Environment, Resource and Ecology Journal (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/erej.2022.060501
ISSN 2616-3756 Vol. 6 Num. 5

Problems and Countermeasures in Prediction Model of Potential Distribution of Plant Species with Extremely Small Populations

Wei Hu, Xin Liang, Yaqi Zhang, Xiansheng Tan, Renyan Duan*

College of Agriculture and Biology Technology, Hunan University of Humanities Science and Technology, Loudi, Hunan, 417000, China
*Corresponding author

Keywords: Plant species with extremely small populations, Suitable habitat, Species distribution model, Prediction model

Abstract: The plant species with extremely small populations (PSESP) is a kind of wild plant population that is on the verge of extinction. It is important to reasonably predict the suitable distribution area of the PSESP for its protection. However, the lack of suitable prediction models, low quality of distribution data and high sensitivity of simulation process can limit the application of species distribution models in the protection of PSESP. It is suggested that a small model niche factor model set should be constructed and optimized according to the different distribution characteristics of PSESP. In order to improve the quality of distribution data, the method of combining real species with virtual species is used to conduct field iterative stratified sampling. Based on the distribution data of iterative stratified sampling lifting and the optimized small model niche factor model, the key factors in the simulation process are optimized to reasonably predict the suitable distribution area of PSESP. Based on the distribution of nature reserves and the potential distribution of PSESP, field investigation and verification, the appropriate in situ or ex situ protection strategies for PSESP can be proposed.

1. Introduction

Plant species with extremely small population (PSESP) is a kind of wild plant species that is endangered, and has important ecological, conservation and economic values. Its distribution area is narrow or intermittent, and its population is degraded and its number continues to decrease under the stress of external factors for a long time. The smallest surviving population whose population and individual number are lower than the stable survival limit is one of the species with extremely high extinction risk[1]. The PSESP have the following characteristics: (1) extremely small and extremely endangered wild population; (2) unique habitat requirements and narrow ecological amplitude; (3) unclear potential gene value with high gene loss and biodiversity reduction if extinction. In view of the fact that PSESP are one of the most easily lost biological resources, protecting them is the most important thing for biodiversity protection[2].

Due to the decline of the viability of wild plants, habitat destruction and over-utilization of resources caused by human activities, etc., plant species with extremely small population are more likely to be endangered[1-2]. With the rapid development of biomathematical models, computational science, big-data technology and biogeography, species distribution modelings (SDMs) can be used to quickly simulate and predict the dynamic changes of the suitable distribution areas of biological populations, which is of great significance to formulate in-situ or ex-situ protection policies for PSESP[3]. Species distribution model mainly quantifies the internal relationship between species (appearance data and missing data) and environment (environmental information characteristics of sites) through a specific mathematical model, evaluates the environmental niche of species, and finally predicts the suitable distribution area of species in the form of occurrence probability.

2. Existing Problems in Predicting the Potential Distribution of PSESP

There are three main problems limiting the prediction of the suitable distribution area of small population plants: 1) Lack of suitable prediction model of PSESP. In order to meet the requirements of certain accuracy and stability of the model, the basic condition of species modeling is that the number of rasterized sites should be more than 10 times of the predicted variables used for modeling[3,4]. However, it is difficult for thePSESP to meet the basic conditions of modeling, and there is a paradox of minimal population modeling. The main results are as follows: ① The distribution points are extremely limited, and some of them are even less than 500 strains. ② Some species are clustered, and few spots appear after grid[3,5-6].2) The data quality of the distribution of PSESP is low. The PSESP are scattered in inaccessible places, and there is sampling deviation, which leads to low representativeness of distribution data[6-7]. In addition, the survey of distribution data involves the identification of plants, the accuracy of GPS and the professional knowledge of collectors. However, most of the obtained distribution data have a long time span and involve a large number of investigators, which leads to the low quality of distribution data. In order to expand the distributed data and improve the data quality, it is necessary to conduct targeted supplementary investigation under the guidance of a specific model. 3) The prediction results are highly sensitive to the key variables selected in the simulation process. There are many programs in the process of model simulation, which affects the stability of the results. The distribution points are mostly concentrated in a relatively narrow area, with high spatial autocorrelation. The prediction results of suitable distribution areas of species are highly sensitive to the key variables (spatial resolution, spatial autocorrelation, missing data selection, threshold criteria, etc.) selected in the simulation process[7-9].

There are three problems in the simulation of the suitable distribution area of PSESP, which will lead to inaccurate simulation results and limit the application of species distribution model in the protection of PSESP. If the commonly used species distribution model is used indiscriminately in PSESP, it may be will lead to wrong prediction conclusions[5,10-11]. Therefore, how to accurately predict the suitable distribution area of PSESP according to the distribution characteristics of PSESP is a scientific problem to be solved urgently.

3. The Solution Method of Predicting the Potential Distribution of PSESP

3.1. A Small Model Set

Small model set is a set of species distribution models that only need a few simulation variables, which can solve the problem of too many simulation variables and insufficient distribution data. Generally speaking, the number of sample occurrence points should be at least 10 times larger than the prediction variables used for modeling, and the more data points, the more accurate it is[3,4,10]. However, for the PSESP, extremely low occurrence data and high sampling deviation will limit the application of species distribution model, leading to "paradox of PSESP"[3,7]. By greatly reducing the number of predictors in the species distribution model, the small model set can avoid the problem that the traditional species distribution model needs a large number of species occurrence points (grid sites) because of the large number of predictors[3]. The basic idea of small model set is to simulate species distribution based on small and simple models. By establishing a series of competing models, instead of the traditional single model, the traditional model is overcome by the limitation of the number of predicted variables, thus providing an ecologically and statistically correct model. The advantage of small model set lies in its high performance and small sample size, which can better meet the data characteristics of PSESP.

3.2. The Niche Factor Model

Niche factor model is a method based on the comparison between the environmental niche of species and the environmental characteristics of the whole study area. It is mainly based on the emergence data (without missing data) and background geographic information system to build a model for prediction[12]. Different from the traditional forecasting methods of species distribution models, the niche factor model mainly converts the original eco-geographical variables into new continuous axes through principal component analysis (PCA) to obtain the correlation factors between the centroid of the observed species and the centroid of the background grid in the study area, and calculates the habitat suitability of each grid through the niche factor model. The biggest advantage of the ecological factor model is that there is no need for missing data in the whole process of the model operation, and the requirements for the quantity and accuracy of data appearing are also low[12]. Most traditional species distribution models need data and missing data. Some models don't need missing data even in the construction stage, but they also need missing data in the evaluation stage. For PSESP, the location and quantity of missing data have great influence on the model.

3.3. Virtual Species

Virtual species is the construction of virtual species, the establishment of virtual and real species databases, and the expansion of species occurrence databases and missing databases. A virtual artificial species based on real or virtual geographic information system is a simplified and abstract species. Virtual species have the advantages of easy data acquisition, controllable data quality and avoiding over-simulation. Virtual species is an indispensable and important tool in large-scale research, which is conducive to solving scientific problems that cannot be solved by real data. The distribution of virtual species can be adjusted and set artificially by adjusting the composition algorithm according to the research needs, so its actual distribution (data appearing/missing) is completely known, which overcomes the shortage of unknown data points or sampling deviation of real species[7].

Iterative sampling can effectively increase the quality of distributed data or the representativeness of sampling. By constructing virtual species, the sample data can be sampled hierarchically to increase the number of distributed data[6]. Based on the prediction results of small model-niche factor model set, stratified sampling (field investigation, verification and stratified simulation) can improve the accuracy of the data of known distribution points and expand the distribution point databases (existing databases and missing databases). Through the small model-niche factor model set simulation, the sample data (real data and virtual data) are sampled by field iterative stratified simulation, which can effectively improve the quality of distributed data or the representativeness of sampling.

3.4. Optimizing the Simulation Process

The main uncertainties in the simulation process of the suitable distribution area of small population plants include: ① the choice of spatial resolution; ② avoiding spatial autocorrelation; ③ selection of missing data; ④ selection of threshold standards, etc. In each step of ecological simulation, a variety of algorithms can be used. The accuracy of the algorithm plays a decisive role in the stability and prediction ability of the model[6,9]. The algorithm and working principle of the model are different with different research objects[8,10]. The proposal of virtual species and the optimization of simulation process are conducive to monitoring the data of PSESP acquired in real environment, avoiding the influence of unknown distribution area of real species, reducing the influence of random sampling, and improving the accuracy and stability of the model.

4. Conclusions

Therefore, in order to study the suitable distribution area of the PSESP, the approximate distribution range of PSESP can be obtained through field investigation, aerial photography by unmanned aerial vehicles and analysis of satellite images. Based on the virtual species with similar distribution to the real minimal population plants constructed in the real habitat, under the guidance of the constructed and optimized small model-niche factor model set, the field investigation is supplemented. Through iterative stratified sampling, the number of distributed data is increased and the representativeness of sampling is improved. By optimizing the simulation process, optimizing the selection of key variables (spatial resolution, spatial autocorrelation, missing data and threshold standards, etc.) of the model, the prediction accuracy and stability of the species distribution model are improved. The optimized simulation process is adopted to predict the suitable distribution area of the PSESP, and the appropriate in-situ or ex-situ protection measures for the PSESP are put forward in combination with the distribution of the existing protected areas.

Acknowledgment

This work was supported by grants from the Natural Science Foundation of Hunan Province (2021JJ30375), the natural science foundation of Hunan Provincial Department of Education (20A275), and Hunan Postgraduate Scientific Research Innovation Project (CX20221256).

References

[1] Ma, Y.P., Chen, G., Grumbine, R.E., Dao, Z.L., Sun, W.B. and Guo, H.J. (2013) Conserving plant species with extremely small populations (PSESP) in China. Biodiversity and Conservation, 22, 803-809.
[2] Qu, H., Wang, C.J. and Zhang, Z.X. (2018) Planning priority conservation areas under climate change for six plant species with extremely small populations in China. Nature Conservation, 25, 89-106.
[3] Breiner, F.T., Guisan, A., Bergamini, A. and Nobis, M.P. (2015) Overcoming limitations of modelling rare species by using ensembles of small models.Methods in Ecology and Evolution, 6, 1210-1218.
[4] Støa, B., Halvorsen, R., Stokland, J.N. and Gusarov, V.I. (2019) How much is enough? Influence of number of presence observations on the performance of species distribution models.Sommerfeltia, 39, 1-28.
[5] Gábor, L., Moudrý, V., Lecours, V., Malavasi, M., Barták, V., Fogl, M., Šímová, P., Rocchini, D. and Václavík, T. (2020) The effect of positional error on fine scale species distribution models increases for specialist species. Ecography, 43, 256-269.
[6] Burrows, M.T., Schoeman, D.S., Richardson, A.J., Molinos, J.G., Hoffmann, A., Buckley, L.B., Moore, P.J., Brown, C.J., Bruno, J.F., Duarte, C.M., Halpern, B.S., Hoegh-Guldberq, O., Kappel, C.V., Kiessling, W., O'Connor, M.I., Pandolfi, J.M., Parmesan, C., Sydeman, W.J., Ferrier, S., Williams, K.J. and Poloczanska, E.S. (2014) Geographical limits to species-range shifts are suggested by climate velocity. Nature, 507, 492-495.
[7] Komori, O., Eguchi, S., Saigusa, Y., Kusumoto, B. and Kubota, Y. (2020) Sampling bias correction in species distribution models by quasi-linear poisson point process. Ecological Informatics, 55, 101015.
[8] Qiao, H.J., Feng, X., Escobar, L.E., Peterson, A.T., Soberón, J., Zhu, G.P. and Papeş, M. (2019) An evaluation of transferability of ecological niche models. Ecography, 42, 521-534.
[9] Tikhonov, G., Duan, L., Abrego, N., Newell, G., White, M., Dunson, D. and Ovaskainen, O. (2020) Computationally efficient joint species distribution modeling of big spatial data. Ecology, 101, e02929.
[10] Duan, R.Y., Kong, X.Q., Huang, M.Y., Varela, S. and Ji, X. (2016) The potential effects of climate change on amphibian distribution, range fragmentation and turnover in China. PeerJ, 4, e2185.
[11] García-Roselló, E., Guisande, C., González-Vilas, L., González-Dacosta, J., Heine, J., Pérez-Costas, E. and Lobo, J.M. (2019) A simple method to estimate the probable distribution of species. Ecography, 42, 1613-1622.
[12] Engler, R., Guisan, A. and Rechsteiner, L. (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology, 41, 263-274.