

Exploration of machine learning models to predict the environmental and remote sensing risk factors of haemonchosis in sheep flocks of Rajasthan, India

Kuralayanapalya Puttahnappa Suresh^a, Pinaki Prasad Sengupta^{a,*}, Siju Susan Jacob^a, Mohan Kumar Garudanagiri Sathyanarayana^a, Sharanagouda Shiddanagouda Patil^a, Chander Prakash Swarnkar^b, Dharendra Singh^b

^a ICAR- National Institute of Veterinary Epidemiology and Disease Informatics (NIVEDI), Yelahanka, Bengaluru 560064, Karnataka, India

^b Animal Health Division, ICAR-Central Sheep and Wool Research Institute (CSWRI), Avikanagar 304501, Rajasthan, India

ARTICLE INFO

Keywords:

Sheep
Haemonchosis
Climatic variables
Basic reproduction number
Risk factors

ABSTRACT

Globally haemonchosis in sheep is a known devastating disease imposing considerable economic loss. Understanding the environmental risk factors and their role is essentially required to manage the disease successfully. In this study, 14 years' disease data was analysed to predict the risk factors responsible for the occurrence of the disease. Season-wise analysis revealed high incidence during monsoon and post-monsoon and least in winter and summer seasons. The linear discriminant analysis (LDA) revealed the significant environmental and remote sensing risk factors contributing to haemonchosis incidence as enhanced vegetation index, leaf area index, potential evapotranspiration and specific humidity. Further, significant ecological and environmental risk factors identified using LDA were subjected to the climate-disease modelling and risk maps were generated. Basic reproduction number (R_0) was estimated and was ranged from 0.76 to 2.08 for >1000 egg per gram of faeces (EPG) in four districts whereas R_0 values of 1.09–1.69 for >2000 EPG in three districts indicating the severity of the infection. The random forest and adaptive boosting models emerged out as best fitted models for both the EPG groups. The results of the study will help to focus on high-risk areas of haemonchosis in sheep to implement the available control strategies and better animal production globally.

Introduction

Haemonchosis in sheep due to *Haemonchus contortus* (*H. contortus*), a nematode helminth, deserves special attention globally due to its exceptionally high biotic potential coupled with short developmental period and high pathogenicity that inflicts substantial economic losses (Amarante, 2014). The disease is highly prevalent across tropical, subtropical and temperate regions of the world, especially under humid conditions (Jabbar et al., 2008). Being a voracious blood feeder, *H. contortus* can cause anaemia, reduced productivity and considerable mortality especially in young animals (Itty et al., 1997; Githigia et al., 2001; Swarnkar et al., 2014). The total annual cost of parasitism in sheep was estimated as \$ 222.0 million in Australia (McLeod, 1995), > \$ 3.0 billion in USA (Smith, 2002), US\$ 41.8 million in Uruguay (Nari et al., 1997), \$ 2.0 billion in sub-Saharan Africa (de Haan and Bekure,

1991), US\$ 400 million in Ethiopia (Gezehegn, 1992), \$ 5.6 million in Indonesia and \$ 103.0 million in India (McLeod, 2004). With the approach of partial farm budgeting, Singh et al. (2011) estimated an annual economic loss of Rs. 1191.0 million due to the natural challenge of gastrointestinal nematode infections in sheep flocks of Rajasthan and the components of losses were reduced mutton production (59.56%), increased susceptibility for mortality (16.57%), premature culling (11.25%), reduced fertility (7.97%) and decreased wool yield (4.65%).

Bimodal seasonal pattern of infection pressure is well established in *H. contortus* (Rose et al., 2016) and also reported from Rajasthan (Swarnkar et al., 1996). The incidence of the disease can be limited by the environmental temperature with temperature below the development threshold during winter months can arrest the parasite development within the host (hypobiosis) thereby the parasite population will be persisted over winter (Waller et al., 2004). On the other hand, the

* Corresponding author.

E-mail address: pinakiprasad_s@rediffmail.com (P.P. Sengupta).

<https://doi.org/10.1016/j.actatropica.2022.106542>

Received 17 September 2021; Received in revised form 11 May 2022; Accepted 24 May 2022

Available online 25 May 2022

0001-706X/© 2022 Elsevier B.V. All rights reserved.

infection pressure will be limited during the summer months due to increased temperature and decreased humidity. The epidemiological profile from Rajasthan also exhibited a similar phenomenon during winter and summer (Swarnkar and Singh, 2015; Singh et al., 1997, 2015, 2018). Further, the effective management of haemonchosis is often challenged by the emergence of anthelmintic resistance in worms which is a growing concern across the world (Almeida et al., 2010). This situation demands for alternative approaches for the control of the disease such as rotational grazing, nutritional supplementation, selective treatment, use of medicinal plants and breeding of sheep for disease resistance (Sahoo and Khan, 2016; Gowane et al., 2019).

In the epidemiology of haemonchosis, translation of exogenous stages and their survival in the environment is climate-dependant. Though the infection can persist throughout the year, the peak usually observes during monsoon season in tropical countries including India (Singh et al., 1997; 2018). Even though haemonchosis was recognised as a disease of tropical countries (Kao et al., 2000; O'Connor et al., 2006), the niche has recently been expanded to northern temperate countries such as the UK (Kenyon et al., 2009; Burgess et al., 2012). The maintenance of the parasite population depends on a continuous cycle of infection between the host and the herbage (Gupta et al., 1987; Gatongi et al., 1988; Barger, 1999). In this scenario, understanding about the spatial risk factors will aid in strategic management of the disease.

Currently Rajasthan ranked first in wool production and fourth in sheep population (7.9 million) in India (DAHD 2019). Addressing the climate changing scenario, the climate-driven expansion of the niche of *H. contortus* poses a serious threat to the sustainability of sheep industry. So, the present study was undertaken in Rajasthan to identify the high-risk areas of *H. contortus*. In the era of increasing anthelmintic resistance (Swarnkar and Singh, 2010), sustainable animal health management strategies need to be developed with optimisation of the use of available resources, underpinned by a thorough understanding about the climate change on the dynamics of parasite as well as disease. Application of machine learning and artificial intelligence (AI) models to predict the occurrence of haemonchosis has not been attempted earlier. It is important to ensure that the selected models have to incorporate environmental stochasticity since the threat of climate change can influence the spatio-temporal dynamics of haemonchosis. The impact of climate change on the disease dynamics of *H. contortus* is of particular interest in different regions of the world and is well documented that heterogeneity in the spatial distribution of the parasite was accredited to variation in mean temperature, rainfall and humidity (Rinaldi et al., 2015; Rose et al., 2016). Identification of risk factors and risk modelling is another emerging area which can be utilized for timely intervention to reduce the disease burden. The development of the prediction models of the disease may also be helpful in this regard. Keeping in view the importance of the disease, the present study was designed to predict the temporal and spatial risk of disease incidence in Rajasthan using artificial intelligence system of algorithm by incorporating meteorological and remote sensing data with disease data to enable the sheep breeders and veterinarians in planning to reduce the disease burden and economic losses.

Materials and methods

Study area

The study was focused on 8 districts (Ajmer, Bhilwara, Bikaner, Jaipur, Jodhpur, Pali, Sikar and Tonk) of Rajasthan state, India, comprising of 20 blocks spreading across 85 villages. In Rajasthan state, there are mainly two agroclimatic zones are prevailing- arid and semi-arid zones. The Sikar, Bikaner, Pali and Jodhpur districts belong to arid zone whereas Ajmer, Bhilwara, Jaipur and Tonk districts belong to semi-arid zone. Arid zone receives less rainfall in comparison to the semi-arid zone. Temperature extremism both in summer and winter are more in arid zone in comparison to semi-arid zone. The climatological

characterization in Rajasthan has already been done by Swarnkar and Singh (2011). Data on the intensity of the infection in terms of eggs per gram (EPG) from eight districts of Rajasthan during 2002–2015 (14 years) were compiled. The EPG data were categorised into two groups based on the cut-off that has been drawn out as (a) >1000 EPG and (b) > 2000 EPG to study the contribution of risk factors for the intensity of the infection. In sheep, treatment is advisable when EPG exceeds 1000. However, more than 2000 EPG is considered as severe infection (Soulsby, 1982).

Disease data

The village level disease data of eight districts based on sheep faecal sample examination was received from the ICAR-Central Sheep and Wool Research Institute (CSWRI), Avikanagar, Rajasthan, India. Out of these, we picked up 8532 data represented > 1000 EPG and 5046 data represented >2000 EPG whereas 11,430 data represented <1000 epg and was considered as control data. The disease data was organised, geo-referenced and cross-checked for coordinates and processed to eliminate errors, if any. These observed coordinates were expressed as longitude (X) and latitude (Y). The data were converted to binary variables, as presence (1) or absence (0) of incidence for 14 years (2002–2015) in a specific location at a particular time.

The data for FECs were not categorized at animal level as in general farmers in study area possessing around 70–80% adult animal in their flock (mainly female). Most of the samples (>95%) evaluated were from adult sheep.

Risk factors data

In the present study, the population density and demographics were collected from 20th Livestock census, Department of Animal husbandry Statistics division, Ministry of Fisheries, Animal husbandry & Dairying, Govt of India. The Meteorological variables like, Air temperature(k), Potential evaporation rate (w/m^2), rainfall ($kg/m^2/s$), soil moisture (kg/m^2), specific humidity (kg/kg), surface pressure(pa) and wind speed(m/s) were extracted from Global Land Data Assimilation System(New and reprocessed GLDAS version 2: <https://ldas.gsfc.nasa.gov/gldas>) which is to ingest satellite and ground-based observational data products using advanced land surface modelling and data assimilation techniques with spatial resolution of $0.25^\circ \times 0.25^\circ$ and temporal resolution of one month extracted in network common data format (netCDF), which holds metadata and data with multidimensional arrays and dimensions of data. The package 'ncdf4' in R was used to read and extract data.

The remote sensing variables were like Normalized Difference vegetative index (NDVI, 16 day interval, 500 metre), Enhanced Vegetation Index (EVI, 16 day interval), Potential Evapotranspiration (PET in mm, 16 day interval, 500 metre), Leaf Area Index (LAI in m^2/m^2 , 16 day interval, 500 metre) and Land Surface temperature (LST in degree Celsius, 8 day interval, 1KM) were extracted with atmospheric corrections from Moderate Resolution Imaging Spectroradiometer(MODIS) satellite image products such as MOD13A1, MOD11A2, MOD16A2 and MOD15A2H. These products are available at various spatial and temporal resolutions in HDF format. The R package 'gdalutils' and 'modis' were used to extract the data from HDF files by converting into GeoTIFF files. All the variables are organized as raster (grid) type files by using R package 'raster' and each predictor should be a raster layer representing a variable of Interest.

Predictor variable layers were obtained from open sources which includes bioclimatic variables. The risk factors data were retrieved from GES DISC "GLDAS_NOAH025_M.2.1" dataset. Various risk factors were environmental variables such as potential evaporation rate (W/m^2), pressure (Pa), specific humidity (g/kg), rainfall (mm), soil moisture (kg/m^2), temperature ($^\circ C$) and wind speed (m/s) and the remote sensing variables like Normalized Difference vegetative index (NDVI) and Land Surface temperature (LST) were extracted from MODIS satellite images.

Atmospherically corrected NDVI was collected on 16-day interval at 250-metre resolution using MODIS product MOD13Q1 and LST was collected on 8-day intervals using MOD11A2 at 1 KM resolution. The village level data on sheep population were collected from the 19th livestock census of India (DAHD, 2019). All the risk factors are organized as raster (grid) type files and each predictor should be a raster representing a variable of Interest. These raster data was typically stored in geo TIFF format. The data was retrieved by specifying the coordinates and time period (dates).

Data processing and feature engineering

Data collection from different sources could be internal and /or external to satisfy the objectives of forewarning requirements, data can be any format, CSV, XML, JSON etc. In this process of data and feature engineering, we focus mainly on understanding the given data set and clean up the dataset, better understanding of features and their relationships, extracting essential variables, handling missing values and human error, identifying outliers, transformation of features if there are outliers, so that either truncate a data above threshold or transform the data using log or any other transformation, scaling the features extracted. This process would be maximising the insights into dataset.

Epidemiological analysis

The pattern of disease occurrence was analysed for a period of 14 years from 2002 to 2015. The study area was categorised into two groups based on the severity of infection in terms of EPG as (a) >1000 EPG and (b) >2000 EPG. The village level data were analysed using R software (Allouche et al., 2006) to obtain the cumulative data on disease incidence for each group every year (2002–2015).

R statistical software

All statistical analyses, risk maps and forecast model development were performed using R statistical software version 3.6.3. R packages viz. rgdal, raster, plyr, dplyr, openxlsx, data.table, openxlsx, spdep, sf, tmap, sp, BMM tools, foreign, geosphere, MASS, biomod2, dismo, mgcv, gbm, randomforest, mda, earth, nnet, rpart, caret, kernlab, inputmissings, e1071, psych, SDMtools, pROC, ada, qdap and R₀ were used for data processing, combining, annotating model fitting, computation of R₀, Getis ord's Index, and risk mapping.

Local spatial autocorrelation

Spatial autocorrelation analysis was conducted using the EPG data of haemonchosis in sheep to detect the pattern of distribution of disease and to identify the presence of significant disease clusters. In the present study, Getis Ord's G_i^{*} statistics (GI value) was used to detect the evidence of any spatial clusters of haemonchosis (Getis and Ord, 1992; Ord and Getis, 1995). R statistical software was used to calculate the Getis Ord's Index (Brazil, 2019). A clustered pattern is displayed when the GI value is > 0, whereas a dispersed pattern is exhibited if the GI value is < 0.

Space-time cluster analysis

In order to detect the space-time clusters of haemonchosis in Rajasthan during 14 years, the Poisson based clustering models using the space-time scan statistics were implemented in SaTScan software v9.6. Clusters are reported for those circles where observed values are greater than expected values. For the SaTScan analyses, village wise latitude and longitude coordinates were considered to perform cluster analysis on dataset where each parameter has a disease status (case vs control), as well as spatial and temporal attributes. The model was run on each year case dataset, using the total number of cases in a particular year per *epi*

unit (village) while adjusting for the underlying population of each *epi* unit. Statistical significance was set at the *p*-value of < 0.05 for all SaTScan clusters.

Identification of potential risk factors using linear discriminant analysis

Linear Discriminant Analysis (LDA) was used to carry out a comprehensive analysis of the risk parameters and to establish the mathematical model to provide a reasonable basis for accurately mastering the influence of the parameter and its estimation. LDA was used to identify differences in risk factors between places that were identified as a persistent space-time cluster in SaTScan. A total of 10 environmental/remotely sensed variables (Table 1) were considered for LDA using a binary response 0/1 with clustered locations = 1 and non-clustered locations = 0. This analysis was carried out in R software with the statistical significance of parameters was set at *p*-value of <0.05.

Function used to perform LDA model in R is:

LDA (function, data) = A classifier function ~ x₁ + x₂ + ... That is, the response is the grouping factor and the right hand side specifies the (non-factor) discriminators. data - Data frame from which variables specified in function are preferentially to be taken.

(Classifier function used in LDA ~ Air Temperature + EVI + LAI + LST + NDVI + PET + Rain precipitation rate + Soil moisture + Specific humidity + Surface Pressure + Potential evaporation rate)

Risk estimation using machine learning (ML) system of modelling

The risk factor data generated and georeferenced for two groups of datasets for a period of 14 years (2002–2015) were aggregated to grid level and year wise for uniform representation at two different conditions (>1000 EPG and >2000 EPG). The spatial occurrence of haemonchosis in Rajasthan was predicted through risk maps by performing climate-disease relationship modelling. The data pertaining to risk variables were obtained, pre-processed and annotated with disease condition along with the respective latitude and longitude. The annotated dataset was subjected to accuracy testing for discriminating the efficiency of different ML models. Two regression models, Generalized Linear Models (GLM) and Generalized Additive Models (GAM) along with nine machine learning models viz., Random Forest (RF), Gradient Boosting Machine (GBM), Artificial Neural Network (NNET), Multiple Adaptive Regression Splines (MARS), Flexible Discriminant Analysis (FDA), Classification Tree Analysis (CT), Support Vector Machine (SVM), Naive Bayes (NB) and Adaptive Boosting (ADA) were employed for disease modelling. Different modelling methods yielded number of 'model object' and all these model objects were used to make future predictions for any combinations of values of independent variables. Response plots were created to explore and understand model predictions.

The fitted models were assessed for their discriminating power using Receiving Operating Characteristic (ROC) curve, Cohen's Kappa (Heidke Skill Score) and True Skill Statistics (TSS). These measures were used to evaluate the quality of predictions based on the presence (1) or absence (0) of disease. Raster Stack was used to combine the results of individual predictions by different model methods. All the models were assessed for over-fitting as it can cause misleading of estimated co-efficient, *p*-values and R-Square values. Over-fitting is suspected when the model accuracy was high with respect to the data used in training the model but drops significantly with new data. In this study, the cross-validation procedure was adopted to assess the over-fitting of models by keeping 30% of data on hold while the rest was used for training. The accuracy of held-out data was used to compare the accuracy derived from data used in the training and the significant variance in these two flags for overfitting.

Prediction results from different models were combined and averaging the prediction scores provided the best prediction (Huppert and Katriel, 2013). The average score was obtained by considering the model

Table 1

a). Results of discriminant analysis for group-I (above 1000 EPG) (2002–2015), b). Results of discriminant analysis for group-II (above 2000 EPG) (2002–2015).

a) 2002–2015 (above 1000 epg)							
Sl. No.	Risk Parameters	Mean (presence)	SD	F value	P-value	LD Coefficients	95% CI
1	Air_Temperature	26.71	0.9	2158.23	< 0.050	0.109	26.25 to 27.16
2	EVI	0.2	0.06	23,039.59	< 0.050	3.75	0.10 to 0.29
3	LAI	0.77	9.16	16,611.34	< 0.050	46.2	-11.96 to 13.50
4	LST	31.66	2.17	239.18	0.11	0.0173	31.29 to 32.02
5	NDVI	0.29	0.1	108.65	0.618	1.16	0.27 to 0.30
6	PET	668.81	984.9	3.19	0.000	0.000237	649.83 to 687.78
7	Potential_evaporation_rate	281.21	20.77	112.36	0.137	-0.0033	278.83 to 283.58
8	Rain_precipitation_rate	1.6	0.54	392.87	0.707	-0.0858	1.48 to 1.71
9	Soil_moisture	16.87	3.31	4043.15	< 0.050	0.0381	14.59 to 19.14
10	Specific_humidity	0.01	0	3522.29	< 0.050	142	0.01
b) 2002–2015 (above 2000 epg)							
Sl. No.	Risk Parameters	Mean (presence)	SD	F value	P-value	LD Coefficients	95% CI
1	Air_Temperature	26.43	0.89	25,672.98	0.083	0.05707959	24.40 to 28.45
2	EVI	0.2	0.06	11,523.96	0.000	3.37611415	0.10 to 0.29
3	LAI	1.14	9.18	2481.1	< 0.050	28.7022	-5.27 to 7.55
4	LST	30.99	2.14	4485.36	0.657	0.05099	28.94 to 33.03
5	NDVI	0.3	0.1	16,062.4	0.345	2.74569	0.12 to 0.47
6	PET	729.84	988.6	510.56	0.000	0.0004946	416.98 to 1042.69
7	Potential_evaporation_rate	277.29	19.86	31.3	0.000	-0.005143	275.65 to 278.92
8	Rain_precipitation_rate	1.52	0.53	26,267.83	0.101	-0.079867	0.28 to 2.75
9	Soil_moisture	16.06	3.30	80.34	0.798	0.0512	15.64 to 16.47
10	Specific_humidity	0.01	0	10,548.9	< 0.050	121.43	0.01

satisfying the criteria with kappa >0.60, ROC >0.90 and TSS >0.80.

Irregularities from usual patterns of meteorological and remote sensing variables in both space and time are associated with varying risks which is represented as the risk map derived from the average prediction obtained by AI modelling.

Basic reproduction number (R₀)

Basic reproductive number (R₀), denotes the average number of secondary cases of infectious diseases generated from the initial case in a completely susceptible population. If this number is ≤ 1, minor outbreaks can be expected in the population with a probability of extinction while if the number is > 1, there is a positive chance of a large outbreak affecting nearly the entire population. R statistical software (version 3.6.3) was used to generate the basic reproduction number (R₀). R₀ reveals the transmissibility of disease that helps to forecast the probable size of the epidemics and to create awareness in terms of control measures (Jung et al., 2020).

$$R_0 = -[\log\{(1 - AR) / S_0\} / AR - (1 - S_0)]$$

where, S₀ = the initial percentage of susceptible population, AR = the percentage of the population eventually infected.

Results

Spatial distribution of haemonchosis

During the study period of 14 years from 2002 to 2015, the data on the incidence of haemonchosis in Rajasthan, India were analysed to identify the distribution of disease and is represented in Fig. 1. Out of the 8 districts analysed for the incidence of haemonchosis, all the districts showed haemonchosis at > 2000 EPG except Jaipur, indicating the endemic nature of the disease in the study region. In group I and II (>1000 and > 2000 EPG), the highest incidence was observed in Sikar district (range >2000–3000) and moderately high in Pali and Tonk districts (range >1000–3000).

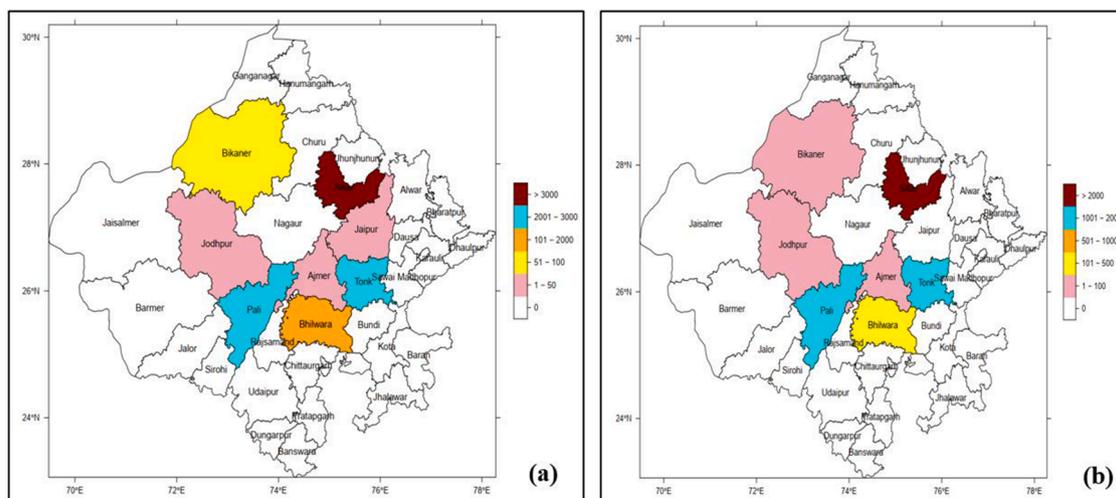


Fig. 1. District wise cumulative incidence of haemonchosis in Rajasthan (2002–2015), (a) above 1000 EPG, b) above 2000 EPG.

Temporal distribution of haemonchosis

Even though, haemonchosis in sheep remained as endemic in all the years studied, the typical seasonal trend was observed. The infection pressure was high during the onset of monsoon (June to October) with the peak noticed in the month of August and September in the two EPG groups (Fig. 2). The lowest incidence of disease was observed during the winter and summer seasons (December to May).

Spatial autocorrelation

Before performing hotspot analysis (Getis-Ord G_i^*), the data was tested to detect the presence of any clusters in the entire dataset by using specific technique involving spatial autocorrelation. A high z-score at >1000 EPG (27.43 to 70.87) and >2000 EPG (32.65 to 56.57) indicated the presence of hotspots. A negative z-score at >1000 EPG (-0.05 to -0.01) and >2000 EPG (-0.04 to -0.01) indicated a cold spot.

Space-Time cluster analysis and geospatial modelling

Space-time cluster analysis revealed the existence of disease cluster in some regions in two situations (>1000 and >2000 EPG) in all the time periods during 2002–2015 (Fig. 3). Two significant clusters of high risk at >1000 EPG (relative risk: 2.08 and 0.232) and at >2000 EPG (relative risk: 1.872 and 0.142) were identified. During the period of 2002 to 2015, the village level disease clustering was identified and incidence was represented by red coloured dots within the significant red circles indicating the villages with high risk of disease incidence whereas purple circles represent villages having disease incidence but are not part of a significant cluster in the two conditions i.e., at >1000 and >2000 EPG.

Linear discriminant analysis

Identification of significant disease clusters using the Space-time cluster model prompted us to identify the variables responsible for the significant cluster formation. Therefore, in the next step, we attempted to use the Linear Discriminant Analysis (LDA) to determine significant risk factors (climatic and remote sensing) responsible for the formation of disease clusters at the village level. The identified risk parameters were further utilized for spatial risk modelling and mapping. The LDA results are presented in Table 1 (a), (b) and revealed the model was fit for the data.

Environmental factors that were found to be significantly associated with the disease incidence at p -value < 0.05 were considered for risk modelling. The study revealed the potential risk factors as EVI, LAI, PET and specific humidity in the two groups whereas in >1000 EPG group air temperature and soil moisture and in >2000 EPG group potential evaporation rate (PER) were also turned to be a significant risk factors [Table. 1(a), (b)].

Risk assessment and estimation

The significant environmental risk factors identified using LDA were subjected to the climate-disease modelling. Maps were generated based on affected (case) and unaffected (control) areas of haemonchosis (Fig. 4). In the map, the case data is represented by red circles indicating the places having the disease incidence and the control data is represented by blue dots indicating the places without incidence of haemonchosis. Out of the two regression models and nine machine learning models tested, RF and ADA models emerged out as best fitted models for both the EPG groups.

Risk prediction and mapping and estimation of basic reproduction number (R_0) and layering on predicted risk maps

Fig 5 (a and b) showed the average score model of RF and ADA (best fit models). In fig.5(a) (>1000 EPG), the risk was observed in the central, eastern, northern and southern part of the state, while the western part of the state was not showing any risk. In fig.5(b) (>2000 EPG), the risk was concentrated on the central and northern part of the state indicating severe disease can be expected in these areas.

Fig. 5(c and d) showed the R_0 values at district level superimposed on the risk prediction map of the state. The R_0 values above 1.00 indicate the places with an increased trend of disease occurrence/ severity/ high risk. The R_0 value was 0.76 to 1.43 for group I (>1000 EPG) with districts like Jodhpur, Ajmer, Sikar and Tonk shows high disease severity (Fig. 5(c)) and R_0 values 1.13–1.54 for group II (>2000 EPG) revealed that the districts at high risk are Ajmer, Jodhpur, Sikar and Tonk (Fig. 5 (d)). Further, the places/region having low R_0 possibly might shift to high R_0 values in near future due to the migration of infected animals from one place to another.

Discussion

Haemonchosis in sheep is regarded as a disease with direct dependency on environmental conditions. In order to maximize the chance of being ingested by the susceptible host, the infective larvae (L3) of *H. contortus* have to exit from the faecal pellet and have to position themselves at the vegetation which is referred as the ‘initial cruising phase’. The movement of the larva is highly influenced by photoperiodicity (Fenton and Rands, 2004; Santos et al., 2012). Once the infective larvae reach the grass blade, they will prefer to remain in the ‘ambushing phase’ awaiting for the host contact on the vegetation. Hence, the initial cruising phase followed by the ambushing phase is a remarkable event in the life cycle of *H. contortus* and the initial cruising phase is highly influenced by the climatic conditions. During the pre-parasitic period (exogenous phase), temperature and humidity were reported to exert maximum effect on the presence of viable and infective L3 on pasture (Iliev et al., 2018).

In the present study, based on disease incidence data, the high-risk

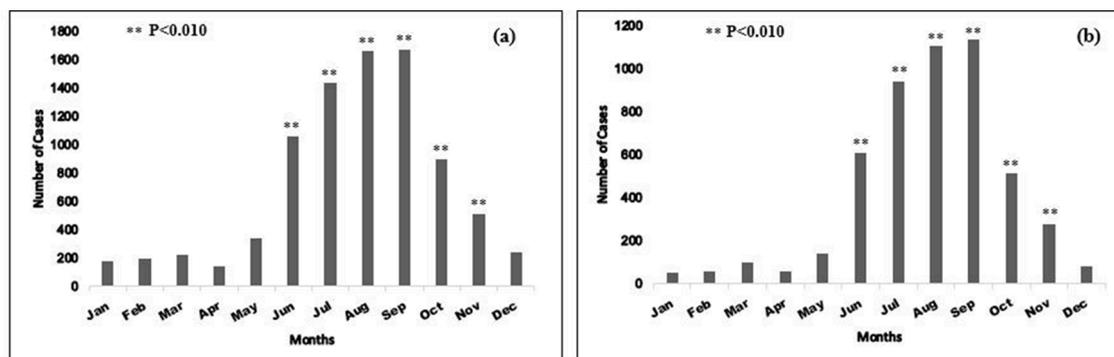


Fig. 2. Month wise cumulative incidence of haemonchosis in Rajasthan (2002–2015), a). above 1000 EPG, b). above 2000 EPG.

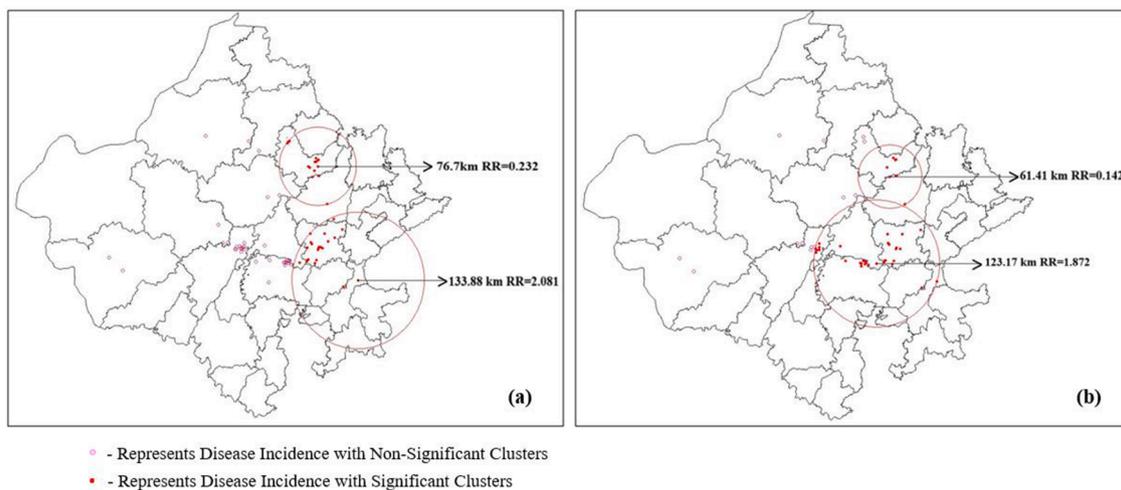


Fig. 3. Space- Time cluster analysis (2002–2015), a). above 1000 EPG, b). above 2000 EPG.

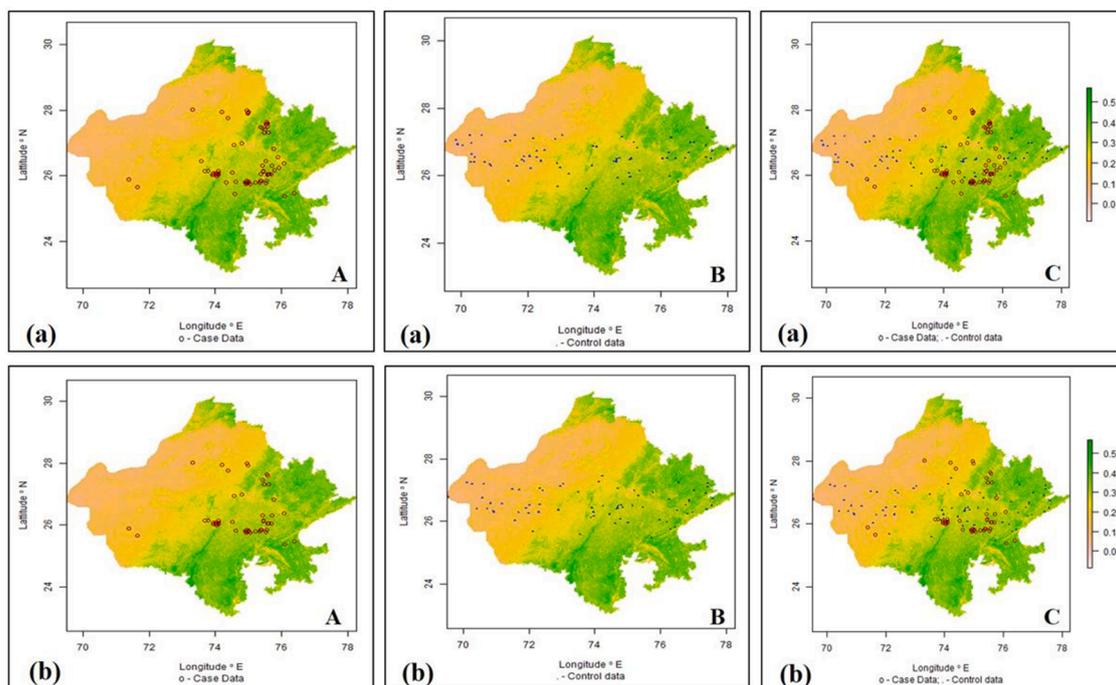


Fig. 4. Rajasthan map depicting the occurrence of haemonchosis at above 1000 EPG (a) and above 2000 EPG (b) over NDVI shown in colour ranging 0 (brown) to 0.5 (green). (A) Case data – red colour circles indicate the places having haemonchosis incidence, (B) Control data – blue dots indicating the places without haemonchosis incidence, and (C) Case- Control data – showing both presence – absence of haemonchosis incidence.

areas for haemonchosis were identified for planning the control strategies. The overall prevalence of haemonchosis in Rajasthan was 34.09% at >1000 EPG and >20.07% at >2000 EPG. The prevalence of haemonchosis observed in the present study was lesser when compared to the disease prevalence in other regions of the world like Middle Guinea (94%) (Barry et al., 2002) and Eastern Ethiopia (60%) (Sissay et al., 2007). A positive correlation persists between faecal egg count (FEC) and presence of worm in abomasum in host (Swarnkar and Singh 2020) with higher FEC indicates higher worm load.

The highest prevalence of haemonchosis was witnessed during the rainy season (July to September), while winter and summer (December to May) recorded the least prevalence. The incidence of haemonchosis usually occurs during monsoon and post monsoon seasons. In Rajasthan, the monsoon months of June, July August, September and October recorded the maximum outbreaks, the highest peak being observed in

August and September. The reason may be during monsoon, the stringent moisture content on the pasture cover favours the free-living stages of *H. contortus* and rapid hatching of eggs when exposed to moisture (O'Connor et al., 2006), but the survival period increases as the relative humidity or faecal moisture content (FMC) increases (Berberian and Mizelle, 1957). From India, Singh et al. (2018) reported higher incidence of strongylosis during monsoon season in sheep in Rajasthan whereas earlier survey in Rajasthan indicated the burden of peak infection of GI nematode during monsoon and remained at very low-level during winter and summer seasons with absence of typical peri-parturient rise in FEC (Singh et al., 1997).

The results of the study are in accordance with the observations of other workers who also reported a high prevalence of haemonchosis during the rainy season (Mushonga et al., 2018; Qamar et al., 2009). They reported that environmental risk factors like pasture cover and

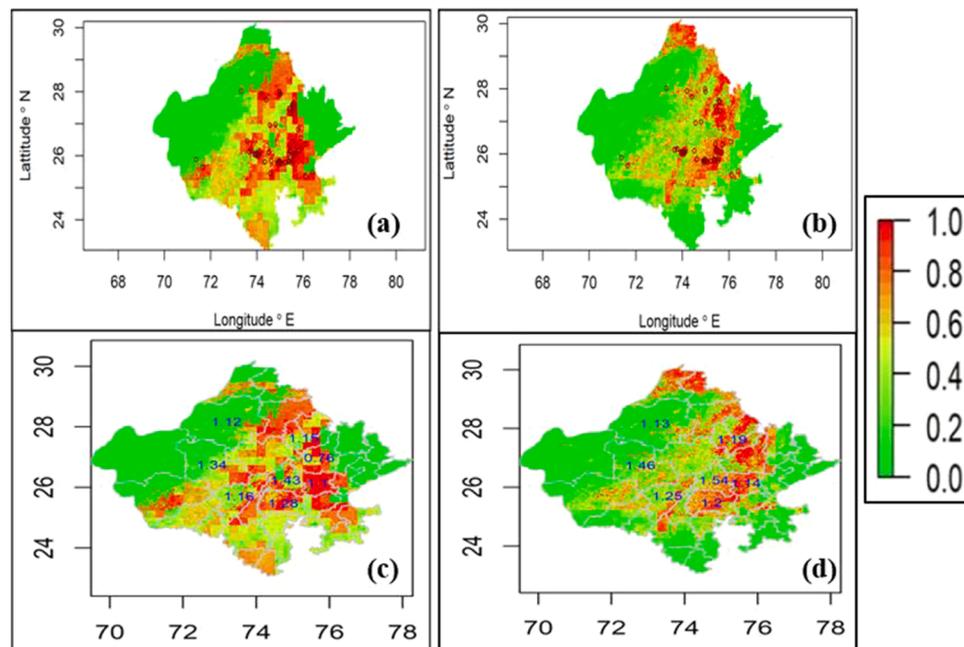


Fig. 5. Haemonchosis risk prediction map (2002–2015): (a) above 1000 EPG, (b) above 2000 EPG. The risk shown in colour ranging from 0 (green) to 1 (red). Small red circles indicate disease prevalence locations. Haemonchosis R_0 values on risk prediction map district wise (2002–2015): (c) above 1000 EPG, (d) above 2000 EPG. The risk shown in colour ranging from 0 (green) to 1 (red).

favourable season are important for the development and survival of the free-living larvae. During the warm and wet grazing season, the animal spends more time on pasture and the practice of inappropriate deworming procedure might have contributed to the high incidence of disease (Lindqvist et al., 2001).

In the present study, the space-time cluster analysis at above 1000 EPG and at above 2000 EPG indicated two significant clusters each of high risk. Such scattered distribution is in accordance with the model studied by Bolajoko et al. (2015) who also showed the spatial heterogenic distribution of haemonchosis in different agro-climatic conditions. Persistent clusters of haemonchosis were recorded from the places where air temperature ranged between 23.77–28.74 °C. It indicates that micro environment conditions within the vegetation cover during the rainy and post monsoon season were quite favourable for the larval development. Factors such as pasture height and density can affect the local microclimate. In addition to the presence of shade from trees and shrubs, such environment-modifying factors are likely to explain differences in the development of infective larvae on a particular pasture. Moderate temperatures (20–29 °C) and moderate humidity (51–81%) cause an increase in larval development (Misra and Ruprah, 1968; Swarnkar and Singh, 2020) further confirmed the crucial role of temperature and moisture for the translation of nematode larva in the environment.

Machine learning models provides the advantage of learning non-linear relationships between the disease incidence and risk variables of interest, and allowing heterogenous variable types in one model, which is not feasible in univariate or multivariate analysis of risk-score prediction models. In this study the RF and ADA models emerged as best fitted models to provide the improved accuracy of risk prediction of haemonchosis in sheep.

In this study, similar to Bekuma and Dufera (2019), we considered logistic regression/discriminant function analysis to identify the significant risk factors associated with disease burden. In the present study EVI, LAI, PET and specific humidity were found to be positively correlated with the persistent occurrence of the infection. Besides, earlier studies also showed rainfall, relative humidity and presence of green vegetation as important factors determining the level of faecal egg count (Keyyu et al., 2005; Sissay et al., 2007). However, exploration of the

interaction amongst different risk variables may reveal better interpretation.

The risk factor analysis highlighted the importance of seasonal cycle while mean air temperature and specific humidity dominated in the variations of *H. contortus* infection from district to district. These results are in agreement with the previous reports that temperature and water availability are dependant factors for the development and survival of the larval stages and transmission is strongly influenced by seasonal changes in the prevailing climate (Van Dijk et al., 2008; Morgan and Van Dijk, 2012).

Predictive models [based either on time series analysis or on R_0 model] of *H. contortus* transmission to sheep have recently been developed by Bolajoko et al. (2015). However, collection of incidence data derived from active surveillance, as with the present survey, are necessary for continued improvement of models, validation and meaningful predictions (Bolajoko et al., 2015).

The basic reproduction number (R_0) is of key importance in the study to evaluate the transmissibility of haemonchosis potential in the identified risk region. The daily incidence and the magnitude of outbreak are largely dependant on the value of R_0 suggesting more population would be infected in future. The R_0 is the key measure in estimating the ability of pathogen to spread. The R_0 values have important implications for disease control and it indicates the level of mitigation efforts needed to bring an epidemic under control (Bolajoko et al., 2015). The R_0 ranging from 0.76 to 1.43 for group I (>1000 EPG), districts like Jodhpur, Ajmer, Sikar and Tonk showed high disease severity and R_0 values 1.13–1.54 for group II (>2000 EPG) revealed that the districts at high risk are Ajmer, Jodhpur, Sikar and Tonk. As the herd immunity goes down the R_0 is likely to be increased. Since the >2000 EPG represents more severe condition and the lower herd immunity, there is a probability that the new secondary host will acquire the infection faster and hence R_0 values are expected to be high in Group II.

This is a first endeavour towards an integrative method to predict the risk of establishment of disease based on mathematical modelling combined with a geographic information system including satellite data, landscape, host parameters and other relevant factors determining risk for haemonchosis. The results of the study will help the sheep farmers, veterinarians and policy makers to strategically implement the control

plan in high-risk areas in vulnerable seasons. As haemonchosis in sheep is a global concern, the findings of the study will help to identify risk factors and to adopt suitable control measures in other part of the world too.

Ethical statement

The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to. Ethical approval was not applicable as per institutional guidelines and recommendations.

Declaration of Competing Interest

The authors declare that there is no conflict of interest.

Acknowledgement

The financial assistance under the AINWP on GIP of ICAR is thankfully acknowledged. The authors are grateful to the Directors of ICAR-NIVEDI and ICAR-CSWRI for supporting and providing necessary help. The technical assistance from Mr. Bharath Prasad is also acknowledged.

References

- Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* 43, 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>.
- Almeida, F.A., Garcia, K.C., Torgerson, P.R., Amarante, A.F.T., 2010. Multiple resistance to anthelmintics by *Haemonchus contortus* and *Trichostrongylus colubriformis* in sheep in Brazil. *Parasitol. Int.* 59, 622–625. DOI: 10.1016/j.parint.2010.09.006.
- Amarante, A.F.T., 2014. Sustainable worm control practices in South America. *Small Rumin. Res.* 118, 56–62. <https://doi.org/10.1016/j.smallrumres.2013.12.016>.
- Barger, I.A., 1999. The role of epidemiological knowledge and grazing management for helminth control in small ruminants. *Int. J. Parasitol.* 29, 41–47 [https://doi.org/10.1016/S0020-7519\(98\)00176-3](https://doi.org/10.1016/S0020-7519(98)00176-3).
- Barry, A.M., Pandey, V.S., Bah, S., Dorny, P., 2002. Epidemiological study of gastrointestinal helminths in goats in Central Guinea. *Rev. Elev. Med. Vet. Pays. Trop.* 55 (2), 99–104.
- Bekuma, F., Dufera, B., 2019. Prevalence of haemonchosis in small ruminants and its associated risk factors in and around Ejere Town, West Shoa, Oromia, Ethiopia. *Am. J. Biomed. Sci.* 3 (5), 409–414. DOI: 10.34297/AJBSR.2019.03.000704.
- Berberian, J.F., Mizelle, J.D., 1957. Developmental studies on *Haemonchus contortus* Rudolphi (1803). *Am. Midl. Nat.* 57, 421–439. <https://doi.org/10.2307/2422408>. DOI:
- Bolajoko, M.B., Rose, H., Musella, V., Bosco, A., Rinaldi, L., Van Dijk, J., Cringoli, G., Morgan, E.R., 2015. The basic reproduction quotient (Q0) as a spatial predictor of the seasonality of ovine haemonchosis. *Geospat. Health* 9, 333–350.
- Brazil, N., 2019. Spatial methods in community research-lab 5: spatial autocorrelation. CRD 298.
- Burgess, C.G.S., Bartley, Y., Redman, E., Skuce, P.J., Nath, M., Whitelaw, F., Tait, A., Gilleard, S., Jackson, F.A., 2012. Survey of the trichostrongylid nematode species present on UK sheep farms and associated anthelmintic control practices. *Vet. Parasitol.* 189 (2–4), 299–307. <https://doi.org/10.1016/j.vetpar.2012.04.009>. DOI: DAHD, 2019. Livestock census. Department of Animal Husbandry and Dairying, 2019 New Delhi <http://dahd.nic.in/about-us/divisions/statistics>.
- de Haan, C., Bekure, S., 1991. Animal Health Services in Sub-Saharan Africa: Initial Experience With Alternative Approaches. World Bank Technical Paper No. 134. World Bank, Washington DC, p. 49.
- Fenton, A., Rands, S.A., 2004. Optimal parasite infection strategies: a state-dependent approach. *Int. J. Parasitol.* 34, 813–821. <https://doi.org/10.1016/j.ijpara.2004.02.003>.
- Gatongi, P.M., Gathuma, J.M., Munyua, W.K., 1988. Influence of weather on population dynamics of infective larvae of cattle nematodes in Nyeri. Kenya. *Bull. Anim. Health Prod. Afr.* 36, 148–154.
- Getis, A., Ord, J.K., 1992. The Analysis of Spatial Association by Use of Distance Statistics. *Geogr. Anal.* 24 (3), 189–206. <https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>.
- Githigia, S.M., Thamsborg, S.M., Munyua, W.K., Maingi, N., 2001. Impact of gastrointestinal helminths on production in goats in Kenya. *Small Rum. Res.* 42 (1), 21–29. 10.1016/S0921-4488(01)00240-1.
- Gowane, G.R., Swarnkar, C.P., Misra, S.S., Kumar, R., Kumar, A., Prince, L.L.L., 2019. Genetic parameter estimates for fecal egg counts and their relationship with growth in Avikalin and Malpara sheep. *Animal* 13, 1788–1796. <https://doi.org/10.1017/S1757131118003749>. DOI:
- Gupta, R.P., Yadav, C.L., Chaudri, S.S., 1987. Epidemiology of gastrointestinal nematodes of sheep and goats in Haryana, India. *Vet. Parasitol.* 24, 117–127. [https://doi.org/10.1016/0304-4017\(87\)90136-1](https://doi.org/10.1016/0304-4017(87)90136-1). DOI:
- Huppert, A., Katriel, G., 2013. Mathematical modelling and prediction in infectious disease epidemiology. *Clin. Microbiol. Infect.* 19 (11), 999–1005. <https://doi.org/10.1111/1469-0691.12308>.
- Iliev, P.T., Ivanov, A., Prelezov, P., 2018. Effects of temperature and desiccation on survival rate of *Haemonchus contortus* infective larval stage. *Trakia J. Sci., Ser. Soc. Sci.* 1, 17–21. DOI:10.15547/tjs.2018.01.004.
- Itty, P.J., Zinsstag, P., Ankers, Nijje, M., Pfister, K., 1997. Returns from Strategic anthelmintic treatments in village Sheep in the Gambia. *Prev. Vet. Med.* 32 (3–4), 299–310.
- Jabbar, A., Iqbal, Z., Saddiqi, H.A., Babar, W., Saeed, M., 2008. Prevalence of multiple anthelmintic resistant gastrointestinal nematodes in dairy goats in a desolated tract (Pakistan). *Parasitol. Res.* 103, 29–35. <https://doi.org/10.1007/s00436-008-0923-6>. DOI:
- Jung, S., Akhmetzhanov, A.R., Hayashi, K., Linton, N.M., Yang, Y., Yuan, B., Kobayashi, T., Kinoshita, R., Nishiura, H., 2020. Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: inference using exported cases. *J. Clin. Med.* 9, E523. <https://doi.org/10.3390/jcm9020523>, 2020.
- Kao, R.R., Leathwick, D.M., Roberts, M.G., Sutherland, L.A., 2000. Nematode parasites of sheep: a survey of epidemiological parameters and their application in a simple model. *Parasitol.* 121 (1), 85–103. <https://doi.org/10.1017/s0031182099006095>. DOI:
- Kenyon, F., Sargison, N.D., Skuce, P.J., Jackson, F., 2009. Sheep helminth parasitic diseases in south eastern Scotland arising as a possible consequence of climate change. *Vet. Parasitol.* 163 (4), 293–297. <https://doi.org/10.1016/j.vetpar.2009.03.027>. DOI:
- Keyyu, J.D., Kyvsgaard, N.C., Monrad, J., Kassuku, A.A., 2005. Epidemiology of gastrointestinal nematodes in cattle on traditional, small-scale dairy and large-scale dairy farms in Iringa district. Tanzania. *Vet. Parasitol.* 127, 285–294. <https://doi.org/10.1016/j.vetpar.2004.10.014>. DOI:
- Lindqvist, A., Ljungstrom, B.L., Nilsson, O., Waller, P.J., 2001. The dynamics, prevalence and impact of nematode parasite infections in organically raised sheep in Sweden. *Acta Vet. Scand.* 42, 377–389.
- McLeod, R.S., 1995. Costs of major parasites to the Australian livestock industries. *Int. J. Parasitol.* 25, 1363–1367. 10.1016/0020-7519(95)00071-9.
- McLeod, R.S., 2004. The economic impact of worm infections in small ruminants of Southeast Asia, India and Australia. In: Worm Control for Small Ruminants in Tropical Asia (R.A. Sani, G.D. Gray and R.L. Baker, Eds.), ACIAR Monograph 113, Australia, pp. 23–33.
- Misra, S.C., Ruprah, N.S., 1968. Incidence of helminths in goats at Hissar. *J. Res.* 5, 276–286.
- Morgan, E.R., Van Dijk, J., 2012. Climate and the epidemiology of gastrointestinal nematode infections of sheep in Europe. *Vet. Parasitol.* 189, 8–14. <https://doi.org/10.1016/j.vetpar.2012.03.028>. DOI:
- Mushonga, B., Habumugisha, D., Kandiwa, E., Madzingira, O., Samyang, A., Segwagwe, B.E., Jaja, I.F., 2018. Prevalence of *Haemonchus contortus* infections in Sheep and Goats in Nyagatare District, Rwanda. *J. Vet. Med. doi.org/10.1155/2018/3602081*.
- Nari, A., Salles, J., Castell, D., Hansen, J.W., 1997. Control of gastro-intestinal nematodes in farming systems of Uruguay. In: Hansen, J.W., Waller, P.J. (Eds.), Proc. Biological Control Workshop, 141. FAO Technical Series Report, pp. 89–94. Eds.
- O'Connor, L.J., Walkden-Brown, S.W., Kahn, L.P., 2006. Ecology of the free-living stages of major trichostrongylid parasites of sheep. *Vet. Parasitol.* 142, 1–15. <https://doi.org/10.1016/j.vetpar.2006.08.035>. DOI:
- Ord, J.K., Getis, A., 1995. Local spatial autocorrelation statistics: distributional issues and an application. *Geogr. Anal.* 27 (4), 286–306. <https://doi.org/10.1111/j.1538-4632.1995.tb00912.x>.
- Qamar, M.F., Maqbool, A., Khan, M.S., Ahmad, N., Muneer, M.A., 2009. Epidemiology of haemonchosis in sheep and goats under different managemental conditions. *Vet. World* 2 (11), 413–417.
- Rinaldi, L., Catelan, D., Musella, V., Cecconi, L., Hertzberg, H., Torgerson, P.R., Mavrot, F., Waal, T.D., Selemetas, N., Coll, T., Bosco, A., Biggeri, A., Cringoli, G., 2015. *Haemonchus contortus*: spatial risk distribution for infection in sheep in Europe. *Geospat. Health* 9 (2), 325–331. <https://doi.org/10.4081/gh.2015.355>. DOI:
- Rose, H., Caminade, C., Bolajoko, M.B., Phelan, P., Dijk, J.V., Baylis, M., Williams, D., Morgan, E.R., 2016. Climate-driven changes to the spatio-temporal distribution of the parasitic nematode, *Haemonchus contortus*, in sheep in Europe. *Glob. Chang. Biol.* 22, 1271–1285. <https://doi.org/10.1111/gcb.13132>.
- Sahoo, A., Khan, F., 2016. Nutritional and biological control synergism against gastrointestinal nematodes in small ruminants. *Int. J. Vet. Sci. Anim. Husb.* 4 (1), 104. DOI:10.15744/2348-9790.4.104.
- Santos, M.C., Silva, B.F., Alessandro, F., Amarante, A.F.T., 2012. Environmental factors influencing the transmission of *Haemonchus contortus*. *Vet. Parasitol.* 188 (3–4), 277–284 <https://doi.org/10.1016/j.vetpar.2012.03.056>.
- Singh, D., Swarnkar, C.P., Khan, F.A., 2018a. Epidemiology of gastrointestinal parasites and impact of worm management schemes in sheep flocks of Rajasthan. *Small Rum. Res.* 164, 22–27. DOI:10.1016/j.smallrumres.2018.04.011.
- Singh, D., Swarnkar, C.P., Sanyal, P.K., 2015. Dynamics of epidemiological intelligence for exploitation in effective worm management in sheep – A Rajasthan experience. *Indian J. Anim. Sci.* 85, 679–694.
- Singh, D., Swarnkar, C.P., Khan, F.A., Srivastava, C.P., Bhagwan, P.S.K., 1997. Epidemiology of ovine gastrointestinal nematodes at an organised farm in Rajasthan. *India. Small Rum. Res.* 26, 31–37. 10.1016/S0921-4488(96)00988-1.
- Singh, D., Swarnkar, C.P., Prince, L.L.L., Pathak, K.M.L., 2011. Economic Analysis and Impact of Gastrointestinal Nematodes On Sheep Production in Rajasthan. Directorate of Knowledge Management in Agriculture. ICAR, New Delhi, pp. 1–84.

- Sissay, M.M., Uggle, A., Waller, P.J., 2007. Prevalence and seasonal incidence of nematode parasites and fluke infections of sheep and goats in Eastern Ethiopia. *Trop. Anim. Health Prod.* 39 (7), 521–531. DOI: 10.1007/s11250-007-9035-z.
- Smith, S.M., 2002. AGRI-FOCUS. July. <http://grant-adams.wsu.edu>.
- Soulsby, E.J.L., 1982. *Helminths, Arthropods and Protozoa of Domesticated Animals*. Bailliere Tindall, London, pp. 1–809, 7th ed.
- Swarnkar, C.P., Singh, D., 2010. Worm control practices, anthelmintic use and its implication on anthelmintic resistance in gastrointestinal nematodes of sheep in Rajasthan. *Indian J. Anim. Sci.* 80, 593–600.
- Swarnkar, C.P., Singh, D., 2011. Role of bioclimatographs in forecasting of strongyle infection in Rajasthan. *Ind. J. Anim. Sci.* 81 (3), 216–223.
- Swarnkar, C.P., Singh, D., 2015. Monthly variation in worm burden exhibiting possibility of hypobiosis of *Haemonchus contortus* in sheep under farm conditions of semi-arid Rajasthan. *J. Vet. Parasitol.* 29, 20–26.
- Swarnkar, C.P., Singh, D., 2020. Rhythmicity in thermal humidity index and regulates of *Haemonchus contortus* in sheep of semi-arid tropical Rajasthan. *Biol. Rhythm Res.* 51, 58–66. <https://doi.org/10.1080/09291016.2018.1515806>.
- Swarnkar, C.P., Singh, D., Chopra, A., Prince, L.L.L., 2014. Influence of advancement of age on intensity of strongyle infection and performance in sheep selected for resistance /resilience to infection. *Indian J. Anim. Sci.* 84, 254–261.
- Swarnkar, C.P., Singh, D., Srivastava, C.P., Bhagwan, P.S.K., Dimri, U., 1996. A retrospective study on ovine gastrointestinal helminthoses under semi-arid condition. *J. Vet. Parasitol.* 10, 15–21.
- Van Dijk, J., David, G.P., Baird, G., Morgan, E.R., 2008. Back to the future: developing hypotheses on the effects of climate change on ovine parasitic gastroenteritis from historical data. *Vet. Parasitol.* 158, 73–84. <https://doi.org/10.1016/j.vetpar.2008.08.006>. DOI: vetpar.2008.08.006.
- Waller, P.J., Ruby-Martin, L., Ljungstrom, B.L., Rydzik, A., 2004. The epidemiology of abomasal nematodes of sheep in Sweden, with particular reference to over-winter survival strategies. *Vet. Parasitol.* 122, 207–220. <https://doi.org/10.1016/j.vetpar.2004.04.007>. DOI: vetpar.2004.04.007.