Spatial modeling of the distribution of *R. microplus* and *Haematobia* sp. flies in Maharashtra state though Inverse Distance Weighting

.

ABSTRACT

Aims: Spatial model of distribution of two important veterinary pest *viz. Rhipicephalus microplus* ticks and *Haematobia*sp flies in Maharashtra state was prepared.
Study design: An Inverse Distance Weighting approach was employed for the data obtained through primary survey of insects under study.
Place and Duration of Study: The study was carried out during January 2023 to June 2023 covering entire state of Maharashtra in India.
Methodology: Total 143 location points for *R.microplus* ticks and 31 locations for *Haematobia* sp. flies in the Maharashtra state were finalized and included in te modeling exercise. These location points were converted to spatial dataframe and imported in R software and processed for Inverse distance weighting by using the precipitation and temperature bioclimatic raster files for the state.
Results: The *R. microplus* distribution studied using inverse distance weighted interpolations, the model's negative correlation of -0.061 was found to have an acceptable AUC of 0.746, indicating a decent fit of the IDW model. On the other hand, the IDW model was shown to be non-significant with a minimal AUC value of 0.579 in the case of *Haematobia* flies. Nonetheless, the zonal statistics indicated that around 30% of the state's total area was unsuitable for *R. microplus* and that 57% was unsuitable for *Haematobia* sp. flies.

Keywords: Rhipicephalus, inverse distance weighting, GIS, Maharashtra, Haematobia

1. INTRODUCTION

Ticks and tick-borne diseases are one of the important medical, veterinary and economic problems globally. Ticks have a strong vector capacity and a blood-sucking behaviour. Ticks are the most significant vectors of disease-causing pathogens in both domestic and wild animals, ranking second only to mosquitoes as worldwide disease vectors due to their capacity to inflict direct and indirect harm on their hosts (de la Fuente et al., 2008). Production loss is directly linked to tick infestations and is expected to cost several million dollars globally (Grisi et al., 2014). The management cost of ticks and tick-borne diseases (TTBDs) in India has been estimated to be approximately US $498.7 million annually, according to Minjauw and McLeod (2003). Considering the importance of ticks in livestock husbandry, its prevalence estimates through spatial and temporal distribution is a major tool for implementing control strategies. Since lot of works have been carried to study the prevalence of ticks on livestock as well as farm premises, the inclusion of geographic information systems for the same is scare in India. Maps depicting the distribution of ticks are frequently used to represent the spread of human diseases or as a proxy for transmission exposure risk; nevertheless, the majority of tick-borne infections have poorly defined vector ranges.

Regarding the spatial distribution pattern of livestock insect pest several evidences suggest utility of GIS and RS techniques in predicting and monitoring of vector borne diseases of livestock (Rogers et al., 2006; Kalluri et al., 2007; Hartemink et al., 2015). As the distribution of livestock pests viz. insect and ticks largely governed by bio-physical characteristics of the area, its spread can be better estimated by these advanced tools (Medlock *et al.,* 2013). Maps are a key tool for communicating potential exposure risk and visualizing spatial information about diseases. In public health, disease maps—which can range in complexity from plotted cases, or dot maps, to projected risk forecasts modelled with machine learning algorithms—have long been used to depict the distribution of vector-borne diseases.

Mapping products, no matter how complicated, depend on geo-referenced datasets being available. We considered that the most of livestock tick and fly species have not been thoroughly mapped in India as on today. Therefore the present study was planned to use one of the geographic information system (GIS) tools *viz.* Inverse Distance Weighting approach for mapping of *R. microplus* ticks and *Haematobia* sp. flies based on the occurrence records collected through an entomological survey.

2. MATERIAL AND METHODS

This study was based on the presence and occurrence points of the *Rhipicephalus microplus* ticks of cattle and *Haematobia*spp flies in the Maharashtra State. Ticks identified as *Rhipicephalus microplus*from 143 locations were geo-referenced in a *".csv"* file based on the latitude-longitude records collected at the time of samples collections itself. For the location data of *Haematobia* sp. flies, samples identified from 18 number of locations and it was also added with the data from earlier work done by Gudewar (2020) which remained to 31 points after removal of duplicates. Therefore, having the above tick presence data, geo-referenced locations, or occurrence points that represent *Rhipicephalus microplus* ticks of cattle and *Haematobia*spp flies were recorded in terms of coordinate pair as decimal latitude/longitude in the WGS84 system where the cattle was sampled for ticks using a GPS app in a smartphone (*NoteCam*). However, the coordinate pair data for the locations for occurrence points from a study by Gudewar (2020) were obtained from GoogleEarth. These geo-coordinate data was transformed into “Latitude” and “Longitude” in WGS 84 Global Projection System in QGIS software. This GIS analysis was limited to the occurrences records for Maharashtra state for both tick and fly distribution. The location file was then saved in “.csv” format to be later used for GIS modeling.

The GIS modeling was performed by using “*dismo*” package (version 1.3-8) in R environment (Hijmans*et al.,* 2022). The approach used was Inverse Distance Weighting *i.e.* IDW; a spatial interpolation method. IDW interpolation assumes that things which are close to one another are more alike than those that are farther apart. To predict a value for any unmeasured location, IDW uses the measured values surrounding the prediction location. The distribution was predicted based on the 3 basic climatic attributes *viz.* annual mean temperature, annual precipitation and elevation. These climatic covariates were obtained from WorldClim database (https://www.worldclim.org/) in the form of raster files which were then cropped to Maharashtra state boundary in QGIS before using in the analysis.

In RStudio, the analysis is first started with loading the required packages *viz*. “dismo”, “raster” and “sp” which helps perform modeling algorithm, read, process and analyze raster files, spatial analysis and producing graphical outputs of data respectively (Hijmans et al., 2015, Hijmans et al., 2017, Pebesma et al., 2012). After loading these library, the data on occurrence records were inputted by “*read.csv*” command. In next step, the spatial covariates i.e. raster files of annual mean temperature, annual precipitation and elevation were called and entered into the RStudio environment. These raster files were then stacked as a RasterStack object. Both these dataset objects were assigned a common projection system *i.e.*WGS84. Subsequently pseudo absence records or background points were obtained by “*RandomPoints*” commands referring to any of raster files and occurrence location file. Subsequently, having all these objects in hand, the model was run with help of “*geoIDW*” command, an algorithm in “*dismo’* package. This run provided a model object whose results were obtained through simple “*summary*” command. The summary was then interpreted and presented in table format while the model object was then subjected to form the prediction through a “*predict*” command and this predicted model was then imported in QGIS for visualization for final model. In QGIS, the predicted model files were loaded and categorized in to five classes assigning the suitability for the presence and spread of the vector species under study.

3. RESULTS AND DISCUSSION

The Inverse distance weighted interpolations for the studying the distribution of *R.microplus*(Figure 1) revealed that the model negative correlation of -0.061 showed an acceptable AUC of 0.746 which indicated moderate fitting of IDW model (Table 1). However the results for *Haematobia* flies revealed that the IDW model was non-significant with a minimal value of AUC *i.e.* 0.579. The zonal statistics however pointed out towards nearly 30% as non-suitable for *R.microplus* and 57% of total area of the state as non-suitable for *Haematobia*sp flies (Table 2 / Figure 2).

Table 1. The IDW model output for *R. microplus* and *Haematobia*sp flies in Maharashtra

R. microplus		***Haematobia*sp. flies**	
class	: ModelEvaluation	class	: ModelEvaluation
n presences	: 143	n presences	: 31
n absences	: 500	n absences	: 50
AUC	: **0.74679**	AUC	: **0.5793548**
cor	: -0.061757	cor	: -0.0516856
max TPR+TNR at	: 0.378463	max TPR+TNR at	: 0.2855463

Table 2. Zonal statistics for *R.microplus* ticks and *Haematobia*sp flies in Maharashtra by IDW method

Suitability	***Rhipicephalus microplus***		***Haematobia*sp flies**	
	Area in sq km	Percent	Area in sq km	Percent
Absence	80476.649	22.97	76493.124	21.84
Non Suitable	27033.06	7.72	124624.63	35.58
Low Suitable	151133.54	43.14	45468.13	12.98
Mod Suitable	55473.05	15.84	53670.84	15.32
High Suitable	36185.60	10.33	39476.30	11.27

In order to estimate the geospatial distribution of both the vectors under this study, an Inverse Distance Weighting approach was employed as a simple GIS tool. Inverse distance weighted (IDW) interpolation is widely used spatial interpolation method in Geographic information systems analysis. It is a deterministic approach that estimates the value of a variable at an un-sampled location based on the values of the surrounding sampled locations. The basic principle behind IDW is that the influence of a known data point on the interpolated value decreases as the distance from the known data point increases (Malieka *et al.,* 2020). IDW assumes that the variable being interpolated is more similar between closer points than between distant points. It gives higher weights to closer points and lower weights to more distant points when estimating the value at an un-sampled location (Lu and Wong, 2008).

The estimated value at an un-sampled location is calculated as a weighted average of the known data points within a specified neighboring area. The weight assigned to each known data point is inversely proportional to the distance between the known data point and the un-sampled location.

The formula for IDW interpolation is:

$$\mathbf{Z(x,y) = \Sigma(Wi * Zi) / \Sigma(Wi)}$$

Where:

Z(x,y) is the estimated value at the un-sampled location (x, y)

Zi is the known value at the i^{th} data point

Wi is the weight assigned to the i^{th} data point, calculated as:

$$Wi = 1 / d^p$$

Where:

d is the distance between the unsampled location and the i^{th} data point

p is the power parameter, typically set between 1 and 3, controlling the rate of decline in the weight as the distance increases.

The power parameter (p) in the IDW formula determines the rate of decline in the weight as the distance increases. A higher power parameter (e.g., p=3) gives more weight to the closest points and reduces the influence of distant points. A lower power parameter (e.g., p=1) results in a more gradual decline in the weight as the distance increases, leading to a smoother interpolated surface (Liu *et al.,* 2021). IDW can be influenced by spatial autocorrelation, and it may be necessary to account for this effect, for example, by adjusting the power parameter or the search radius used in the interpolation.

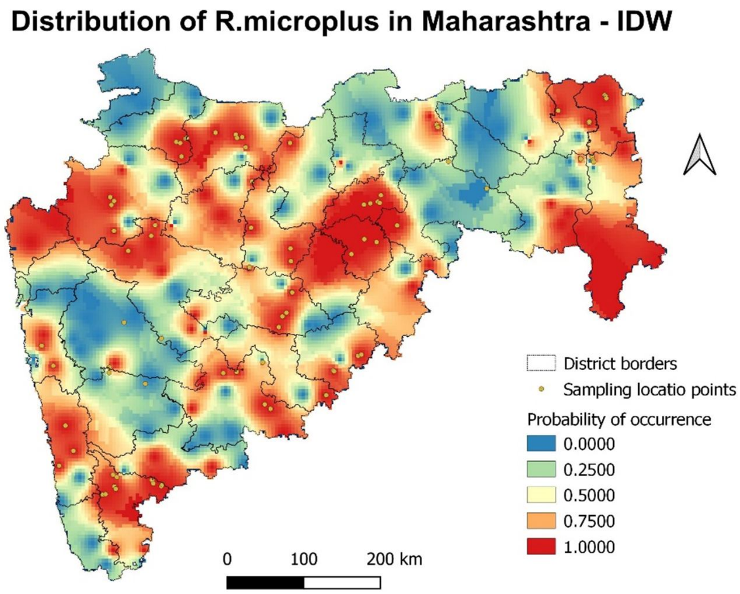

Fig. 1.Distribution of *R.microplus* ticks inn Maharashtra State, India

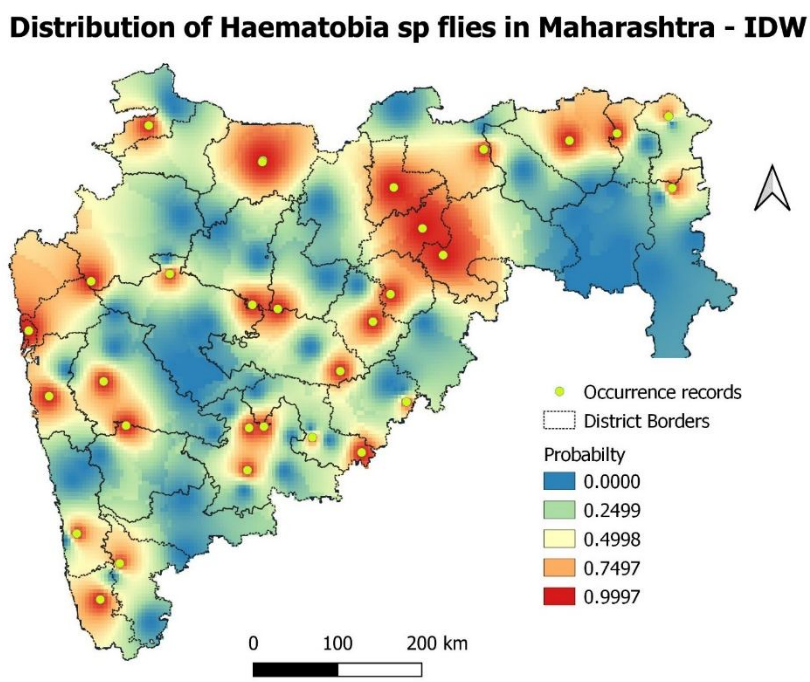

Fig. 2.Distribution of Haematobia sp. flies in Maharashtra State, India

As more often the IDW approach is used when the spatial data points have a quantifiable attribute as against binary ones (Auchincloss *et al.,* 2012), its use in presence-absence models is seldom reported. However in some of the studies focusing on appraising spatial dependencies between the disease events, it has given reliable outcomes. One such study was conducted by Sharma *et al*. (2019) wherein they studied the spatial sero-epidemiology of bovine trypanosomiasis in low lying areas of Punjab, India. They reported that the prevalence of bovine trypanosomiasis as with a Pearson's correlation with environmental factors influencing *T. evansi* prevalence *viz*. mean of average temperature, precipitation, potential evapotranspiration and cloud cover was done and results were then interpolated with IDW. Their results indicated the spatial aggregation of infective zones in Punjab. Another study from Punjab itself by Sumbria*et al.* (2016) found that the prevalence of *Theileriaepui* was more inclined towards south-west of the state as compared to other zones and the prevalence trend diagnosed by nested PCR exhibited strong correlation with temperature while showing an inverse relationship with precipitation and cloud cover. The similar observations in terms of the importance of IDW approach in determining the zones of *R.microplus* presence was noted in current study. Also along with the results of *Haematobia* sp. flies interpolation maps showed "bulls eye" pattern which may be due the limited number of presence points, its utility in studying species distribution with given environmental variables was found non-significant with AUC values in the range of 50s indicated that the models output was just a random prediction.

4. CONCLUSION

It was concluded that for determining the geographical spread of *R.microplus* in Maharashtra the Inverse Distance weighting is a suitable approach and for *Haematobia* sp. flies, it has trivial outcomes. It is therefore recommended to use the resultant projected maps of *R.microplus* obtained through this research for its further study in the state.

CONSENT (WHEREEVER APPLICABLE)

Not Applicable

ETHICAL APPROVAL (WHEREEVER APPLICABLE)

Not Applicable

REFERENCES

Auchincloss, A.H., S.Y. Gebreab, C. Mair, A.V. Diez Roux, 2012. A review of spatial methods in epidemiology, 2000–2010. Annual Review of Public Health, 33,107-122

De la Fuente, J., Estrada-Pena, A., Venzal, J.M., Kokan, K.M. and Soneshine, D.E. 2008. Overview: Ticks as vector of pathogens that cause disease in humans and animals. *Front. Biosci*. **13**: 6938-46.

Grisi, L., Leite, R.C., Martins, J.R., Barros, A.T., Andreotti, R., Cançado, P.H., León, A.A., Pereira, J.B. and Villela, H.S. 2014. Reassessment of the potential economic impact of cattle parasites in Brazil. Braz. J. Vet. Parasitol. **23**(2): 150-156

Gudewar, J.G. 2020. Bio-Ecological Aspects and Integrated Management of Lyperosia-Haematobia Complex (Diptera: Muscidae) in Maharashtra State. PhD thesis submitted to Maharashta Animal and Fishery Sciences University Nagpur, India

Hartemink, N., Vanwambeke, S.O., Purse, B.V., Gilbert, M. and Van Dyck, H., 2015. Towards a resource-based habitat approach for spatial modelling of vector-borne disease risks. Biological Reviews, 90(4), pp.1151-1162.

Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J. and Hijmans, M.R.J., 2017. Package ‘dismo’. Circles, 9(1), pp.1-68.

Hijmans, R.J., S. Phillips, J. Leathwick, J.Elith, 2022. Dismo: Species Distribution Modeling. R Package Version 1.3-8.http://CRAN.R-project.org/package=dismo

Hijmans, R.J., Van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J.A., Lamigueiro, O.P., Bevan, A., Racine, E.B., Shortridge, A. and Hijmans, M.R.J., 2015. Package ‘raster’. R package, 734, p.473.

Kalluri, S., p. Gilruth, D. Rogers, M. Szczur, 2013. Surveillance of arthropod vector-borne infectious diseases usinf remote sensing techniques: a review. PLoS Pathogens, **3,** 1361-1371. https://doi.org/10.1371/journal.ppat.0030116

Liu, Z.N., X.Y. Yu, L.F. Jia, et al., 2021. The influence of distance weight on the inverse distance weighted method for ore-grade estimation. Scientific Reports, **11,** 2689. https://doi.org/10.1038/s41598-021-82227-y

Lu, G.Y., Wong, D.W., 2008. An adaptive inverse-distance weighting spatial interpolation technique Computational Geosciences, 34(9), 1044 – 1055. https://doi.org/10.1016/j.cageo.2007.07.010

Maleika, W. 2020. Inverse distance weighting method optimization in the process of digital terrain model creation based on data collected from a multibeam echo sounder. Applied Geomatics, 12, 397–407. https://doi.org/10.1007/s12518-020-00307-6

Medlock, J. M., K. M. Hansford, A. Bormane, M. Derdakova, A. Estrada-Peña, J. C. George, I. Golovljova, T. G. Jaenson, J. K. Jensen, P. M. Jensen and M. Kazimirova, 2013. Driving forces for changes in geographical distribution of *Ixodesricinus* ticks in Europe. Parasites & vectors, 6, 1-11

Minjauw, B. A. McLeod, 2003. Tick-borne diseases and poverty. The impact of ticks and tick-borne diseases on the livelihood of small scale and marginal livestock owners in India and eastern and southern Africa. Research report, DFID Animal Health Programme, Centre for Tropical Veterinary Medicine, University of Edinburgh, UK

Pebesma, E., Bivand, R., Pebesma, M.E., RColorBrewer, S. and Collate, A.A.A., 2012. Package 'sp'. The Comprehensive R Archive Network, 9.

Rogers, D.J., 2006. Models for vectors and vector-borne diseases. Advances in parasitology, 62, pp.1-35.

Sharma, A., Singla, L.D., Kaur, P., Bal, M.S., Sumbria, D. and Setia, R., 2019. Spatial seroepidemiology, risk assessment and haemato-biochemical implications of bovine trypanosomiasis in low lying areas of Punjab, India. Comparative immunology, microbiology and infectious diseases, 64, pp.61-66.

Sumbria, D., Singla, L.D., Kumar, S., Sharma, A., Dahiya, R.K. and Setia, R., 2016. Spatial distribution, risk factors and haemato-biochemical alterations associated with Theileriaequi infected equids of Punjab (India) diagnosed by indirect ELISA and nested PCR. Acta Tropica, 155, pp.104-112.