

MAXIMUM ENTROPY MODELING USING CITIZEN SCIENCE: USE CASE ON JACOBIN CUCKOO AS AN INDICATOR OF INDIAN MONSOON

Priyanka Singh^{1,*}, Sameer Saran¹

¹ Department of Geoinformatics, Indian Institute of Remote Sensing, ISRO, Dept. of Space, Govt. of India, 4 Kalidas Road, Dehradun – 248001, Uttarakhand, India - priyankaiirs@gmail.com, saran.iirs@gmail.com

Commission V, WG V/3

KEY WORDS: Citizen Science, Maximum Entropy Approach, Biodiversity, Range Expansion

ABSTRACT:

Due to a growing revolution of the citizen science era with the involvement of non-professionals in scientific tasks such as species observation, yields an opportunistic data for modeling and planning purposes. Such citizen science based scientific observations can be a sustainable option to answer many research questions. Here, citizen science data of the *Clamator jacobinus* bird is taken from Global Biodiversity Information facility to predict its habitat suitability through maximum entropy approach. The distribution data is divided into two monthly sets – June to October and November to May for critically analysing the probable climatic reasons for its migration and understanding the influence of climatic variables in its suitability during the Indian monsoon season and Southern Africa rainy season. Also, the influential role of different bioclimatic variables in determining the bird's suitability is described in this paper and to predict how this bird will react to different climate change scenarios in 2050 year. The maximum entropy modeling is performed on both sets of data and results are surprisingly interesting, which verifies an Indian myth that this bird is harbinger of the monsoon in India. This study concluded that the precipitation during warmest and wettest quarter, and isothermality are the major factors in determining the migration of *Clamator jacobinus*, but, hot, dry and cold climate is not suitable for this bird suitability. Such study using the citizen science data can be used in biodiversity planning as well as in improving the agricultural economy because monsoon is considered as an auspicious season for functioning of biodiversity and agricultural tasks.

1. INTRODUCTION

The idea of studying climatic variations and its strong influence on species distribution started back in the literature to around 5th century BCE (Woodward, 1987). In twentieth century, this study leads to general suite of powerful maximum entropy models that can be used in generating the predictive suitability of species occurrences by occupying the detailed environmental variables (Franklin, 2010; Richardson, Whittaker, 2010). Such a maximum entropy approach of estimating the different environmental requirements of geographically distributed species using presence/absence data are known as a climate envelope modeling, ecological niche modeling, species distribution modeling, niche-based modeling or habitat suitability modeling. With such a widespread interest, niche-based models and species distribution models have always faced some conflicting views on what they truly represent. Some of the views on niche-based is that this model provides an approximation to the species' fundamental niche (Soberón, Peterson, 2005), while others illustrated that the species distribution model provides an spatial depiction of the realized niche on the grounds of spatial distribution patterns of observed species which are utilized to estimate the frequencies among species and environment relationships, inhibited by non-climatic factors (for e.g. Austin et al. 1990; Guisan, Zimmermann, 2000; Pearson, Dawson, 2003).

In recent decades, countless studies are carried out on maximum entropy method (SDM) to evaluate its accuracy of predicting species range using presence only data (some of the studies are Elith* et al. 2006; Hernandez et al. 2006; Hu, Jiang, 2011; Kadmon et al. 2003; Skidmore et al. 1996; Stockwell, Peterson, 2002; Wisz et al. 2008). Hence, these studies are now serving as a guideline on what would be the minimum sample size, how many species occurrences are required, what would be the

prediction on selection of random samples from voluminous data set and if the random subsampling process would be opted then what range predictions model will on per species at per sample size (Elith* et al. 2006; Hernandez et al. 2006; Kadmon et al. 2003; Stockwell, Peterson, 2002; Wisz et al. 2008). Due to its increasing popularity, maximum entropy model (Maxent) becomes widespread for species distribution modeling and papers/articles familiarized it had received more than 5,000 citations in the Web of Science Core Collection, and more than 60% of the distribution modelers report uses it (Ahmed et al. 2015). The species information from herbaria and museum, theoretically observed data and survey data, can deliver a substantial amount of resource information (Chapman, Grafton, 2008) for modeling the species habitat suitability.

But, the main challenge that exists in these datasets is the location uncertainty (the error in geotagging), might have been caused due to incorrect geotagging of places or setting of GPS to improper datum. (Graham et al. 2004; Wieczorek et al. 2004). Afore popularization of citizen science and geospatial technology, species data were collected and recorded as a textual description in a form of names and places which changed over a while (Wieczorek et al. 2004). However, digitization of these textual descriptions provided substantial errors in positional uncertainty (~ several kilometres) (Feeley, Silman, 2010). As the geographical coordinates played a major role in extracting the co-located environmental variables and such erroneous occurrences would provide inaccurate species-environment relationship (Feeley, Silman 2010). Nevertheless, researchers have developed various techniques which estimates and documents the positional uncertainty among species occurrence records which removes the highest uncertainties prior to suitability modeling (Guo et al. 2008; Wieczorek et al. 2004) but this reduction in sample size causes model inaccuracy (Graham et al. 2008; Hernandez et al. 2006). After assessment

of such uncertainty in data collection process, citizen science, with trained or untrained participants, and information technology are considered as a robust and rigor data collection efforts ensuring qualitative and quantitative distribution records.

Citizen science ‘is a process where concerned citizens, government agencies, industry, academia, community groups, and local institutions collaborate to monitor, track and respond to issues of common community concern’ (Whitelaw et al. 2003) “where citizens and stakeholders are included in the management of resources” (Cooper et al. 2007; Howe, 2000). In this paper, study is carried out on a current and future habitat suitability of bird, *Clamator jacobinus* (known as jacobin cuckoo or pied cuckoo) by implementing using the maximum entropy model. The migratory movements of this bird are considered as good indicators of monsoon in Southern Africa and India (Urvi, 2011). As per the Indian myth, arrival of Pied Cuckoo is known for beginning of monsoon, therefore, the purpose of this study is to disparagingly analyze the likely climatic reasons for its migration and which environmental variables play an influential role in its suitability. As, this species are sighted in March and April from southern India, where the species is known for its extant year-round until the middle of May. When the monsoon hits the Andaman, the first birds in northern India are seen with summer monsoon winds, afterward these birds are sighted across the West, North and East. In first week of June, the monsoon reaches Kerala and these birds are seen everywhere, except the extreme North-West and West. The extant range of *Clamator jacobinus* with its resident, breeding and non-breeding (downloaded from The IUCN Red List of Threatened Species) is shown in Figure 2, which is used in validating the results of this study, later in this paper.

2. NEED FOR CITIZEN SCIENCE IN BIODIVERSITY

The automated way of collecting geographical information on the species occurrence (Damoulas et al. 2010; Stoeckle, 2003; Turner et al. 2003) by human observers are considered as a reliable source. In biodiversity domain, many organizations have designed citizen science projects as per their scientific needs to make fine resolution data with added rigorous sampling locally and globally, such as bioblitzes (Novacek, 2008), shell polymorphism survey (Silvertown et al. 2011), water quality survey (Conrad, Hilchey, 2011), and breeding bird surveys (Freeman et al. 2007). However, many informatics challenges occur in biodiversity data collection which have been tackled in recent years are as follows (i) apposite planning required for management of voluminous set of data by handling the data-management infrastructure which also helps in motivating the volunteers, and (ii) data quality and handling of certain observational biases essential to sighting variation among volunteers (Cooper et al. 2007), ‘false absences’ that yields inadequate sightings (McClintock et al. 2010) and often uneven data distributions (Boakes et al. 2010). Thereby, the need of addressing such challenges prompted the rise of data intensive science (Hey et al. 2009), which is being applied to large-scale citizen-science based research (Kelling et al. 2009). After reviewing accuracies of citizen observatory data, the observation data Global Biodiversity Information Facility (GBIF) repository is used in this paper to apply a data intensive science approach in maximum entropy modeling. The motivation and solution for using occurrences data of GBIF because GBIF has a process of managing the voluminous as well as continual data, acquired from thousands of volunteers using informatics and social networking. Also, GBIF is an

authentic repository where various organizations/ institutes share their data by ensuring its data quality and, particularly data quantity which are relevant to modeling and decision-making purposes.

3. METHODS

3.1 Distribution Data

The distribution data obtained from the GBIF repository for *Clamator jacobinus* are total of 36,365 records, observed on the basis of specimen data, human and machine observations, holding very high number of NA values and duplicate records. To clean distribution records, NA values and repeated geographical records are eliminated and final set are left with 10,292 presence records among which majority is of citizen science data (10,160 records are human observations). These citizen observations are incorporated into the GBIF database by various organizations such as eBird, Kenya Bird Map, naturgucker, Safring, National Biodiversity Data Bank (1900–2000) and Southern African Bird Atlas Project2 (SABAP2). These data are continuously in use for developing of suitability models and for planning the safeguarding actions (Coxen et al. 2017; Pacifici et al. 2017; Robinson et al. 2018; Sullivan et al. 2017). For distribution data, this study targeted GBIF datasets because this is the most updated catalogue on species distributions, having occurrence datasets from resourceful and authentic citizen science databases, systematic surveying stations and ad hoc observations from experts. To the best of our knowledge, GBIF maintains a proper balance between quality and quantity of citizen science data at a broad scale. After accounting the null and duplicate records from raw data, 10,292 independent presence records are used in constructing the suitability model for *Clamator jacobinus* species. For the study researched in this paper, the distribution data is categorized into two sets on the basis of favourable climatological seasons as per target species extant - November to May set (mostly suitable for bird resident in Southern Africa) and June to October set (Indian monsoon).

3.2 Environmental Variables

3.2.1 Current Climatic Data

To study and determine the eco-physiological responses of *Clamator jacobinus* with respect to rapid changes in the environment, nineteen bioclimatic variables for the period 1970–2000 are obtained from the WorldClim database at a spatial resolution of 2.5 arc minutes (~4.5 km) catalogue (Fick, Hijmans, 2017; Hijmans et al. 2005; Smeraldo et al. 2018). These bioclimatic variables are built using monthly data from the 1st of January 1970 to the 31st of December 2000 at very high resolution, and contains more meaningful information than simple precipitation.

3.2.2 Future Climatic Data

For predicting the habitat suitability conditions from current to future, bioclimatic layers for the year 2050 (averaged for 2041–2060) analogous to the climatic responses of Representative Concentration Pathways (RCP) 8.5. These datasets are obtained on the basis of mean ensemble of various Coupled Model Intercomparison Project (CMIP5) models (Amman et al. 2003; Amman et al. 2007; Sato et al. 1993; Stenchikov et al. 1998) at spatial resolution of 2.5 arc minutes (~4.5 km). The RCP 8.5 is used to estimate a little migration by observing a pessimistic scenario, in which atmospheric CO₂ levels of 2100 are 2.5 times

higher than current levels (Riahi et al. 2011; Sharma et al. 2017).



Figure 1. *Clamator jacobinus* (Source: Gulshan 2019)

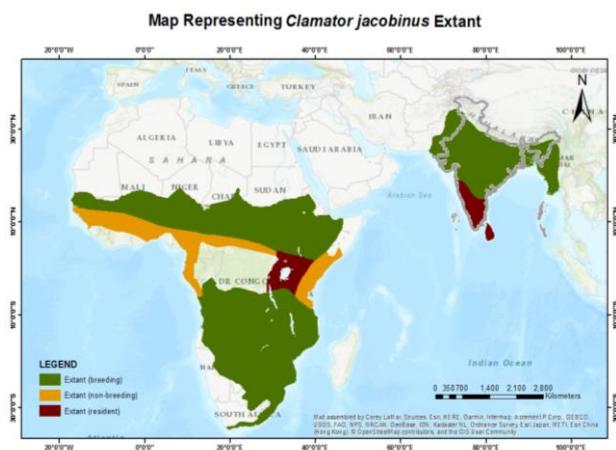


Figure 2. Map showing extant of *Clamator jacobinus*

3.3 Model Construction and Validation

The focal objective of this research study is to predict the scenarios of future suitability from *Clamator jacobinus* current suitability. For model construction related to this research, machine learning based Maxent approach is used to predict habitat suitability by establishing the spatial relationship between presence records and most significant bioclimatic variables, which further provides high extrapolation accuracies, even for absence or less presence records (Phillips et al. 2004; Phillips et al. 2006).

Before constructing the model, it is required to estimate reasonable geographic boundaries because the Maxent map outputs are scale dependent and the range of climates predict the most suitable habitat by relating the environmental conditions in which species are consistently absent. Therefore, if the range is too wide, then the conditions for favourable environments may be lost in the prevalent poor environments. For this necessarily primary approach, a two-degree buffer is considered around the extremes of the species extant and range of bioclimatic variables geographical boundaries. This model starts with uniformly distributed data of *Clamator jacobinus* and runs in multiple iterations with maximum substantial environmental variables until no improved predictions are built.

To assess the behaviour of model, a k-fold cross validation approach (Kohavi, 1995) is applied on occurrence data, which separated the data into two sets, each containing five random observation group (because five case fold is considered here) – train and test data. The train data holds 75% of total records for

model construction and remaining 25% is with test data to validate the model results. The second step was the selection of parameters - (i) betamultiplier - a regularization multipliers value is set to '1' because the smallest value makes the projected distribution more close fit to the training data, its value ranges from 1 to 15; (ii) Explanatory Variables (EVs) - in model, four variable transformation types are used as a calculation features for continuous EVs such as hinge, linear, quadratic and threshold (Phillips et al. 2017; Phillips, Dudík, 2008); and (iii) threshold features used are as 'lq2lqptthreshold', 'l2lqthreshold' and 'hingethreshold'.

The statistical assessment of model is done by the area under receiver operating characteristic (ROC) curve (AUC) which gives the probability of randomly chosen presence and absence site (Fielding, Bell, 1997). The ROC analysis is responsible for evaluating the model performance by two factors - sensitivity (absence of omission error) and specificity (absence of commission error) are used to test the predictions. The AUC process involves setting of thresholds on prediction to generate false positive rate (prediction of presence for site where species is absent) and then evaluates the true positive rate (successful presence prediction) as a function of false positive rate. The random ranking of sites has an average AUC of 0.5 and a perfect ranking of sites shows the best possible AUC of 1.0, also the models with AUC above 0.75 are considered possibly suitable with noble discriminatory command (Elith 2002; Phillips, Dudík, 2008).

4. RESULTS

Maxent is a machine learning process that goes through multiple iterations to convert a training model into acceptable model. Due to this stochastic nature, multiple replicates of the model must be run in order to bring the average of the outputs to a suitable result. Therefore, the model was set to ten replicate runs for two sample sets, in each set 75% of random samples are used for standardizing of model and remaining 25% are for assessing the model's implementation. The first set of data consists of citizen observation records from June to October month, which is Indian summer monsoon season with more than 90% of total annual precipitation in central and western parts of India, and 50%-75% of total annual rainfall in southern and north-western parts of India. The second set of data holds distribution records from November to May month (sighted by citizen scientists), covering four Indian climatological seasons – post-monsoon, winter and summer, and rainy season of Southern Africa where *Clamator jacobinus* is known for its resident.

4.1 Model Evaluation

After execution of model, its performance is measured by the plots shown in Figure 3 and 5. These plots deliver information on AUC for both training and test data, the red (training) line denotes the model is performing well for the training data and blue (testing) line indicates the performance of model with respect to the testing data, and it is good for model that training AUC will be higher than test AUC. If the model performs worse, then the blue line falls below the turquoise line, whereas, if blue line sets on top left of the graph, the model is better in predicting the present records on test data. If the sensitivity value is closer to 1, the prediction is better. It can be seen in both the plots the model predicted better than the random one, with AUC higher than 0.5 i.e., 0.909 for first set of data and 0.834 for second set.

Figure 4 and 6, showed the separate AUC for both sample sets - 0.97 and 0.9 which is closer to 1, thus this model's prediction can be trusted.

4.2 Contribution and Importance of Environmental Variables

Of the 19 variables used in modeling (Table 1), the significant environmental conditions affecting the spatial distribution of *Clamator jacobinus* are isothermality (16.8%), mean temperature of warmest quarter (15.7%), annual precipitation, precipitation of warmest quarter (13.6%) and precipitation of wettest month (11.3%) during Indian summer monsoon season, i.e. June to October. The permutation importance involves method of computing feature importance by measuring the decrease when feature is not present. The variables having highest permutation importance are precipitation of warmest quarter and isothermality.

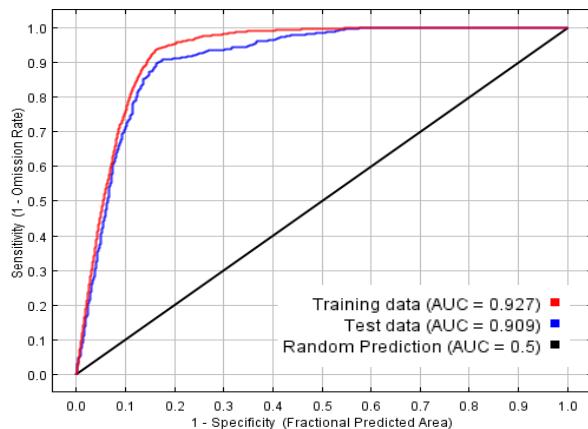


Figure 3. Training and test AUC plot to measure performance of model against first set of data

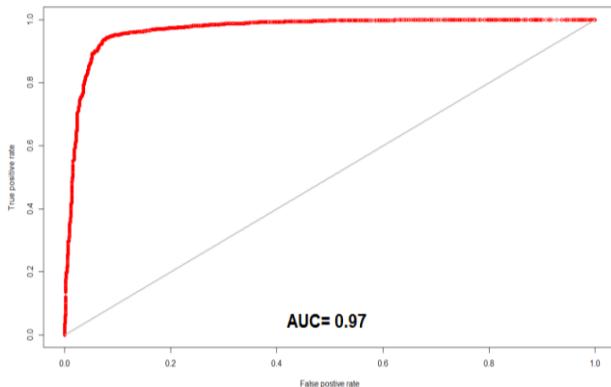


Figure 4. AUC plot against test data of first set

For another seasonal set (November to May), the environmental variables favourable for suitability of *Clamator jacobinus* are precipitation of warmest quarter (31.9% of variation), isothermality (16.1% of variation), temperature seasonality (14.9% of variation) and precipitation of wettest quarter (8.9% of variation) (Table 2). This model gives precipitation of warmest quarter as a highest permutation importance in computing the current site suitability of species which explains that the preferable season for this bird's suitability is monsoon season.

4.3 Current and Future Suitability of *Clamator jacobinus*

Based on the presence records of June to October month, the model computed the suitable habitat of *Clamator jacobinus* (Figure 7) depicting potential suitability of this bird is in India due to monsoon season and remaining parts of its resident is found unsuitable or low suitable for its distribution.

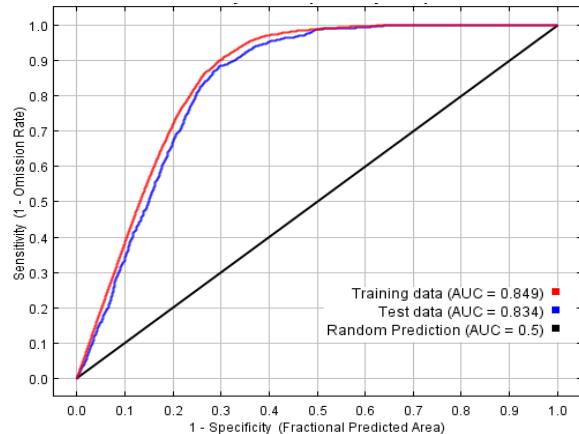


Figure 5. Training and test AUC plot to measure performance of model against second set of data

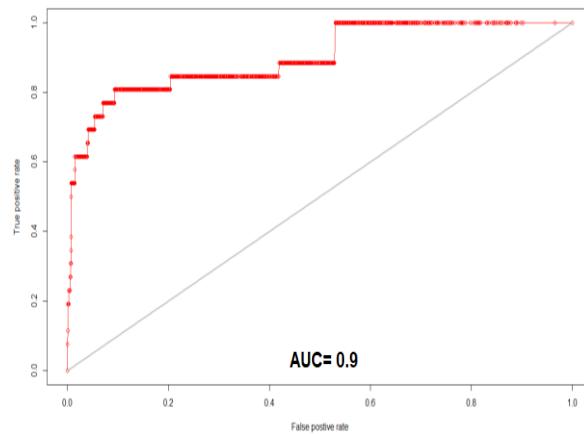


Figure 6. AUC plot against test data of second set

The colour scale shows the probability of output models ranging from 0 to 1 denoting the suitability ranges - high (light and dark green colour), medium (yellow and dark brown), low suitability (light brown colour), and unsuitable (grey colour). In Figure 7, the southern part of India such as Andhra Pradesh, Goa, Karnataka, Kerala, Maharashtra, and Tamil Nadu shows high and medium site suitability, whereas western part – Gujarat is having medium suitability, and northern parts of India is found to be suitable at high and medium range. And the Southern Africa is found unsuitable for this bird because its June to October months are having dry and hot climate which is not favourable for this bird's habitat. The area covered by grey colour depicted that the species suitability is not available in those parts, which is true when compared with its extant layer of IUCN, as shown in Figure 2. Another suitability for the months November to May is shown in Figure 8 with a scale bar depicting the good suitability range of *Clamator jacobinus* in South Africa and southern parts of India. Figure 8 illustrated that the Southern Africa is having the medium and low range suitability whereas the South Africa is found with good suitable extant of this bird. As Indian monsoon season typically lasts from June to September in North and Central India, therefore,

model provides a low or no suitability in these parts, whereas, in South India, especially Tamil Nadu gets most of its annual precipitation in October and November, thus, Figure 8 is satisfying this favourable season which proved that model is predicting the good suitability for this species.

Code	Description	Percent contribution	Permutation importance
bio3	Isothermality	16.8	19
bio10	Mean Temperature of Warmest Quarter	15.7	2.5
bio12	Annual Precipitation	13.6	4.2
bio18	Precipitation of Warmest Quarter	13.2	37.1
bio13	Precipitation of Wettest Month	11.3	4.8
bio2	Mean Diurnal Range	9.3	3.6
bio17	Precipitation of Driest Quarter	4.7	1.6
bio15	Precipitation Seasonality	3.6	1.9
bio19	Precipitation of Coldest Quarter	3.1	4.9
bio16	Precipitation of Wettest Quarter	1.9	1.6
bio4	Temperature Seasonality	1.4	8.9
bio9	Mean Temperature of Driest Quarter	1.3	1.4
bio6	Min Temperature of Coldest Month	1.2	2.3
bio11	Mean Temperature of Coldest Quarter	0.9	0.7
bio5	Max Temperature of Warmest Month	0.9	1.1
bio8	Mean Temperature of Wettest Quarter	0.7	2.4
bio7	Temperature Annual Range	0.3	1.9
bio1	Annual Mean Temperature	0.1	0.2
bio14	Precipitation of Driest Month	0	0

Table 1. Environmental variables used for modeling of June to October data and its percent contribution & permutation importance

Code	Description	Percent contribution	Permutation importance
bio18	Precipitation of Warmest Quarter	31.9	34.7
bio3	Isothermality	16.1	9.1
bio4	Temperature Seasonality	14.9	7.5
bio16	Precipitation of Wettest Quarter	8.9	0.8
bio14	Precipitation of Driest Month	6.4	2
bio12	Annual Precipitation	6.1	17
bio17	Precipitation of Driest Quarter	3.8	0.9

bio8	Mean Temperature of Wettest Quarter	2.7	11.5
bio13	Precipitation of Wettest Month	2.3	4.6
bio5	Max Temperature of Warmest Month	2.1	2
bio2	Mean Diurnal Range	1.7	3
bio11	Mean Temperature of Coldest Quarter	1.7	0.9
bio10	Mean Temperature of Warmest Quarter	0.5	2.1
bio19	Precipitation of Coldest Quarter	0.3	0.9
bio1	Annual Mean Temperature	0.2	0.1
bio6	Min Temperature of Coldest Month	0.1	0.6
bio7	Temperature Annual Range	0.1	0.7
bio15	Precipitation Seasonality	0.1	0.3
bio9	Mean Temperature of Driest Quarter	0	1.4

Table 2. Environmental variables used for modeling of November to May data

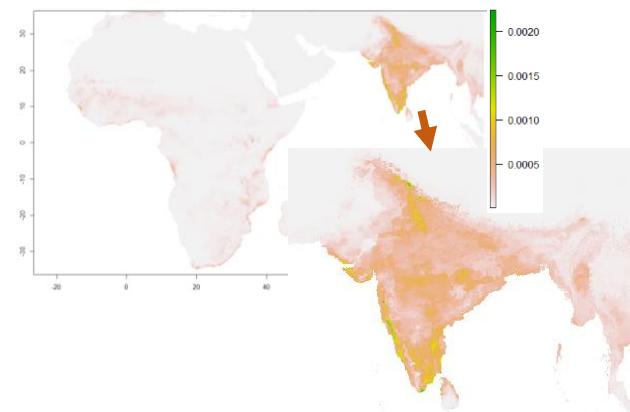


Figure 7. Current suitability of *Clamator jacobinus* for Indian monsoon season (June to October)

This study also predicted future suitability map of 2050 (Figure 9) from current suitability conditions by projecting changes in movement of bird and this revealed that the range contraction would happen in all parts of India except southern parts of Tamil Nadu which would experience little bit of improvement. Also, the quantiles (at 5% and 95%) of relevant environmental variables (occupied by bird) are calculated to get a comparative range in climate change scenarios of current and 2050 suitability for *Clamator jacobinus* with respect to Indian monsoon season (Table 3).

5. DISCUSSION

This is one of the first studies to explore the habitat suitability of *Clamator jacobinus* in Southern Africa and India by efficaciously demonstrating the efficacy of a wide-scale citizen science observations data in developing distribution models for habitat availability across an entire extant of species. Based on the presence records and environmental variables, species distribution modeling is done for the current distribution of *Clamator jacobinus* in Indian and Southern Africa. The results of model are validated against the extant layers of The IUCN Red List of Threatened Species.

According to suitability results of Indian monsoon season, seems surprisingly interesting to verify the old age belief that this bird is a sign of monsoon in India. Also, the study presented in this paper proved that the bird is not favourable in other climatological seasons such as hot, dry and cold, therefore when the season changes in Africa other than monsoon, the bird makes its arrival to North and Central India through South India from Southern Africa by the journey of Arabian Sea, along with the monsoon winds and thereby, monsoon starts in India from June and lasts till September.

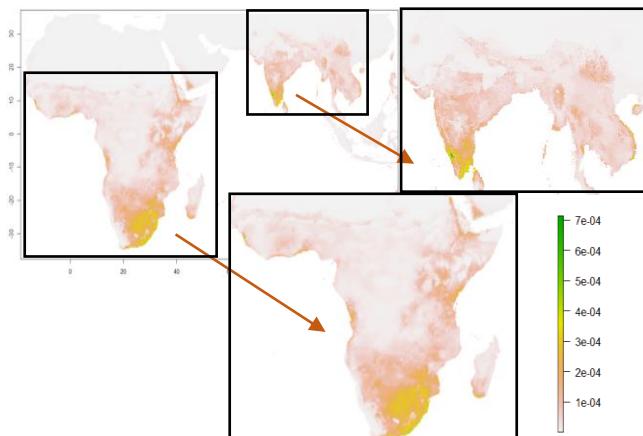


Figure 8. Current suitability of *Clamator jacobinus* for November to May month

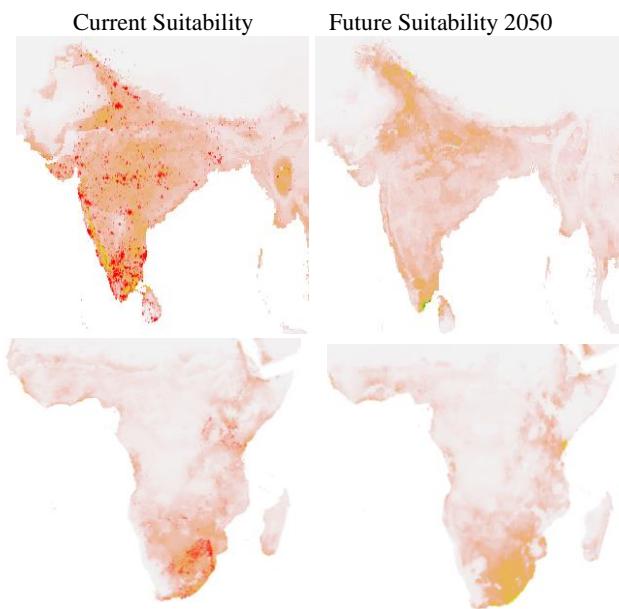


Figure 9. Comparison between current (with observations in red points) and 2050 suitability of *Clamator jacobinus* in India and Southern Africa

To evaluate the effect of climate change, this study showed the future suitability condition of this bird in 2050. From the suitability prediction of 2050, it is fairly clear that there would be no arrival of *Clamator jacobinus* in the districts of Maharashtra, Goa, Karnataka, Kerala, Tamil Nadu and Andhra Pradesh, Gujarat and northern parts of India, whereas the southern part of Tamil Nadu would be a suitable habitat for *Clamator jacobinus* in all climatic scenarios. This kind of study may be considered for the planning of agricultural activities and biodiversity planning because monsoon is considered as an

auspicious season for having a high impact on agricultural economy and biodiversity functioning (Urifi, 2011).

Variable Name	Current Suitability Quantile (5%-95%)	Future Suitability (2050) Quantile (5%-95%)
Isothermality	0.39- 0.64	0.36-0.62
Mean Temperature of Warmest Quarter [°C]	25.1- 33.2	28.2-36.6
Annual Precipitation [mm/year]	494- 2904	585.8-3225.8
Precipitation of Warmest Quarter [mm/quarter]	79-744	69-683
Precipitation of Wettest Month [mm/month]	128-968.3	151-985.6
Mean Diurnal Range [°C]	7.7-14.2	7.4-13.7
Precipitation of Driest Quarter [mm/quarter]	1-97	1-78
Precipitation Seasonality [coefficient of variation]	0.7-1.5	0.74-1.57
Precipitation of Wettest Quarter [mm/quarter]	282.1-2218	347-2482
Precipitation of Coldest Quarter [mm/quarter]	2-347.15	2-367.8

Table 3. Comparative climate change scenarios of current and 2050s

The maximum entropy approach gives a natural probabilistic interpretation on utmost to tiniest suitable habitat conditions and its results can be easily inferred by the domain experts. In Maxent, a set of distribution estimates the target distribution by finding the suitability using maximum entropy, which is closest to uniform, in such constraints that the expected value of each feature under this estimated distribution set should matches with its empirical average. The inexpensive citizen science data remains a vital source for biodiversity monitoring and keeps encouraging public participation in the process of scientific findings as well as for country well-being (Bonney et al. 2009; Ward et al. 2015).

6. CONCLUSIONS

Correlating bioclimatic variables with species distribution for suitability analysis is a needed stride towards implementing conservation plans. Citizen science catalogues can provide a vast set of distribution records in achieving better results for modeling of species. As a matter of fact, the old-age belief and myth is true in context of Jacobin Cuckoos that this bird arrives in most parts of central and northern India with monsoon winds but due to climatological changes in 2050 its suitability may decrease as compared to current one. Citizen science and maximum entropy are valuable suitability prediction techniques for ecological requirements, allowing several climatic variables to be assimilated with maximum entropy modeling. The study presented in this paper could help decision makers in defining conservation plans and places *Clamator jacobinus* inhabit in future, as a crucial perspective for any strategy on biodiversity preservation.

REFERENCES

- Ahmed, S.E., McInerny, G., O'Hara, K., Harper, R., Salido, L., Emmott, S., Joppa, L.N., 2015. Scientists and software—

- surveying the species distribution modelling community. *Divers. Distrib.*, 21(3), 258-267. doi.org/10.1111/ddi.12305
- Ammann, C.M., Meehl, G.A., Washington, W.M., Zender, C.S., 2003. A monthly and latitudinally varying volcanic forcing dataset in simulations of 20th century climate. *Geophys. Res. Lett.*, 30(12). doi.org/10.1029/2003GL016875
- Austin, M.P., 1992. Modelling the environmental niche of plants: implications for plant community response to elevated CO₂ levels. *Aust. J Bot.*, 40(5), 615-630.
- Boakes, E.H., McGowan, P.J., Fuller, R.A., Chang-qing, D., Clark, N.E., O'Connor, K. and Mace, G.M., 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS biology*, 8(6). 10.1371/journal.pbio.1000385
- Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V., Shirk, J., 2009. Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience*, 59(11), 977-984. doi.org/10.1525/bio.2009.59.11.9
- Chapman, A.D., Grafton, O., 2008. Guide to best practices for generalising sensitive/primary species occurrence-data. Version 1.0. Global Biodiversity Information Facility, Copenhagen, Denmark. dx.doi.org/10.25607/OPB-168
- Conrad, C.C., Hilchev, K.G., 2011. A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environ. Monitoring Assess.*, 176(1-4), 273-291.
- Cooper, C.B., Dickinson, J., Phillips, T., Bonney, R., 2007. Citizen science as a tool for conservation in residential ecosystems. *Ecol. Soc.*, 12(2).
- Coxen, C.L., Frey, J.K., Carleton, S.A., Collins, D.P., 2017. Species distribution models for a migratory bird based on citizen science and satellite tracking data. *Global ecology and conservation*, 11, 298-311. doi.org/10.1016/j.gecco.2017.08.001.
- Damoulas, T., Henry, S., Farnsworth, A., Lanzone, M., Gomes, C., 2010. December. Bayesian classification of flight calls with a novel dynamic time warping kernel. In *2010 Ninth International Conference on Machine Learning and Applications*, 424-429. IEEE. 10.1109/ICMLA.2010.69
- Elith, J., 2000. Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. In *Quantitative methods for conservation biology*. 39-58. Springer, New York, NY.
- Elith*, J., H. Graham*, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129-151. doi.org/10.1111/j.2006.0906-7590.04596.x
- Freeman, S.N., Noble, D.G., Newson, S.E., Baillie, S.R., 2007. Modelling population changes using data from different surveys: the Common Birds Census and the Breeding Bird Survey. *Bird Study*, 54(1), 61-72.
- doi.org/10.1080/00063650709461457
- Feeley, K.J., Silman, M.R., 2010. Modelling the responses of Andean and Amazonian plant species to climate change: the effects of georeferencing errors and the importance of data filtering. *J Biogeogr.*, 37(4), 733-740. doi.org/10.1111/j.1365-2699.2009.02240.x
- Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.*, 37(12), 4302-4315. doi.org/10.1002/joc.5086.
- Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.*, 24(1), 38-49.
- Graham, C.H., Ferrier, S., Huettmann, F., Moritz, C., Peterson, A.T., 2004. New developments in museum-based informatics and applications in biodiversity analysis. *Trends Ecol. Evol.*, 19(9), 497-503. doi.org/10.1016/j.tree.2004.07.006
- Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Townsend Peterson, A., Loiselle, B.A., NCEAS Predicting Species Distributions Working Group, 2008. The influence of spatial errors in species occurrence data used in distribution models. *J Appl. Ecol.*, 45(1), 239-247. doi.org/10.1111/j.1365-2664.2007.01408.x
- Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecol. Model.*, 135(2-3), 147-186. doi.org/10.1016/S0304-3800(00)00354-9
- Gulshan, A., 2019. Rain catcher: on Jacobin Cuckoo. <https://www.thehindu.com/sci-tech/energy-and-environment/rain-catcher-on-jacobin-cuckoo/article28359143.ece> (30 January 2020).
- Guo, Q., Liu, Y., Wieczorek, J., 2008. Georeferencing locality descriptions and computing associated uncertainty using a probabilistic approach. *Int. J Geogr. Inf. Sci.*, 22(10), 1067-1090. doi.org/10.1080/13658810701851420
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 25(15), 1965-1978. doi.org/10.1002/joc.1276
- Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modelling methods. *Ecography*, 29(5), 773-785. doi.org/10.1111/j.0906-7590.2006.04700.x
- Hey, T. ed., 2009: *The fourth paradigm: data-intensive scientific discovery* (Vol. 1). Redmond, WA: Microsoft research.
- Howe, J., 2006. The rise of crowdsourcing. *Wired magazine*, 14(6), 1-4.
- Hu, J., Jiang, Z., 2011. Climate change hastens the conservation urgency of an endangered ungulate. *PloS one*, 6(8). 10.1371/journal.pone.0022873

- Kelling, S., Hochachka, W.M., Fink, D., Riedewald, M., Caruana, R., Ballard, G. and Hooker, G., 2009. Data-intensive science: a new paradigm for biodiversity studies. *BioScience*, 59(7), 613-620.
- Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, 14(2), 1137-1145.
- McClintock, B.T., Bailey, L.L., Pollock, K.H., Simons, T.R., 2010. Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections. *Ecology*, 91(8), 2446-2454. doi.org/10.1890/09-1287.1
- Novacek, M.J., 2008. Engaging the public in biodiversity issues. *Proceedings of the National Academy of Sciences*, 105 (Supplement 1), 11571-11578. doi.org/10.1073/pnas.0802599105
- Pacifici, K., Reich, B.J., Miller, D.A., Gardner, B., Stauffer, G., Singh, S., McKerrow, A., Collazo, J.A., 2017. Integrating multiple data sources in species distribution modeling: a framework for data fusion. *Ecology*, 98(3), 840-850. doi.org/10.1002/ecy.1710
- Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful?. *Global Ecol. Biogeogr.*, 12(5), 361-371. doi.org/10.1046/j.1466-822X.2003.00042.x
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecological modelling*, 190(3-4), 231-259. doi.org/10.1016/j.ecolmodel.2005.03.026
- Phillips, S.J., Anderson, R.P., Dudík, M., Schapire, R.E., Blair, M.E., 2017. Opening the black box: An open-source release of Maxent. *Ecography*, 40(7), 887-893. doi.org/10.1111/ecog.03049
- Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31(2), 161-175. doi.org/10.1111/j.0906-7590.2008.5203.x
- Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., Rafaj, P., 2011. RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Climatic change*, 109(1-2), p.33.
- Richardson, D.M., Whittaker, R.J., 2010. Conservation biogeography—foundations, concepts and challenges. *Divers. Distrib.*, 16(3), 313-320. doi.org/10.1111/j.1472-4642.2010.00660.x
- Robinson, O.J., Ruiz-Gutierrez, V., Fink, D., Meese, R.J., Holyoak, M., Cooch, E.G., 2018. Using citizen science data in integrated population models to inform conservation. *Biol. Conserv.*, 227, 361-368. doi.org/10.1016/j.biocon.2018.10.002
- Sato, M., Hansen, J.E., McCormick, M.P. and Pollack, J.B., 1993. Stratospheric aerosol optical depths, 1850–1990. *J. Geophys. Res.*, 98(D12), 22987-22994. doi:10.1029/93JD02553
- Sharma, J., Upgupta, S., Jayaraman, M., Chaturvedi, R.K., Bala, G., Ravindranath, N.H., 2017. Vulnerability of forests in India: a national scale assessment. *Environ. Manage.*, 60(3), 544-553.
- Silvertown, J., Cook, L., Cameron, R., Dodd, M., McConway, K., Worthington, J., Skelton, P., Anton, C., Bossdorf, O., Baur, B., Schilthuizen, M., 2011. Citizen science reveals unexpected continental-scale evolutionary change in a model organism. *PloS one*, 6(4). 10.1371/journal.pone.0018927
- Skidmore, A.K., Gauld, A., Walker, P., 1996. Classification of kangaroo habitat distribution using three GIS models. *Int. J. Geogr. Inf. Syst.*, 10(4), 441-454.
- Smeraldo, S., Di Febbraro, M., Bosso, L., Flaquer, C., Guixé, D., Lisón, F., Meschede, A., Juste, J., Prüger, J., Puig-Montserrat, X., Russo, D., 2018. Ignoring seasonal changes in the ecological niche of non-migratory species may lead to biases in potential distribution models: lessons from bats. *Biodivers. Conserv.*, 27(9), 2425-2441. doi.org/10.1007/s10531-018-1545-7.
- Stenchikov, G., Delworth, T.L., Ramaswamy, V., Stouffer, R.J., Wittenberg, A., Zeng, F., 2009. Volcanic signals in oceans. *J. Geophys. Res.*, 114(D16). doi:10.1029/2008JD011673.
- Stockwell, D.R., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. *Ecol. Model.*, 148(1), 1-13. doi.org/10.1016/S0304-3800(01)00388-X
- Stoeckle, M., 2003. Taxonomy, DNA, and the bar code of life. *BioScience*, 53(9), 796-797. doi.org/10.1641/0006-3568(2003)053[0796:TDATBC]2.0.CO;2
- Sullivan, B.L., Phillips, T., Dayer, A.A., Wood, C.L., Farnsworth, A., Iliff, M.J., Davies, I.J., Wiggins, A., Fink, D., Hochachka, W.M., Rodewald, A.D., 2017. Using open access observational data for conservation action: A case study for birds. *Biol. Conserv.*, 208, 5-14.
- Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M., 2003. Remote sensing for biodiversity science and conservation. *Trends. Ecol. Evol.*, 18(6), 306-314. doi.org/10.1016/S0169-5347(03)00070-3
- Ward, E.J., Jannot, J.E., Lee, Y.W., Ono, K., Shelton, A.O. and Thorson, J.T., 2015. Using spatiotemporal species distribution models to identify temporally evolving hotspots of species co-occurrence. *Ecological Applications*, 25(8), pp.2198-2209.
- Wieczorek, J., Guo, Q., Hijmans, R., 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *Int. J. Geogr. Inf. Sci.*, 18(8), 745-767. doi.org/10.1080/13658810412331280211
- Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., NCEAS Predicting Species Distributions Working Group, 2008. Effects of sample size on the performance of species distribution models. *Divers. Distrib.*, 14(5), 763-773. doi.org/10.1111/j.1472-4642.2008.00482.x
- Woodward, F.I., Woodward, F.I., 1987: *Climate and plant distribution*. Cambridge University Press.