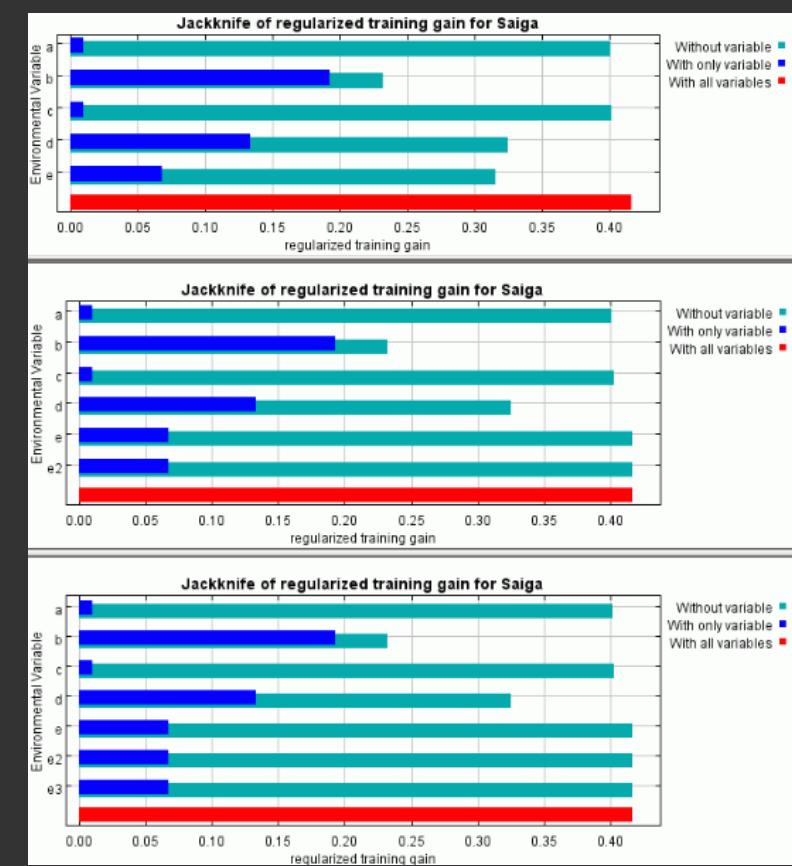
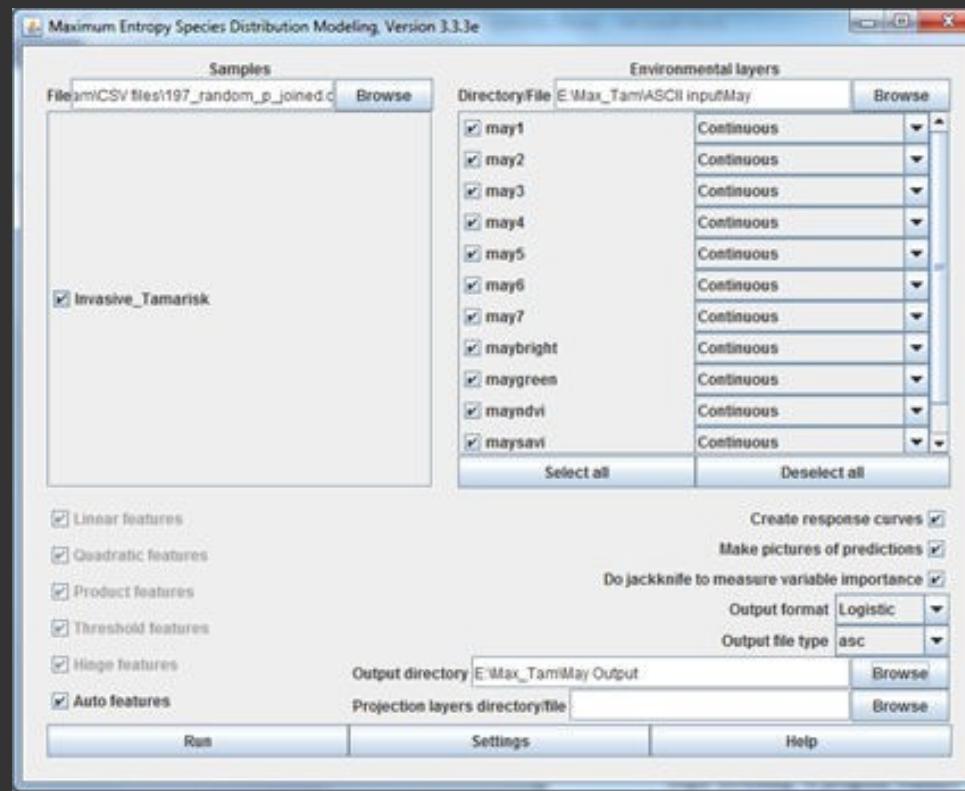


# INTRODUCTION TO MAXENT

- Software for Maximum Entropy algorithm.
- Written on Java.
- Cross-platform.
- Species coordinates as csv files.
- Environmental variables as ESRI ASCII files.
- <http://www.cs.princeton.edu/~schapire/maxent/>



---

## A Maximum Entropy Approach to Species Distribution Modeling

---

**Steven J. Phillips**

AT&T Labs – Research, 180 Park Avenue, Florham Park, NJ 07932

PHILLIPS@RESEARCH.ATT.COM

**Miroslav Dudík**

**Robert E. Schapire**

Princeton University, Department of Computer Science, 35 Olden Street, Princeton, NJ 08544

MDUDIK@CS.PRINCETON.EDU

SCHAPIRE@CS.PRINCETON.EDU

Appearing in *Proceedings of the 21<sup>st</sup> International Conference on Machine Learning*, Banff, Canada, 2004. Copyright 2004 by the authors.



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)



**Ecological Modelling 190 (2006) 231–259**

---

**ECOLOGICAL  
MODELLING**

---

[www.elsevier.com/locate/ecolmodel](http://www.elsevier.com/locate/ecolmodel)

## Maximum entropy modeling of species geographic distributions

Steven J. Phillips<sup>a,\*</sup>, Robert P. Anderson<sup>b,c</sup>, Robert E. Schapire<sup>d</sup>

<sup>a</sup> AT&T Labs-Research, 180 Park Avenue, Florham Park, NJ 07932, USA

<sup>b</sup> Department of Biology, City College of the City University of New York, J-526 Marshak Science Building,  
Convent Avenue at 138th Street, New York, NY 10031, USA

<sup>c</sup> Division of Vertebrate Zoology (Mammalogy), American Museum of Natural History, Central Park West at 79th Street,  
New York, NY 10024, USA

<sup>d</sup> Computer Science Department, Princeton University, 35 Olden Street, Princeton, NJ 08544, USA

Received 23 February 2004; received in revised form 11 March 2005; accepted 28 March 2005

Available online 14 July 2005

Journal of Machine Learning Research 8 (2007) 1217-1260

Submitted 4/06; Revised 4/07; Published 6/07

## Maximum Entropy Density Estimation with Generalized Regularization and an Application to Species Distribution Modeling

**Miroslav Dudík**

*Princeton University*

*Department of Computer Science*

*35 Olden Street*

*Princeton, NJ 08540*

MDUDIK@CS.PRINCETON.EDU

**Steven J. Phillips**

*AT&T Labs – Research*

*180 Park Avenue*

*Florham Park, NJ 07932*

PHILLIPS@RESEARCH.ATT.COM

**Robert E. Schapire**

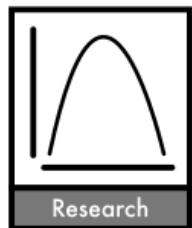
*Princeton University*

*Department of Computer Science*

*35 Olden Street*

*Princeton, NJ 08540*

SCHAPIRE@CS.PRINCETON.EDU



**Ecography 31: 161–175, 2008**  
doi: 10.1111/j.2007.0906-7590.05203.x  
© 2007 AT&T, Inc. Journal compilation © 2007 Ecography  
Subject Editor: Miguel Araújo. Accepted 13 December 2007

## Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation

**Steven J. Phillips and Miroslav Dudík**

*S. J. Phillips (phillips@research.att.com), AT&T Labs Research, 180 Park Avenue, Florham Park, NJ 07932, USA. – M. Dudík, Computer Science Dept, Princeton Univ., 35 Olden Street, Princeton, NJ 08540, USA.*

*Ecological Applications*, 19(1), 2009, pp. 181–197  
© 2009 by the Ecological Society of America

## Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data

STEVEN J. PHILLIPS,<sup>1,8</sup> MIROSLAV DUDÍK,<sup>2</sup> JANE ELITH,<sup>3</sup> CATHERINE H. GRAHAM,<sup>4</sup> ANTHONY LEHMANN,<sup>5</sup>  
JOHN LEATHWICK,<sup>6</sup> AND SIMON FERRIER<sup>7</sup>

<sup>1</sup>AT&T Labs—Research, 180 Park Avenue, Florham Park, New Jersey 07932 USA

<sup>2</sup>Computer Science Department, Princeton University, 35 Olden Street, Princeton, New Jersey 08544 USA

<sup>3</sup>School of Botany, University of Melbourne, Parkville, Victoria 3010 Australia

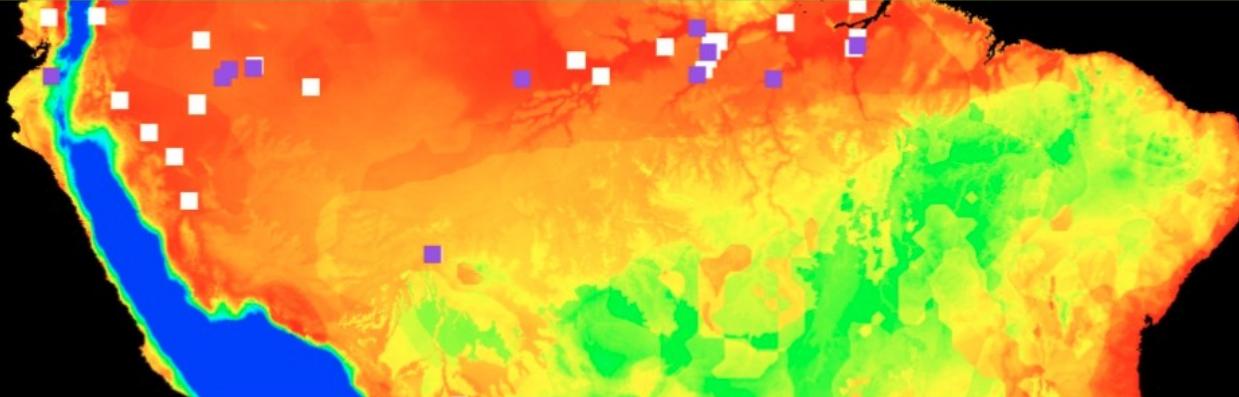
<sup>4</sup>Department of Ecology and Evolution, 650 Life Sciences Building, Stony Brook University, New York 11794 USA

<sup>5</sup>Climatic Change and Climate Impacts, University of Geneva, 7 Route de Drize, 1227 Carouge, Switzerland

<sup>6</sup>NIWA, Hamilton, New Zealand

<sup>7</sup>New South Wales Department of Environment and Climate Change, P.O. Box 402, Armidale 2350 Australia

Maxent software for modeling species niches and distributions



**Maxent is now open source!**

Use this site to download Maxent software for modeling species niches and distributions by applying a machine-learning technique called maximum entropy modeling. From a set of environmental (e.g., climatic) grids and georeferenced occurrence localities, the model expresses a probability distribution where each grid cell has a predicted suitability of conditions for the species. Under particular assumptions about the input data and biological sampling efforts that led to occurrence records, the output can be interpreted as predicted probability of presence (cloglog transform), or as predicted local abundance (raw exponential output).

Here you can download the open-source release of Maxent (under an MIT license; suggested citation below). See below for key changes in the current version.

The idea for Maxent was first conceived of here at the Center for Biodiversity and Conservation at the American Museum of Natural History (AMNH) through a public-private partnership between the AMNH and AT&T-Research. Steven Phillips and the other developers of Maxent are still engaged in its development and maintenance, and the [Google group](#) will remain the main mechanism for user questions. Much additional information can be found in the Google group, software tutorials, and other resources listed below.

**Main changes in Version 3.4.1**

- Minor bug fixes to the 3.4.0 release

**Main changes in Version 3.4.0**

- Released under a MIT license\*\*
- Threshold features are now turned off by default
- A cloglog transform has been added and now constitutes the default transformation for model output (formerly the logistic transform)

[http://biodiversityinformatics.amnh.org/open\\_source/maxent/](http://biodiversityinformatics.amnh.org/open_source/maxent/)

00000100000  
00001010000  
00010001000  
00100000100  
01000000010  
10000000001  
  
Software notes

**Ecography 40: 887–893, 2017**

doi: 10.1111/ecog.03049

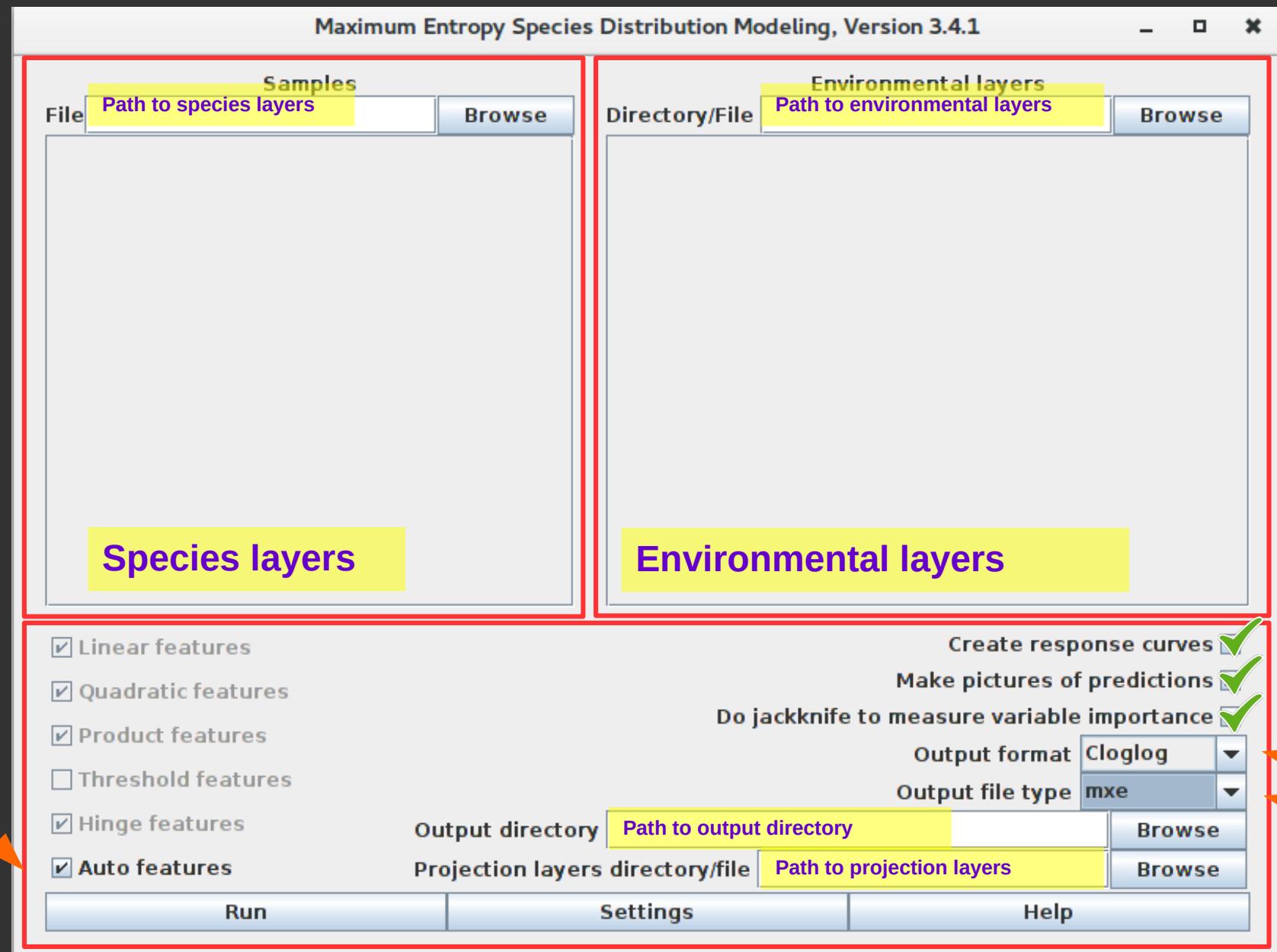
© 2017 The Authors. Ecography © 2017 Nordic Society Oikos

Subject Editor: Michael Borregaard. Editor-in-Chief: Miguel Araújo. Accepted 9 March 2017

## Opening the black box: an open-source release of Maxent

**Steven J. Phillips, Robert P. Anderson, Miroslav Dudík, Robert E. Schapire and Mary E. Blair**

*S. J. Phillips (<http://orcid.org/0000-0002-6991-608X>) and M. E. Blair, Center for Biodiversity and Conservation, American Museum of Natural History, New York, NY, USA. – R. P. Anderson, Dept of Biology, City College of New York, City Univ. of New York, New York, NY, USA, and Program in Biology, Graduate Center, City Univ. of New York, New York, NY, USA, and Div. of Vertebrate Zoology (Mammalogy), American Museum of Natural History, New York, NY, USA. – M. Dudík and R. E. Schapire, Microsoft Research, New York, NY, USA.*



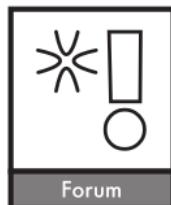
# HOW TO USE MAXENT: FEATURES AND OUTPUTS

Diversity and Distributions, (Diversity Distrib.) (2011) 17, 43–57



## A statistical explanation of MaxEnt for ecologists

Jane Elith<sup>1\*</sup>, Steven J. Phillips<sup>2</sup>, Trevor Hastie<sup>3</sup>, Miroslav Dudík<sup>4</sup>,  
Yung En Chee<sup>1</sup> and Colin J. Yates<sup>5</sup>



EDITOR'S  
CHOICE

Ecography 36: 1058–1069, 2013

doi: 10.1111/j.1600-0587.2013.07872.x

© 2013 The Authors. Ecography © 2013 Nordic Society Oikos  
Subject Editor: Niklaus E. Zimmermann. Accepted 27 March 2013

## A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter

Cory Merow, Matthew J. Smith and John A. Silander, Jr

C. Merow ([cory.merow@gmail.com](mailto:cory.merow@gmail.com)) and J. A. Silander, Jr, Univ. of Connecticut, Ecology and Evolutionary Biology, 75 North Eagleville Rd., Storrs, CT 06269, USA. – M. J. Smith and CM, Computational Ecology and Environmental Science Group, Computational Science Laboratory, Microsoft Research Ltd., 21 Station Road, Cambridge, CB1 2FB, UK.

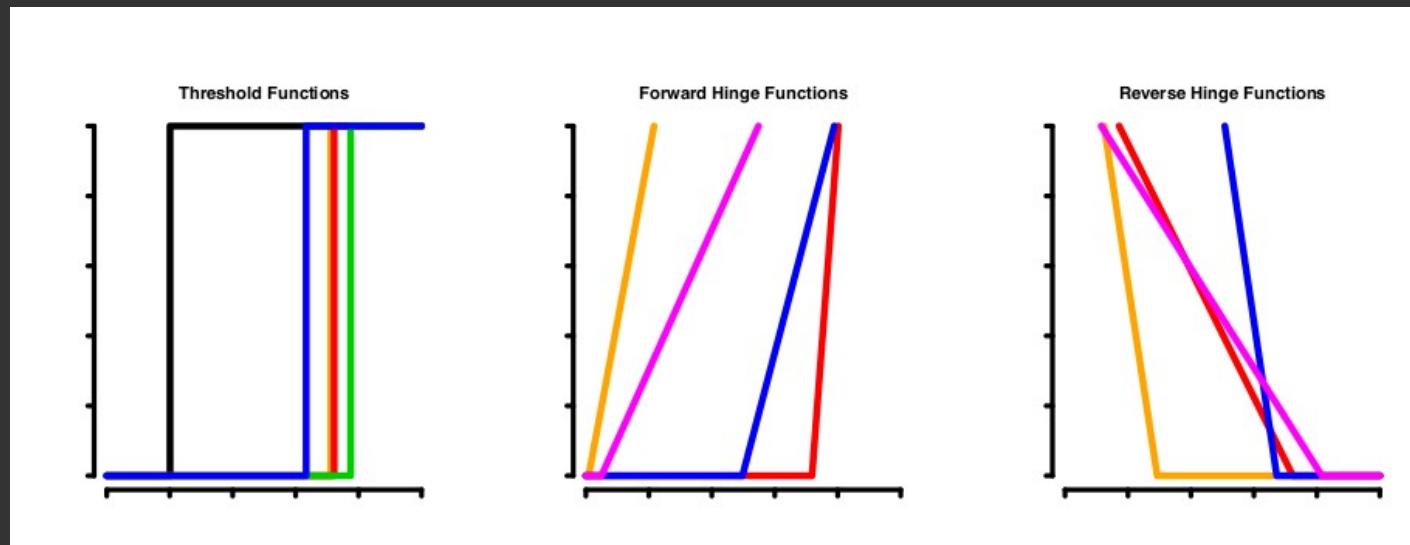
**Linear** - continuous variables should be close to their observed values (their mean at occurrence localities)

**Quadratic** - variance of continuous variables should be close to observed values

**Product** - covariance of two continuous variables should be close to observed values

**Threshold** - proportion of model that has values above a threshold for a continuous variable should be close to observed proportion

**Hinge** - linear feature truncated at threshold



# Methods in Ecology and Evolution



*Methods in Ecology and Evolution* 2014, 5, 1192–1197

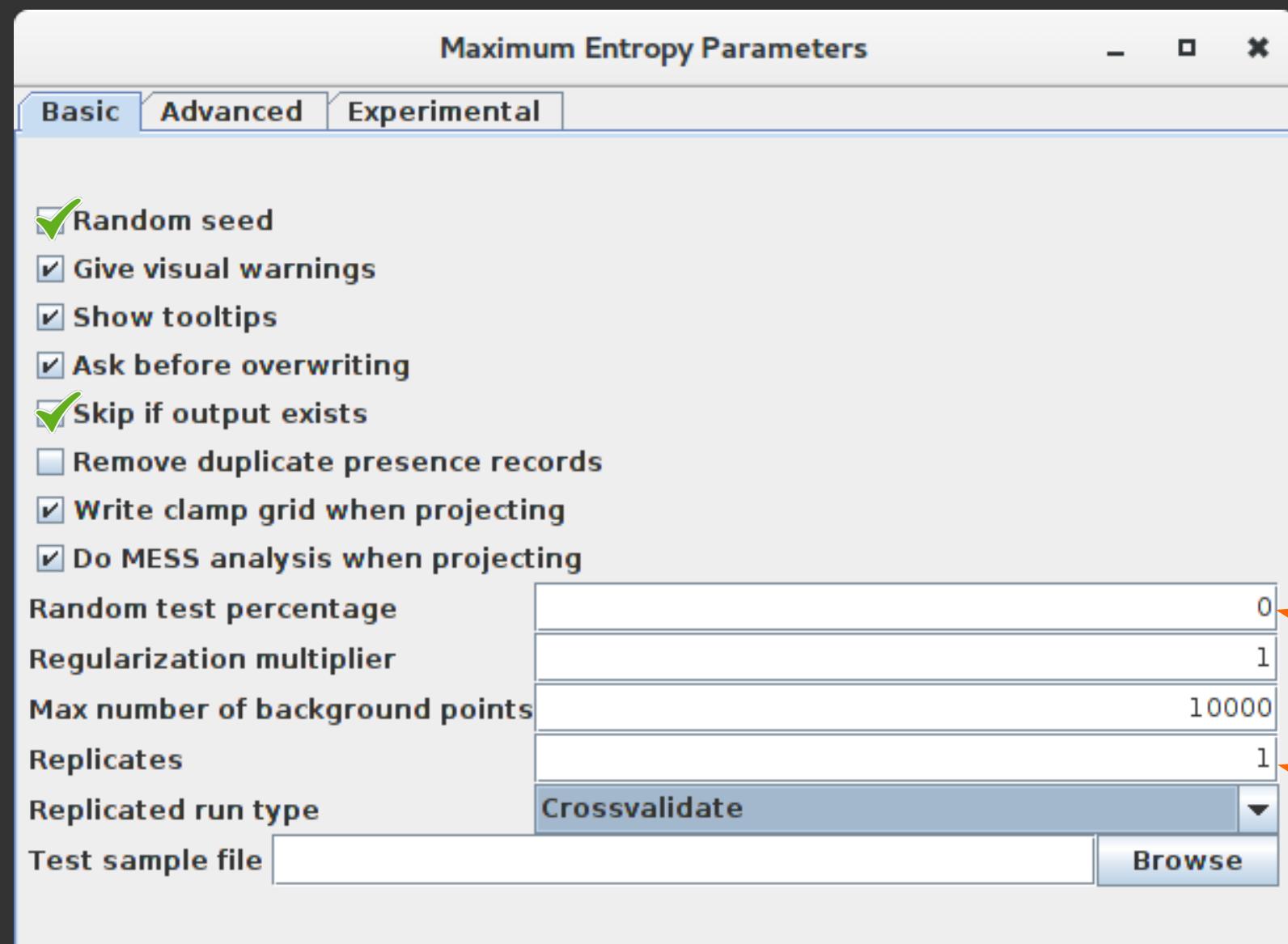
doi: 10.1111/2041-210X.12252

FORUM

## Maxent is not a presence–absence method: a comment on Thibaud *et al.*

Gurutzeta Guillera-Arroita\*, José J. Lahoz-Monfort and Jane Elith

School of Botany, University of Melbourne, Parkville, Victoria, Australia



Diversity and Distributions, (Diversity Distrib.) (2011) 17, 43–57



## A statistical explanation of MaxEnt for ecologists

Jane Elith<sup>1\*</sup>, Steven J. Phillips<sup>2</sup>, Trevor Hastie<sup>3</sup>, Miroslav Dudík<sup>4</sup>,  
Yung En Chee<sup>1</sup> and Colin J. Yates<sup>5</sup>

## Methods in Ecology and Evolution

Methods in Ecology & Evolution 2010, 1, 330–342

doi: 10.1111/j.2041-210X.2010.00036.x



## The art of modelling range-shifting species

Jane Elith<sup>1\*</sup>, Michael Kearney<sup>2</sup> and Steven Phillips<sup>3</sup>

<sup>1</sup>School of Botany, The University of Melbourne, Parkville 3010, Australia; <sup>2</sup>Department of Zoology, The University of Melbourne, Parkville 3010, Australia and <sup>3</sup>AT&T Labs – Research, 180 Park Avenue, Florham Park, NJ 07932, USA

# MESS: MULTIVARIATE ENVIRONMENTAL SIMILARITY SURFACE

- MESS measures the similarity of any given point to a reference set of points, with respect to the chosen predictor variables.
- MESS reports the closeness of the point to the distribution of reference points
- MESS gives negative values for dissimilar points and maps these values across the whole prediction region

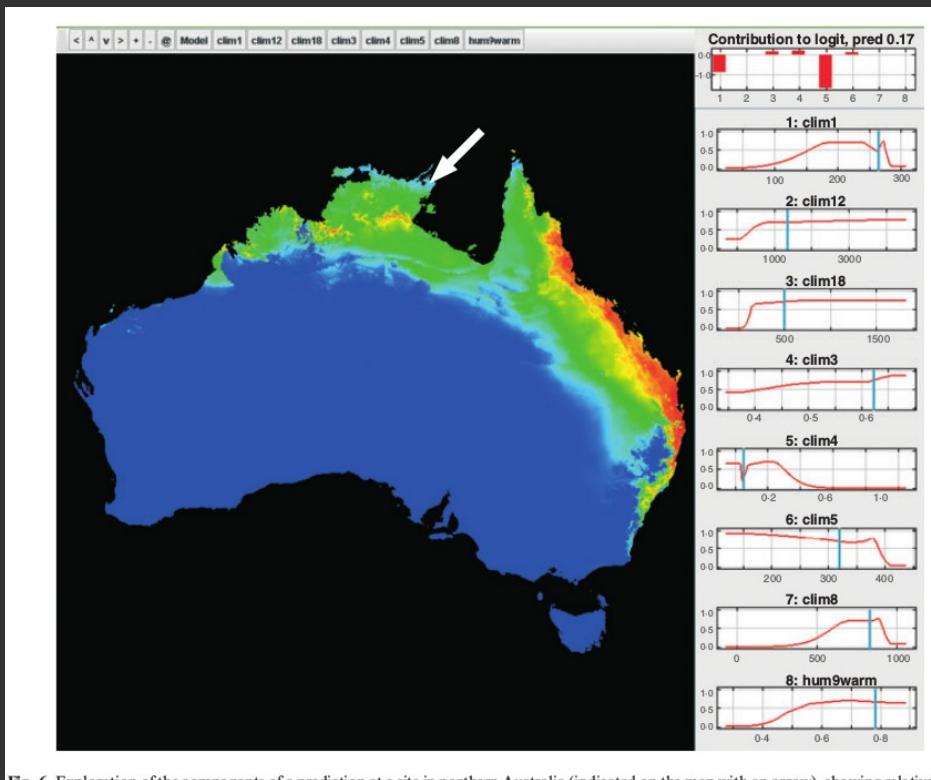


Fig. 6. Exploration of the components of a prediction at a site in northern Australia (indicated on the map with an arrow), showing relative influence of each variable at that site (top right) and fitted functions (right, other panels) with vertical blue lines showing conditions at the selected site.

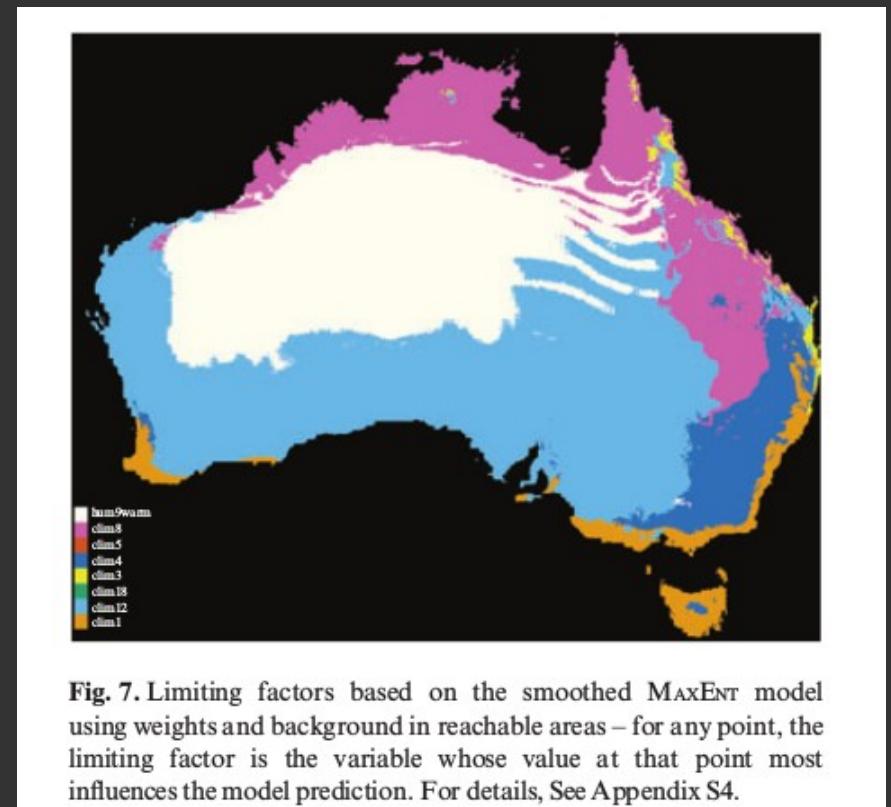


Fig. 7. Limiting factors based on the smoothed MAXENT model using weights and background in reachable areas – for any point, the limiting factor is the variable whose value at that point most influences the model prediction. For details, See Appendix S4.

- Parameter to restrict the model
- <1 more restricted
- >1 less restricted
- You should test different regularisation parameters to find the best

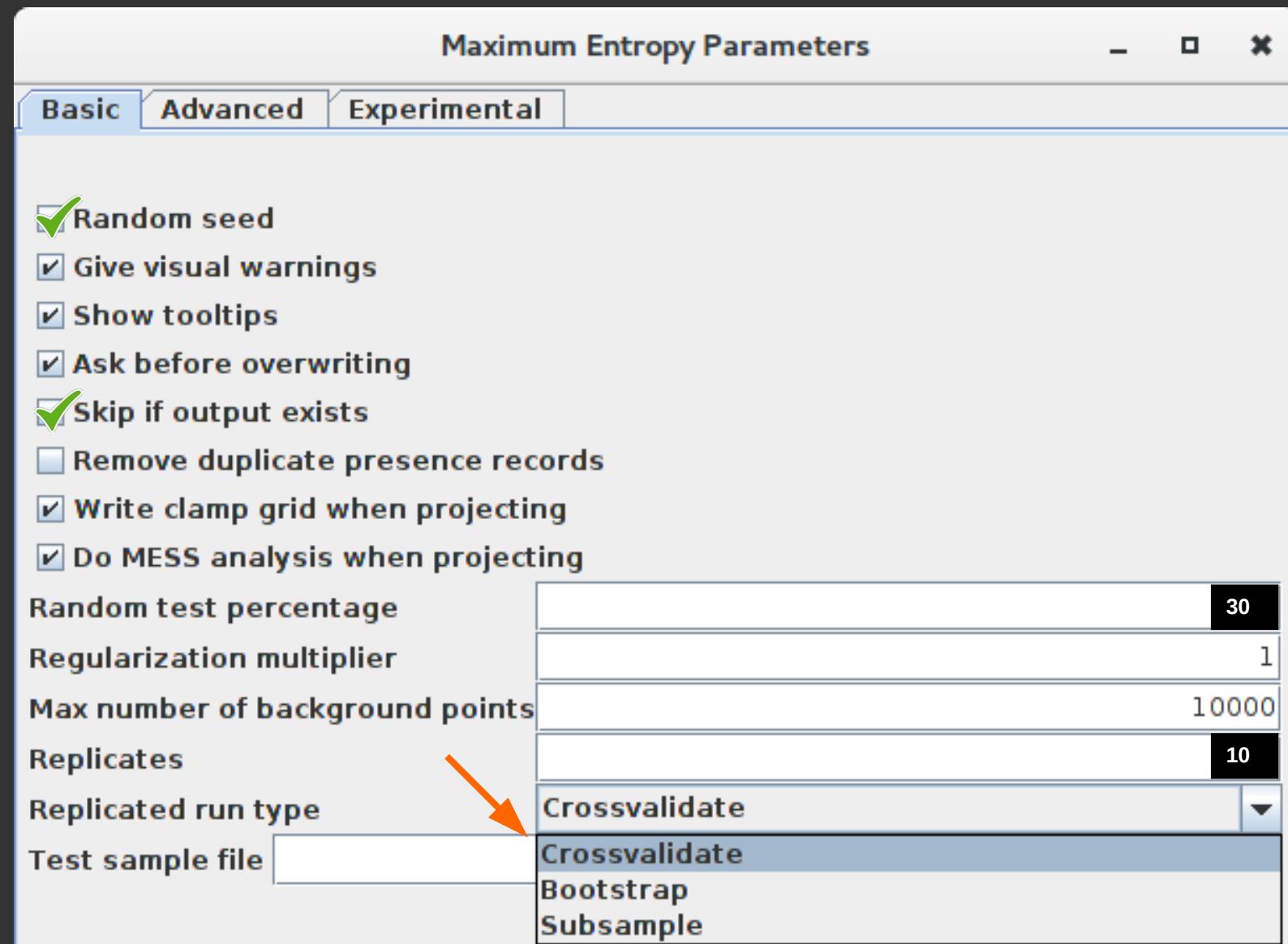
*Ecological Applications*, 21(2), 2011, pp. 335–342  
© 2011 by the Ecological Society of America

## Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria

DAN L. WARREN<sup>1,3</sup> AND STEPHANIE N. SEIFERT<sup>2</sup>

<sup>1</sup>Section of Integrative Biology, University of Texas at Austin, Austin, Texas 78712 USA

<sup>2</sup>Department of Entomology, University of California, Davis, California 95616 USA



## Crossvalidate

- samples divided into replicates folds
- each fold in turn used for test data

## Bootstrap

- replicate sample sets chosen by sampling with replacement

## Subsample

- replicate sample sets chosen by removing random test percentage without replacement to be used for evaluation

# DATA PARTITIONING METHODS

Table 1 Data partitioning methods for the allocation of cases to training and testing data sets.

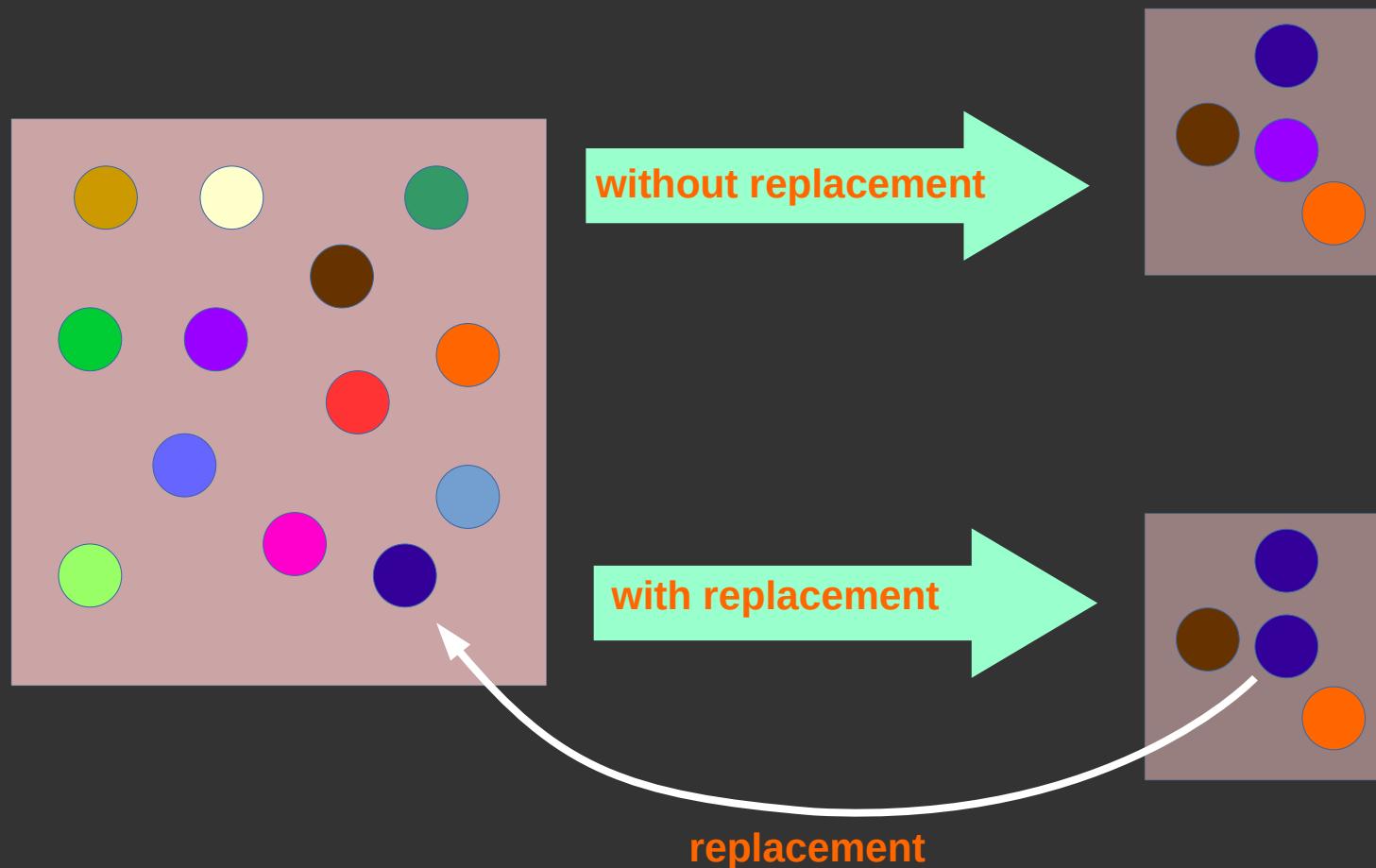
Method	Examples	Notes
Resubstitution	Stockwell (1992) Osborne & Tigar (1992)	No partitioning is carried out, the same data are used for training and testing. This tends to provide optimistic measures of prediction success.
Bootstrapping	Buckland & Elston (1993) Verbyla & Litaitis (1989)	Bootstrap samples (sampling with replacement) are used to assess prediction success. Accuracy is usually reported as a mean and confidence limits.
Randomization	Capen <i>et al.</i> (1986)	Random samples are obtained by sampling without replacement. Accuracy is usually reported as a mean and confidence limits.
Prospective sampling	Capen <i>et al.</i> (1986) Fielding & Haworth (1995) Morrison <i>et al.</i> (1987)	A new sample of cases is obtained after the model has been developed. These could be from a different region or time.
<i>k</i> -fold partitioning	Stockwell (1992)	The data are split into $k$ ( $k > 2$ ) sets, only one of which is used for training. The remaining $k - 1$ sets are pooled for testing purposes. Also known as the hold-out or external method. Accuracy is usually reported as a mean and confidence limits.
Special cases of <i>k</i> -fold partitioning		
Leave-One-Out (L-O-O)	Capen <i>et al.</i> (1986) Osborne & Tigar (1992)	Also known as jackknife sampling, $n$ samples of 1 case are tested sequentially, the remaining $n - 1$ cases forming the training set.
$K = 2$	Smith (1994)	Data are split into one training set and one testing set. A variety of strategies may be employed to determine the split.

Fielding & Bell 1997  
Manel *et al* 2001

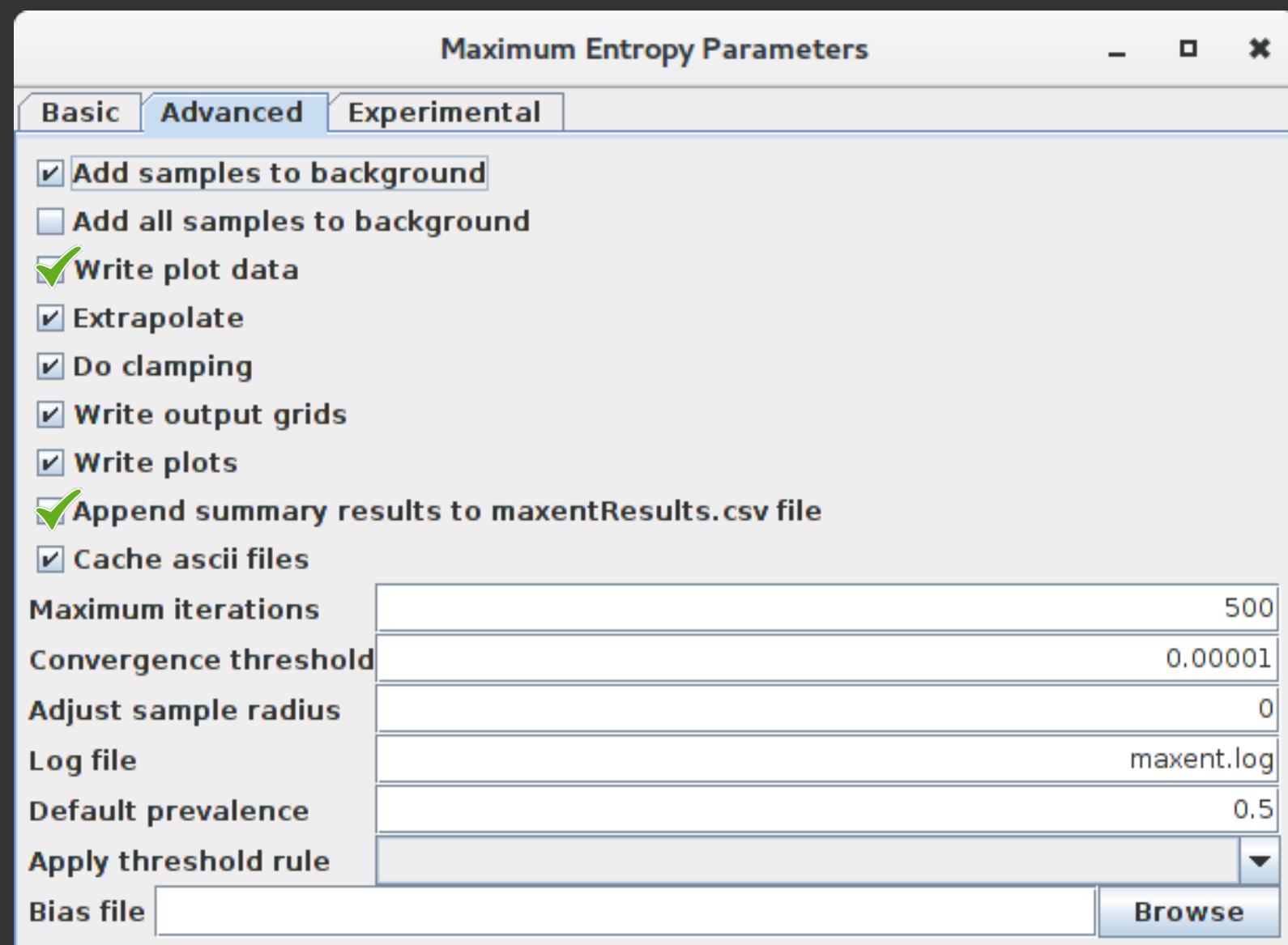
# DATA PARTITIONING METHODS

**Sampling without replacement** → no element can be selected more than once in the same sample.

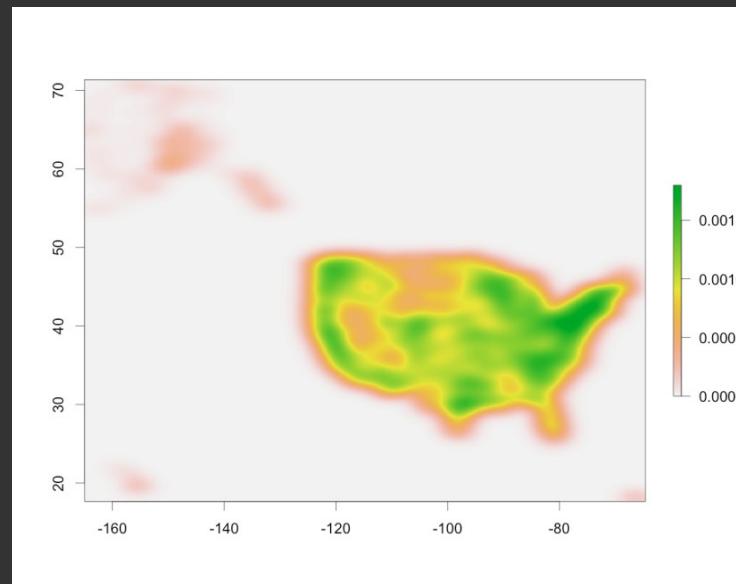
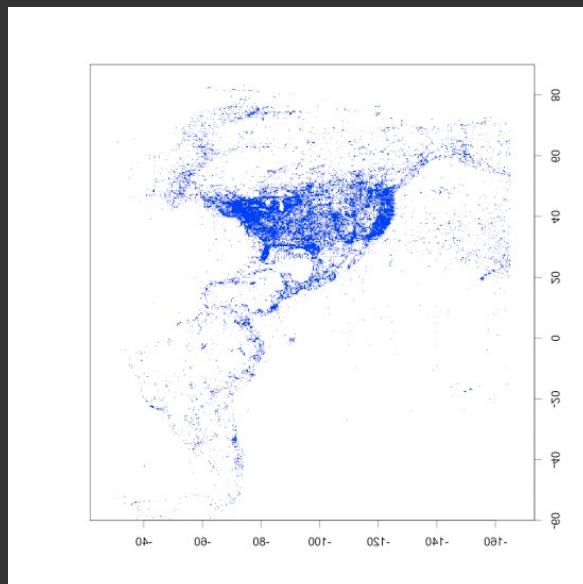
**Sampling with replacement** → an element may appear multiple times in the one sample.

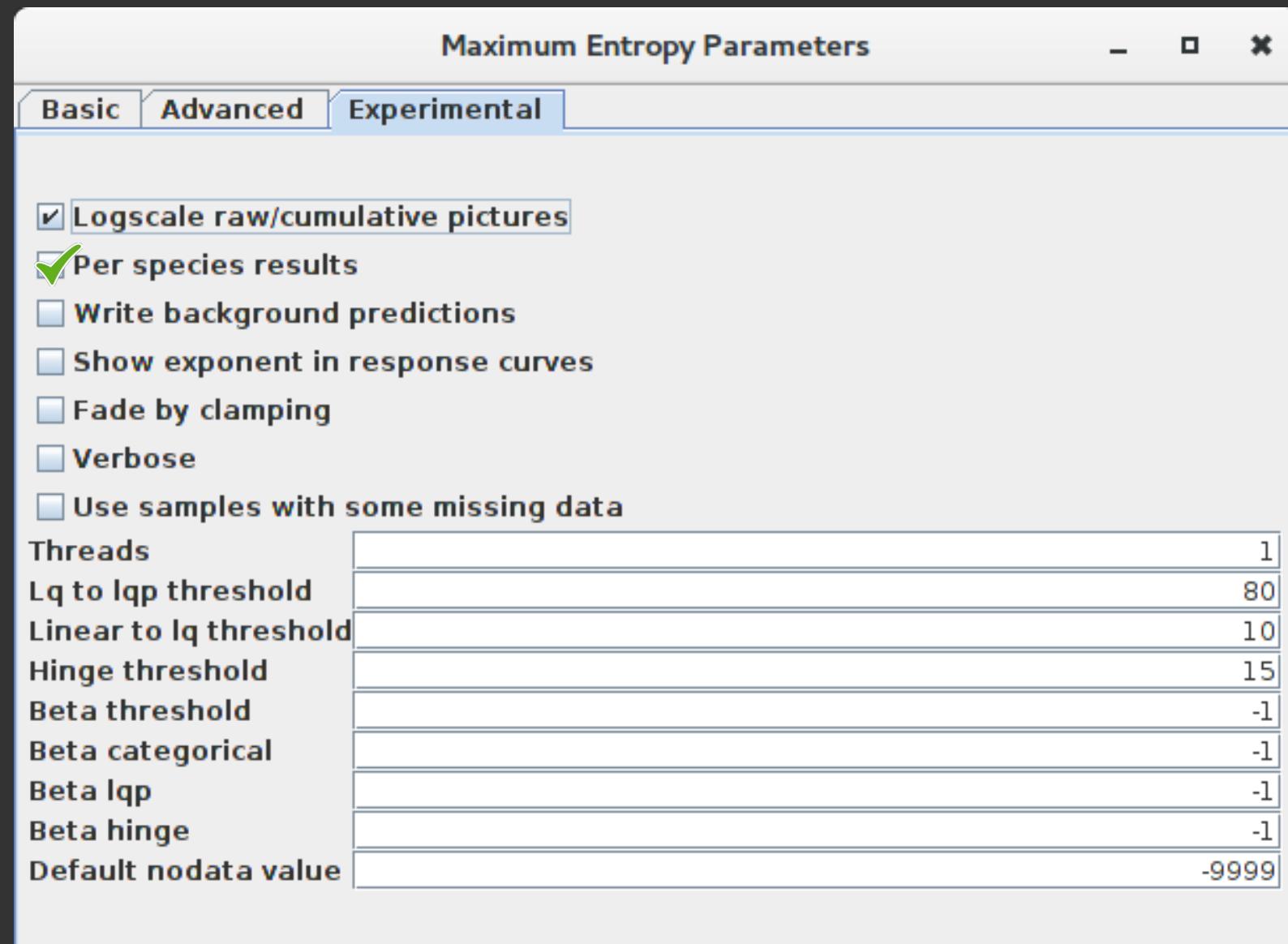


Fielding & Bell 1997  
Manel et al 2001



- A bias file describes the relative sampling intensity across the whole study area
- The sampling bias file needs to be a grid, typically in .asc format
- Values must not be zero or negative
- Distance raster
- Mask raster without zeros





1. Add the path to the environmental layer folder.
2. Maxent only accepts ESRI ASCII formats.
3. Add the path to the species' dataset.
4. The species' records must be a txt or csv file.

Species,Coord X,Coord Y

5. Fields must be separated by COMMA.
6. Add the path to the projection layer folder if required.
7. Choose your settings.
8. Export model output layers to MXE format. This format is faster.
9. Run the model.
10. Open the html output file.
11. Export MXE files to other formats using the command line.
12. You can run also the models by command line in the bat file.

## .CSV FILE

**species,x,y**

**Rana,-6.990315,38.60562**

**Rana,-7.427625,39.603442**

**Rana,-7.317719,39.331613**

**Rana,-7.315555,39.421679**

**Rana,-7.201751,39.329879**

**Rana,-7.199438,39.41994**

**Rana,-6.04569,39.576408**

```
java -mx1512m -jar maxent.jar
environmentallayers=path_to_environmental_layers
samplesfile=path_to_species_dataset
outputdirectory=path_to_output_models
projectionlayers=path_to_projection_layers randomseed jackknife
pictures responsecurves removeduplicates skipifexists
writebackgroundpredictions writeclampgrid writeplotdata writemess
outputgrids=FALSE nowarnings autorun mxe -X 30
replicatetype=Bootstrap replicates=10
```

```
java -cp maxent.jar density.Convert path_to_maxent_models mxe  
path_to_exported_maxent_models asc
```

It is possible to run also Maxent in R using several packages:

- **Dismo**
- **Biomod2**

# QUESTIONS?