International Society for Environmental Information Sciences 2010 Annual Conference (ISEIS)

Evaluation of different aspects of maximum entropy for niche-based modeling

Elisângela S. da C. RODRIGUES[a,b,*], Fabrício A. RODRIGUES[a], Ricardo L. de A. da ROCHA[b], Pedro L. P. CORRÊA[a] and Tereza C. GIANNINI[c]

[a] *Laboratory of Agricultural Automation – Engineering School of the University of São Paulo. Av. Prof. Luciano Gualberto, travessa 3, 158, sala C2-56, Edifício de Engenharia Elétrica, Cidade Universitária – São Paulo/SP – Brazil. CEP 05508–900.*
[b] *Laboratory of Languages and Adaptive Techniques – Engineering School of the University of São Paulo. Av. Prof. Luciano Gualberto, travessa 3, 158, sala C2-30, Edifício de Engenharia Elétrica, Cidade Universitária – São Paulo/SP – Brazil. CEP 05508–900.*
[c] *Laboratory of Bees – Institute of Biosciences of the University of São Paulo. Rua do Matão, 321, Cidade Universitáira – São Paulo/SP – Brazil. CEP 05508-900*

Abstract

Biodiversity conservation is a world challenge that needs attention and efficient strategies for its success. Modeling of geographic distributions of species is used in assorted applications related to biodiversity conservation. Maximum entropy (maxent) is a technique recently applied to modeling of geographic distributions of species and is being largely used by biologists. The aim is to evaluate different viewpoints of this technique. The first evaluation is concerned with the performance of the algorithm. A parallel version of the maxent-based algorithm available in openModeller is presented. openModeller is a set of tools provided for researchers interested in modeling of geographic distributions of species. The second evaluation is focused on tuning the regularization parameter, since it can severely affect the performance of the algorithm and can take a long time to be adjusted. In addition, the algorithm was evaluated without the use of a regularization parameter and with an adaptive maximum entropy approach. This approach was evaluated as a replacement of the regularization parameter. The validation of the assessments was based on a dataset with 20 species. The results show: an improvement in the algorithm performance using parallelism, considering only the running time; the regularization parameter does not depend on the number of samples, or on the number of iterations in training; species with the same number of samples fit better with different values of the regularization parameter (different magnitude order); the adaptive approach cannot replace the regularization parameter.

Keywords: Biodiversity conservation; Niche-based modeling; Maximum entropy; Parallelism; Adaptive technology; Regularization

---

* Corresponding author.
E-mail address: elisangela.rodrigues@poli.usp.br.

1. Introduction

Biodiversity conservation is a challenge with various world impacts. The term biodiversity refers not only to the diversity of the biological world, as the word suggests, but it also refers to the complex interactions among the different forms of life and the ecological functions performed by them. Thus, biodiversity conservation includes the preservation of all species in the world, not only of an individual species.

Modeling of geographic distributions of species has been used in assorted applications related to biodiversity conservation. For example, it has been applied in monitoring the spatio-temporal variation in habitat suitability of species [1], impacts of climate change [2], and study of species delimitation [3].

There are many techniques that have been used to model environmental systems [4]. Maximum entropy (maxent) [5] is a technique recently applied to the modeling of geographic distributions of species and is being largely used by biologists. The results of some evaluations concerned with different viewpoints of this technique are presented herein.

The performance of a technique in modeling of geographic distributions of species depends on the number of data and mainly how it was implemented. Some techniques demand a long running time, requiring some hours or even days to finish the modeling. The first evaluation of maxent, presented here, concerns the performance of the algorithm. A parallel version of the maxent-based algorithm, available in openModeller [6], which is a set of tools provided for researchers interested in modeling of geographic distributions of species, is presented. This algorithm improves the running time without efficiency loss.

Modeling techniques require the tuning of parameters to be used as default values. This is important because to tune parameters can take a long time and default values are always validated with a large number of experiments. In maxent, the regularization parameter prevents overfitting, that is, the model loses the ability of generalizing beyond the fitting data. This behavior can severely affect the performance of the algorithm. Thus, the second evaluation is focused on tuning the regularization parameter. Two other evaluations were made of the regularization parameter: its removal of the algorithm and the possibility of replacing it by another strategy, the adaptive approach of maxent [7]. The adaptive approach tries to find the best set of features at each step instead of using the same data all along the training.

2. Basic Concepts

This section presents some basic concepts needed for the complete understanding of the work presented here. This is a multidisciplinary work, including niche-based modeling, maximum entropy, parallelism and adaptive technology.

2.1. Modeling of geographic distributions of species – an overview

The main concept related to modeling of geographic distributions of species is the ecological niche [8]. It represents the set of resources and ecological conditions essential for species to sustain a population for a long period of time. A modeling tool produces an approximation of the ecological niche of a species known as niche-based model. Before beginning the modeling, the problem should be delineated, that is, queries that the model should answer.

Fig. 1 shows the general modeling process. The set of data is selected based on the delineation of the problem. The selected set of data should be treated, that is, possible errors such as wrong digitization and unreasonable values should be cleaned up. The model performance depends on the data quality. After the selection and treatment of data, an algorithm should be chosen and it is required to set its parameters. A reasonable understanding of the algorithm and its parameters is recommended. Executing the chosen algorithm generates a model. Generally, the modeling tools offer some statistics that allows the post analysis of the model. The last step is the validation of the model. At this level, the model can be accepted or one may decide to return to any of the previous level. There are more details about the modeling process in Santana et al.[9] .
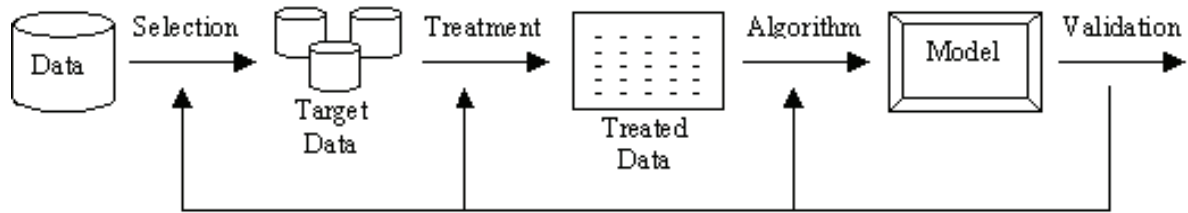
Fig. 1. General modeling process.

The model generated is a probability function that represents the relation between the suitable environmental conditions for the species and the given environmental/ecological data [5]. The selected data are occurrence points and environmental layers. Occurrence data are georeferenced points – latitude and longitude – recorded where the species were observed. They can be divided into presence and absence points, indicating the existence or not of the species. Environmental layers represent the ecological niche of the species [10], that is, the environmental conditions needed for the survival of the species. Some environmental layers commonly used in modeling are temperature and precipitation. They are also georeferenced data and all of them must belong to the same study region [5]. After the model generation, it is applied to a geographic region of interest, resulting in a georeferenced map that contains the occurrence probabilities of the species.

2.2. Modeling of geographic distributions of species – an overview

The increasing amount of occurrence data of species available in databases over the Internet [11] stimulates the efforts towards studies of tools to deal with them. However, presence data are plentiful and absence data are rarely available. Thus, the use of presence-only modeling methods which require just presence data, such as maxent [5,11], is very important.

According to Phillips et al. [5], in a maxent method, the resulting model is not an occurrence probability function of the species, but a probability distribution, that is, the problem is to estimate the distribution with maximum entropy. Suppose X is a finite set of pixels over the geographic study area. The points $x_1,...,x_m$ in X represent the occurrence points of a species. The goal is to estimate the probability distribution p^* over X that approximates p, the probability distribution that represents the true distribution of the species.

A set of real-valued functions $f_1,...,f_n$, where $f_j: X \rightarrow$, known as features, represents the environmental layers or functions derived from them. The notation $p[f_j]$ denotes the expectation of the feature f_j under the probability distribution p. The expectation is for each feature to be equal to its empirical average, denoted by $\tilde{p}[f_j]$ and defined by Eq. (1) [12].

$$\tilde{p}[f_j] = 1/m \sum_{i=1}^{m} f_j(x_i). \tag{1}$$

Commonly, there are many probability distributions satisfying such constraints, but the maximum entropy principle suggests that, among all of them, the chosen probability distribution is that with maximum entropy. The entropy of a probability distribution p on X [12] is defined as

$$\tag{2}$$

The probability distribution with maximum entropy, defined in Eq. (3) can be proved to be equivalent to Gibbs distribution, that is, an exponential distribution with a vector of feature weights that parameterizes it [13]. This probability distribution is defined as

$$q_\lambda(x_i) = \exp(\sum_{j=1}^{n} \lambda_j f_j(x_i)) / Z_\lambda \tag{3}$$

where λ is a vector of the feature weights, with real values, and Z_λ is a normalizing constant that guarantees that the probability distribution sums to one over the area of interest. The probability distribution with maximum entropy is also equivalent to minimizing the log loss, that is, the negative log likelihood [5,12,13]. The log loss is defined in Eq. (4). Thus, both Eq. (3) and Eq. (4) are used as objective functions.

$$\tag{4}$$

The work presented here was performed in openModeller, a framework for modeling of geographic distributions of species with a lot of resources [6]. The maxent algorithm available in openModeller is alike to the sequential algorithm used in MaxEnt program [5,12]. Unlike most algorithms, the sequential one chooses one weight at a time to update them. This procedure is executed until either the convergence is reached or a specific number of rounds are performed.

The expectation for each feature to be equal to its empirical average may cause overfitting. This is because the empirical average generally just approximates the true mean. Regularization parameter is used in maxent to avoid this problem [5]. Thus, instead of the expectation for each feature to be equal to its empirical average, the expectation is for it to be close to them. The use of regularization replaces Eq. (4) with Eq. (5)

$$\tag{5}$$

where β_j is an estimate of how close the empirical average is from the true mean.

3. Methodology

According to Phillips and Dudík [13], the best choice of the regularization parameter depends on the class of features and the number of occurrence records. However, the best value for β_j is still under study and it may interfere severely in the performance of the algorithm. Besides that, it can take a long time to be adjusted. Therefore, an evaluation of different values for β_j with several species is presented herein. The values for β_j were ranged from 0.00001 to 0.9 considering just the variation of the first number different from zero in each magnitude order.

The removal of regularization was also evaluated for the same set of species and an alternative technique, known as adaptive approach, was considered for replacing regularization. The adaptive maxent approach [7] searches for a new set of environmental layers at each round, changing the data set used in the training during the algorithm running.

The modeling process uses a massive number of data and the algorithm processing may consume a lot of time. Besides this, some algorithms are considered very computationally expensive. Consequently, the project in which the openModeller tool was inserted bought a cluster. Therefore, the more algorithms are parallelized, the more efficient the tool can be. This was one of the motivations for the parallelization of the maxent algorithm available in openModeller. The parallel version of the maxent algorithm presented here was developed using the MPI library (Message Passing Interface).

The master-slave approach presented in Fig. 2 was carried out in the development of this work. In this approach, the master process distributes the tasks for slave processes and after the processing it receives the result from each of them. The strategy adopted was the dynamic scheduling of the tasks. In the dynamic scheduling, the distribution of processes to the processors is carried out during the program running in agreement with some standard. The measure used in this work was the on-demand scheduling, that is, when a slave process finishes the execution of a task, the master process attributes a new task to this slave process. This procedure occurs until all the tasks are concluded.
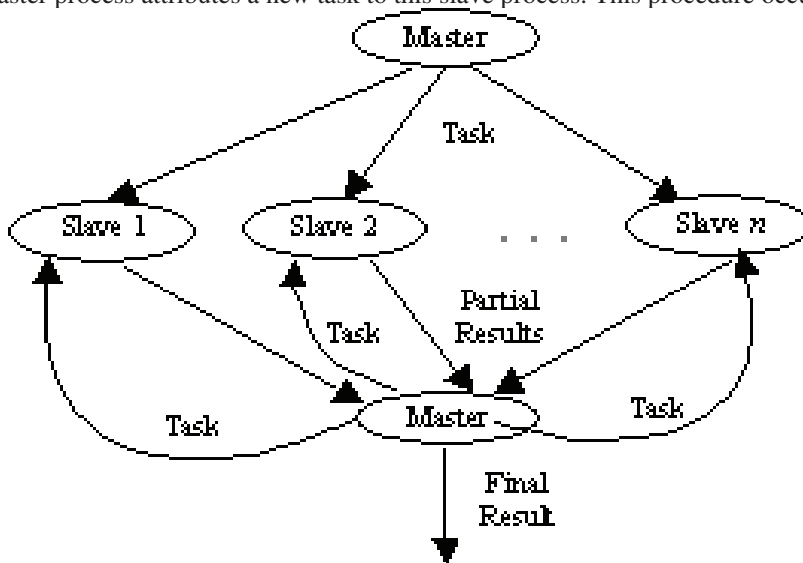

Fig. 2. Master-slave model applied in the implementation of the parallel version of maxent – extracted from [14]

The parallelization was carried out in the minimization of the Eq. (4). The master process is responsible for determining the layer j that each slave process will calculate in Eq. (4). Each slave process returns to the master process the (j, α) values, in which α is the value for updating the weight of this layer. When the master process receives a result, it verifies if the α value is the smallest. The master process distributes the tasks to the slave

processes and receives the processed result from each of them. If the number of layers in a given experiment is greater than the number of processes, when the master process receives a result, it sends a new job to the slave process from which it received the result. This procedure is repeated until all the tasks have been finished.

4. Data Processing

Occurrence data of two species were used to validate the parallel algorithm and 18 species to evaluate the regularization parameter. The species used in the experiments of parallelization validation were *Byrsonima intermedia* and *Xylopia aromatica*. The first one is a shrub of the *Malpighiaceae* family. It is popularly known as small-murici and is a native medicinal species of the Brazilian cerrado. The second one is a small tree of the *Annonaceae* family. It is popularly known as malagueto and is found in the Brazilian cerrado. It is commonly used for firewood. These two species are derived from SinBiota – environmental information system for the program Biota/Fapesp[b]. The experiments were carried out with 38 records of *Byrsonima intermedia* species and 33 records of *Xylopia aromatica* species. These species were chosen because they had been used in other experiments in openModeller.

A dataset with species of *Krameria* genus was used for evaluating the regularization parameter. The *Krameria* (*Krameriaceae*) genus presents 18 species with geographical distribution encompassing the Neotropical region. There are two peculiar characteristics of this genus: 1) the *Krameria* species presents a hemi parasitic habit – their leaves are green and able to carry their own photosynthesis but they need hosts (other non-specific plants) to grow; 2) their flowers present special glands that produce floral oil – this oil attracts specialized bees (Centris, Apidae) that pollinate the flowers [15,16]. The North American *Krameria* species occur in deserts in Mexico and in USA, and the South American ones occur in seasonally dry regions (mainly in the Brazilian *cerrado* and *caatinga*) and at high altitudes near the Andes [16]. The species diversity center is located in Mexico and presents eleven species; the secondary one is located in Brazil (5 species) [16].

The occurrence points of 18 *Krameria* species were obtained at biodiversity portals of the Internet, mainly in the Global Biodiversity Information Facility[c], Southwest Environmental Information Network[d], *Comisión Nacional para el Conocimiento y uso de la Biodiversidad*[e] and speciesLink[f]. The papers by Simpson [16] and by Simpson et al. [17] were also used. The geographical coordinates obtained were checked, converted and corrected, whenever necessary, using the available tools of speciesLink website. The final database used presented 2748 occurrence points. Fig. 3 shows the *Krameria* occurrence points.

Since the occurrence points of the species used in the validation of the algorithm parallelization were from the Brazilian *cerrado*, the models were generated for the São Paulo state, Brazil. Thus the environmental layers were from the same region with spatial resolution of 30 arc-second provided by WorldClim – Global Climate Data [18]. The layers used were annual mean temperature, mean diurnal range, maximum temperature of the warmest month, minimum temperature of the coldest month, annual precipitation, precipitation of the wettest month and precipitation of the driest month, totaling 7 layers.

For the experiments with *Krameria*, 37 environmental layers were used including the American continent with spatial resolution of 5 arc-min. The layers used were maximum temperature for twelve months, minimum temperature for twelve months, precipitation for twelve months and altitude. WorldClim also provided these layers.

The sequential and the parallel version of maxent were run with the aim of comparing the performance regarding the running time between them. Two measurements were carried out: the total running time of the application and the running time of the algorithm to estimate the maxent parameters. Both running were carried out in the openModeller project cluster. The cluster is a SGI Altix XE 1300 system composed by an input node Altix XE 210

---

with two Xeon quad Core processors of 2.00GHz, 8 GB of RAM, hard disk of 500 GB, 24-port InfiniBand switch, 24-port Gigabit ethernet switch, SGI Propack 5, SUSE Linux 10, and 10 Altix XE 310 nodes, each one with two Xeon quad Core processors of 2.00GHz, 8 GB of RAM and hard disk of 250 GB, totaling 80 cores. The total running time of the application was measured with timeg command available in Linux. The running time of the algorithm to estimate the maxent parameters was measured through code instrumentation.

The sequential version of the algorithm was run 10 times for each species and the values considered were the averages of the two times measured. In the parallel version of the algorithm, the number of processes varied from 3 to 8. The number of processes was upper limited according to the number of layers, thus one process run as master and seven processes run as slaves, one for each layer. In the parallel version, for each species, 10 models were generated for each number of processes and the averages were considered.
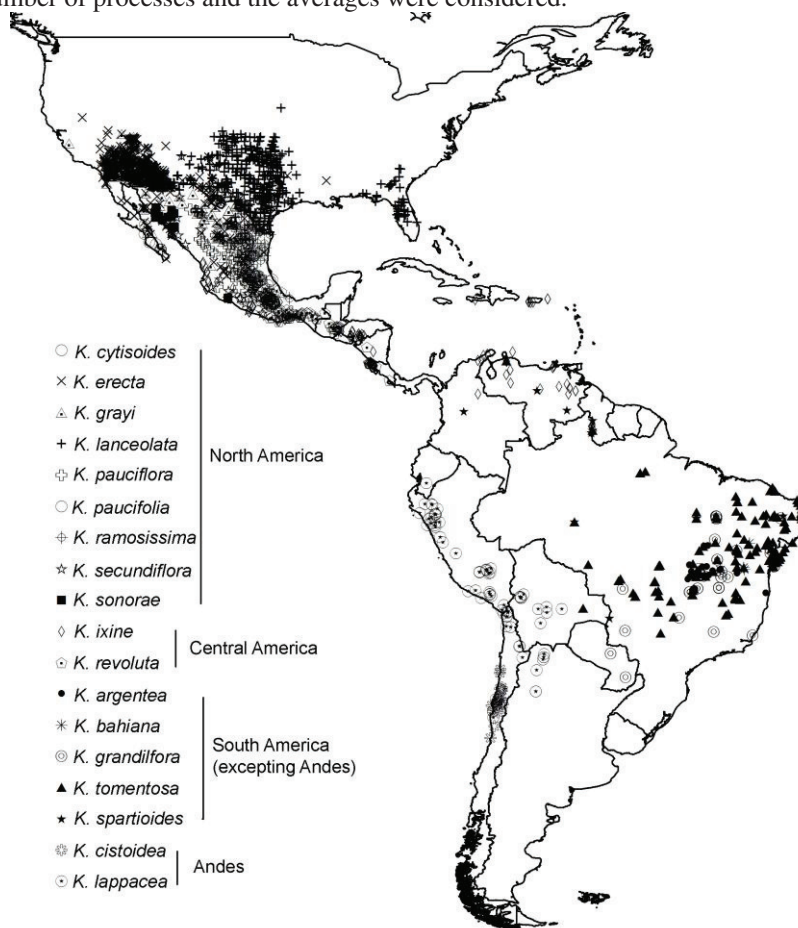

Fig. 3. Krameria occurrence points.

The experiments with regularization parameter were carried out in a computer with Core 2 Duo Intel processor of 1.66 GHz and 2 GB of RAM. The Operational System used in this architecture was Ubuntu 8.04, a Linux distribution. Since the aim of this part of the work was to evaluate the influence of the regularization parameter

---

f http://www.ncsa.uiuc.edu/~gbauer/time.1.html

choice over the models, the measures collected were model accuracy, AUC (Area Under the ROC curve) and number of iterations.

5. Results and Discussion

5.1. Parallelization

In the sequential version run, the total running time average of the application was 11.18 seconds for modeling the *Xylopia aromatica* species and 11.17 seconds for modeling the *Byrsonima intermedia* species. The running time of the algorithm to estimate the maxent parameters was 5.03 for modeling the first species and 5.02 for the second one. Fig. 4 shows the parallel version running time for modeling the *Xylopia aromatica* and *Byrsonima intermedia* species. The greater columns represent the total running time of the application and the smaller columns represent the running time of the algorithm to estimate the maxent parameters.
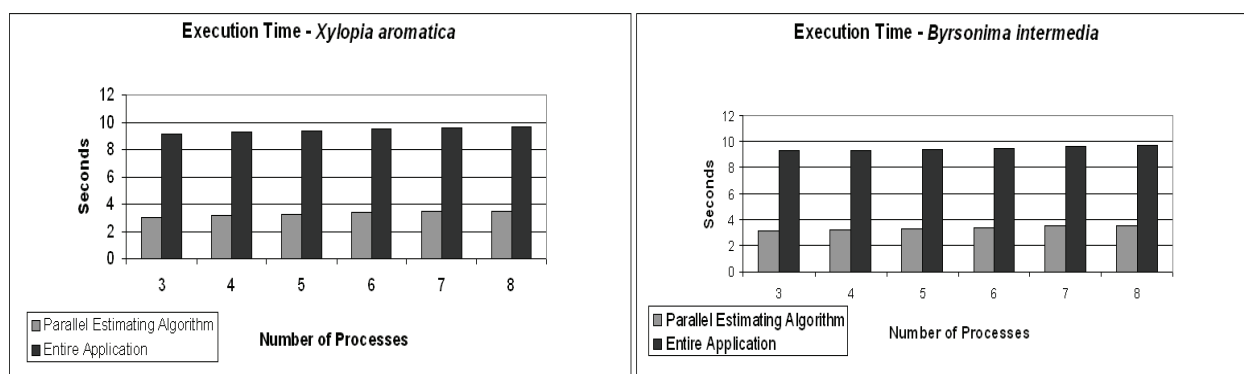

Fig. 4. Running time of the parallel maxent algorithm with different number of processes for *Xylopia aromatica* and *Byrsonima intermedia* species.

The best running times were achieved when the algorithm ran using 3 processes, in both modeling. For *Xylopia aromatica*, the best times were 5.03 and 11.18 seconds and, for *Byrsonima intermedia*, the best times were 5.02 and 11.17 seconds. The increase in the number of processes did not decrease the running time. Conversely, the running time increased as the number of processes was extended, although the time variation had been very small.

It is important to point out that even though the running time had increased when the number of processes was extended, there was a considerable performance gain of the parallel version with respect to the sequential version. The performance gain in modeling *Xylopia aromatica* species was almost 40% for the estimating parameters algorithm and approximately 18% for the entire application. The performance gain in modeling Byrsonima intermedia species was about 37% for the estimating parameters algorithm and approximately 17% for the entire application.

Analyzing the results obtained, it is possible to observe that the parallelized procedure had a significant contribution to the performance improvement in the estimating parameters algorithm. However, the performance gain for the entire application did not occur in the same proportion, indicating that other procedures must also be parallelized for the best exploitation of the openModeller project cluster.

5.2. Regularization Parameter

Table 1 shows the best results attained in the experiments carried out for tuning the regularization parameter. The first and the second columns contain the name and the number of samples of each species. The third column has the

best value of the regularization parameter for each species considering the collected statistics. The other columns have the best value of each measure considered in the evaluation of the regularization parameter. The choice of the best value for regularization parameter was based on the best values of accuracy, AUC and number of iterations, respectively. The Lappacea species was discarded from the data set because its AUC was less than 0.75 and just the values above 0.75 are considered useful [13,19].

Table 1. Best results of regularization parameter.

Species	Number of samples	Regularization Parameter	Accuracy (%)	AUC	Number of iterations
Paucifolia	4	0.07	100	1.0	3752
Bahiana	14	0.9	100	0.88	1036
Cistoidea	19	0.05	100	1.0	9632
Sonorae	19	0.9	100	0.94	1250
Spartioides	27	0.9	100	0.93	1369
Grandiflora	35	0.9	100	0.89	1408
Argentea	45	1.0	100	0.93	1362
Ramosissima	66	0.9	100	0.98	1357
Secundiflora	66	1.0	100	0.91	1257
Revoluta	77	1.0	89.61	0.95	1804
Pauciflora	80	1.0	86.25	0.80	392
Cytisoides	134	1.0	97.01	0.93	684
Ixine	162	1.0	79.01	0.97	1816
Tomentosa	163	0.07	68.71	0.95	5542
Lanceolata	465	0.7	96.56	0.96	1614
Grayi	511	0.9	89.04	0.98	1863
Erecta	755	1.0	86.49	0.97	1978

The results shown in Table 1 indicate that the regularization parameter does not depend on the number of samples or on the number of iterations in training. Moreover, species with the same number of samples or with a number of samples close to each other, fit better with different values of the regularization parameter (different in magnitude order), such as *Cistoidea* and *Sonorae*, *Ixine* and *Tomentosa*. This behavior reinforces the difficulty in tuning this parameter. As an alternative, this parameter was removed from the algorithm and an adaptive approach was considered for their replacement. The best results are presented in Table 2.

Table 2. Maxent without regularization parameter and with an adaptive approach.

Species	Maxent without regularization			Adaptive approach		
	Accuracy (%)	AUC	Iterations	Accuracy (%)	AUC	Iterations
Paucifolia	100	1.0	9007	100	1.0	5421
Bahiana	78.57	0.97	2588	78.57	0.97	2410
Cistoidea	94.74	1.0	10000	94.74	1.0	10000
Sonorae	68.42	0.99	4290	68.42	0.99	3208
Spartioides	62.96	0.97	1262	66.67	0.97	1408
Grandiflora	54.29	0.94	3628	57.14	0.95	3341
Argentea	37.78	0.94	200	37.78	0.94	213
Ramosissima	96.97	0.99	1762	96.97	0.99	1854
Secundiflora	27.66	0.74	49	27.66	0.74	54
Revoluta	55.84	0.95	377	63.64	0.96	446
Pauciflora	0.00	0.43	17	0.00	0.45	18
Cytisoides	0.75	0.59	23	0.75	0.56	26
Ixine	59.26	0.97	484	59.26	0.97	502
Tomentosa	54.60	0.95	4055	57.67	0.94	4042
Lanceolata	86.88	0.95	903	87.96	0.96	972
Grayi	74.17	0.98	3155	74.17	0.98	2738
Erecta	62.12	0.97	904	61.46	0.97	896

The results presented in Table 2 show that removing the regularization parameter causes a decline in the measurements. The adaptive approach presented almost the same results as the algorithm without regularization parameter. Therefore, it cannot replace the regularization parameter. Fig. 5 shows an example of a generated model. Hot colors represent more suitable environmental conditions for the species.

6. Conclusion

The aim was to evaluate different aspects of the maximum entropy algorithm for modeling of geographic distributions of species. The first evaluation was to compare the performance regarding running time of a parallel maxent algorithm and its sequential version available in openModeller. The MPI library was used to implement the parallelism proposed. Some experiments were carried out with occurrence data of two species. For both data sets, the parallel version achieved a significant performance gain in relation to the sequential version, about 40% for one species and 37% for the other. These results motivate the research and development of high-performance methods for modeling of geographic distribution of species.

The second evaluation was about tuning the regularization parameter. The results show that there is no relation between regularization parameter and the number of samples, or with the number of iterations in the training process. Moreover, species with the same or a similar number of samples fit better with different values of the regularization parameter (different magnitude order). The adaptive approach could not replace the regularization parameter because it causes a decline in statistics, that is, worse models.
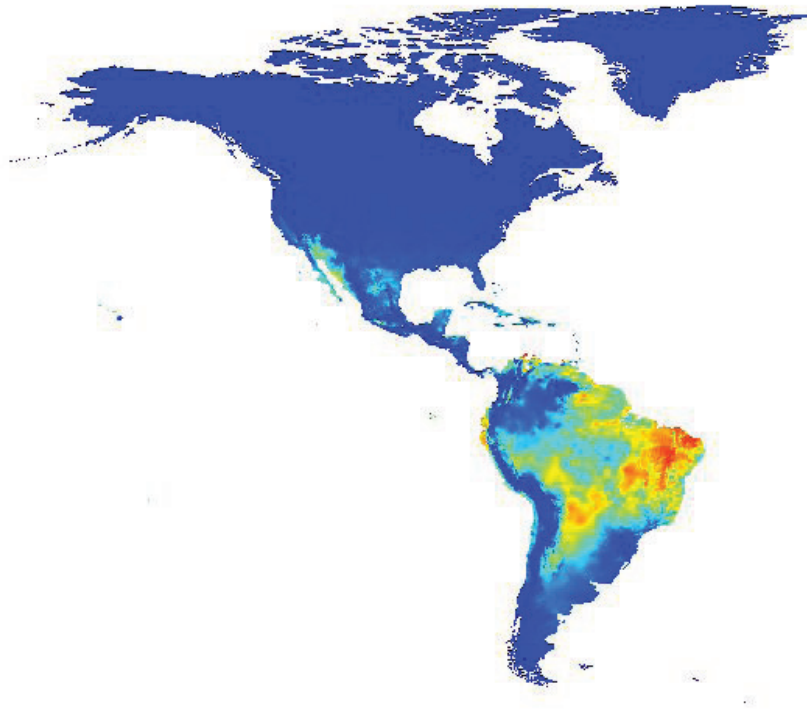

Fig. 5. Best model for Tomentosa species.

The algorithm to estimate the maxent parameters available in openModeller is based on the sequential algorithm used in the MaxEnt program. Therefore, one of the future works suggested is to research and develop another parallel version from an algorithm that updates all parameters at each step. Hence, the modeling of geographic

distribution of species with maxent would benefit more from the parallel programming methodology. Another promising future work is the selection of a larger data set to study the behavior of the regularization parameter.

Acknowledgements

References

[1] Bartel RA, Sexton JO. Monitoring habitat dynamics for rare and endangered species using satellite images and niche-based models. *Ecography* 2009; **32**: 888–896.

[2] Fitzpatrick MC, Gove AD, Sanders NJ, Dunn RR. Climate change, plant migration, and range collapse in a global biodiversity hotspot: the Banksia (Proteaceae) of Western Australia. *Global Change Biology* 2008; **14**: 1337–1352.

[3] Raxworthy CJ, Ingram CM, Rabibisoa N, Pearson RG. Applications of ecological niche modeling for species delimitation: A review and empirical evaluation using day Geckos (Phelsuma) from Madagascar. *Systematic Biology* 2007; **56**(6): 907–923.

[4] Chen SH, Jakeman AJ, Norton JP. Artificial Intelligence techniques: An introduction to their use for modelling environmental systems. *Mathematics and Computers in Simulation* 2008; **78**(2–3) 379–400.

[5] Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecol Model* 2006; 190: 231–259.

[6] Muñoz MES, Giovanni R, Siqueira MF, Sutton T, Brewer P, Pereira RS, Canhos DAL, Canhos VP. OpenModeller: a generic approach to species' potential distribution modeling. *GeoInformatica*; 2009.

[7] Rodrigues ESC, Rodrigues FA, Rocha RLA, Corrêa PLP. An adaptive maximum entropy approach for modeling of species distribution. *Memories of WTA* 2010 – IV Workshop of Adaptive Technology. 2010; ISBN 978-85-86686-56-6.

[8] Hutchinson GE. *Introduction to ecology of populations*. Barcelona, Editorial Blume, 1981; 492p (In Spanish).

[9] Santana FS, Siqueira MF, de Saraiva AM, Corrêa PLP. A reference business process for ecological niche modelling. *Ecol Inf* 2008; **3**: 75–86.

[10] Siqueira MF. *Use of fundamental niche modeling in the pattern evaluation of vegetal species geographic distribution*. PhD Thesis. Department of Ambient Engineering of University of São Carlos. São Carlos/SP – Brazil; 2005 (In Portuguese).

[11] Graham CH, Ferrier S, Huettman F, Moritz C, Peterson AT. New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution* 2004; **19**(9): 497–503.

[12] Phillips SJ, Dudík M, Schapire RE. A maximum entropy approach to species distribution modeling. *Proceedings of the 21st International Conference on Machine Learning*, Banff, Canada; 2004.

[13] Phillips SJ, Dudík M. Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography* 2008; 31: 161–175.

[14] Rodrigues FA, Rodrigues ESC, Sato LM, Midorikawa ET, Corrêa PLP, Saraiva AM. Parallelization of the jackknife algorithm applied to a biodiversity modeling system. *Proceedings of the 7th International Information and Telecommunication Technologies Symposium* – I2TS'2008, 2008; 58–65.

[15] Vogel S. Ölblumen und ölsammelnde Bienen. Tropische und Subtropische Pflanzenwelt 1974; **7**: 285-547.

[16] Simpson BB. Krameriaceae. Flora Neotropica. *Monograph* 1989; **49**, 1-109.

[17] Simpson BB, Weeks A, Helfgott DM, Larkin LL. Species relationships in Krameria (Krameriaceae) based on ITS sequences and morphology: Implications for character utility and biogeography. *Systematic Botany* 2004; **29**: 97-108.

[18] Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Clim* 2005; **25**: 1965–1978.

[19] Elith J. Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. In: Ferson S and Burgman M (eds), *Quantitative methods for conservation biology*. Springer, 2002; pp. 39-58.