LaMaB: Laboratório de Mastozoologia e Biogeografia

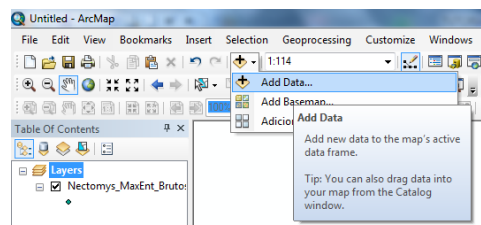
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



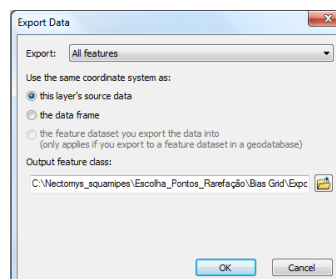
distribuição real? Provavelmente esses pontos fazem parte da distribuição real, lembre-se que a espécie não segue os limites que nós delimitamos para os biomas, bacias e etc. A zona de amortecimento criada por um *buffer* “alarga” as bordas da sua máscara abarcando esses dados. O tamanho do *buffer* depende da escala do projeto (continental, local, etc.) normalmente utiliza-se entre 100 e 500 km de *buffer* em escalas continentais, esse valor depende também da capacidade de dispersão da espécie-alvo.

O primeiro passo é acrescentar uma projeção espacial (*datum*) à sua máscara se ela já não tiver. Comece um projeto novo no ArcGIS . Escolha a projeção do ArcGIS como sendo **WGS84** como foi explicado no **tópico I.4** e abra a máscara nesse projeto. Depois exporte-a com a projeção **WGS84**.

Importe a máscara em **ADD DATA**.

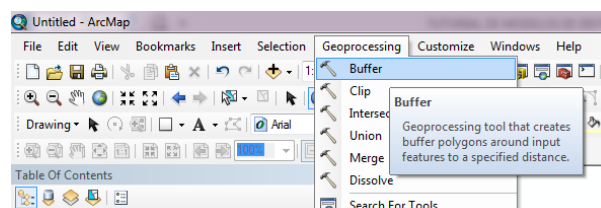


Carregue a máscara **TABLE OF CONTENTS**, e depois clique com o botão direito na máscara e escolha **DATA >> EXPORT DATA**.



Em **EXPORT** deixe **ALL FEATURES**. Como **COORDINATE SYSTEM** mude para **THE DATA FRAME**. Escolha o nome e onde salvar em **OUTPUT FEATURE CLASS** e depois clique **OK**. O programa perguntará se você quer importar essa máscara criada. Importe-a.

Com a máscara e com projeção definida, vá a **GEOPROCESSING >> BUFFER** na barra de menus ou vá no **ARCTOOLBOX >> ANALYSIS TOOLS >> PROXIMITY >> BUFFER**.

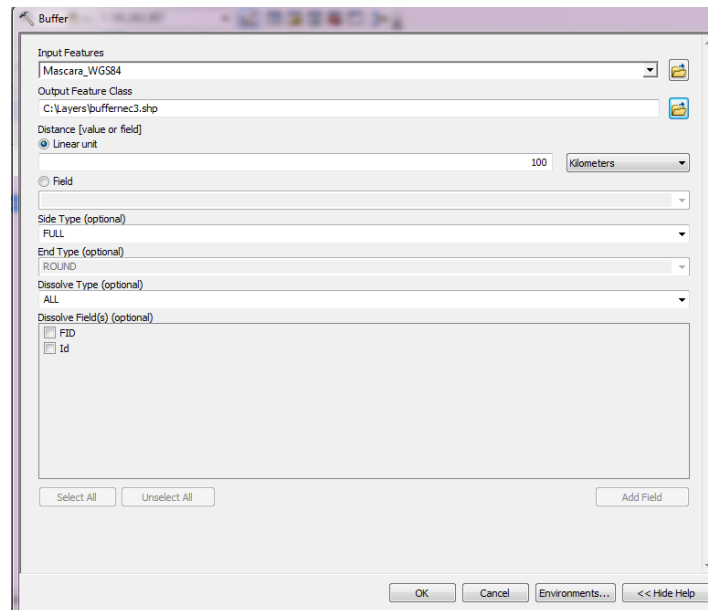




LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



A seguinte janela se abrirá:



Input Features: escolha a máscara com a projeção em **WGS84**.

Output Features Class: indique o local e o nome da nova máscara com **buffer**.

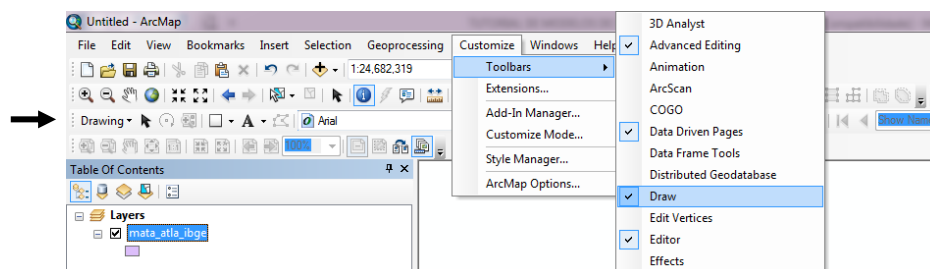
Distance: escolha **LINEAR UNIT** e digite o valor das unidades. Na caixa de diálogo ao lado escolha a unidade.

Side Type: escolha **FULL**.

Dissolve Type: **ALL**. E depois clique **OK**.

Essa máscara com buffer e com projeção que será usada para cortar as camadas ambientais.

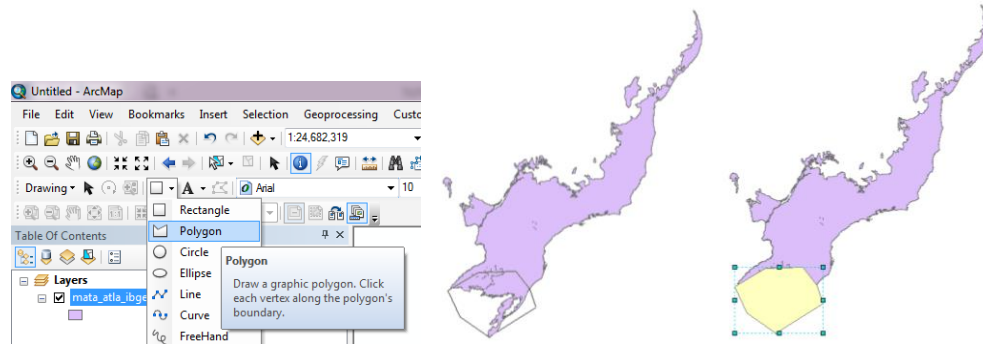
c) *Desenhar à mão a própria máscara:* Às vezes não existe um *shapefile* com os limites que almejamos e algumas vezes existe algo parecido, mas ainda falta algumas áreas para serem acrescentadas ou retiradas. Nesse caso é possível desenhar ou editar um *shapefile* no ArcGIS. Habilite a ferramenta de desenho, na barra de menus: **CUSTOMIZE >> TOOLBARS >> DRAW**.



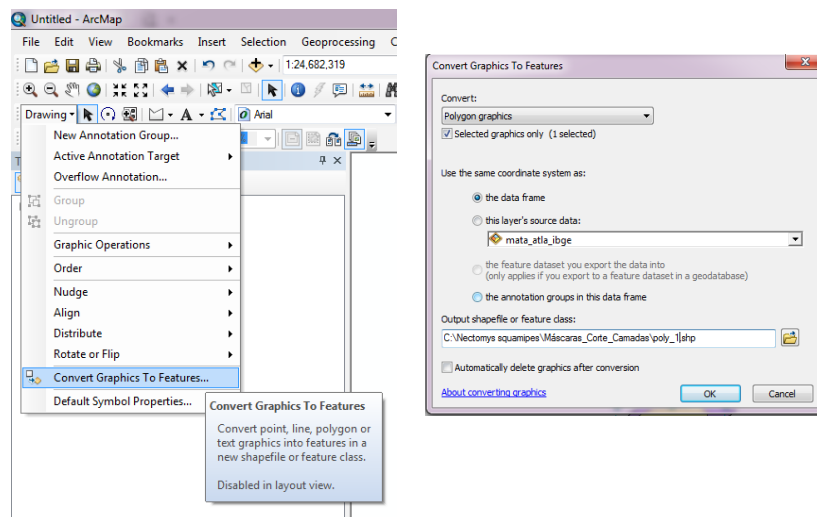
Há várias ferramentas para desenhar, o objetivo é criar a máscara usando essa *toolbar*. Você pode adicionar um outro *shapefile* para usar como modelo ou simplesmente para ter uma referência geográfica para a máscara. Se o objetivo é apenas acrescentar ou retirar uma área de um *shapefile* pré-existente, adicione esse *shapefile* ao projeto e com as ferramentas disponíveis, desenhe essa área extra. No exemplo abaixo eu tinha o *shapefile* da Mata Atlântica e queria

Dalapicolla, J. 2016. Tutorial de modelos de distribuição de espécies: guia prático usando o MaxEnt e o ArcGIS 10. Laboratório de Mastozoologia e Biogeografia, Universidade Federal do Espírito Santo, Vitória. Disponível em: <http://blog.ufes.br/lamab/tutoriais>

acrescentar a porção do sul do Rio Grande do Sul a ele. Usei a ferramenta **POLYGON** para essa tarefa. Note que eu não me preocupei com o contorno do litoral, pois as camadas ambientais não tem dados para o oceano, então não trará diferenças para o modelo o formato da costa. **ATENÇÃO!** Isso muda se você quiser modelar a área da plataforma continental no LGM, sua máscara deverá contemplar uma boa parte do oceano!



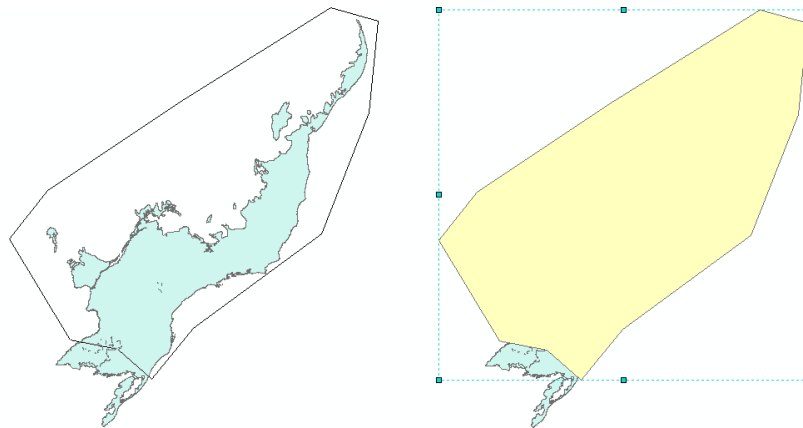
Depois de desenhar a área vá à barra de desenho e clique **DRAWING >> CONVERT GRAPHICS TO FEATURES** e uma janela se abrirá.



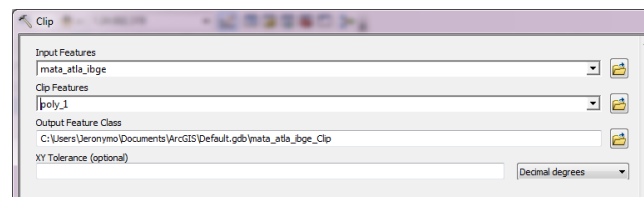
Salve a nova área com um nome apropriado na opção **OUTPUT SHAPEFILE OR FEATURE CLASS** e escolha a projeção para ela, de acordo com o projeto aberto no ArcGIS (**THE DATA FRAME**) ou com a projeção de outro arquivo (**THIS LAYER'S SOURCE DATA**). Clique em **OK** e o programa perguntará se você quer importar esse *shapefile* criado. Importe-o.

A polígono que você criou com a barra de desenho ainda continuará no projeto e talvez por isso você não veja a nova área importada. Apague o desenho: selecione-o e aperte **DEL**, abaixo dele você verá o *shapefile* salvo. Se essa área/polígono/quadrado já for a sua máscara, não se esqueça de fazer *buffer* e usar o **DISSOLVE** como no **tópico III.3.a** e no **tópico III.3.b**, se for necessário. Se quiser adicionar a área a um outro *shapefile* use as ferramentas **MERGE** e depois **DISSOLVE** como mostrado no **tópico III.3.a**.

Se se quiser retirar uma área de um *shapefile* utilize a ferramenta **CLIP**. Essa ferramenta recorta um *shapefile* de acordo com o formato de outro *shapefile*. No caso eu quero eliminar a parte sul do mesmo *shapefile* da Mata Atlântica. O método mais fácil é eu criar um polígono sobre a área que quero preservar, a área que quero manter na máscara. Salvo esse polígono como *shapefile*, como descrito acima e depois de importar esse polígono ao projeto do ArcGIS eu uso a ferramenta **CLIP**.



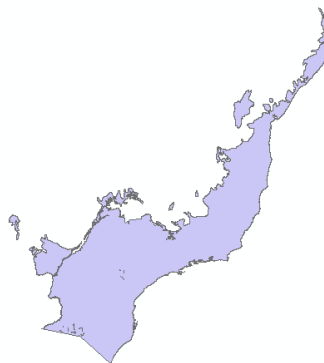
Vá a **GEOPROCESSING >> CLIP** na barra de menus ou vá no **ARCTOOLBOX >> ANALYSIS TOOLS >> EXTRACT >> CLIP**.



Input Features: escolha o *shapefile* que será cortado.

Clip Features: indique o *shapefile* com o molde do recorte (o polígono que cobriu toda a área de interesse).

Output Features Class: indique o local e o nome do novo *shapefile*. Clique em **OK** e o programa perguntará se você quer importar esse *shapefile* criado. Importe-o e veja se está no formato adequado.





LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



Você pode criar quantos polígonos, círculos e quadrados forem necessários para criar sua máscara e pode também usar quantas vezes quiser a função **MERGE** e **CLIP** e outras (as ferramentas apresentadas aqui são as mais usadas, mas existem outras no ArcGIS). Não se esqueça de usar na versão final da máscara a função **DISSOLVE**, acrescentar um *buffer* e a projeção se for necessário.



4) Padronizar os dados abióticos

Para avançar a partir desse ponto você deve ter em mãos as camadas ambientais de interesse, a máscara que representa a área de estudo e saber qual a resolução das camadas que você quer (30 s, 2.5 min, 5 min ou 10 min). Se tiver dúvidas do que representa esse número, sempre veja esse esquema da pesquisadora Flávia Pinto.

Precisão	graus decimais	graus	distância
0	1.0	1°0'0"	111.319 km
1	0.1	0°6'0"	11.132 km
2	0.01	0°0'36"	1.113 km
3	0.001	0°0'3.6"	111.3 m
4	0.0001	0°0'0.36"	11.13 m
5	0.00001	0°0'0.036"	1.11 m
6	0.000001	0°0'0.0036"	11.1 cm
7	0.0000001	0°0'0.00036"	1.11 cm

kms	graus decimais	min/seg
1km	0.008333 graus	30 seg
5km	0.041666 graus	2.5 min
10km	0.083333 graus	5 min
20km	0.166666 graus	10 min
50km	0.500000 graus	30 min
111km	1 grau	60 min

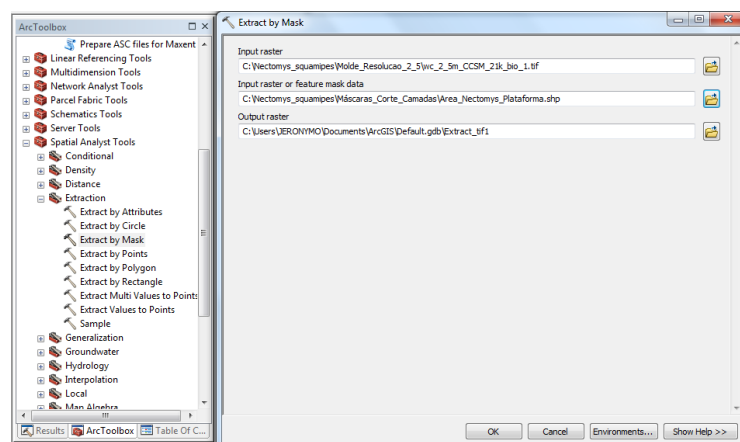
ERRO ESCALA	
250m	escala 1:500.000
30m	escala 1:100.000
10m	escala 1:50.000
1 - 5 m	escala 1:25.000

escala=resolução/0.0002

Nesse ponto nós trabalharemos com as camadas ambientais. Vamos deixá-las no formato correto para o *input* no programa MaxEnt.

a) *Cortar variáveis ambientais*: O primeiro passo é cortar as variáveis no mesmo formato da máscara que você criou no **tópico III.3** utilizando a ferramenta **EXTRACTION BY MASK**.

No **ARCTOOLBOX >> SPATIAL ANALYST TOOLS >> EXTRACTION >> EXTRACTION BY MASK**



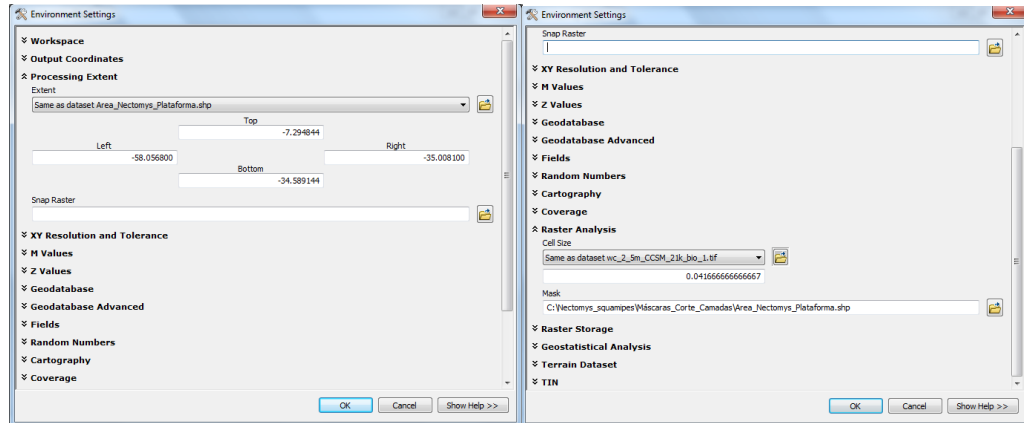
Input raster: Colocar a variável que quer cortar no formato raster (*.GRID*, *.TIFF* ou *.IMG*);

Input raster or mask data: Colocar a máscara que você criou no **tópico III.3**;



Output raster: Escolher o nome do arquivo cortado e onde salvá-lo;

Depois de escolher tudo isso vá em **ENVIRONMENTS** ao lado do **OK** e do **CANCEL**!



Em **PROCESSING EXTENT**: escolha a máscara que você está usando, assim a nova camada será realinhada com base na máscara;

Em **RASTER ANALYSIS >> CELL SIZE**: Escolha um **RASTER** (.GRID, .TIFF ou .IMG) e **NÃO .ASC**! de referência com a resolução que você quer. Você também pode editar manualmente o tamanho do pixel.

Em **RASTER ANALYSIS >> MASK**: escolha novamente a máscara de corte.

Dê **OK** em **ENVIRONMENTS SETTINGS** e outro **OK** em **EXTRACT BY MASK**.

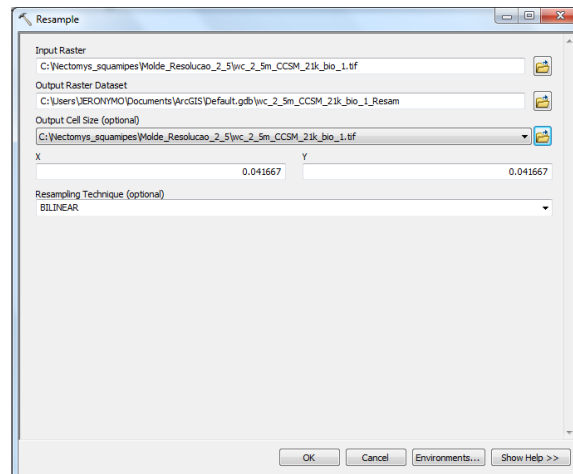
Neste procedimento você fez cortar uma variável ambiental para a área de estudo (máscara) e reamostrou o *raster* cortado ao mesmo tempo, deixando ele na resolução de interesse. Contudo só é possível fazer isso com uma camada de cada vez. Para fazer o mesmo procedimento com várias variáveis, veja os tópicos adiante.

b) Reamostrar a resolução de uma camada: No passo anterior você reamostrou uma camada ao mesmo tempo em que a cortava com a máscara. Porém se você se esqueceu de reamostrar a variável cortada, ainda há um jeito de fazer só isso.

ARCTOOLBOX >> DATA MANAGEMENT TOOLS >> RASTER >> RASTER PROCESSING >> RESAMPLE



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



Input raster: *Raster* que você quer reamostrar.

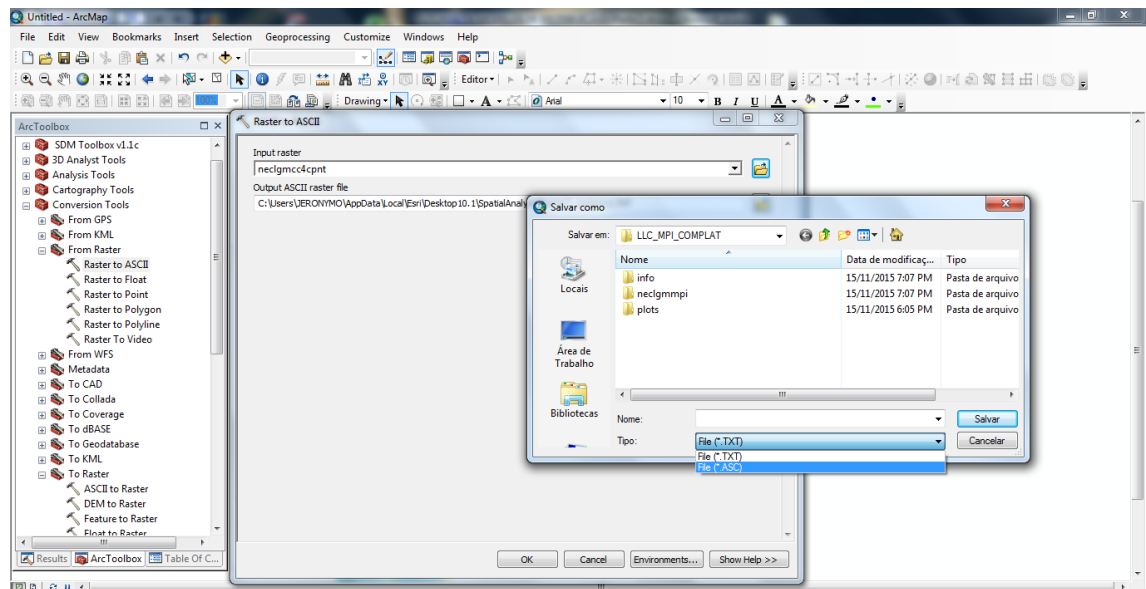
Output raster dataset: Onde e com que nome vai salvar o *raster*.

Output cell size (optional): Escolha um **RASTER** (não *.asc*) de referência com a resolução que você quer. Você também pode editar manualmente o tamanho do pixel em X e Y.

Resampling technique (optional): **BILENEAR**.

c) *Converter as camadas ambientais:* O MaxEnt só lê as variáveis ambientais no formato **.ASC**, por isso é necessário convertê-las do formato **RASTER** para **ASCII**.

No **ARCTOOLBOX >> CONVERSION TOOLS >> FROM RASTER >> RASTER TO ASCII**



Input raster: Qual *raster* você quer converter, escolha a camada ambiental já cortada e reamostrada.

Output ASCII raster file: Onde e com que nome vai salvar o arquivo. **IMPORTANTE!** Pode-se salvar o arquivo **ASCII** em duas extensões, **.TXT** e **.ASC**. Escolha sempre **.ASC**, porque o outro formato não funciona no MaxEnt.

Depois clique **OK**.

Dalapiccola, J. 2016. Tutorial de modelos de distribuição de espécies: guia prático usando o MaxEnt e o ArcGIS 10. Laboratório de Mastozoologia e Biogeografia, Universidade Federal do Espírito Santo, Vitória. Disponível em: <http://blog.ufes.br/lamab/tutoriais>

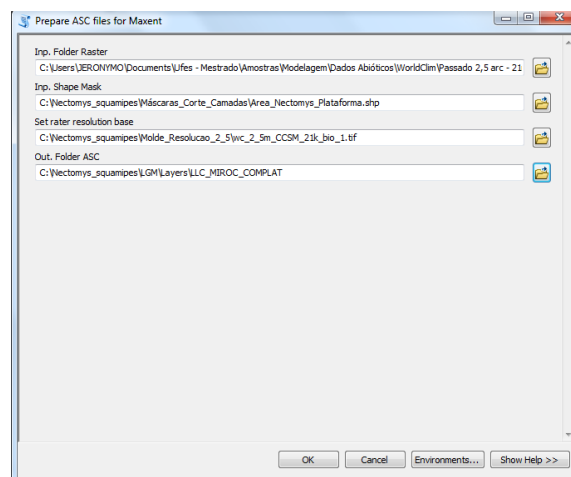


d) *Cortar, reamostrar e converter várias variáveis ambientais ao mesmo tempo*: Há uma forma de fazer os três passos anteriores (**tópicos III.4.a, III.4.b, III.4.c**) de uma única vez e para várias camadas ao mesmo tempo. Instale a Toolbox *LEEC_SDM_Toolboxes*:

No **ARCTOOLBOX** >> clique com o botão direito do *mouse* e escolha **ADD TOOLBOX** >> escolha o arquivo **LEEC_SDM_Toolbox.tbx** dentro da pasta **LEEC_SDM_Toolbox** que você extraiu da pasta compactada em que esse tutorial estava.

Essa toolbox foi desenvolvida por alunos do Laboratório de Ecologia de Espacial e Conservação (LEEC) da UNESP de Rio Claro coordenado pelo professor Milton Cezar Ribeiro. Para usá-la é necessário autorização prévia.

ARCTOOLBOX >> LEEC_SDM_Toolbox >> PREPARE ASC FILES FOR MAXENT



Inp. Folder Raster: Indique a pasta onde estão todas as suas camadas originais, no mesmo formato *raster*. **IMPORTANTE!** Há vários tipos de *raster* (.GRID ou ESRI GRID, .TIFF, .IMG) e para essa ferramenta funcionar bem todas as variáveis dentro da pasta devem ser do mesmo tipo. Por exemplo, vou modelar com 20 variáveis, 15 delas no formato *.tiff* e outras 5 no formato *.GRID*. Neste caso devo separar os formatos em duas pastas e repetir o processo duas vezes. Quando se mistura vários tipos de *raster* na mesma pasta a chance de ocorrer um erro na reamostragem das variáveis é muito grande.

Inp. Shape Mask: Indique a máscara para corte.

Set raster resolution base: Indique um arquivo **RASTER** que servirá de molde para a resolução.

Out. Folder ASC: Onde você quer salvar os novos arquivos raster.

Dê **OK**.

Há uma segunda forma de preparar várias variáveis ambientais para o MaxEnt ao mesmo tempo e sem usar a Toolbox *LEEC_SDM*. Você deve instalar a **SDM Toolbox** o link para *download* se



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo

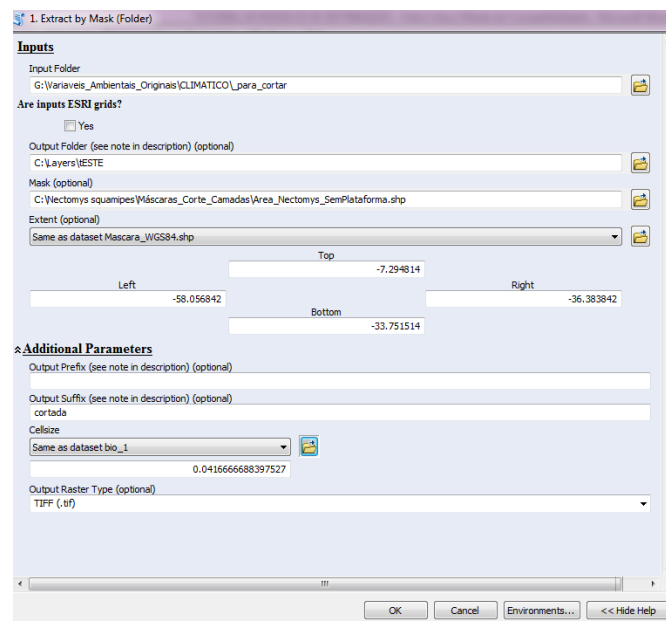


encontra no **tópico I**. Para usar essa toolbox separe todas as variáveis ambientais no formato *raster* em uma pasta, a toolbox trabalhará com o conteúdo de toda uma pasta, como o LEEC fazia. **IMPORTANTE!** Há vários tipos de *raster* (.GRID ou ESRI GRID, .TIFF, .IMG) e para essa ferramenta funcionar bem todas as variáveis dentro da pasta devem ser do mesmo tipo. Por exemplo, vou modelar com 20 variáveis, 15 delas no formato *.tiff* e outras 5 no formato *.GRID*. Neste caso devo separar os formatos em duas pastas e repetir o processo duas vezes. Quando se mistura vários tipos de *raster* na mesma pasta, a chance de ocorrer um erro na reamostragem das variáveis é muito grande.

No **ARCTOOLBOX** >> clique com o botão direito do *mouse* e escolha **ADD TOOLBOX** >> escolha o arquivo **.tbx** dentro da pasta **SDM Toolbox** que você extraiu do arquivo *.zip* do **link do tópico I**.

Para cortar e reamostrar as variáveis ao mesmo tempo:

ARCTOOLBOX >> SDM Toolbox >> BASIC TOOLS >> RASTER TOOLS >> EXTRACT BY MASK (FOLDER)



Input Folder: Indique a pasta onde estão todas as suas camadas originais, no mesmo formato *raster*.

Are inputs ESRI grids? se o formato *raster* for o GRID, o formato do ArcGIS, marque essa opção.

Output folder: Onde você quer salvar os novos arquivos *raster* cortados.

Mask: Indique a máscara para corte.

Extent: Indica como as variáveis serão alinhadas. Indique a máscara com uma projeção definida no **tópico III.3.b** na pasta à direita da caixa de diálogo. Você pode indicar o alinhamento manualmente por meio da opção **AS SPECIFIED BELOW** na caixa de diálogo.



Output prefix: digitar uma palavra que será adicionada como prefixo em cada uma das variáveis cortadas.

Output suffix: digitar uma palavra que será adicionada como sufixo em cada uma das variáveis cortadas.

Cellsize: Indica a resolução das variáveis cortadas, é o tamanho do pixel. Indique uma variável *raster* que tenha o tamanho de pixel do interesse na pasta à direita da caixa de diálogo. Você pode indicar o tamanho do pixel manualmente por meio da opção **AS SPECIFIED BELOW** na caixa de diálogo.

Output Raster Type: escolhe o formato de saída.

Caso precise redefinir o tamanho dos pixels e reamostrar as variáveis você pode usar:

ARCTOOLBOX >> SDM Toolbox >> BASIC TOOLS >> RASTER TOOLS >> RESAMPLE GRIDS (FOLDER)

Input Folder: Indique a pasta onde estão todas as suas camadas originais, no mesmo formato *raster*.

Output folder: Onde você quer salvar os novos arquivos *raster* reamostrados.

Cellsize: Indica a resolução das variáveis cortadas, é o tamanho do pixel. Indique uma variável *raster* que tenha o tamanho de pixel do interesse na pasta à direita da caixa de diálogo. Você pode indicar o tamanho do pixel manualmente por meio da opção **AS SPECIFIED BELOW** na caixa de diálogo.

Resampling method: BILINEAR.

Output prefix: digitar uma palavra que será adicionada como prefixo em cada uma das variáveis cortadas. **ATENÇÃO!** Nomes de arquivos GRID só podem ter 13 caracteres.

Output suffix: digitar uma palavra que será adicionada como sufixo em cada uma das variáveis cortadas. **ATENÇÃO!** Nomes de arquivos GRID só podem ter 13 caracteres.



Extent: Indica como as variáveis serão alinhadas. Indique a máscara com uma projeção definida no **tópico III.3.b** na pasta à direita da caixa de diálogo. Você pode indicar o alinhamento manualmente por meio da opção **AS SPECIFIED BELOW** na caixa de diálogo.

Para converter arquivos *raster* para o formato **ASCII** aceito pelo MaxEnt:

ARCTOOLBOX >> SDM Toolbox >> BASIC TOOLS >> RASTER TOOLS >> RASTER TO ASCII (FOLDER)

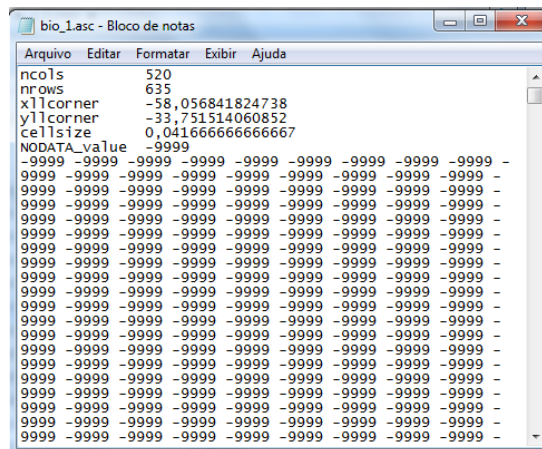


Input Folder: Indique a pasta onde estão suas camadas cortadas no formato *raster*.

Input Raster Type: Escolha o formato dos *raster*.

Output folder: Onde você quer salvar os novos arquivos **ASCII**.

e) *Alinhamento das variáveis ambientais:* Para saber se todas as camadas que você criou estão alinhadas e com o mesmo número de pixel, você deverá abrir cada arquivo *.asc* no bloco de notas (se o arquivo for pesado demorará alguns segundos para a informação aparecer).



Confira todas as camadas, diferenças no número de linhas, colunas, *xll* ou *yll corner* ou no *cell size* podem fazer com que o MaxEnt dê erro. Diferença no número de casas decimais depois da vírgula no *cell size* já é o suficiente para o MaxEnt não funcionar. **FIQUE ATENTO!**

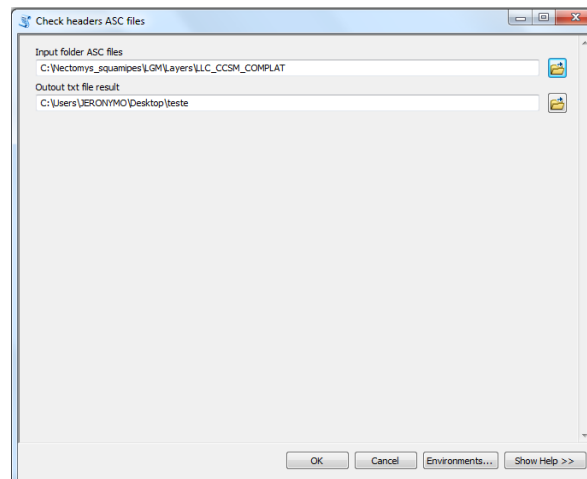
Uma maneira fácil de conferir todas de uma vez é usar o **LEEC_SDM_Toolbox**:

ARCTOOLBOX >> LEEC_SDM_Toolbox >> CHECK HEADERS ASC FILES



LaMaB: Laboratório de Mastozoologia e Biogeografia

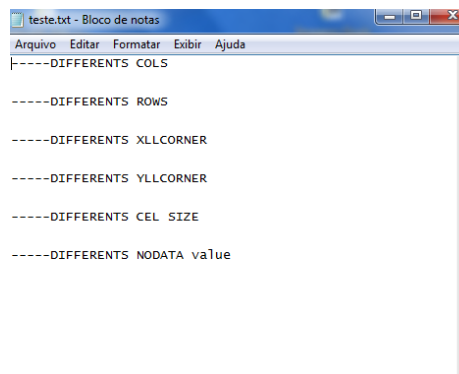
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



Input Folder ASC files: Indique a pasta onde estão todas as suas camadas cortadas, reamostradas e prontas para o MaxEnt.

Output txt file result: Onde e com qual nome salvar o arquivo .txt com o resultado. Dê **OK**.

Se tudo der certo o arquivo de resultado será como abaixo. Se tiver algo errado ele indicará o que está errado e em qual das camadas.





5) Eliminar dados abióticos redundantes

Nessa etapa vamos reduzir a autocorrelação ambiental entre as variáveis ambientais e escolher quais camadas devem ser usadas nos modelos para diminuir a complexidade e assim o sobreajuste ou *overfitting* (ver a parte 1 tutorial).

Normalmente se modela com poucas variáveis ambientais (entre quatro e sete) para diminuir o número de parâmetros e aumentar os graus de liberdade da análise. A regra geral é que o tamanho amostral (no caso os pontos de ocorrências) deve ser 10 vezes maior do que o número de variáveis ambientais utilizados para modelagem (Breiner et al. 2015). Isso significa que se você tem 80 pontos de presença, pode utilizar até 8 variáveis sem aumentar muito o *overfitting*. Não é recomendado modelar com menos de quatro ou cinco variáveis, se você tem menos de 40 pontos tem que compensar o sobreajuste de outra forma, dentro do MaxEnt (ver adiante).

Com isso o primeiro passo é escolher quais das variáveis são mais importantes para o conjunto de dados, há três formas de fazer isso: intuitivamente (escolhendo variáveis relacionadas ao modo de vida da espécie-alvo), por um modelo prévio utilizando a validação cruzada para seleção de variáveis ou por meio de análises multivariadas como a Análise de Fatores ou Análise de Componente Principal (PCA). Aqui demonstrarei como utilizar o PCA para isso. Independentemente do método de escolha de variáveis, o primeiro passo é calcular as correlações entre elas.

a) *Calcular a correlação entre as variáveis ambientais*: Há duas formas de você fazer a correlação: no **R** ou **RStudio** e no ArcGIS utilizando o **SDMToolbox**.

No **R** ou **RStudio**:

Crie uma pasta contendo apenas as camadas no formato **ASCII** já cortadas para a área de interesse e reamostradas na resolução correta. Copie e cole na mesma pasta o arquivo chamado **cor.camadas.R** que você encontra dentro do arquivo *.zip* deste tutorial. Abra o R, instale e carregue os pacotes “raster”, “rgdal” e “vegan”, mude o diretório do R para a pasta com as camadas e o arquivo do R e copie o script abaixo no console.

```
source("cor.camadas.R")
```

Uma função com o mesmo nome aparecerá no *Workspace*. Rode a função sem nenhum argumento.

```
cor.camadas()
```

Se você seguiu os passo-a-passo corretamente a função não dará erro. **ATENÇÃO!** Para a função ser executada, o *Workspace* do R deve estar vazio, sem objetos, só com a função *cor.camadas* carregada. Se há outros objetos delete-os antes de executar a função. O script completo está no **Apêndice 1** caso não encontre o arquivo na pasta. É só copiar o código para o R e executar para a função ser criada. Após o término da função uma pasta será criada dentro do diretório e dentro dela haverá um arquivo *.csv* com os valores de correlação entre as camadas.

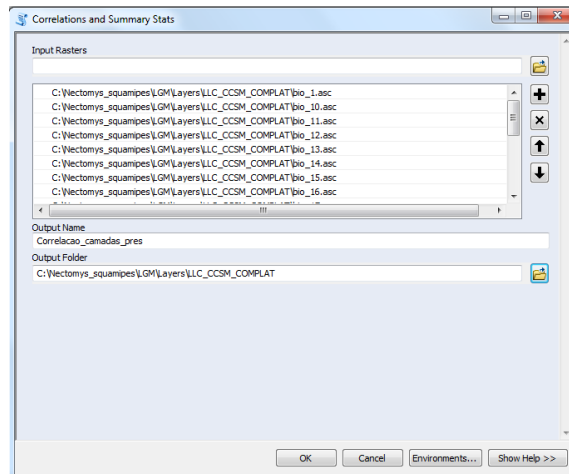


LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



Outra forma de fazer a tabela de correlação é no ArcGIS:

**ARCTOOLBOX >> SDM Toolbox >> SDM TOOLS >> UNIVERSAL TOOLS >> EXPLORE
CLIMATE DATA >> CORRELATIONS AND SUMMARY STATS**



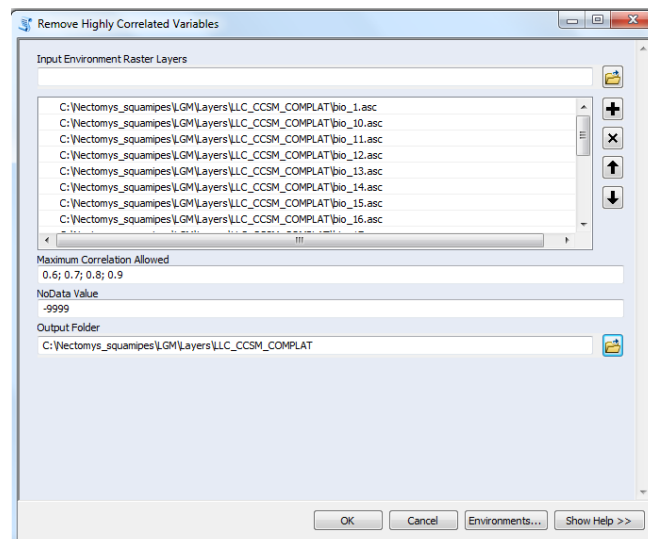
Input Rasters: Seleciona várias camadas **ASC** para calcular a correlação.

Output Name: Qual nome terá o arquivo.

Output Folder: Onde vai salvar. O arquivo salvo é em .csv.

Há uma outra forma de fazer a correlação no ArcGIS e ao mesmo tempo remover as camadas correlacionadas. No fim da função ele gera uma lista das camadas escolhidas para usar no MaxEnt:

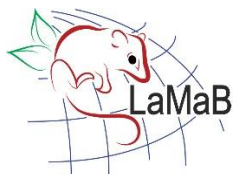
**ARCTOOLBOX >> SDM Toolbox >> SDM TOOLS >> UNIVERSAL TOOLS >> EXPLORE
CLIMATE DATA >> REMOVE HIGHLY CORRELATED VARIABLES**



Input Environment Rasters Layers: Seleciona várias camadas **ASC** para calcular a correlação.

Maximum Correlation Allowed: Nível máximo de correlação permitida. É permitido colocar vários valores para analisar o melhor ponto de corte. Não há consenso de qual valor seria o melhor, mas a maioria dos trabalhos aceita até 0.7 de correlação máxima entre as camadas

Dalapicolla, J. 2016. Tutorial de modelos de distribuição de espécies: guia prático usando o MaxEnt e o ArcGIS 10. Laboratório de Mastozoologia e Biogeografia, Universidade Federal do Espírito Santo, Vitória. Disponível em: <http://blog.ufes.br/lamab/tutoriais>



ambientais. Você pode usar outro corte, mas terá que explicar o porquê. Se a correlação entre duas camadas for abaixo de 0.7, as duas camadas são selecionadas pelo algoritmo. Se a correlação for maior que 0.7 então uma das duas camadas será descartada e a outra será selecionada pelo algoritmo.

NoData Value: -9999.

Output Folder: Onde vai salvar. O arquivo salvo é em .csv. E nele terá uma lista das camadas que não são correlacionadas (ou tem pouca correlação) e que podem ser usadas para a modelagem.

b) Escolher as variáveis ambientais:

No **R** ou **RStudio**:

Crie uma pasta contendo apenas as camadas no formato **ASCII** já cortadas para a área de interesse e reamostradas na resolução correta. Essa etapa pode ser feita na mesma pasta que você calculou a correlação no tópico anterior. Copie e cole na mesma pasta o arquivo chamado **pca.camadas.R** que você encontra dentro do arquivo .zip deste tutorial. Abra o R, instale e carregue os pacotes “raster”, “rgdal” e “vegan”, mude o diretório do R para a pasta com as camadas e o arquivo do R e copie o script abaixo no console.

```
source("pca.camadas.R")
```

Uma função com o mesmo nome aparecerá no *Workspace*. Rode a função sem nenhum argumento.

```
pca.camadas()
```

Se você seguiu os passo-a-passo corretamente a função não dará erro. **ATENÇÃO!** Para a função ser executada, o *Workspace* do R deve estar vazio, sem objetos, só com a função **pca.camadas** carregada. Se há outros objetos delete-os antes de executar a função. O script completo está no **Apêndice 2** caso não encontre o arquivo na pasta. É só copiar o código para o R e executar para a função ser criada. Após o término da função uma pasta será criada dentro do diretório e dentro dela haverá dois arquivos .csv com os resultados do PCA: um deles com o valor da contribuição de cada componente em porcentagem (**pca_percent_PC.csv**) e outro arquivo com a contribuição de cada variável para cada componente (**contribuicao_variavel_PCA.csv**).

Abra a **pca_percent_PC.csv**:

Você terá que escolher quais PC são mais importantes e abarcam a maior parte da variação. No exemplo abaixo, são os quatro primeiros PC representam mais que 70% da variação (70,31%) e já o quinto PC contribui bem menos que quatro (queda de 11,28% para 6,89%). Alguns autores falam que você deve escolher um número de componentes que juntos somem 95% da variação. No caso eu escolhi os quatro primeiros componentes.



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



E16			
	A	B	C
1		x	
2	1	28.44	
3	2	17.62	
4	3	12.97	
5	4	11.28	
6	5	6.89	
7	6	5.58	
8	7	4.02	
9	8	2.56	
10	9	2.28	
11	10	2.17	
12	11	1.56	
13	12	1.11	
14	13	1.09	
15	14	0.77	
16	15	0.57	
17	16	0.46	
18	17	0.37	
19	18	0.15	
20	19	0.1	
21			

Com os PC definidos, abra a **contribuicao_variavel_PCA.csv**. Será algo parecido com isso:

Rotation:				
	PC1	PC2	PC3	PC4
bio_1_val	-0.29196092C	0.07201024	-0.19423054	0.15708419
bio_2_val	-0.06301427	-0.27221226*	-0.38279522*	-0.43772870*
bio_3_val	-0.27310035C	-0.13367482	0.09900619	-0.09489273
bio_4_val	0.27589914C	0.10038847	-0.22019744	-0.06593020
bio_5_val	-0.20524098	0.12656635	-0.48515068C	-0.03023372
bio_6_val	-0.27306690C	0.18681052	0.04358578	0.25974602
bio_7_val	0.17114501	-0.12871344	-0.45872673*	-0.35070419*
bio_8_val	-0.24639109	-0.03280118	-0.18247818	0.13602580
bio_9_val	-0.24766347	0.15844950	-0.10473740	0.16978128
bio_10_val	-0.22067910	0.17452470	-0.38119291C	0.17662028
bio_11_val	-0.30579633*	0.03409508	-0.08132404	0.14353413
bio_12_val	0.23286241	-0.17580859	-0.20017344	0.36160410*
bio_13_val	0.01180980	-0.45375052*	-0.07817573	0.30730947
bio_14_val	0.27146128C	0.19617936	-0.13233857	0.13867832
bio_15_val	-0.24593063	-0.27454268*	0.08623499	-0.13638370
bio_16_val	0.02502718	-0.45487194C	-0.07897055	0.31483519
bio_17_val	0.27222871C	0.19511507	-0.13981155	0.13584254
bio_18_val	0.15501861	-0.34764875*	-0.04786005	0.22678247
bio_19_val	0.25008874	0.22579660	-0.11982153	0.21870583

Primeira coisa a fazer é analisar o PC1 (o de maior contribuição para a variação) e ver qual variável ambiental mais contribui para ele, o que vale nesse caso é o valor modular do número, o sinal só indica o sentido da variação no eixo do componente. No caso é a bio_11 com 0.30579. Então essa é a camada que mais contribui com a variabilidade ambiental na minha área de estudo. Ela com certeza entra na modelagem. A segunda com maior contribuição para o PC1 é a bio_1 com 0.29196, mas ela só entrará na modelagem se tiver uma correlação com bio_11 (que já está selecionada para entrar na modelagem) menor que 0.7. Para saber isso abra a **tabela de correlação**:

I23						
	A	B	C	D	E	F
1		bio_1_val	bio_2_val	bio_3_val	bio_4_val	bio_5_val
2	bio_1_val	-	0.15	0.7	-0.7	
3	bio_2_val	0.15	-	0.39	-0.12	
4	bio_3_val	0.7	0.39	-	-0.91	
5	bio_4_val	-0.7	-0.12	-0.91	-	
6	bio_5_val	0.82	0.39	0.37	-0.29	-
7	bio_6_val	0.9	-0.24	0.63	-0.73	
8	bio_7_val	-0.44	0.62	-0.47	0.67	
9	bio_8_val	0.82	0.23	0.63	-0.59	
10	bio_9_val	0.84	-0.01	0.57	-0.62	
11	bio_10_val	0.9	0.1	0.36	-0.34	
12	bio_11_val	0.98	0.14	0.8	-0.84	
13	bio_12_val	-0.57	-0.02	-0.6	0.58	
14	bio_13_val	-0.06	0.31	0.13	-0.14	
15	bio_14_val	-0.66	-0.33	-0.84	0.83	
16	bio_15_val	0.57	0.44	0.81	-0.79	
17	bio_16_val	-0.09	0.3	0.08	-0.11	
18	bio_17_val	-0.65	-0.32	-0.84	0.84	
19	bio_18_val	-0.46	0.13	-0.3	0.35	
20	bio_19_val	-0.57	-0.4	-0.79	0.77	
21						



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



No caso a bio_1 é explicada em 98% pela bio_11, elas estão correlacionadas em mais de 0.7, então só uma das duas poderá entrar na análise. Como a bio_11 contribui mais para o PC1, ela fica e a bio_1 é descartada. Pronto analisei duas variáveis, faltam as outras 17.

Depois a terceira maior contribuição é da bio_4 com 0.27589. A única variável selecionada até agora é a bio_11, veremos a correlação entre elas: é de 84%, acima do permitido que é 70%. Então ela não entra na análise. Você não precisa fazer isso com todas as variáveis do PC, só com as que mais contribuem, normalmente não mais de 7 e não menos que 3. No PC1 analisei as variáveis com contribuição até 0.27, no PC2 até 0.27, PC3 até 0.38 e PC4 até 0.35. No fim fiquei com sete variáveis que serão usadas no MaxEnt e para os próximos passos.

Se você escolheu um número de componentes que juntos somam 95% da variação, provavelmente você terá que analisar entre 20 e 30 componentes. Nesse caso escolha apenas a variável que mais contribui para cada componente e não três como no exemplo acima. Isso agiliza o trabalho.

Se o número de variáveis ambientais ainda for muito grande para o número de pontos que você tem é recomendado eliminar algumas variáveis. Você pode criar um modelo prévio no MaxEnt com os parâmetros no *default* e analisar as contribuições de cada variável e escolher somente aquelas que mais contribuem para o modelo (ver nos tópicos adiante).

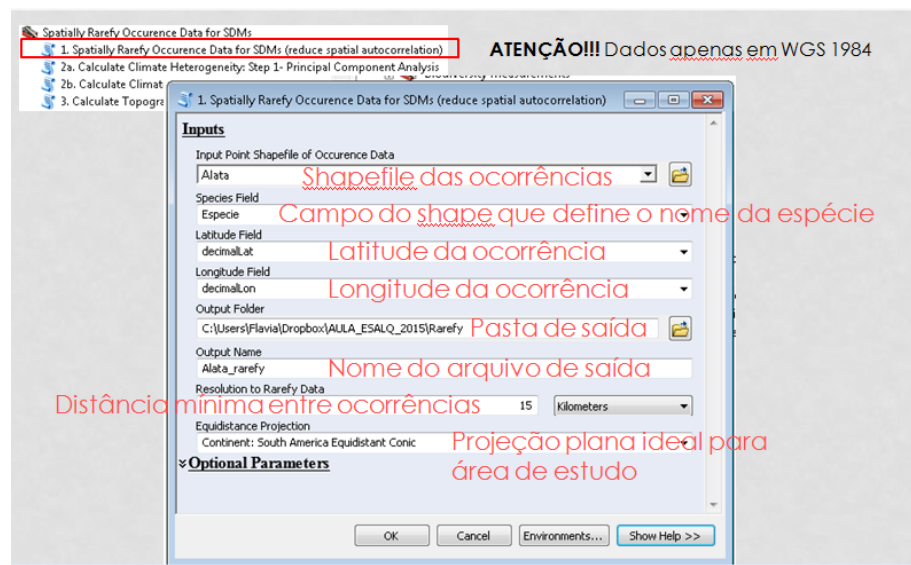


6) Eliminar dados bióticos redundantes

Há duas formas de reduzir o viés amostral ou autocorrelação espacial : **Rarefação Simples**, **Rarefação por Heterogeneidade Ambiental**.

a) *Rarefação Simples*:

ARCTOOLBOX >> SDM Toolbox >> SDM TOOLS >> UNIVERSAL TOOLS >> SPATIALLY RAREFY OCCURRENCE DATA FOR SDM >> 1.SPATIALLY RAREFY OCCURRENCE DATA FOR SDM (REDUCE SPATIAL AUTOCORRELATION)



Normalmente a distância de ocorrência é 5 Km ou 10Km, seguindo a maioria dos trabalhos, não há um porquê desse valor. A projeção plana deve ser **CONTINENT: SOUTH AMERICA EQUIDISTANCE CONIC**. Se ocorrer erro na análise, pode ser por um desses motivos:

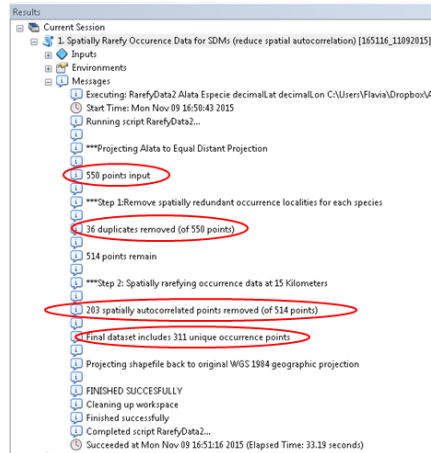
- 1) Projeção deve estar em WGS1984: Escolha o *shapefile* de pontos em WGS84 que você criou no **tópico II.1**. Ou vá ao **ARCTOOLBOX >> DATA MANAGEMENT TOOLS >> PROJECTION AND TRANSFORMATIONS >> RASTER >> DEFINE PROJECTION >>** Acrescente o *shapefile* de pontos e escolha a projeção em: **GEOGRAPHIC COORDINATE SYSTEMS >> WORLD >> WGS1984**.
- 2) Os pontos de ocorrência dever estar salvos em *shapefile*;
- 3) **IMPORTANTE**: O *shapefile* dos pontos só deve ter três colunas (long, lat e species) se tiver mais colunas como localidade e os nomes tiverem acentos gráficos ou cedilha pode dar erro na análise;
- 4) Se a pasta onde o ArcGIS for puxar os arquivos ou salvá-los for muito longe do C: também pode dar erro, melhor opção é salvar na raiz, isto é, no C: ou próximo a ele.

Depois de escolher os parâmetros dê **OK**. Se deu erro, você pode ir à aba **RESULTS** lá terá um comentário sobre o erro, jogue no *Google* e veja se resolve. Vá à **BARRA DE MENUS >>**



GEOPROCESSING >> RESULTS >> CURRENT SESSION >> 1.SPATIALLY RAREFY OCCURRENCE DATA FOR SDM (REDUCE SPATIAL AUTOCORRELATION) >> MESSAGES.

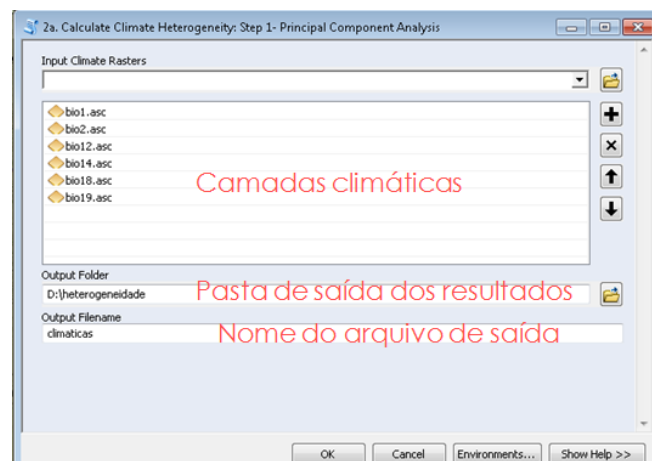
Nela também fala quantos pontos foram retirados, quantos eram duplicados, autocorrelacionados e quantos restaram para a modelagem.



Essa ferramenta criará um *buffer* do tamanho pedido por você e se houver dois ou mais pontos dentro desse *buffer* ele escolherá aleatoriamente um deles, o restante será descartado. O .csv final terá os pontos que você colocará no *input* do MaxEnt, os pontos descartados podem ser usados para a Validação Externa do Modelo, **APENAS OS CORRELACIONADOS, NÃO OS DUPLICADOS!** Acrescente-os na planilha de Validação Externa que você criou no **tópico III.1**.

b) Rarefação por Heterogeneidade Ambiental:

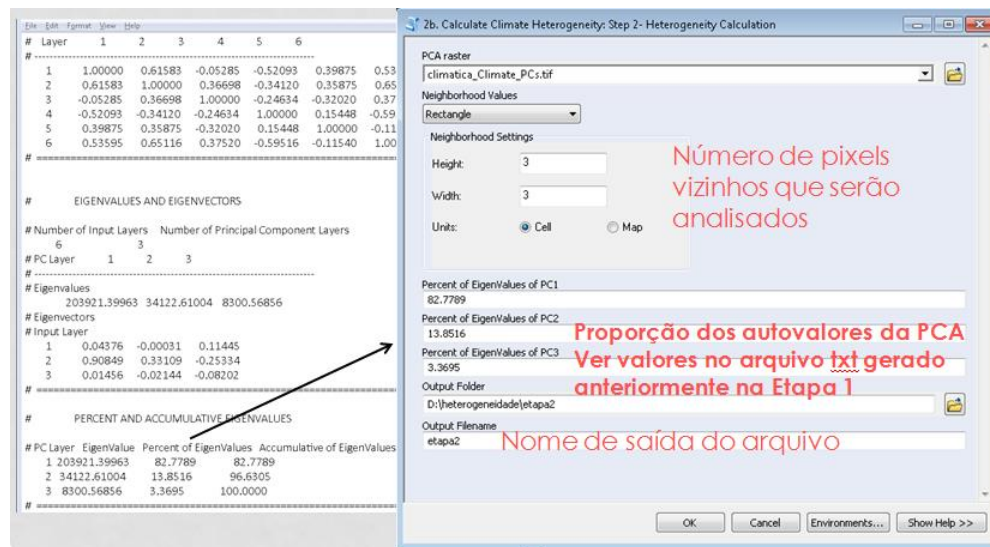
ARCTOOLBOX >> SDM Toolbox >> SDM TOOLS >> UNIVERSAL TOOLS >> SPATIALLY RAREFY OCCURRENCE DATA FOR SDM >> 2a. CALCULATE CLIMATE HETEROGENITY STEP 1 - PRINCIPAL COMPONENTS ANALYSIS



As camadas climáticas que você vai acrescentar aqui são aquelas escolhidas no fim **tópico III.5**, camadas independentes (não correlacionadas), aquelas escolhidas para ir ao MaxEnt. Depois de escolher a pasta de saída e o nome do arquivo dê **OK**.

O mapa pronto tem 3 bandas que representam os 3 primeiros componentes principais, só que o ArcGIS não consegue visualizá-las. Antes de seguir ao passo seguinte, você deve remover o mapa do projeto no **TABLE OF CONTENTS** (Botão direito do *mouse* >> **REMOVE**) e carregá-lo de novo. Agora você verá as 3 bandas do mapa (**RED**, **GREEN** e **BLUE**). Só quando vir as três bandas continue.

ARCTOOLBOX >> SDM TOOLBOX >> SDM TOOLS >> UNIVERSAL TOOLS >> SPATIALLY RAREFY OCCURRENCE DATA FOR SDM >> 2b. CALCULATE CLIMATE HETEROGENITY STEP 2 – HETEROGENITY CALCULATION



PCA raster: Colocar o mapa gerado na PCA do **STEP1**, com as três bandas.

Neighborhood Values: **RECTANGLE** (pode ser circular ou outro tipo).

Neighborhood Settings: Só aceita números ímpares. Tamanho da reamostragem. Se sua resolução é 1 km, 3 pixels representam 3 km. Isto significa que a criação do mapa usará 3 km ao redor do pixel central para o cálculo do valor de heterogeneidade ambiental desse pixel central. Quanto maior o valor dos pixels mais homogêneo será o mapa (ambientes pouco complexos), quanto menor o valor dos pixels mais heterogêneo será o mapa (ambientes mais complexos). Para *Nectomys squamipes* para 5 bacias hidrográficas usei, 3 pixels, e o mapa tinha resolução de 5 km, então minha heterogeneidade foi calculada com 15 km em torno no pixel central.

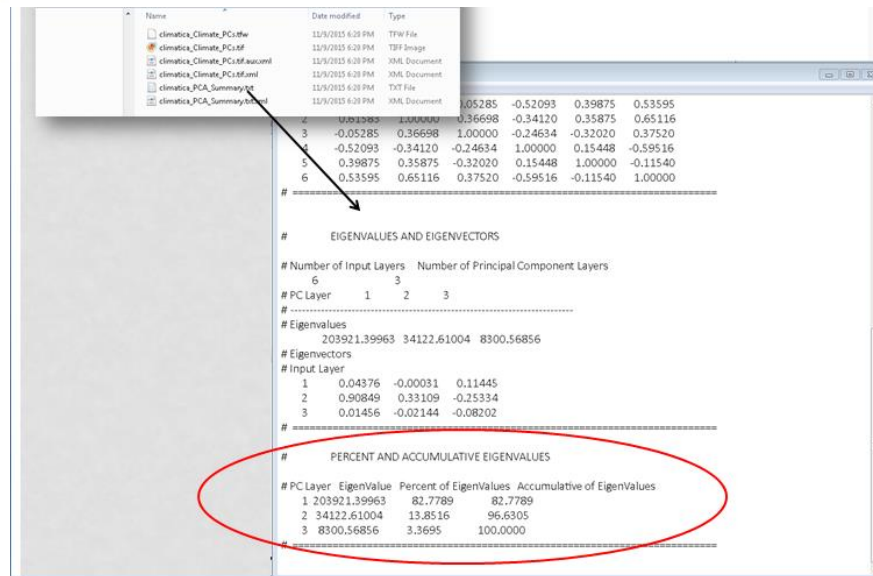
Percent of EigenValues of PC1, PC2 e PC3: Vá no *output* do **STEP1** "**NOME_QUE_VOCÊ_DEU_PCA_SUMMARY.TXT**". No final do arquivo tem a contribuição de cada componente para o PCA. Coloque esses números. **NÃO ESQUEÇA QUE O SEPARADOR DE DECIMAL É PONTO!**



LaMaB: Laboratório de Mastozoologia e Biogeografia

Departamento de Ciências Biológicas

Universidade Federal do Espírito Santo



Escolha a pasta de saída e nome do arquivo gerado (mapa de heterogeneidade) e dê **OK**. Se der erro tente mudar a pasta pra mais perto da raiz **C:**, se não adiantar jogue o erro no Google que sempre tem um fórum de discussão para ajudar com isso.

Agora vamos usar a mesma ferramenta da rarefação simples:

Vá no **ARCTOOLBOX >> SDM Toolbox >> SDM TOOLS >> UNIVERSAL TOOLS >> SPATIALLY RAREFY OCCURRENCE DATA FOR SDM >> 1.SPATIALLY RAREFY OCCURRENCE DATA FOR SDM (REDUCE SPATIAL AUTOCORRELATION)**

Preencha os campos da mesma forma que na etapa de rarefação simples. E depois clique em **OPTIONAL PARAMETERS**.

Inputs

Input Point Shapefile of Occurrence Data: Nectomys_MaxEnt_Bruto

Species Field: Species

Latitude Field: Latitude

Longitude Field: Longitude

Output Folder: C:\Nectomys_squamipes\Escolha dos Pontos - Rarefação\Rarefeitos_Heterogeneidade_10km

Output Name: Nectomys_Rarefacao_10km_hetero

Resolution to Rarefy Data: 10 Kilometers

Equidistance Projection: South America Equidistant Conic

Optional Parameters

The following options rarefy at different distances based on input habitat heterogeneity

☐ Check to use multiple rarefying distances

Input Heterogeneity Raster (optional): C:\Nectomys_squamipes\Vect_Hetero_Climate_Hetero.tif

Number of Heterogeneity Classes: 3

Classification Type: NATURAL_BREAKS

Maximum Distance: 25 Kilometers

Minimum Distance: 5 Kilometers

Buttons: OK, Cancel, Environments..., Show Help >>

Input Heterogeneity Raster (Optional): o mapa de heterogeneidade que você gerou no passo anterior.

Dalapicolla, J. 2016. Tutorial de modelos de distribuição de espécies: guia prático usando o MaxEnt e o ArcGIS 10. Laboratório de Mastozoologia e Biogeografia, Universidade Federal do Espírito Santo, Vitória. Disponível em: <http://blog.ufes.br/lamab/tutoriais>



Number of heterogeneity Classes: 5. Quantas zonas (classes) ambientais diferentes o algoritmo vai criar. Esse número deve ser o maior possível, mas o máximo que o programa permite é 5.

Classification Type: NATURAL_BREAKS. Ele cria histogramas com o valores dos pixels do mapa e procurar 5 classes “naturais” diferentes neles.

Maximum Distance: 25 km (default).

Minimum Distance: 5 km (default).

Dê OK.

Se ocorrer erros verifique o tópico sobre a Rarefação Simples que nele há dicas que podem resolver esses problemas.

Essa ferramenta criará um *buffer* do tamanho pedido por você e se houver dois ou mais pontos dentro desse *buffer* e na mesma zona ambiental ele escolherá aleatoriamente um deles, o restante será descartado. Se esses pontos que estão dentro do mesmo *buffer* estiverem em zonas (classes) ambientais diferentes (aquelas 5 classes criadas pelo NATURAL_BREAKS) os dois pontos serão mantidos. Isso porque mesmo estando próximos, esses pontos representam regiões ambientais diferentes e eles são importantes para o modelo. O .csv final terá os pontos que você colocará no *input* do MaxEnt, os pontos descartados podem ser usados para a Validação Externa do Modelo, **APENAS OS CORRELACIONADOS, NÃO OS DUPLICADOS!** Acrescente-os na planilha de Validação Externa que você criou no Tópico 1.



7) Análise do Espaço Ambiental

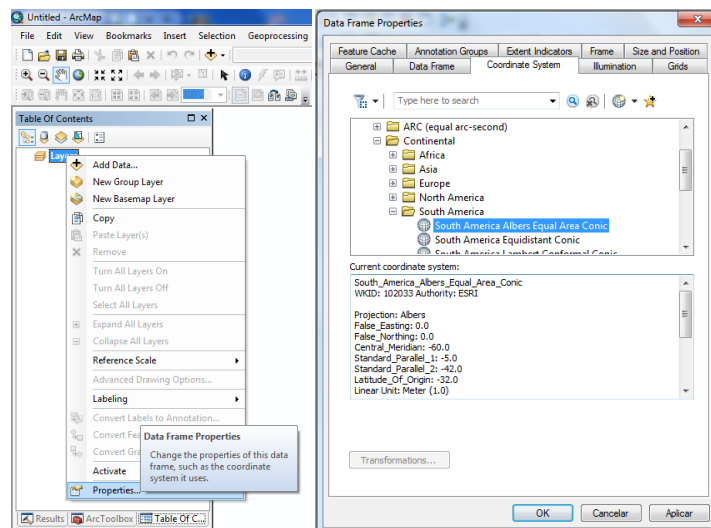
Nessa etapa vamos controlar o sobreajuste (*overfitting*) dos modelos selecionando os pontos de background para uma melhor avaliação do espaço ambiental ocupado pela espécie (ver parte 1 desse tutorial). Além disso, avaliaremos se os pontos de presença conseguem representar bem o espaço ambiental da espécie-alvo.

a) Seleção dos pontos de background por Bias Grid ou Bias File:

O MaxEnt seleciona aleatoriamente os pontos de *background* (pseudoausência) dentro da sua área de estudo (ver a parte 1 desse tutorial). Com o *Bias Grid* é possível indicar áreas dentro da sua região de estudo para que sejam mais amostradas na hora do MaxEnt selecionar esses pontos. Isso significa que com o *Bias Grid* a seleção de pontos do *background* deixa de ser aleatório e passa a ocorrer com maior frequência na região onde se tem mais pontos de ocorrência, a fim de amostrar ambientalmente melhor essa região. Essa é uma técnica muito criticada, mas vários trabalhos publicados em revistas importantes já usaram. Você pode criar modelos com e sem o *Bias Grid* e ver qual é o melhor para gerar um modelo estatisticamente e biologicamente mais robusto.

O *Bias Grid* só funciona em uma projeção plana e em metros, por isso devemos redefinir a projeção do *shapefile* dos pontos, ou criar o *shapefile* em Albers como no **tópico III.1**. Nesse caso, você usará os pontos já filtrados, sem dados redundantes.

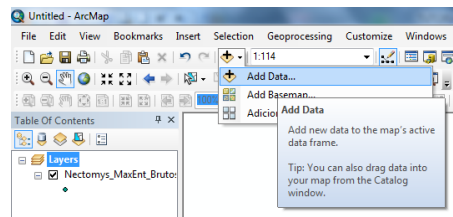
Abra um novo projeto >> aba **TABLE OF CONTENTS** >> clique com o botão direito do mouse em **LAYERS** >> **PROPERTIES** >> aba **COORDINATE SYSTEM** >> **PROJECTED COORDENATE SYSTEMS** >> **SOUTH AMERICA ALBERS EQUAL AREA CONIC** >> **OK**.



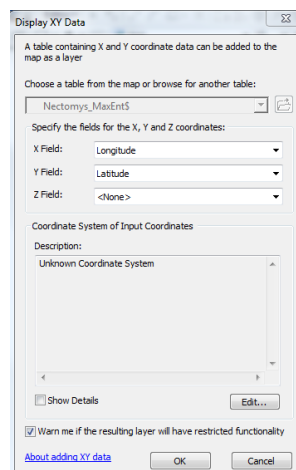
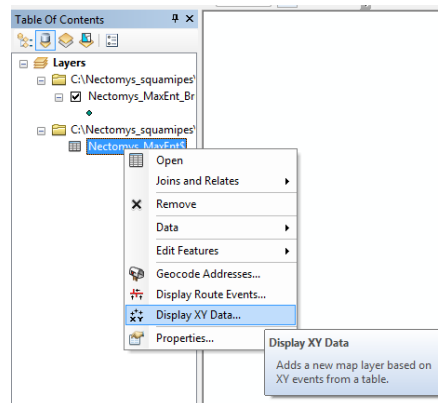
Importe a tabela de pontos feito no **tópico III.6.a** ou **III.6.b** em **ADD DATA**.



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



Depois de carregado no **TABLE OF CONTENTS**, clique com o direito do *mouse* na tabela com o nome de seus pontos e escolha **DISPLAY XY DATA**.

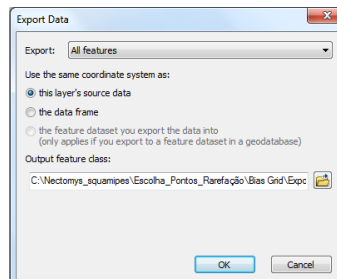


Escolha os campos de latitude e longitude e clique em **EDIT** no campo da coordenada geográfica. Escolha a projeção em: **PROJECTED COORDENATE SYSTEMS >> CONTINENTAL >> SOUTH AMERICA ALBERS EQUAL AREA CONIC >> OK >> OK**.

Salve o *shapefile* com a projeção em **ALBERS** (use isso no nome do *shapefile*). Vá ao **TABLE OF CONTENTS >>** clique com o botão direito do *mouse* no *shapefile* de pontos em Albers e escolha **DATA >> EXPORT DATA**.



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo

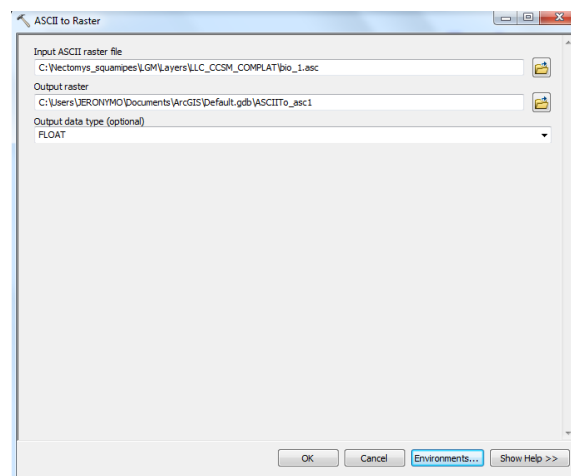


Em **EXPORT** deixe **ALL FEATURES**. Como **COORDINATE SYSTEM** pode deixar como **THIS LAYER'S SOURCE DATA**. Escolha o nome e onde salvar em **OUTPUT FEATURE CLASS** e depois clique **OK**.

Para o *Bias Grid* é necessário terá que criar um molde em *raster* com a mesma projeção do *shapefile* dos pontos e com a mesma resolução (tamanho de pixel e alinhamento) das variáveis ambientais que você vai colocar no MaxEnt.

Escolha uma variável em **ASCII** que você gerou no fim do **tópico III.4** e a transforme em *raster*:

ARCTOOLBOX >> CONVERSION TOOLS >> TO RASTER >> ASCII TO RASTER



Escolha um *.ASC* que você criou e um lugar e nome para o *raster* de molde e no **Output data type** escolha **FLOAT**.

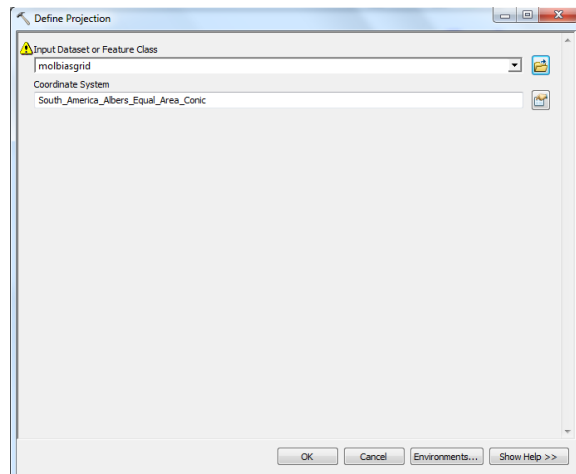
Mude a projeção do *raster* de molde criado no passo anterior:

ARCTOOLBOX >> DATA MANAGEMENT TOOLS >> PROJECTION AND TRANSFORMATIONS >> RASTER >> DEFINE PROJECTION.

Em **Coordinate System** estará **UNKNOWN**, clique do lado e procure: **PROJECTED COORDINATE SYSTEMS >> CONTINENTAL >> SOUTH AMERICA ALBERS EQUAL AREA CONIC >> OK >> OK.**

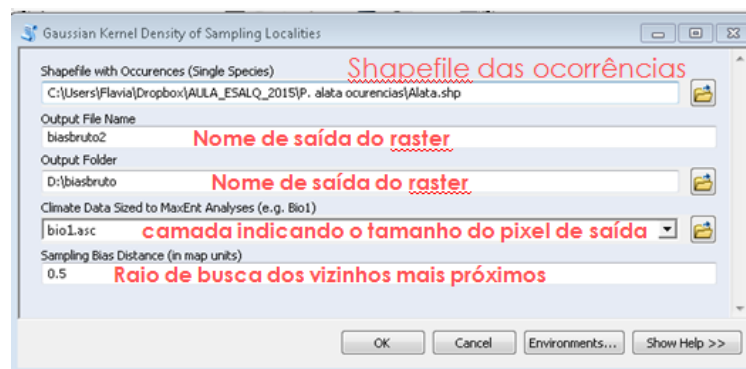


LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



Agora vamos criar o *Bias File*:

**ARCTOOLBOX >> SDM TOOLBOX >> SDM TOOLS >> MAXENT TOOLS >> BACKGROUND
SELECTION VIA BIAS FILES >> GAUSSIAN KERNEL DENSITY OF SAMPLING LOCALITIES**



O *shapefile* de ocorrências é aquele projetado em Albers, a camada indicando o tamanho do pixel é aquela em *raster* que você definiu uma projeção no passo anterior, escolha o nome e o local de saída do output. O raio de busca coloque 0.5 ou 1.0. Esse valor representa o valor em Km de 1º de latitude, que no Equador é em torno de 110 Km. 0.5 assim representará algo em torno de 50-55 Km de raio e o valor de 1.0 cerca de 100 Km. A maioria dos trabalhos que usam seleção de *background* usam 100 Km de raio.

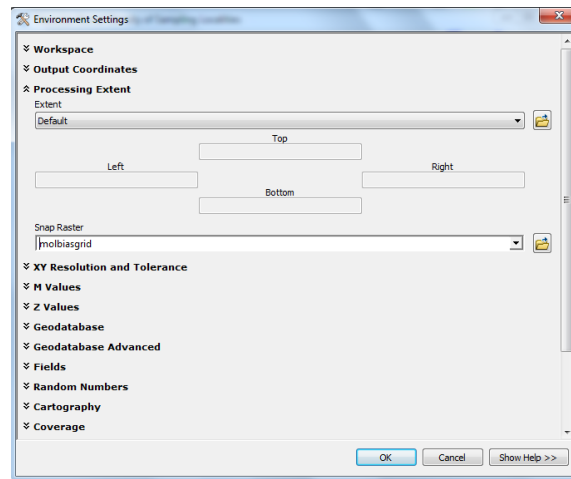
Depois de escolher tudo isso vá a **ENVIRONMENTS...** e **NÃO EM OK!**

Vá a **PROCESSING EXTENT**: em **SNAP RASTER** escolha o molde de pixel que você definiu a projeção, os pixel serão redefinidos a partir dela. Dê **OK** e **OK**.



LaMaB: Laboratório de Mastozoologia e Biogeografia

Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo

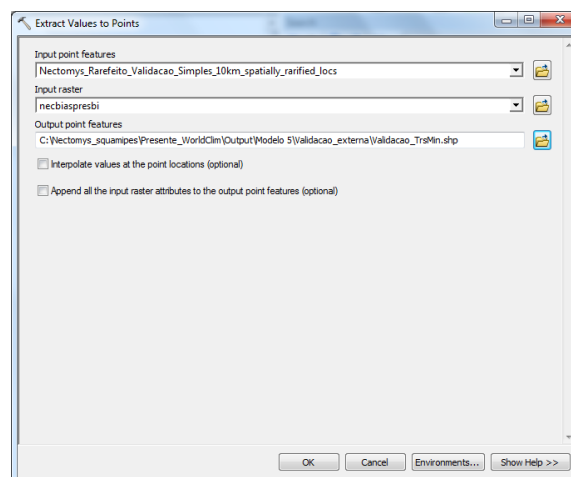


Um arquivo *.asc* será criado. Ela será o *Bias File* da sua modelagem.

b) Estimar o Espaço Ambiental:

Com o PCA você descobriu quais variáveis são as que mais contribuem para a variação ambiental daquela região. Selecione as três variáveis mais importantes (as três primeiras da lista). No meu caso foram *bio_8*, *bio_2*, *bio_5*. Abra o ArcGIS, carregue essas camadas no formato *.asc* ou *raster* e carregue também os pontos de ocorrência que você possui para criar os modelos no MaxEnt (sem viés amostral, após o **tópico III.6**). Você vai extrair os valores das variáveis ambientais dos pontos de ocorrência para construir gráficos *biplot* para cada combinação de variável (*bio_8* x *bio_2*; *bio_8* x *bio_5*; *bio_2* x *bio_5*).

ARCTOOLBOX >> SPATIAL ANALYST TOOLS >> EXTRACTION >> EXTRACT VALUES TO POINTS



Input point features: selecionar o *shapefile* de pontos do Maxent;

Input raster: selecionar a camada ambiental. Se der erro com a camada no formato *.asc*, tente um outro formato *raster* (*.img* ou *.tif*);

Output point features: onde salvar e com qual nome.

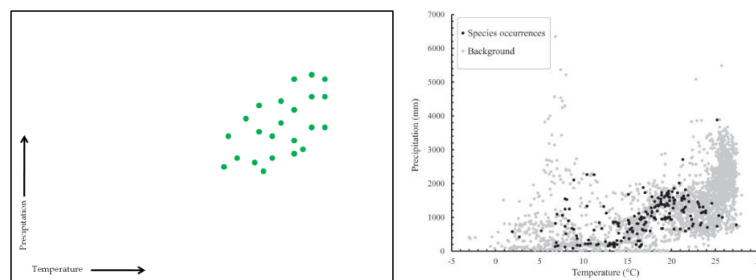
Deixe as opções desmarcadas e dê **OK**.



Será criado outro *shapefile* e ele será carregado no **TABLE OF CONTENTS**. Vá para onde o novo *shapefile* foi salvo, lá terá uma tabela executável no *MSEXcel* ou similar no formato **dBASE table (.dbf)**. Clique nela e busque uma coluna de valores que foi adicionada após a última coluna, uma coluna chamada **RASTERVALUE**. Nessa coluna tem valores dos pontos para essa camada ambiental. Salve uma cópia dessa tabela como arquivo do *MSEXcel*, pois você modificará esse arquivo.

É comum o tamanho das colunas se modificarem no formato *.dbf*, especialmente se você o abrir no *MSEXcel*. Parece que essa coluna **RASTERVALUE** não foi adicionada. Só diminua o zoom e tamanho das colunas que você a achará.

Faça isso para as outras camadas. Com esses valores será possível construir os gráficos no *MSEXcel*, *PAST*, *R* ou outro programa. Você pode extrair os valores de todas as camadas e criar um gráfico para alguma análise multivariada como PCA, análise canônica, para visualizar os seus pontos. Veja como se comporta os seus dados. O ideal é os pontos nos gráficos se comportem de duas formas: (1) fiquem próximos uns dos outros, formando um único *cluster*, sem *outliers* e sem agrupamentos separados por espaços vazios; ou (2) se espalhem por todo o gráfico equidistantemente, sem a formação de grupos aparentes. A área desse gráfico representa o espaço ambiental e se houver grandes espaços vazios separando os pontos (um ponto ou um agrupamento de pontos) significa que seus pontos de ocorrência não conseguiram contemplar todo espaço ambiental da espécie-alvo. Em outras palavras: seus pontos de ocorrência não são bons, há lacunas de conhecimento e você precisa conseguir mais pontos ou seu modelo pode ficar ruim por causa disso. Analise se ele precisa de mais pontos que abarquem os espaços vazios, lembrem-se que espaço ambiental não amostrado corretamente por superestimar ou subestimar a validação dos modelos.





IV) Modelagem

1) O programa MaxEnt

a) *Checklist de arquivos necessários:* Antes de começar a mexer nos parâmetros do programa, cheque se você possui todos os arquivos necessários.

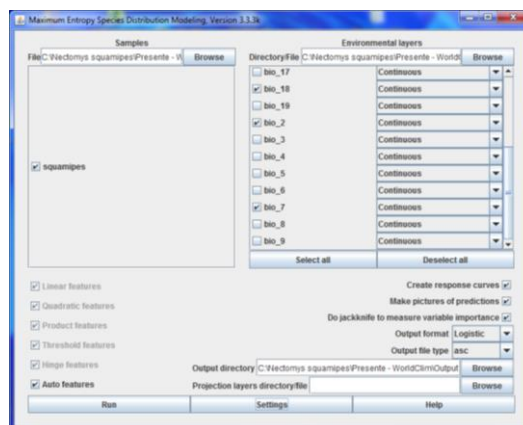
i) *Pontos de ocorrência no formato .csv:* Esse arquivo deverá ter 3 colunas: 1ª coluna nomeada como **species** com o nome da espécie (os epítetos não podem ser separados por espaço, use sublinhado/underline), **long** (Longitude) e **lat** (Latitude). As coordenadas precisam estar em formato de graus decimais, e o separador deverá ser o ponto, e devem ser precedidas de um “-” caso estejam em latitude sul e/ou longitude oeste. Esse pontos devem ter as coordenadas conferidas e ter pouca ou nenhuma autocorrelação espacial.

ii) *Variáveis ambientais em .asc:* As variáveis deverão estar cortadas para a área de estudo/*background*, todas reamostradas para a mesma resolução de interesse e com pouco ou nenhuma autocorrelação ambiental. Deixe todos as camadas que serão usadas na mesma pasta.

iii) *Bias File:* Se for usar o Bias Grid como forma de selecionar os pontos de *background*.

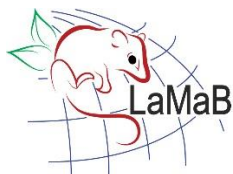
iv) *Variáveis ambientais em .asc com outro recorte temporal ou espacial:* Se você for trabalhar com mudanças climáticas (modelagem no futuro), paleomodelagem (modelagem no passado), espécies invasoras (modelar em outras áreas) ou projetar em uma área maior (pontos concentrados em uma determinada área, uso de *background* diferente) é necessário ter as mesmas camadas usadas no item ii, **COM O MESMO NOME E RESOLUÇÃO**, só que em pastas diferentes. Se algumas das variáveis estiverem com nomes ou resoluções diferentes o programa dará erro dizendo que não conseguiu achar a camada tal.

b) *Parâmetros do MaxEnt:* Abra o arquivo **maxent.jar** (você deve ter a última versão do Java instalado no seu computador para ele executar perfeitamente).



Em **SAMPLES** escolha e carregue os pontos de ocorrência clicando no arquivo .csv.

Em **ENVIRONMENTAL LAYERS** carregue **SEMPRE** a pasta das variáveis ambientais do **PRESENTE** em **ASCII**, mesmo em paleomodelagem ou modelagem do futuro. O MaxEnt vai carregar todos os arquivos .asc que estiverem na pasta. Somente as variáveis ambientais selecionadas serão



usadas nos modelos, você desmarcar uma variável se for o caso. Você deve indicar indique quais variáveis são contínuas e quais são categóricas (ver parte 1 do tutorial).

Clique em **CREATE RESPONSE CURVES** e **DO JACKKNIFE TO MEASURE VARIABLE IMPORTANCE**. Com isso o MaxEnt criará um gráfico para cada variável mostrando o comportamento da probabilidade de ocorrência da espécie com o aumento do valor das variáveis e além disso calculará a importância de cada variável para o modelo.

OUTPUT DIRECTORY: escolha a pasta de saída, onde serão salvos os modelos.

PROJECTION LAYERS DIRECTORY/FILE: Escolha a pasta onde estão as camadas ambientais do passado/futuro/outra área na qual você quer projetar, se for o caso.

FEATURE CLASS (FC): AUTO FEATURES.

Há uma forma de diminuir a complexidade dos modelos, selecionando o um **FEATURE CLASS (FC)** e um valor **REGULARIZATION MULTIPLIER (RM)** específico levando em conta o número de pontos de ocorrência e as variáveis ambientais. Há um pacote do R “ENMeval” que calcula esses valores automaticamente (Muscarella et al. 2014). Merow et al. 2014 e os tutoriais e os arquivos de ajuda oficiais do MaxEnt explicam o que seriam e como atuam os FC e o RM nos modelos. Um outro tutorial para o uso do pacote “ENMeval” está em fase final de edição.

Ao usar **AUTO FEATURES**, todos os FC são usados quando há pelo menos 80 amostras para a etapa de treino (parte 1 do tutorial). De 15 a 79 amostras, são utilizadas FC linear, quadratic e hinge; De 10 a 14 amostras, são utilizadas FC linear e quadratic. Abaixo de 10 amostras, apenas FC linear são usadas. Os valores *default* para beta (que será multiplicado pelo RM antes de ser aplicado nos modelos) de vários FC são dados nas tabelas a seguir:

Linear	
Tamanho Amostral	Beta
0	1
10	1
30	0.2
100+	0.05

Linear + Quadratic	
Tamanho Amostral	Beta
0	1.3
10	0.8
17	0.5
30	0.25
100+	0.05

Linear + Quadratic + Product	
Tamanho Amostral	Beta
0	2.6
10	1.6
17	0.9
30	0.55
100+	0.05

Categorical	
Tamanho Amostral	Beta
0	0.65
10	0.5
17+	0.25

Hinge	
Tamanho Amostral	Beta
0+	0.5

Threshold	
Tamanho Amostral	Beta
0	2
100+	1

Depois de escolher o FC clique em **SETTINGS**. Uma caixa de diálogo irá se abrir. Se você tiver 14 pontos de ocorrência ou mais para construir os modelos, siga esses passos:



Na aba **BASIC**:

Clique em todas as opções, **EXCETO** em **Skip if output exists**.

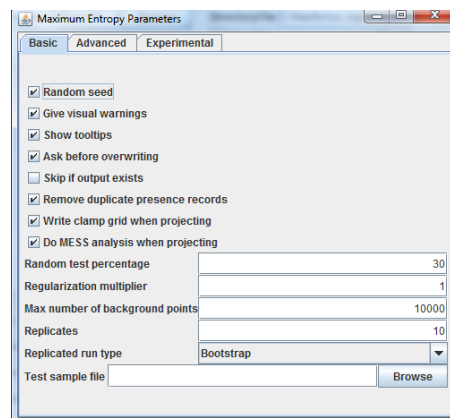
RANDOM TEST PERCENTAGE: 30. Porcentagem dos pontos que serão usados na etapa de teste.

REGULARIZATION MULTIPLIER (RM): 1. Altera a complexidade do modelo a fim de aumentar ou diminuir o sobreajuste (*overfitting*), quanto maior o número, menor será o sobreajuste e erro de omissão, em contrapartida aumentará o erro de sobreprevisão. Há trabalhos que testaram os modelos com os números 2 e 4. E há pacotes no R que testam a partir de seus pontos e camada qual seria o valor ideal para modelos com complexidade e sobreajuste ideais (Muscarella et al. 2014).

MAX NUMBER OF BACKGROUND POINTS: 10000.

REPLICATES: 10. Alguns trabalhos fazem com 15, 30 ou 50 replicações.

REPLICATED RUN TYPE: BOOTSTRAP. Alguns trabalhos usam **CROSSVALIDATE** que seria equivalente ao *Jackknife*. Assim o número de replicações seria $n-1$. Assim se você tem 45 pontos terá 44 replicações e deve digitar 45 ao invés de 10 em **REPLICATES**.



Se você tiver entre 8 e 14 pontos de ocorrência:

Na aba **BASIC**:

Clique em todas as opções, **EXCETO** em **Skip if output exists**.

RANDOM TEST PERCENTAGE: 0.

REGULARIZATION MULTIPLIER: 1.

MAX NUMBER OF BACKGROUND POINTS: 10000.

REPLICATES: COLOCAR O NÚMERO DE AMOSTRA, se tem 9 pontos, coloque 9.

REPLICATED RUN TYPE: CROSSVALIDATE (Equivalente ao *Jackknife*).

Na aba **ADVANCED**: igual para quem tem muitos ou poucos pontos de ocorrência.

Clicar em todas as opções, **EXCETO** em **Add all samples to background** e **Append summary results to maxentResult.csv file**.

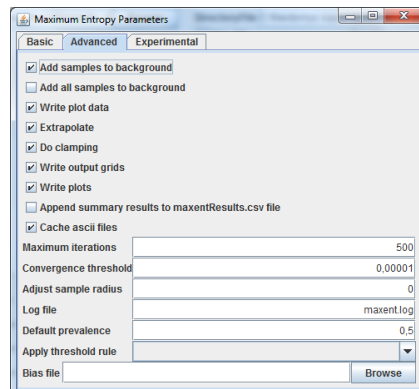


LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



MAXIMUM ITERATIONS: 5000.

Se você criou um *Bias Grid*, acrescente-o na última caixa de diálogo, chamada **BIAS FILE**.



Na aba **EXPERIMENTAL**, deixar o *default*, não mexer nela, nela tem os valores de beta para serem editados. **NÃO** tem botão de **OK**, só precisa fechar a caixa de diálogo que as configurações serão salvas.

Quando voltar para a janela inicial do MaxEnt clique em **RUN**. A análise começará, mas normalmente acontecem alguns erros e avisos.



2) Erros mais comuns ao executar o MaxEnt

a) *Tamanho do pixel (Cell size) fora do padrão:* é necessário reamostrar as camadas (**RESAMPLE**). O MaxEnt não diz qual das camadas está errada, tem que verificar cada uma delas ou ir no **CHECK HEADERS ASC FILES**;

b) *Número de linhas e colunas ou alinhamento diferentes:* Mesma coisa do passo i;

c) *Pontos fora das camadas ambientais:* Mesmo que você tenha verificado todos os pontos, um a um, sempre é possível que um ponto na borda das camadas, devido à reamostragem, vá para fora da área de estudo/*background*, um aviso aparecerá dizendo que ele será descartado. Só dizer que sim, que a análise continuará, anote a posição e o número de pontos fora, esses pontos não serão usados na construção do modelo e por isso deverão ser excluídos do material e métodos e do *input samples* que será disponibilizado para outros pesquisadores.

d) *Caminhos longo demais:* Se o aviso de erro diz que tal pasta não existe, é porque o local onde ela está guardada é longo demais e o programa não consegue achá-la, tente mover a pasta para o **C:**. Esse aviso de que a pasta não existe também aparece quando você usa vírgula, pontos, espaço ou outro acento gráfico no nome das pastas que servem de caminho para chegar às camadas. Exemplo, ter uma pasta chamada: **lgm_2,5min** referente a resolução em minutos. Essa pasta com vírgula não será lida pelo MaxEnt.

e) *Não foi possível achar uma variável ambiental quando está projetando:* O nome ou a extensão das variáveis usadas na projeção não são os mesmos que as do presente. Tente renomear as variáveis da projeção com os mesmos nomes das variáveis do presente, o erro pode ser em um espaço a mais no fim do nome.

ATENÇÃO: Sempre que der erro ao executar o MaxEnt **NUNCA** use a opção **REDO** ou **REDO ALL**, que irá reescrever os arquivos novos por cima dos velhos. Se você fizer isso alguns arquivos importantes ficarão em branco. **SEMPRE** apague o output e a pasta antiga e crie uma nova.

No fim, se tudo der certo, a barra que marca as porcentagens das análises irá sumir em determinado momento e a janela do MaxEnt com os mesmos parâmetros que você escolheu continuará aberta. É nessa hora que a análise acabou. Feche e vamos à última etapa, a de validação dos modelos.



V) Pós-Análise

1) Verificar a qualidade do modelo estatisticamente

Depois que a análise terminar, não vá ver os mapas, primeiramente é necessário avaliar o arquivo **maxentResults.csv** na pasta onde você salvou o *output*. Nesse arquivo estarão os dados estatísticos que demonstrarão que seu modelo é melhor que o acaso. Se a estatística não for significativa, descarte o modelo de imediato. Se for significativo aí sim, você analisará o mapa e verá se ele tem sentido biológico. Se o mapa não fizer sentido biológico (muitos erros do tipo II - omissão), descarte-o também.

Primeiramente abra o arquivo **maxentResults.csv**, ele está na pasta onde você salvou o *output* do modelo. Nesse arquivo há várias colunas referentes às análises dos modelos que foram criados. Lembre-se que se você pediu para que fossem criados 10 réplicas, cada linha representará uma dessas réplicas e a última linha será o modelo médio (*avg* ou **average**), a média de todas as 10 réplicas. Essa é a linha mais importante, só ela que nos interessa para a verificação do modelo.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Species	#Training samples	Regularized training gain	Unregularized training gain	Iterations	Training AUC	#Test samples	Test gain	Test AUC	AUC Standard Deviation	#Background points	bio_11 cor
2	squamipes_0	45	0.7087	0.9743	500	0.8784	19	1.1128	0.8678	0.0395	10038	
3	squamipes_1	45	1.0995	1.4201	500	0.9218	19	1.1113	0.8664	0.043	10033	
4	squamipes_2	45	0.993	1.4444	500	0.9363	19	1.4603	0.9231	0.0223	10040	
5	squamipes_3	45	1.2061	1.5412	500	0.9202	19	1.0313	0.8729	0.0306	10038	
6	squamipes_4	45	1.0003	1.253	500	0.8986	19	1.0329	0.8836	0.0279	10038	
7	squamipes_5	45	0.8147	1.0759	480	0.8805	19	1.2214	0.8623	0.0534	10039	
8	squamipes_6	45	0.8508	1.0496	300	0.8845	19	1.52	0.9243	0.0325	10040	
9	squamipes_7	45	0.9261	1.2565	500	0.9041	19	1.6551	0.9373	0.0202	10036	
10	squamipes_8	45	1.0191	1.3526	500	0.9139	19	1.1461	0.8749	0.04	10036	
11	squamipes_9	45	1.0063	1.3164	500	0.8903	19	1.1853	0.8528	0.043	10035	
12	squamipes (average)	45	0.9625	1.2684	478	0.9029	19	1.2476	0.8865	0.0351	10037.3	
13												
14												
15												

As colunas que você precisa verificar e recuperar os valores são:

#Training samples: Quantos pontos de ocorrência foram usados para criar os modelos.

#Test samples: Quantos pontos de ocorrência foram usados para validar os modelos.

Test AUC: Valores abaixo de 0.7 são ruins, não podem ser usados e valores próximos a 0.7 também tem que ser melhorados, maior AUC, melhor o modelo.

AUC Standard Deviation: desvio-padrão da AUC.

bio_1 contribution: A contribuição de cada variável para a criação dos modelos, terão tantas colunas quanto forem as camadas ambientais usadas para a criação do modelo.

bio_12 contribution

bio_13 contribution

bio_14 contribution

bio_18 contribution

bio_2 contribution

bio_7 contribution

Minimum training presence logistic threshold: Valor do *threshold* mínimo.



Minimum training presence test omission: erro de omissão do modelo para esse *threshold* (erro tipo II – até 0.2, é aceitável, melhor próximo de 0.15 ou menor que isso).

Minimum training presence binomial probability: p-valor do modelo, se esse valor for maior que 0.05, descarte esse *threshold*. Um p-valor não-significativo pode indicar que você tem poucos pontos de ocorrência. Se nenhum *threshold* for significativo não poderá usar o modelo. O melhor modelo é aquele que tiver os três *thresholds* mais usados com valores significativos. O *threshold* mínimo é sempre significativo e com baixa omissão, por isso ele é pouco utilizado. Normalmente se utiliza o *threshold* de 10%, veja a parte 1 do tutorial para entender o significado dos *threshold*.

10 percentile training presence logistic threshold: Valor do *threshold* de 10%.

10 percentile training presence test omission: erro de omissão do *threshold*.

10 percentile training presence binomial probability: p-valor do *threshold*.

Maximum test sensitivity plus specificity logistic threshold: Valor do *threshold* máximo.

Maximum test sensitivity plus specificity test omission: erro de omissão do *threshold*.

Maximum test sensitivity plus specificity binomial probability: p-valor do *threshold*.

Se nenhum dos *threshold* deu significativo, com baixa omissão ou apenas o *threshold* mínimo satisfaz essas premissas, descarte o modelo, ele não é bom. Nem o projete no ArcGIS para você não se sentir culpado na hora de deletá-lo. **NUNCA** fique com o primeiro modelo criado, mesmo que ele tenha AUC maior que 0.7, baixa omissão e p-valor significativo. Sempre é possível melhorá-lo, siga as dicas do abaixo tópico e crie vários modelos e veja qual satisfaz você estatisticamente e biologicamente. Eu rodei cerca de 60 modelos diferentes com camadas ambientais diferentes, com os três tipos de rarefação de dados bióticos, com aumento no número de replicações, alterando os valores de RM e FC para conseguir um modelo mais robusto. Faça uma tabela com esses valores estatísticos, assim fica mais fácil compará um grande número de modelos. Algumas opções para melhorar os modelos que estatisticamente ou biologicamente não são bons:

i) *Mudar a forma de seleção de variáveis ambientais:* Nesse tutorial mostrei duas formas, mas existem outras. É só pesquisar e justificar o uso dessa nova forma de seleção;

ii) *Acréscitar mais variáveis:* adicionar variáveis que você não usou no conjunto inicial e que sejam importantes para a espécie que você está estudando. Crie novos *shapefiles* que podem ser usados como variáveis ambientais. Por exemplo, pegue o *shapefile* de hidrografia e crie um *raster* de distância da água a partir da ferramenta **EUCLIDEAN DISTANCE**, ou modele a principal presa/alimento da espécie-alvo e entre com esse dado como nova variável ambiental;

iii) *Aumente número de replicações:* na aba **BASIC** do MaxEnt, na opção **REPLICATES** colocamos 10 replicações, mas se a AUC for muito baixa, aumente para 15 ou 20 replicações e veja se melhora, normalmente essa técnica aumenta o valor da AUC;

iv) *Mudar a forma de rarefação:* modele com os dois tipos de controle de viés amostral e veja qual deles tem a melhor resposta para o modelo;



v) *Evite o uso de Bias Grid*: O *Bias Grid* normalmente diminui o valor da AUC e como é algo bem criticado na literatura pode ser uma opção não utilizar a técnica.

vi) *Acrescente pontos de ocorrência novos*: Verifique os gráficos do espaço ambiental (**tópico III.7**) e descubra se novos pontos de ocorrência da espécie-alvo são necessários para melhorar o amostragem do espaço ambiental. Tente adquirir pontos novos em áreas que você ainda não tem registros de ocorrência;

vii) *Altere o FC e o RM*: use o pacote do R descrito por Muscarella et al. (2014) e ache o melhor valor de FC e RM para o seu conjunto de dados.

viii) *Troque de algoritmo*: Se o MaxEnt, depois de todas essas mudanças, não conseguiu um modelo bom, tente outro programa com outro algoritmo. A forma de criar o *input* de camadas ambientais e de pontos de ocorrência são os mesmos desse tutorial, no outro algoritmo só muda a etapa de modelagem (**tópico IV**), os parâmetros que são necessários para o algoritmo executar.

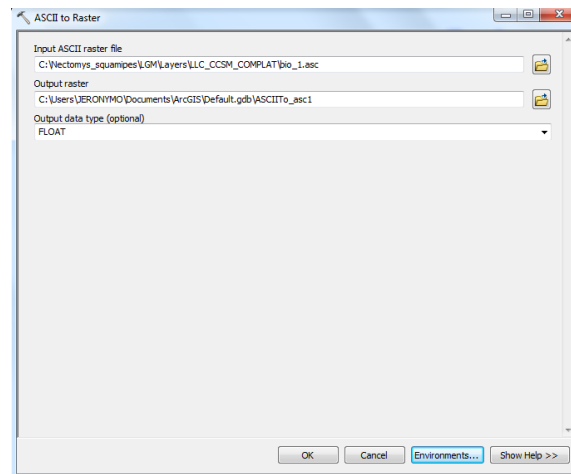
ix) *Desista momentaneamente*: Se mesmo trocando de algoritmo o modelo não é bom, é porque você não tem dados (variáveis ambientais ou pontos de ocorrência) que representem o nicho/distribuição da espécie-alvo. Desista da modelagem enquanto novos dados não surgirem.



2) Cortar o modelo pelo threshold

Uma vez escolhido o modelo, você deverá cortá-lo com os três valores de *threshold*. Existem vários outros valores de *threshold* na planilha **maxentResults.csv**, mas esses três são os mais usados em publicações. O *threshold* é um limite de corte que separará os valores de pixels do seu modelo em **ADEQUADO (SUITABLE)** e **INADEQUADO (UNSUITABLE)**. Para fazer isso você tem que converter o seu modelo para o formato *raster*. Abra o ArcGIS:

ARCTOOLBOX >> CONVERSION TOOLS >> TO RASTER >> ASCII TO RASTER



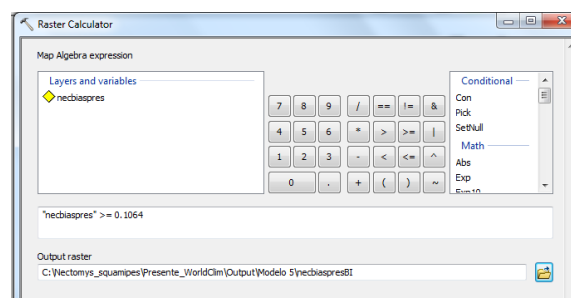
Input ASCII raster file: escolha o modelo médio gerado na pasta do *output* do MaxEnt, ele vai ter o “**nome_da_especie_avg.asc**”. Quando se projeta o modelo para o passado, futuro ou outra área, há dois arquivos *.avg*, um só com o **nome_da_especie_avg.asc** e outro com **Nome_da_especie_Nome_da_pasta_das_camada_projetadas_avg.asc** você deve usar o com nome da pasta, o primeiro é o modelo gerado no presente ou na área original.

Output raster: nome para o modelo e em formato *raster* e onde vai salvá-lo.

Output data type: FLOAT.

Com o modelo em *raster* projetado no ArcGIS, abra o **RASTER CALCULATOR**:

ARCTOOLBOX >> SPATIAL ANALYST TOOLS >> MAP ALGEBRA >> RASTER CALCULATOR



Você criará uma fórmula: “**MODELO_RASTER**” $\geq$ **VALOR_THRESHOLD**.

Escolha o nome e onde salvar o mapa, acrescente ao nome se esse será o *threshold* mínimo, máximo ou de 10%. Dê **OK**. Esse é o **MAPA BINÁRIO**, com pixels inadequados para a presença da

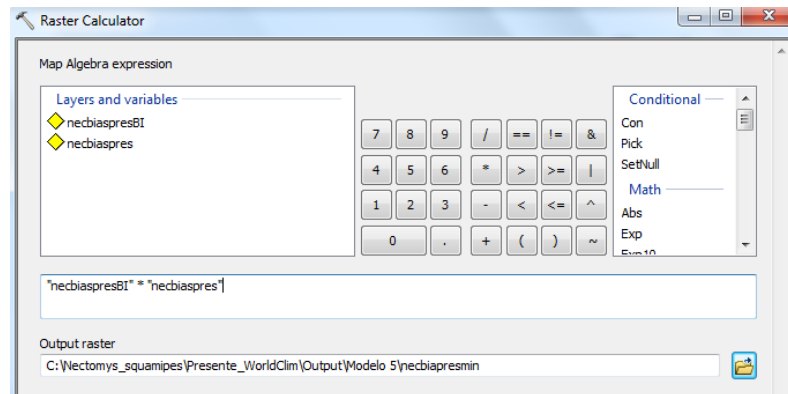


LaMaB: Laboratório de Mastozoologia e Biogeografia

Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



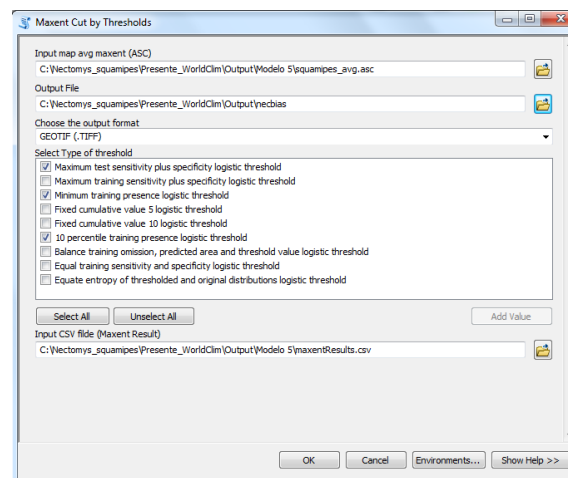
espécie com valor 0 e com pixels adequados com valores 1. Abra o **RASTER CALCULATOR** de novo e agora multiplique o mapa binário pelo original em *raster*, escolha um nome e salve.



Esse mapa será o mapa final, que você usará nas publicações. Acrescente os pontos que foram usados no MaxEnt no mapa e veja quais são os casos de omissão no modelo. Faça esse processo para os três *thresholds* e veja qual *threshold* é melhor para a pergunta que você quer responder.

Há uma maneira mais fácil de obter o mapa dos 3 *threshold* de uma vez. Usando a LEEC_SDM_Toolbox:

ARCTOOLBOX >> LEEC_SDM_Toolbox >> MAXENT CUT BY THRESHOLDS >> MAXENT CUT BY THRESHOLDS



Input map avg maxent (ASC): Selecionar o modelo médio do modelo.

Output file: Nome e onde salvar os mapas cortados

Choose the output format: Selecione .TIFF ou .IMG.

Select Type of threshold: Selecione os *thresholds* para o corte do modelo.

Input CSV field (Maxent Result): selecionar a planilha **maxentResults.csv** do modelo, da onde o programa tirará os valores de *thresholds*.

Nesse processo não será criado o mapa binário. Apenas um mapa com o pontos adequados para um dos *thresholds*, que terá o sufixo "**Lixo**".

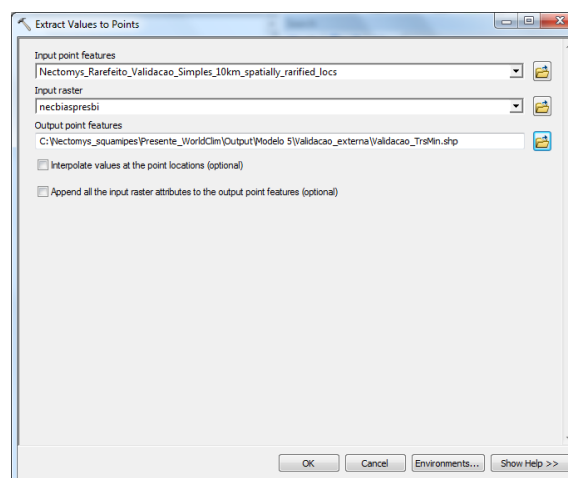


3) Validação externa

Carregue os pontos de validação externa que você separou no início da modelagem. Transforme esses pontos em *shapefile* da mesma forma que no final do **tópico III.1**. Depois faça a **RAREFAÇÃO SIMPLES**, o **tópico III.6.a** para eliminar dados redundantes da validação externa. Crie um *shapefile* dos pontos rarefeitos da validação, é com esse *shapefile* que você fará a validação.

Escolha um *threshold*, no caso eu escolhi o mínimo porque ele responde a minha pergunta. Carregue o **MAPA BINÁRIO** do *threshold* escolhido feito no **tópico V.2** e carregue também o *shapefile* dos pontos rarefeitos da validação.

ARCTOOLBOX >> SPATIAL ANALYST TOOLS >> EXTRACTION >> EXTRACT VALUES TO POINTS



Input point features: selecionar o *shapefile* de pontos de validação;

Input raster: selecionar o mapa binário;

Output point features: onde salvar e com qual nome. Deixe as opções desmarcadas e dê **OK**.

Será criado outro *shapefile* e ele será carregado no **TABLE OF CONTENTS**. Vá para onde o novo *shapefile* foi salvo, lá terá uma tabela que dá pra abrir no Excel no formato **dBASE table (.dbf)**. Clique nela e busque a coluna de valores que foi adicionada após a última coluna, uma coluna chamada **RASTERVALUE**. Nessa coluna tem valores de 0 ou 1. Salve uma cópia dessa tabela como arquivo do *MSEXcel*, você modificará esse arquivo. É comum o tamanho das colunas se modificarem no formato *.dbf*, especialmente se você o abrir no *MSEXcel*. Parece que essa coluna **RASTERVALUE** não foi adicionada. Só diminua o zoom e tamanho das colunas que você a achará.

Abra a cópia no *MSEXcel* e contabilize os valores de **0**, **1** e **NODATA (-9999)**. Faça a porcentagem de **1** (taxa de acerto) e de **0** (taxa de erro/omissão), descontando os **NODATA**:

Taxa de acerto = $[1 / (\text{N}^\circ \text{ de pontos} - \text{N}^\circ \text{ de NODATA})] * 100$

Taxa de omissão = $[0 / (\text{N}^\circ \text{ de pontos} - \text{N}^\circ \text{ de NODATA})] * 100$

Veja se ele está de acordo com as taxas que você achou na validação do MaxEnt. Se você tem um bom valor de acerto (>80%) e baixa omissão (<20%) também no teste externo, parabéns, seu modelo é muito bom!



4) Categorização da área adequada

Às vezes mais interessante do que saber se o ponto na validação caiu em área adequada ou não, é saber se ela caiu em uma área de baixa, média ou alta adequabilidade ambiental. Normalmente essas classes de adequabilidade são definidas a partir do valor dos pixels da seguinte forma:

Abaixo do valor do Threshold: **INADEQUADO**;

Do threshold até 0.5 = **BAIXA** adequabilidade;

0.5 a 0.75 = **MÉDIA** adequabilidade;

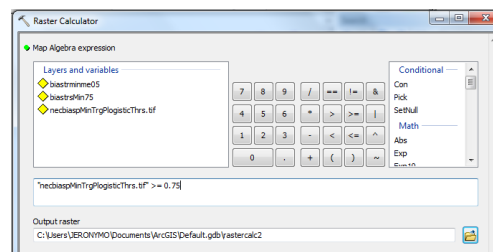
Acima de 0.75 = **ALTA** adequabilidade;

Esse é um exemplo de classificação por classes, você pode criar uma classificação própria que responda melhor a sua pergunta.

No tópico **tópico V.3** já foi calculado o número de pontos que caem acima e abaixo do *threshold*. Os valores abaixo do *threshold* são os que representam a classe **INADEQUADO**, e os valores acima correspondem a soma das classes **BAIXA**, **MÉDIA** e **ALTA**.

Para descobrir o valor específico de cada uma das classes adequadas abra no ArcGIS o mapa final para publicação (**tópico V.2**), nesse mapa a classe **INADEQUADA** está definida como **NODATA** porque o mapa já está cortado pelo *threshold*. Abra também o *shapefile* com os pontos de validação. Abra o **RASTER CALCULATOR** e faça a fórmula:

"MAPA FINAL >= 0.75". Salve o mapa.



Ele será um mapa binário onde o valor 0 dos pixels representa adequabilidade abaixo de 0.75 e valor 1 dos pixels acima ou igual a 0.75. Siga o passo-a-passo do tópico anterior (**tópico V.3**), extraindo os valores dos pontos e calculando as porcentagens de acertos da validação para o mapa. No fim você terá o número de pontos e a porcentagem caindo em pixels acima de 0.75 (**ALTA** adequabilidade).

Se você subtrair o número de pontos da categoria **ALTA** do número de pontos acima do valor de *threshold* (que seria a soma das classes **BAIXA**, **MÉDIA** e **ALTA**) você terá o número de pontos que caem dentro das categorias **BAIXA** e **MÉDIA**.

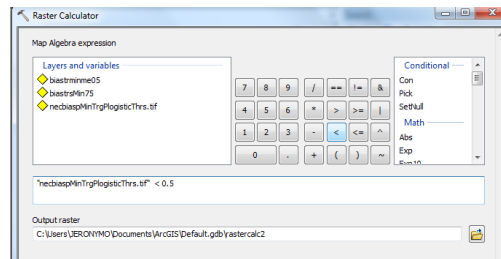
Abra novamente o **RASTER CALCULATOR** para uma segunda fórmula:

"MAPA FINAL < 0.5" e salve esse mapa também.



LaMaB: Laboratório de Mastozoologia e Biogeografia

Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



Ele será um mapa binário onde o valor 0 dos pixels representa adequabilidade abaixo ou igual a 0.5 e valor 1 dos pixels acima de 0.5. Siga o passo-a-passo do tópico anterior (**tópico V.3**), extraindo os valores dos pontos e calculando as porcentagens de acertos da validação para o mapa. No fim você terá o número de pontos e a porcentagem caindo em pixels acima de 0.5 (classe **ALTA + MÉDIA**).

Como você já calculou a classe **ALTA** anteriormente, basta subtrair o número de pontos da classe **ALTA** do número de pontos com valores de pixel iguais a 1, que você descobre o número de pontos que caem na classe **MÉDIA**. Com o número de pontos das classes **MÉDIA** e **ALTA**, basta subtrair esses números do número de pontos que caem na área adequada (**tópico V.3**) que se encontra o número de pontos da classe **BAIXA**. A partir daí é só calcular as porcentagens.



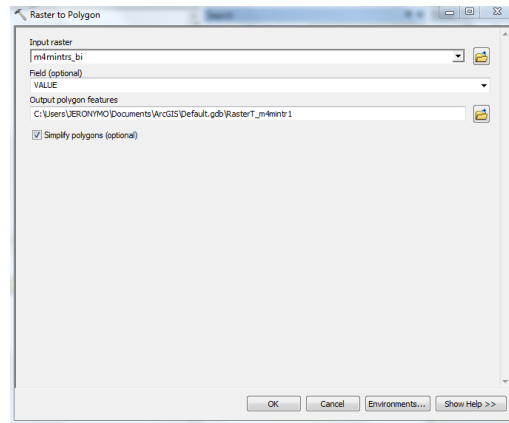
5) Calcular o tamanho da área adequada de um modelo

Crie um novo projeto no ArcGIS e mude a projeção das camadas:

Clique com o botão direito em **LAYERS >>PROPERTIES >> COORDINATED SYSTEM GEOGRAPHIC COORDINATE SYSTEMS >> WORLD >> WGS1984 >> OK**

Carregue o **MAPA BINÁRIO** cortado pelo *threshold* de preferência e o transforme em polígono:

ARCTOOLBOX >> CONVERSION TOOLS >> FROM RASTER >> RASTER TO POLYGON



Input raster: MAPA BINÁRIO.

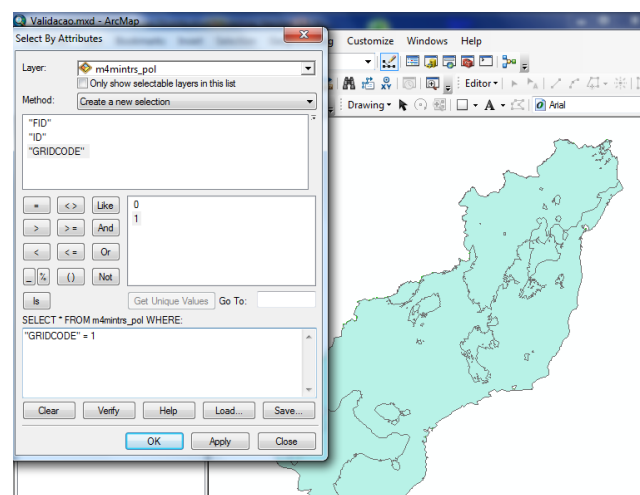
Field (optional): VALUE.

Output polygon features: local e nome salvar o polígono. Dê **OK**.

O polígono criado terá 2 valores: 0 e 1.

TABLE OF CONTENTS >> clique com o botão direito no *shapefile* criado e escolha **OPEN**

ATTRIBUTE TABLE >> ir ao ícone  e **SELECT BY ATTRIBUTE**, e a janela abaixo irá abrir:



Method: Selecione **CREATE A NEW SELECTION**;



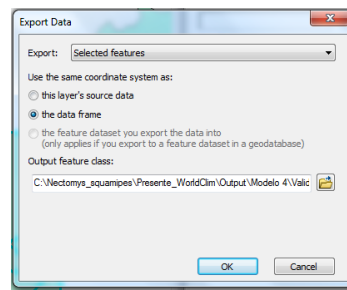
LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



Na caixa de diálogo abaixo selecione o cabeçalho da coluna que contenha os valores dos pixels (provavelmente “GRIDCODE”) e clique em **GET UNIQUE VALUES**. Os valores dos pixels (0 e 1) serão adicionadas. Crie uma fórmula explicando que você quer selecionar só os valores iguais a 1: “GRIDCODE” = 1 e clique em **APPLY**.

No polígono as áreas com valor 1 foram selecionadas. Essas áreas apresentam agora com um contorno diferente. No **TABLE OF CONTENTS** >> clique com o botão direito do *mouse* no polígono e escolha **DATA** >> **EXPORT DATA**.

Exporte a seleção (**SELECTED FEATURES**) como um novo *shapefile* com o mesmo sistema de coordenada que o **DATA FRAME**, queremos em **WGS1984**. Uma pergunta surgirá, pedindo se você quer projetar esse *shapefile* exportado. Dê **OK**. Veja se só a área adequada foi realmente usada para criar o novo polígono.

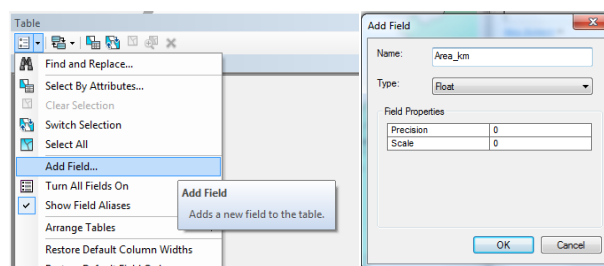


Crie um novo projeto no ArcGIS e mude a projeção das camadas:

Clique com o botão direito do *mouse* em **LAYERS** >> **PROPERTIES** >> **COORDINATED SYSTEM** >> **PROJECTED COORDENATE SYSTEM** >> **CONTINENTAL** >> **SOUTH AMERICA** >> **SOUTH AMERICA ALBERS EQUAL AREA CONIC** >> **OK**

Abra o polígono da área adequada (de valor 1) que você fez no passo anterior.

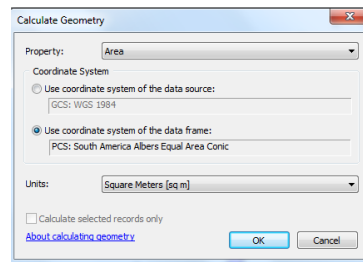
Vá ao **TABLE OF CONTENTS** >> clique com o botão direito do *mouse* no polígono e escolha **OPEN ATTRIBUTE TABLE** >> ir ao ícone  e **ADD FIELD**, e a janela abaixo irá abrir:



Dê um nome à nova coluna (sem acentos ou espaços), por exemplo, *Area_Km* e em **TYPE** escolha **FLOAT** e depois dê **OK**. Clique com o botão direito do *mouse* no cabeçalho da nova coluna (em cima do nome *Area_Km*) e vai a **CALCULATE GEOMETRY**.



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



Property: AREA.

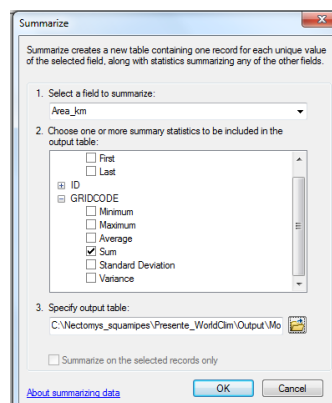
Coordinate System: Selecione o Albers que é em metros.

Units: SQUARE KILOMETERS [SQ KM].

Dê **OK**.

Veja se a coluna que você adicionou e nomeou de *Area_Km* mudou e agora tem valores numéricos. Se há números e não apenas 0 está tudo bem. Se só apareceu 0 nas células é porque você não colocou **WGS1984** como a projeção do polígono ou o **ALBERS** como projeção para o **LAYERS**.

Clique com o direito no cabeçalho da coluna da área e clique em **SUMARIZES**.



Escolha o campo **Area_Km** para sumarizar. Escolha a opção **SUM** nas opções estatísticas e escolha um nome, local e o formato **dBASE table (.dbf)**. Dê **OK**. Aparecerá uma pergunta sobre adicionar o resultado no mapa, você escolha **NÃO**.

Abra o arquivo *.dbf* criado no *MSEXcel* ou similar e lá terá a área total adequada para o mapa binário cortado com o seu *threshold*. Você pode calcular a área dos mapas do passado e do futuro para saber se as áreas adequadas variaram, ou irão variar, em tamanho com o tempo. Ou com a mesma ferramenta você pode descobrir quantos km² de área adequada do modelo se encontram dentro de Unidades de Conservação.



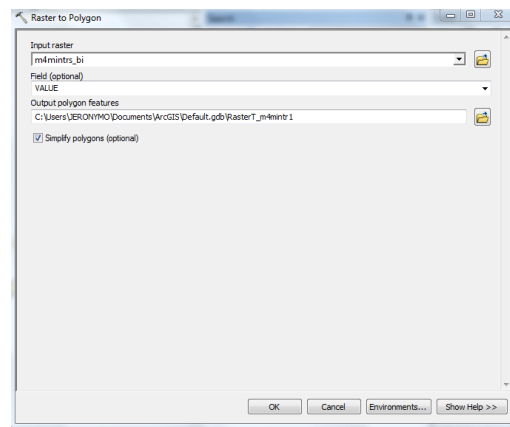
6) Cálculo da matriz de confusão e de outros índices estatísticos

A matriz de confusão será construída a partir dos pontos de validação externa (ou se não tiver, dos pontos que você usou no MaxEnt mesmo - os pontos de presença) e de pontos tomados de forma aleatória (os pontos de “pseudoausência”, já que não temos ausências verdadeiras). Depois de construir a matriz de confusão podemos verificar se o modelo prevê a distribuição da espécie melhor que o acaso.

Carregue o **MAPA BINÁRIO** cortado pelo *threshold* de preferência e clique com o botão direito do *mouse* nele e escolha **OPEN ATTRIBUTE TABLE** e anote os valores de 0 e 1 que estão na tabela.

Transforme o mapa binário em polígono:

ARCTOOLBOX >> CONVERSION TOOLS >> FROM RASTER >> RASTER TO POLYGON



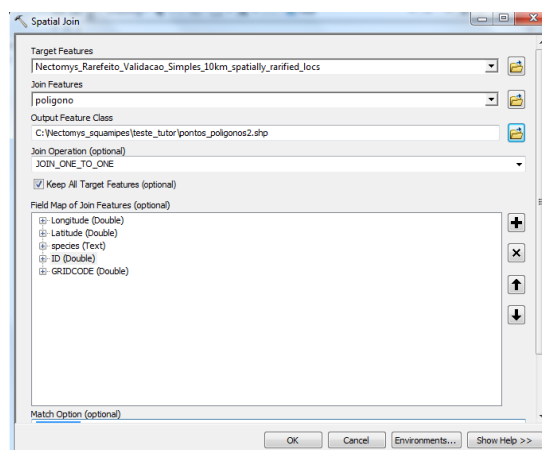
Input raster: MAPA BINÁRIO.

Field (optional): VALUE.

Output polygon features: local e nome salvar o polígono. Dê OK.

O polígono criado terá 2 valores: 0 e 1. Por meio da função **SPATIAL JOIN**, combine o *shapefile* dos pontos de ocorrência utilizados no MaxEnt ou da validação externa com o polígono do mapa binário.

ARCTOOLBOX >> ANALYSIS TOOLS >> OVERLAY >> SPATIAL JOIN





Target Features: o *shapefile* dos pontos de ocorrência ou validação.

Join Features: Polígono do mapa binário.

Output Feature Class: onde salvar e com que nome.

Demais campos deixar como o estão. Dê **OK**.

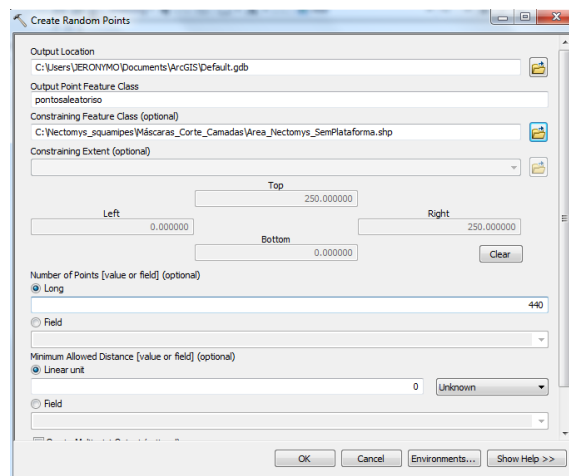
ATENÇÃO! Não se esqueça de retirar os pontos que deram **NODATA** quando você fez a validação externa do modelo (**tópico V.3**) ou os pontos que o MaxEnt não usou no criação do modelo por eles caírem fora das camadas (**tópico IV.2**). Se você colocar esses pontos a taxa de erro aumentará.

TABLE OF CONTENTS >> clique com o botão direito no *shapefile* criado e escolha **OPEN**

ATTRIBUTE TABLE >> ir ao ícone  e **EXPORT**

Exporte **ALL RECORDS** no formato **dBASE table (.dbf)**. Conte quantos registros foram 0 e quantos foram 1. Pronto agora você tem o número de acertos e erros com os pontos presença. Agora vamos calcular os acertos e erros com pontos aleatórios de ausência ("pseudoausência").

ARCTOOLBOX >> DATA MANAGEMENT TOOLS >> FEATURE CLASS >> CREATE RANDOM POINTS



Output Location: o local de saída do *shapefile* dos pontos aleatórios.

Output Point Feature Class: o nome do arquivo a ser criado.

Constraining Feature Class (optional): utilize o *shapefile* da máscara que você usou para cortar as camadas.

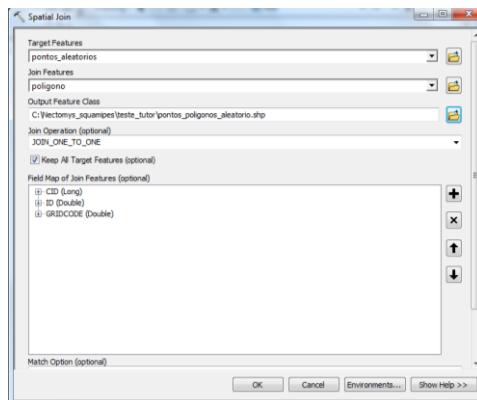
Number of Points: marque a opção **LONG** e coloque um valor na proporção **10:1** de seus pontos verdadeiros. Por exemplo, eu tenho 44 pontos de validação externa, então eu tenho que usar **440** pontos.

Deixe no *default* os outros campos e dê **OK**.



Agora você usará o **SPATIAL JOIN** de novo e repetirá os mesmos passos feitos anteriormente com os pontos de presença, mas agora com esses pontos aleatórios.

ARCTOOLBOX >> ANALYSIS TOOLS >> OVERLAY >> SPATIAL JOIN



Target Features: o *shapefile* dos pontos aleatórios.

Join Features: Polígono do mapa binário.

Output Feature Class: onde salvar e com que nome.

Demais campos deixar como o estão. Dê **OK**.

Exporte a tabela da mesma forma que anteriormente e contabilize o número de 0 e 1 desses 440 pontos. No fim desse passo você deverá estar com esses dados em mãos:

VALOR	MAPA BINÁRIO	PRESENÇA (VALIDAÇÃO)	AUSÊNCIA (ALEATÓRIOS)
0	47948	3	171
1	77058	41	269

Com esses valores, podemos agora avaliar estatisticamente o modelo gerado. Os valores de 1 e 0 obtidos dos pontos validação/presença serão, respectivamente, os valores de A e C da matriz de confusão, enquanto aqueles dos pontos aleatórios serão D e B. Agora podemos calcular a Sensibilidade, Especificidade, Acurácia e TSS do modelo.

Matriz de Confusão

	Presença real	Ausência real
Presença prevista	a	b
Ausência prevista	c	d

a = verdadeiro positivo
b = falso positivo
c = falso negativo
d = verdadeiro negativo

a e d são acertos
b e c são erros

taxa de falso positivo = erro de sobreprevisão = $b/(b+d)$

taxa de falso negativo = erro de omissão = $c/(a+c)$

sensitivity = $a/(a+c)$

specificity = $d/(b+d)$



Substituindo os valores fica assim:

	PRESENÇA REAL	AUSÊNCIA REAL
PRESENÇA PREVISTA	<i>a</i>	<i>B</i>
AUSÊNCIA PREVISTA	<i>c</i>	<i>D</i>

	PRESENÇA REAL	AUSÊNCIA REAL
PRESENÇA PREVISTA	41	171
AUSÊNCIA PREVISTA	3	269

TAXA DE ACERTOS OU ACURÁCIA: $(a + d)/(a+b+c+d)$

TAXA DE ERROS: $(b + c)/(a+b+c+d)$

ERRO DE SOBREPREENSÃO: $b/(b+d)$

ERRO DE OMISSÃO: $c/(a+c)$

SENSIBILIDADE: $a/(a+c)$

ESPECIFICIDADE: $d/(b+d)$

TRUE SKILL STATISTICS (TSS): (Sensibilidade + Especificidade) – 1.

A Sensibilidade representa a ausência do Erro de Omissão (erro tipo II). A Especificidade representa a ausência do Erro de Sobreprevisão (erro tipo I). Para mais detalhes sobre a matriz de confusão e seus índices leia a parte 1 desse tutorial.

A Acurácia representa a capacidade do modelo de classificar corretamente tanto as ausências quando presenças verdadeiras. O **TRUE SKILL STATISTICS** (TSS) é uma medida alternativa à Acurácia. É similar ao **ÍNDICE KAPPA**, mas este é muito questionado pela sua dependência de outros fatores e geração de artefatos estatísticos. O TSS carrega as vantagens do Kappa sem os problemas de dependência. Valores de TSS acima de 0.4 são aceitáveis, quanto mais próximo de 1 melhor.

O teste binomial será o último teste realizado para a validação do modelo. Tome os valores de 1 e 1+0 dos pontos verdadeiros, e nomeie-os de **x1** e **x2**, respectivamente. No exemplo acima: **x1=41** e **x2=44**. Tome também os valores de 1 e 1+0 do mapa binário, e os chame de **y1** e **y2**. No exemplo acima: **y1=77058** e **y2=125006**. De posse desses valores, abra o R e digite o comando substituindo os X e Y pelos números:



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



```
prop.test(c(x1,y1), c(x2,y2))  
prop.test(c(41,77058), c(44,125006))
```

O resultado do R virá assim:

```
2-sample test for equality of proportions with continuity correction  
data:  c(41, 77058) out of c(44, 125006)  
X-squared = 17.1955, df = 1, p-value = 3.372e-05  
alternative hypothesis: two.sided  
95 percent confidence interval:  
 0.2294905 0.4012771  
sample estimates:  
   prop 1    prop 2  
0.9318182 0.6164344
```

Um p-valor abaixo de 0.05 indica que o modelo difere significativamente do acaso. Os índices do meu modelo foram:

TAXA DE ACERTOS OU ACURÁCIA: 0.6405 (melhor próximo de 1);

TAXA DE ERROS: 0.3595 (melhor próximo de 0);

ERRO DE SOBREPREENSÃO: 0.3886 (melhor próximo de 0);

ERRO DE OMISSÃO: 0.073 (melhor próximo de 0);

SENSIBILIDADE: 0.9318 (melhor próximo de 1);

ESPECIFICIDADE: 0.6114 (melhor próximo de 1);

TRUE SKILL STATISTICS (TSS): 0.5432 (melhor próximo de 1);

TESTE BINOMIAL: <0.01 (melhor abaixo de 0.05).

O modelo está pronto e validado internamente pela estatística do MaxEnt e externamente pela matriz de confusão, além da de biologicamente pelo seu conhecimento sobre a espécie-alvo. Mesmo que um modelo seja ótimo estatisticamente não significa que ele faz sentido biológico. Em suma, um modelo só será aceito para as publicações se fizer sentido biológico e estatístico.



7) Gráficos de probabilidade de ocorrência para as variáveis ambientais

O MaxEnt cria gráficos que mostram como a probabilidade de ocorrência da espécie se comporta com o aumento de cada variável. Para as explicações biológicas é necessário analisar esses gráficos. Na pasta onde você salvou o *output* dos modelos tem uma pasta chamada **PLOT**. Dentro dela há várias imagens e arquivos. Entre as imagens estão estes gráficos para cada variável do modelo, mas essas imagens estão em baixa resolução e não servem para publicações, teses e dissertações. É preciso criar novas imagens desses gráficos e os arquivos com os dados brutos são os arquivos com extensão **.dat**.

Contudo há vários desses arquivos, uma para cada replicação e para cada variável utilizada. Você deve usar o arquivo do modelo médio, só das variáveis que mais contribuíram para o modelo e que tenha a terminação **_only.dat**. Exemplo: *bio_8* foi a minha variável mais importante, então vou procurar o arquivo *.dat* dela na pasta **PLOT**. Há várias como, por exemplo:

```
squamipes_0_bio_8_only.dat - Oriundo da replicação 0.  
squamipes_0_bio_8.dat  
squamipes_1_bio_8_only.dat - Oriundo da replicação 1.  
squamipes_1_bio_8.dat  
squamipes_2_bio_8_only.dat Oriundo da replicação 2.  
squamipes_2_bio_8.dat
```

Aquele que você precisa é o oriundo do modelo médio, sem numeração após o nome da espécie, e com **_only.dat** na terminação.

```
squamipes_bio_8_only.dat
```

O arquivo é separado por vírgula, como um *.csv* e pode ser aberto no bloco de notas ou no *MSExcel*. E tem das informações abaixo:

```
variable,x,y  
bio_8,110.4,0.15879391357302666  
bio_8,110.57520000000001,0.15879391357302666  
bio_8,110.7504,0.15879391357302666  
bio_8,110.9256,0.15879391357302666  
bio_8,111.1008,0.15879391357302666  
bio_8,111.27600000000001,0.15879391357302666  
bio_8,111.4512,0.15879391357302666  
bio_8,111.6264,0.15879391357302666
```

Copie os dados para um programa que faça gráficos de dispersão (*MSExcel*, *PAST*, *R*) e edite da forma que quiser. Faça isso, pelo menos para as 3 variáveis que mais contribuem para o modelo. Abaixo tem um exemplo de código no *R* para exportar os gráficos em dois formatos. Não é preciso instalar nenhum pacote.



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



```
#importar os dados do arquivo .dat
bio02=read.csv("squamipes_bio_2_only.dat")
bio18=read.csv("squamipes_bio_18_only.dat")
bio13=read.csv("squamipes_bio_13_only.dat")

#verificar se a importação foi correta
head(bio02)
head(bio18)
head(bio13)

#criar os gráficos em .jpeg
#para o código funcionar lembre de trocar os nomes das variáveis
#e de trocar o nome e unidades dos eixos e o titulo
jpeg(filename = "nome_da_imagem.jpeg", width = 2500, height = 1000, units
= "px", pointsize = 12, quality = 100, bg = "white", res = 300, family = "")
par(bty="l", mfrow= c(1,3))
plot (bio02$x, bio02$y, xlab = "Temperature (°C*10)", ylab =
"Suitability", type="l", main = "Mean Diurnal Range\n(BIO 02)", ylim = c(0,1))
plot (bio18$x, bio18$y, xlab = "Precipitation (mm)", ylab =
"Suitability", type="l", main = "Precipitation of Warmest Quarter\n(BIO 18)",
ylim = c(0,1))
plot (bio13$x, bio13$y, xlab = "Precipitation (mm)", ylab =
"Suitability", type="l", main = "Precipitation of Wettest Month\n(BIO 13)",
ylim = c(0,1))
dev.off()

#criar os gráficos em .jpeg
#para o código funcionar lembre de trocar os nomes das variáveis
#e de trocar o nome e unidades dos eixos e o titulo
pdf(file = "vari_proba.pdf", width=9, height=3, onefile=T, title=
"Probability of occurrence", paper= "special", colormodel = "gray")
par(bty="l", mfrow= c(1,3))
plot (bio02$x, bio02$y, xlab = "Temperature (°C*10)", ylab =
"Suitability", type="l", main = "Mean Diurnal Range\n(BIO 02)", ylim = c(0,1))
plot (bio18$x, bio18$y, xlab = "Precipitation (mm)", ylab =
"Suitability", type="l", main = "Precipitation of Warmest Quarter \n(BIO 18)",
ylim = c(0,1))
plot (bio13$x, bio13$y, xlab = "Precipitation (mm)", ylab =
"Suitability", type="l", main = "Precipitation of Wettest Month\n(BIO 13)",
ylim = c(0,1))
dev.off()
```



VI) Referências

Breiner, F. T., Guisan, A., Bergamini, A., & Nobis, M. P. (2015). Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, 6(10), 1210-1218.

Brown, J. L. (2014). SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods in Ecology and Evolution*, 5(7), 694-700.

Giannini TC, Siqueira MF, Acosta AL, Barreto FCC, Saraiva AM, Alves-dos-Santos I (2012) Desafios atuais da modelagem preditiva de distribuição de espécies. *Rodriguésia-Instituto de Pesquisas Jardim Botânico do Rio de Janeiro*, 63.

Merow, Cory et al. What do we gain from simplicity versus complexity in species distribution models? *Ecography*, v. 37, n. 12, p. 1267-1281, 2014.

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, 5(11), 1198-1205.

Varela, S., Lobo, J. M., & Hortal, J. (2011). Using species distribution models in paleobiogeography: a matter of data, predictors and concepts. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 310(3), 451-463.

Varela, S., Lima-Ribeiro, M. S., & Terribile, L. C. (2015). A short guide to the climatic variables of the Last Glacial Maximum for biogeographers. *PloS one*, 10(6), e0129037.



VII) Apêndices

Apêndice 1: Função para cálculo da correlação entre as variáveis ambientais em formato ASCII.

```
cor.camadas= function (x=cor.camadas){

  if("raster" %in% rownames(installed.packages()) == FALSE){install.packages("raster")
  } else {print (paste0("'raster' ja instalado na biblioteca"))}
  library(raster)

  if("vegan" %in% rownames(installed.packages()) == FALSE){install.packages("vegan")
  } else {print (paste0("'vegan' ja instalado na biblioteca"))}
  library(vegan)

  if("rgdal" %in% rownames(installed.packages()) == FALSE){install.packages("rgdal")
  } else {print (paste0("'rgdal' ja instalado na biblioteca"))}
  library(rgdal)

  rlist=list.files(getwd(), pattern="asc$", full.names=FALSE)
  for(i in rlist) { assign(unlist(strsplit(i, "[.]")[1], raster(i)) )

  rm(x)
  rm(i, rlist)
  rlist=list(ls())
  ncam=length(rlist[[1]])

  obj=get(c(as.matrix(rlist[[1]][1])))
  obj=c(as.matrix(obj))
  assign(paste("var", "1", sep = ""), obj)
  tabela=var1

  for (i in 2:ncam){
    obj=get(rlist[[1]][i])
    obj=c(as.matrix(obj))
    tabela=data.frame(cbind(tabela,assign(paste("var", i, sep = ""), obj)))
  }

  names(tabela)=(list.files(getwd(), pattern="asc$", full.names=FALSE))

  valores_soma<-rowSums(tabela)
  valores_soma_validos<- 1:nrow(tabela)
  valores_soma_validos<-ifelse(is.na(valores_soma), NA, valores_soma_validos)
  valores_soma_validos<-subset(valores_soma_validos, valores_soma_validos>0)
  tabela_validos <- tabela [valores_soma_validos , ]
  correlacao=round(cor(tabela_validos),2)
  if (dir.exists(file.path(getwd(), "Coef_Correlacao")) == FALSE)
  {
    dir.create("Coef_Correlacao", showWarnings = FALSE)
  }
  setwd(file.path(getwd(), "Coef_Correlacao"))

  write.csv(correlacao, file = "correlacao_camadas_R.csv")
  setwd('.')
}
```



Apêndice 2: Função para a Análise de Componente Principal das variáveis ambientais em formato ASCII.

```
pca.camadas= function (x=pca.camadas) {

  if("raster" %in% rownames(installed.packages()) == FALSE){install.packages("raster")
} else {print (paste0("'raster' ja instalado na biblioteca"))}
library(raster)

  if("vegan" %in% rownames(installed.packages()) == FALSE){install.packages("vegan")
} else {print (paste0("'vegan' ja instalado na biblioteca"))}
library(vegan)

  if("rgdal" %in% rownames(installed.packages()) == FALSE){install.packages("rgdal")
} else {print (paste0("'rgdal' ja instalado na biblioteca"))}
library(rgdal)

  rlist=list.files(getwd(), pattern="asc$", full.names=FALSE)
  for(i in rlist) { assign(unlist(strsplit(i, ".")[1], raster(i)) }

  rm(x)
  rm(i, rlist)
  rlist=list(ls())
  ncam=length(rlist[[1]])

  obj=get(c(as.matrix(rlist[[1]][1])))
  obj=c(as.matrix(obj))
  assign(paste("var", "1", sep = ""), obj)
  tabela=var1

  for (i in 2:ncam){
    obj=get(rlist[[1]][i])
    obj=c(as.matrix(obj))
    tabela=data.frame(cbind(tabela,assign(paste("var", i, sep = ""), obj)))
  }

  names(tabela)=(list.files(getwd(), pattern="asc$", full.names=FALSE))

  valores_soma<-rowSums(tabela)
  valores_soma_validos<- 1:nrow(tabela)
  valores_soma_validos<-ifelse(is.na(valores_soma), NA, valores_soma_validos)
  valores_soma_validos<-subset(valores_soma_validos, valores_soma_validos>0)
  tabela_validos <- tabela [valores_soma_validos , ]

  if (dir.exists(file.path(getwd(), "PCA_camadas")) == FALSE)
  {
    dir.create("PCA_camadas", showWarnings = FALSE)
  }
  setwd(file.path(getwd(), "PCA_camadas"))

  tabela_validos<-decostand(tabela_validos, method="standardize")
  summary(tabela_validos)

  #CRIAR OS COMPONENTES PRINCIPAIS:
  tabela_validos_prcomp<-prcomp(tabela_validos)
```



LaMaB: Laboratório de Mastozoologia e Biogeografia
Departamento de Ciências Biológicas
Universidade Federal do Espírito Santo



```
#porcentagem dos componentes
std_list<-tabela_validos_prcomp$sdev
std_list_pct <- std_list / sum (std_list) * 100
std_list_pct_sum=round(std_list_pct, 2)
write.csv(std_list_pct_sum, file = "pca_percent_PC.csv")

###Exportar os valores
print(tabela_validos_prcomp)
contr= as.data.frame(tabela_validos_prcomp$rotation)
write.csv(contr, file = "contribuicao_variavel_PCA.csv")
setwd('..')
}
```