

RESEARCH ARTICLE

Modelling the distribution of the tick *Ixodes ricinus* in England and Wales using passive surveillance data from citizen science reports

Mark Gideon Burdon^{1*}, Maximilian Ayling¹, Nyall Jamieson², Julie Day¹, Jolyon Medlock¹, Kayleigh Hansford¹, G. R. William Wint³, Thomas Ward¹

1 UK Health Security Agency, London, United Kingdom, **2** Department of Mathematics, University of Manchester, Manchester, United Kingdom, **3** Department of Biology, University of Oxford, Oxford, United Kingdom

* mark.burdon@ukhsa.gov.uk



Abstract

Background

Ticks are a significant cause of illness globally. The tick *Ixodes ricinus* is commonly found across Europe and is a significant vector of Tick-Borne Encephalitis virus (TBEv), *Borrelia burgdorferi* s.l. (causative agent of Lyme borreliosis), *Babesia divergens*, *Anaplasma phagocytophilum*, and several *Rickettsia* bacteria, among others.

Methods

The Tick Surveillance Scheme (TSS) administered by the UK Health Security Agency (UKHSA) contains validated reports from the general public of tick encounters over the last twenty years. We modelled the probability of *I. ricinus* tick presence across England and Wales using the locations of TSS reports from 2013 to 2023 and a combination of biotic and abiotic factors. An ensemble of statistical and machine learning models was trained to classify points as presence (true tick report locations) or background (points generated randomly and by target-group sampling).

Results

The ensemble model had a continuous Boyce index of 0.99 and area under the receiver-operator curve (ROC AUC) of 0.84 on out-of-sample 2024 data. Variables relating to roe deer (*Capreolus capreolus*) distribution and land cover type were most important. Most of southern England, as well as other areas with known tick populations such as the New Forest and the Lake District, are modelled as highly probable tick presence areas.

OPEN ACCESS

Citation: Burdon MG, Ayling M, Jamieson N, Day J, Medlock J, Hansford K, et al. (2025) Modelling the distribution of the tick *Ixodes ricinus* in England and Wales using passive surveillance data from citizen science reports. PLoS Negl Trop Dis 19(10): e0013520. <https://doi.org/10.1371/journal.pntd.0013520>

Editor: Travis J Bourret, Creighton University, UNITED STATES OF AMERICA

Received: June 9, 2025

Accepted: September 2, 2025

Published: October 9, 2025

Copyright: © 2025 Burdon et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The Tick Surveillance Scheme data used in the study cannot be made public in full as it contains confidential location information from those who submitted the data. Access can be requested from UKHSA by organisations

looking to use protected data for public health purposes. To request an application pack or discuss a request for the Tick Surveillance Scheme data, contact DataAccess@ukhsa.gov.uk. The figures and analysis in this manuscript also use geological data and national boundaries which are freely available under Open Government Licenses. Use of these data requires the following attribution statements: contains British Geological Survey materials, copyright owned by UKRI (2025), and contains OS data used under Crown copyright and database right (2025).

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Interpretation

Unstructured citizen science data was suitable for creating a high-performing species distribution model for *I. ricinus* after addressing spatial and demographic biases. This model is now being used to inform local public health awareness showing the advantage of passive surveillance through to modelling and public health awareness.

Author summary

Ticks are concerning for public health because their bite can pass on a wide range of diseases to humans. Understanding which areas have ticks allows local public health agencies to target their work more effectively, but manually searching for ticks takes considerable time and resources. One alternative is to record where ticks have been found by members of the public (including vets, health professionals and scientists), ensuring that the tick specimens are collected and verified. We collected locations of verified deer tick encounters in England and Wales from the UK Health Security Agency's Tick Surveillance Scheme. We then collated information about the environment surrounding the tick locations, including average temperatures, whether deer are thought to be present, how much of the area is woodland, etc. This allowed us to create models that estimate how likely each part of England and Wales is to have ticks and create a detailed risk map. The main predicted area for the deer tick was in southern England, while the Midlands and much of Wales were less likely to have ticks; this generally agrees with previous research. On average, the model classified areas with deer and more woodland as more likely to have ticks.

Introduction

Ticks play a significant role in the transmission of zoonotic disease globally, transmitting a range of pathogenic micro-organisms - including protozoa, rickettsiae, spirochaetes and viruses - to humans [1]. Novel tick-borne pathogens continue to emerge in the 21st century, including the Heartland and Bourbon viruses in 2009 and 2014 [2]. Understanding the distribution and spread of ticks is therefore important for public health globally.

The tick *Ixodes ricinus* is commonly found across Europe and is a significant vector of Tick-Borne Encephalitis virus (TBEv), *Borrelia burgdorferi* s.l. (causative agent of Lyme borreliosis), *Babesia divergens*, *Anaplasma phagocytophilum*, and several *Rickettsia* species, among others. In the UK, *I. ricinus* has been confirmed as a vector of TBEv since 2019, with now a small number of probable and confirmed cases recorded having been acquired locally [3]. From a veterinary perspective, *I. ricinus* is a known vector of louping ill virus [4], tick-borne fever (caused by *Anaplasma phagocytophilum*) [5] and bovine babesiosis (caused by *Babesia divergens*) [6].

Ixodes ricinus ticks appear to have also spread into new areas of England since the early 2000s [7,8]. A central component to understanding and mitigating the risk ticks pose to human health is a reliable understanding of where tick populations are likely to be found and where humans might come into contact with them. Targeted field surveillance for ticks carried out by scientifically trained professionals can be resource intensive and therefore have limited spatial and temporal coverage [9,10]. Citizen science projects to record species observations are therefore highly valuable for use in passive surveillance of infectious disease vectors (see also [11]), which can be enhanced when combined with expert verification and validation processes. The TSS identification validation process is critical to prevent morphologically similar species (for example *Ixodes hexagonus*), with varying ecologies, being misclassified. The ability to estimate spatial vector distributions from unstructured citizen science data is important for public health and, at the time of writing, no detailed tick risk map for England and Wales was accessible online.

This study will be of interest to public health agencies globally who aim to use passively collected citizen science data to model vector species distributions. It models the spatial distribution of *I. ricinus* across England and Wales using the Tick Surveillance Scheme (TSS) data collected by the Medical Entomology group in the UK Health Security Agency (UKHSA, previously in Public Health England and Health Protection Agency). The TSS contains expert validated reports of tick encounters sent in by human and animal health professionals and the public (with tick submissions after a bite of a human or animal). Submissions must include a live or dead tick specimen, and include a form that provides details of how and where the tick was encountered. UKHSA's Medical Entomology and Zoonoses Ecology team identify each specimen and record geographic coordinates for the likely acquisition location, noting how these were derived.

Methodology

Data

Vector presences. We used 4,083 TSS records from England and Wales that were verified by UKHSA as *I. ricinus* from between 2013 and 2023 inclusive, with partial records for 2024 held back for use only in model testing to avoid spatial autocorrelation resulting in inflated performance statistics [12]. With most of the reports in England and Wales, data from Scotland and Northern Ireland was excluded from the study. Duplicate reports from the same recorder in the same location and year were also excluded, as this study is focused on modelling presence.

The TSS is based on passive surveillance data from citizen science reports, which introduces spatial biases. There is heterogeneous sampling effort across the study area due to differences in where individuals with a higher propensity to submit reports live or visit - such as urban areas or popular nature parks, as opposed to inaccessible private woodland or farmland [13,14]. Because this pro-urban spatial heterogeneity was apparent in the raw data, we addressed it by a) varying importance weights for presences based on population density at the tick report location; and b) using target-group sampling to generate background points.

To reduce the spatial bias arising from higher observer presence in populous areas, we used importance weights that assigned more influence to occurrences in sparsely populated areas, using normalised log population density of the tick location's 2021 Middle layer Super Output Area [15]. Weighting for sampling effort has previously been shown to increase accuracy in modelling bird presence [16]. The differences in these weights can be seen graphically in the presences subplot of Fig 1.

Background points. Previous modelling research has noted that obtaining true absence data for ticks is challenging because a species' niche changes over time [17] and with weather conditions [18] and seasonality, and sampling via dragging is resource-intensive [10]. Standard practice is to generate either "pseudo-absence" points, often stratified by how confident the researcher is of absence, or "background" points that represent the overall environmental space, including areas where the species may be present. However, neither of these methods addresses the issue of spatial bias due to differential sampling effort; therefore, we combined random background point generation with target-group

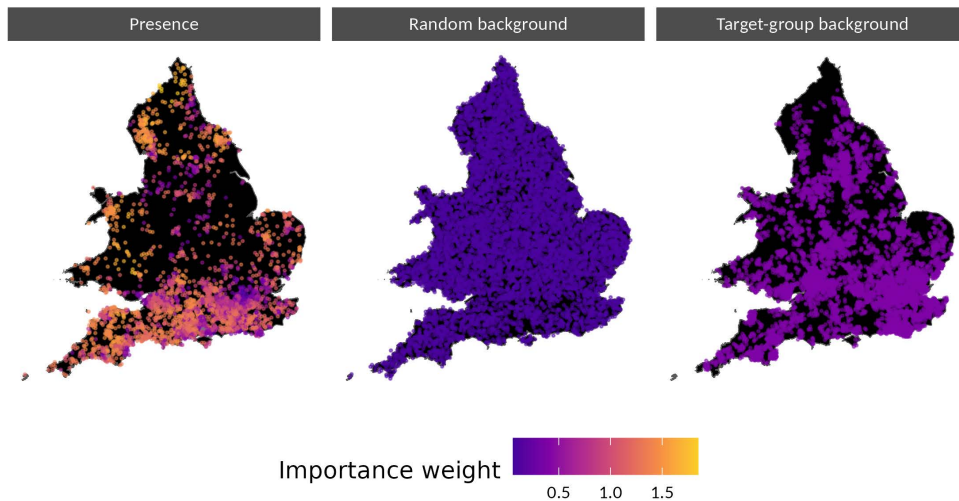


Fig 1. Maps showing the locations and relative weights given to presence points, randomly generated background points and target-group sampled background points. Areas with no points are filled in black. In the plot, all point locations are jittered for pseudonymisation. Country boundaries source: Office for National Statistics licensed under the Open Government Licence v.3.0.

<https://doi.org/10.1371/journal.pntd.0013520.g001>

background sampling. Target-group sampled background points are generated based on presence reports of other species in the same taxonomic group as the one targeted by the SDM and has been shown to be effective at reducing spatial bias [19,20]. Species of the same taxonomic group are likely to have similar detectability and sample biases. We anticipate this may also reduce biases arising from differential propensity to report ticks among different human demographics. We generated two target-group sampled background points for each presence point, distributed according to the density of other tick species reports from the TSS. A 2:1 ratio was selected because of recommendations from Liu et al. (2019) to create background points as a small multiple of the total presences.

However, iterative model development showed that relying entirely on target-group sampling resulted in a lack of background points in some areas (e.g., the north Pennines) that had very few TSS reports for any species of tick. To address these gaps, geographically random background points were also generated within the boundaries of England and Wales, with a minimum distance of 5km from presences (to reduce overlap between presence and background points). Two randomly generated points were created for each presence point, resulting in a total of four background points per presence point.

The effect of this combination is that background points were distributed throughout the geographical and environmental space of England and Wales, but are more likely to be produced in areas where other ticks have been reported. To increase the influence of the target-group sampled background points, these were assigned a weight of 0.8 (compared to 0.2 for randomly generated points). The combined target-group and random background points were then re-weighted such that the total weight of all background points was approximately equal to the total weight of all presence points, following the finding from [21] that a combination of low background-to-presence ratio and equal total group weight between presences and background points performs well in SDMs.

A version of the model without randomly generated background points, and a version with more background points, are demonstrated in the Sensitivity Testing section, as are alternative weighting schemas (e.g., placing more emphasis on target-group sampled points) and a version of the model without buffers around the randomly generated background points. Fig 1 shows the location of both presence and background points in the training set and their relative weighting; Fig A in S1 Text compares the total weight assigned to each type of point (presence, target-group sampled background, random background).

Environmental data. A set of environmental predictor variables, which have previously been associated with the presence of *I. ricinus*, were identified and collated from publicly available data sources. These predictor variables fall into the following categories: climatic conditions, land cover, soil type, geology, host prevalence (cattle, pig, sheep and six species of deer), NDVI and elevation.

Although the *I. ricinus* tick can only crawl a few metres [22], it has a wide range of hosts including deer, livestock, rodents, birds and small mammals, all of which can help with tick dispersal and it can adapt to live in a range of habitats (e.g., [deciduous, coniferous, mixed] woodland, woodland edge, moorland, heathland, grazed grassland, urban parks) provided there is a suitable microclimate to support off-host survival [23]. As a result, many different factors may contribute in complex ways to determine whether if the species is introduced to an area, and if it is able to thrive in that locality. The increase in deer numbers and their expanding range across the UK, along with as land-use changes and climatic variation, may have contributed to an enlarged geographic range for *I. ricinus* [8,24,25]. We therefore aimed to include a diverse range of factors that have some *a priori* ecological justification as to their impact on *I. ricinus*, and are not confounded with spatial bias.

Existing modelling indicates that temperature has a non-linear effect on *I. ricinus* presence and abundance [7]. We therefore include data on temperature ranges, sunshine hours, ground frost and snow lying days from HadUK-Grid 1km v1.3.0 [26]. HadUK-Grid rainfall and humidity were also extracted as ticks are more prevalent in humid environments and risk desiccation in drier climates [27], though we recognise that there are concerns about the relevance of macroclimatic humidity to the microclimate experienced by ticks [10,28]. Soil type has been shown to be predictive of tick presence [29,30]; we included WRBLV1 from the European Soil Data Centre's European Soil Database v2.0 [31,32]. We used land cover type from the UKCEH Land Cover Map for 2021 [33] as we expect woodland and scrubby, grazed grassland areas to be high-risk areas for tick presence [10,34–37]; and to give a broader view of the likely ecology and biodiversity of the area, we also use superficial deposit type from the British Geological Survey [38], as superficial geology and soil permeability has been linked to tick presence [39]. NDVI is often important in predicting *I. ricinus* presence and abundance [10,40–43]; the median NDVI for each pixel is calculated across April to August in each year after masking the “cloud” and “cloud shadow” layers. This time period was chosen as the peak in TSS reporting activity [36], and in line with previous SDMs' exclusion of winter NDVI [42]. NDVI was calculated based on satellite images from USGS Landsat 8 Level 2, Collection 2, Tier 2, accessed via the *rgee* interface to Google Earth Engine [44]. Modelled estimates for presence of and environmental suitability for six deer species (fallow, roe, red, Chinese water deer, Chinese muntjac and Japanese sika) were included [45], as were estimated densities for cattle [46], pigs [47] and sheep [48]. Finally, because the absence of hosts at high altitudes can affect tick distribution [7], we also include elevation from the NASA Shuttle Radar Topography Mission (SRTM) digital elevation data [49].

To supply the model with an approximate picture of the composition of the area around the tick report (for example 56% built-up areas and gardens, 28% arable, etc.), categorical variables were calculated as proportions of a 1km-wide grid square. Each tick report or background point was therefore linked to the environmental conditions of the grid square in which it fell. All spatial data was re-projected to 1km x 1km resolution and cropped to the extent of England and Wales. To reduce inference problems caused by high collinearity between covariates, pairs of variables with the highest Pearson correlation above 0.7 were identified, and one variable was removed from each pair using the *step_corr* function from the *recipes* R package [50].

Models

A set of four base SDMs (two statistical models and two machine learning models) were trained on the 2013–2023 training data to distinguish *I. ricinus* presences from background points, using spatial block cross-validation from the *tidysdm* R package [51]. Twelve potential hyperparameter combinations for each model were tuned using a grid search with racing [52]. For each of the four model types, the hyperparameter combination with the highest average continuous Boyce index

was then included in the ensemble. The ensemble model took the average of the predicted presence probability for each data point from the four finalised base models.

Statistical models. A penalised generalised linear model (pGLM) and a generalised additive model (GAM) were estimated using the *parsnip* package as part of a broader *tidymodels*-based workflow.

The pGLM was specified as a penalised logistic regression model with alpha and lambda terms selected by cross-validation. After applying regularisation and penalisation, the model equation is that of a standard logistic regression model:

$$\text{logit}(p) = \log\left(\frac{p(\text{Presence})}{1-p(\text{Presence})}\right) = \beta_0 + \beta_1 \text{groundfrost}_{\min} + \beta_2 \text{groundfrost}_{\max} + \beta_3 \text{rainfall}_{\min} + \dots + \beta_p x_p.$$

The Generalised Additive Model (GAM) was specified as a binomial-family GAM. The equation that defined the GAM was written as follows:

$$\begin{aligned} \text{logit}(p) = \log\left(\frac{p(\text{Presence})}{1-p(\text{Presence})}\right) = & \beta_0 + s_1(\text{groundfrost max, sun min, sun max, rainfall mean}) \\ & + s_2(\text{coniferous woodland, broadleaf woodland,} \\ & \text{built up areas and gardens, arable}) \\ & + s_3(\text{mountain, heath and bog}) \\ & + s_4(\text{red suitability, red current distribution, roe current distribution,} \\ & \text{Chinese water deer current distribution}) \\ & + s_5(\text{sheep population}) \end{aligned}$$

where

$$i \in \{1, 2, 3, 4, 5\}$$

were thin-plate regression splines.

Key variables were combined thematically into splines (weather, land cover type, deer presence and suitability, and sheep) to allow the model to use combinations of conditions (e.g., areas with a mix of forest and grassland habitats). “Mountain, heath and bog” was estimated in its own spline due to cross-validation issues related to its sparsity. The number of knots in each spline was determined iteratively to maximise model accuracy. Splines s_1 , s_2 and s_3 had five knots, s_4 had seven and s_5 had six knots.

Machine learning models. An extreme gradient boosted trees (XGBoost) model from the *xgboost* package [53] and a Random Forests model from the *ranger* R package [54] were estimated in classifier mode via *parsnip* with early stopping. Details of model hyperparameters are shown in [S1 Text](#).

Results

Model performance

Predictions on the 2024 testing set were compared with the true class to assess the ensemble model’s ability to distinguish risk of *I. ricinus* presence. As climate data for 2024 were not available at the time of analysis, these variables were set as their mean values for the 2021–2023 period, in effect creating a naive forecast for 2024. Performance metrics were calculated on unweighted counts. 63% of presences and 86% of background points were correctly classified. [Fig 2](#) shows the modelled probabilities that the ensemble model assigned to the 2024 testing data.

The median prediction for true presence points was 60%, and for background points was 24%. This indicates that the model was more skilled at classifying background points than presence points in the 2024 testing data. [Table 1](#) shows a range of model performance metrics for the simple ensemble model and the four base models.

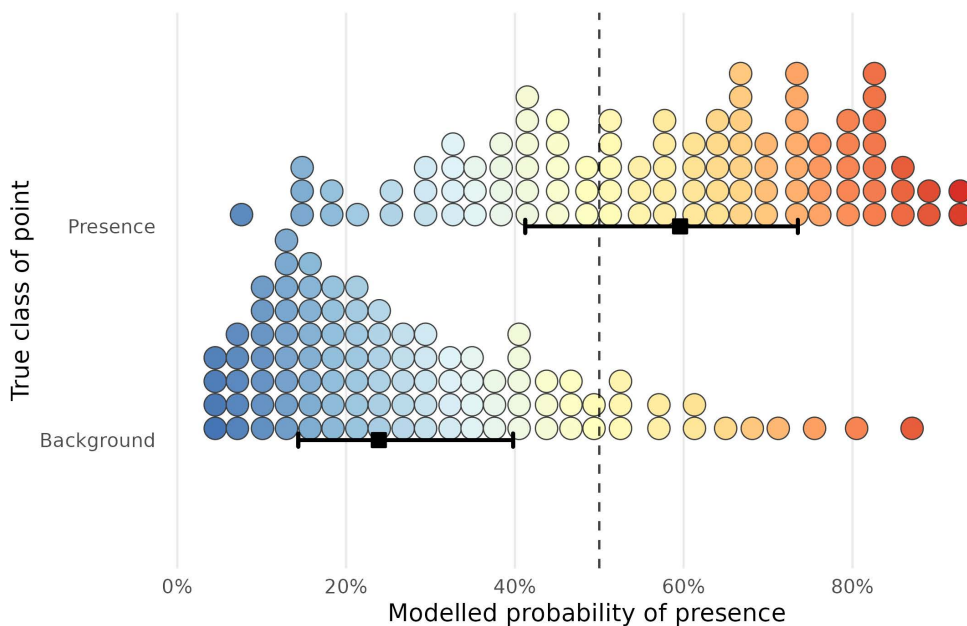


Fig 2. Wilkinson dot plots and intervals showing the distribution of the probability assigned to points in the 2024 testing data by the ensemble model, split by class (whether the point was a true tick presence point or a background point). Each bin represents 1% of the distribution, and each bin represents an equal number of observations (Kay 2023). The square point shows the median, and the thick black horizontal line shows the central two quartiles of the distribution. The dashed vertical line represents the 50% threshold.

<https://doi.org/10.1371/journal.pntd.0013520.g002>

Table 1. Out-of-sample predictive performance metrics for the simple ensemble and base models.

Metric	Simple ensemble	pGLM	GAM	XGBoost	Random Forest
Accuracy	0.82	0.71	0.72	0.83	0.84
Kappa	0.47	0.29	0.28	0.48	0.50
MCC	0.47	0.31	0.30	0.48	0.50
Sensitivity	0.63	0.63	0.58	0.62	0.58
Specificity	0.86	0.73	0.76	0.88	0.91
Boyce Continuous	0.99	0.99	0.97	0.98	0.99
ROC AUC	0.84	0.76	0.74	0.85	0.87
Brier score	0.14	0.19	0.19	0.12	0.11
Max TSS	0.53	0.41	0.39	0.56	0.57

<https://doi.org/10.1371/journal.pntd.0013520.t001>

The machine learning-based models (XGBoost and Random Forest) generally scored higher on measures of overall performance. The statistical models' sensitivity (ability to correctly detect presence points) was very similar to the machine learning models, but their specificity was lower (they incorrectly assigned more background points as presences). The simple ensemble's scores on these metrics were closer to those of the machine learning models than the statistical models, and in some cases were equal to the XGBoost model.

To check how sensitive the model's outputs are to different researcher choices, we carried out a range of sensitivity tests, each time making only one change from the baseline specification described in the main body of the paper. These are described in [S1 Text](#).

Prediction maps

As well as statistical model performance assessment, it is important to sense-check model predictions against expert opinion and other available evidence. [Fig 3](#) shows the modelled presence probabilities and out-of-sample TSS tick report locations for 2024.

Areas with high modelled probability of *I. ricinus* presence include parts of northern England including notably Cumbria and the North Yorkshire Moors. In Wales, the highest modelled probabilities are in parts of Eryri in the north-west. In the Midlands and Anglia, there is currently lower suitability for *I. ricinus*, especially in Lincolnshire, largely on account of lower deer densities, although areas such as Thetford Forest and the Cotswolds are exceptions. The main predicted area for *I. ricinus* is in southern England, particularly Dorset, the New Forest, the South Downs, Exmoor and Dartmoor. The only exception is parts of east Kent; however, this may change as deer populations spread to occupy this area.

The map was inspected by UKHSA medical entomology co-authors for expert review, checking the maps aligned with the understanding of *I. ricinus* distribution. In general, the maps were in line with expectations. Two areas where the prediction maps did not align with the co-authors' understanding of the distribution were in the sheep farming areas of North Wales where ticks are expected to be more prevalent than the maps suggest; conversely, in the sheep-grazed uplands of the Lake District, there is anecdotally little evidence of ticks at high altitudes (as opposed to the forested valleys).

Variable importance and interpretability

We used post-hoc explainability techniques to allow some non-causal interpretation of machine learning model predictions [\[55\]](#). Variable importance from model-agnostic permutation methods [\[56,57\]](#) from the *DALEX* R package [\[58\]](#) were used to produce [Fig 4](#), which shows the ten most important variables across 25 permutation iterations.

Partial dependence plots (PDP) show how average predictions vary across the range of a predictor when all other predictors are held at their mean. [Fig 5](#) contains a PDP for each of the ten most important variables.

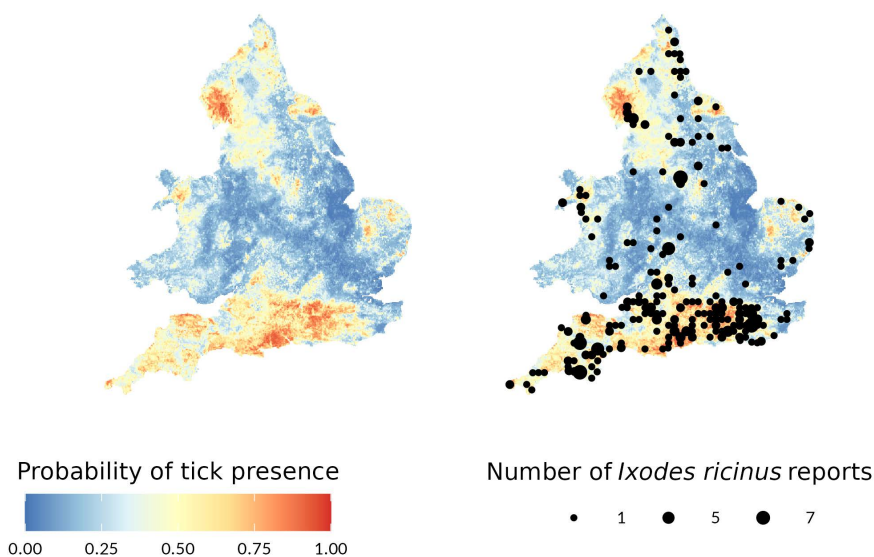


Fig 3. Modelled presence probabilities and actual tick reports from the 2024 testing data. To prevent reidentification of exact locations, reports were aggregated in a 10km x 10km grid and are shown at the centroid of the grid square. Larger dots represent multiple reports in the same grid square. Country boundaries source: Office for National Statistics licensed under the Open Government Licence v.3.0.

<https://doi.org/10.1371/journal.pntd.0013520.g003>

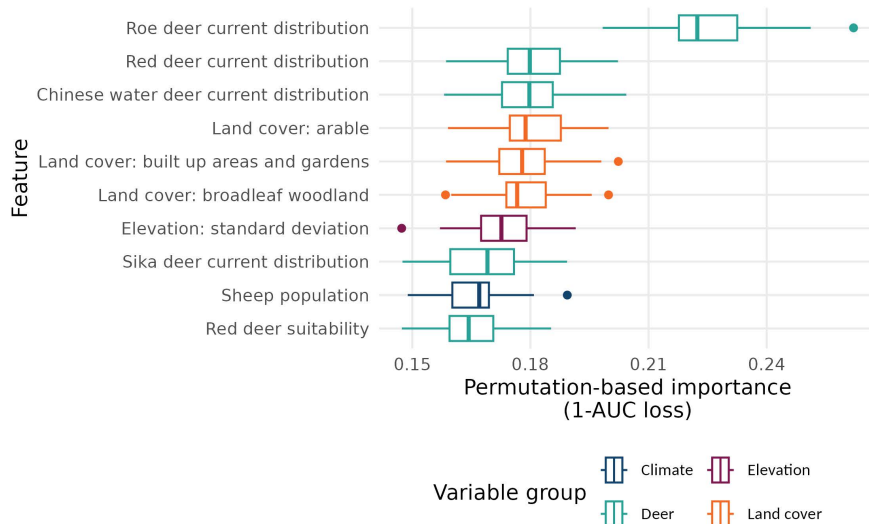


Fig 4. Permutation-based variable importance for simple ensemble predictions on the testing set. The plot shows the ten most important variables on average across 25 iterations, along with confidence intervals. Loss is measured as impact on 1-ROC AUC.

<https://doi.org/10.1371/journal.pntd.0013520.g004>

Discussion

This paper produces, to our knowledge, the first published species distribution model for *I. ricinus* in England and Wales at the 1km resolution or similar. Previously published risk maps for *I. ricinus* in England have been modelled at a coarser resolution than this study [59,60], making them unsuitable for local health interventions. Other UK studies investigating the risk of tick-borne diseases have focused on particular locations within the country, such as national parks [61] and urban green spaces [62], both of which were focused on the variability of infection rates and the drivers for risk. We were interested in determining whether passive surveillance datasets, such as the UKHSA Tick Surveillance Scheme [36], are suitable for species distribution modelling, and what steps public health agencies can take to improve the reliability of models based on these data collection schemes. To enable greater reproducibility, we used open-source software, and in particular chose to use software packages that are compatible with the general purpose *tidymodels* modelling framework [51].

We found that applying less weight to presence data from densely populated areas and using target-group background sampling was effective in reducing presence predictions in towns and cities, and we expect this to also apply to other human biases introduced by the citizen science sampling process. Using an ensemble of statistical and machine learning models resulted in a 'risk map' showing the modelled probability of tick presence that was consistent with expert knowledge of *I. ricinus* ecology and distribution, as well as with previous small-scale survey-based models. Consistent with the literature, areas with broadleaf woodland and deer presence were modelled as higher risk for *I. ricinus*. Some areas with moderately high presence probabilities may be as yet unrealised parts of the species' environmental niche. Further work is needed to validate or challenge predictions in areas such as North Wales and the Lake District where some granular predictions did not match expectations based on the ecology. In the absence of an alternative hypothesis to explain the discrepancy in North Wales, this may arise due to differential awareness of the scheme and therefore lower propensity to report ticks in the area. The Lake District issue may arise due to there being sharp differences in habitat between the valleys and peaks in the Lake District that are effectively smoothed over by the model's grid system; in other words, within the same square kilometre you may have a suitable habitat (forested valley) and an unsuitable peak. The model may then have extrapolated that areas with highly variable elevation (rugged landscapes) are suitable habitats for *Ixodes ricinus* due to ticks picked up in the forested areas. In addition, the use of population density as a proxy for sampling effort is

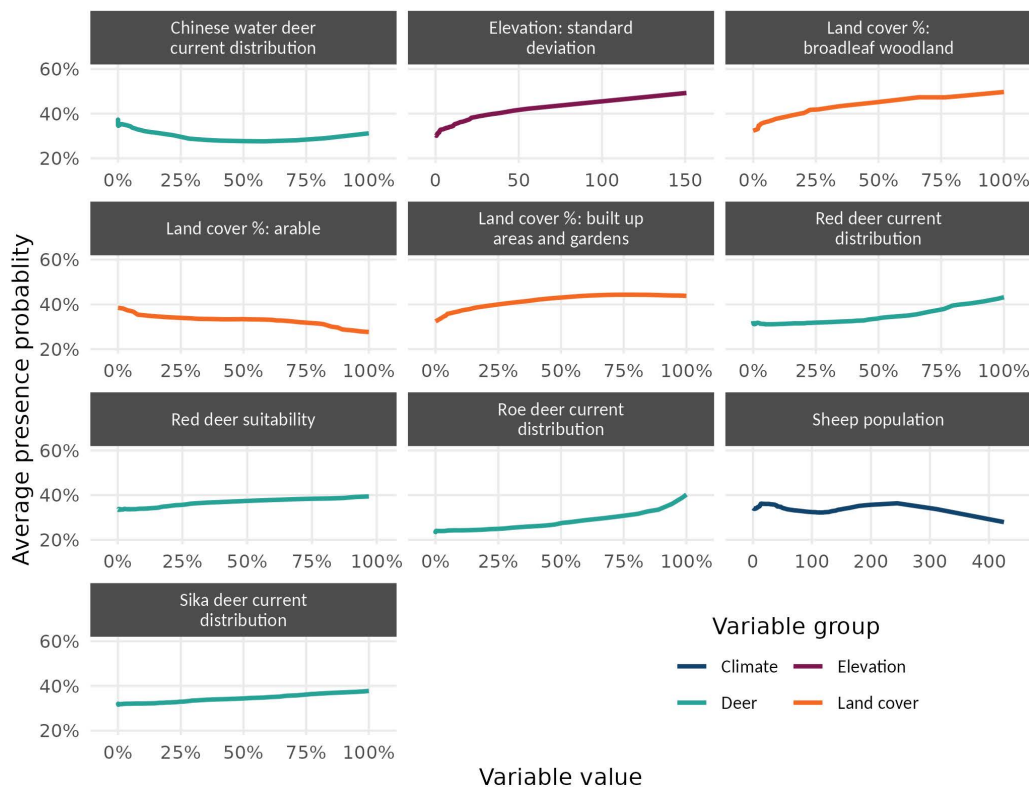


Fig 5. Partial dependence plots (PDPs) showing the average presence probability from the simple ensemble model, for the full range of the ten most important predictors in the testing data. All other predictors are held at their mean in each plot.

<https://doi.org/10.1371/journal.pntd.0013520.g005>

likely to understate the amount of outdoor activity undertaken in this region, and therefore the propensity for ticks to be found and reported.

However, we recognise that our modelling approach has limitations and makes simplifying assumptions. Probably the most significant assumption is that we cannot be sure that we have adequately addressed biases in the dataset because we do not have a ground truth for comparison [63]. Without true absence data, the model's accuracy is heavily dependent on the number and location of background points [64]. However, not having true absence points also results in model accuracy statistics being unreliable [65]. Because the TSS data is acting as a proxy for the real underlying ecological and epidemiological processes in both the training and testing data, the model may not generalise to data gathered using a different sampling strategy (either in the UK or elsewhere) despite high performance metrics on unseen TSS data. Adding sampling study data could improve the model by providing these true absence points, particularly if under-sampled areas of geographical and environmental space were included, resulting in better model performance [66].

Further information on samplers could potentially be used to mitigate demographic differences in propensity to report vectors, and therefore further reduce bias in the model. We also recognise that population density is not necessarily indicative of human activity in popular rural destinations, and a more direct proxy for outdoor footfall would be more effective in reducing sampling bias. We have also taken at face value the point estimates from the animal density inputs, and not accounted for uncertainty in these estimates. The deer density variables were derived from previous modelling [45], and as they are very influential on the models, this is a significant limitation: less certain estimates are being given the same weight in the model as more certain estimates, for the sake of parsimony. Excluding animal density variables would

remove this uncertainty, but result in much poorer predictions. Ideally, the deer estimates would be updated to use more recent underlying data, and a fully Bayesian approach used to propagate uncertainty from the deer model into the tick SDM's posterior predictions.

Another significant limitation is that we have made many choices as researchers that have influenced the modelling process and therefore the distribution maps. For example, the choice of background point generation methods had a significant impact on the model outputs. We have aimed to blend data-first machine learning methods with some more directed methods (in particular, the GAM) to reduce our reliance on one model type, or on our priors, or on the dataset itself, but this prevented us quantifying uncertainty in as direct a manner as a Bayesian approach. We opted to ensemble the base models using a simple average across the two statistical models and the two machine learning models, despite the overall higher performance of the machine learning models. Without true absence data, we were concerned that the machine learning models' extra flexibility to non-linear patterns implied a higher risk of overfitting to the data without generalising to unrealised parts of the species' niche, or even to areas where the species is present but no presence records yet exist. Giving equal weight to simpler statistical models in the ensemble mitigates that risk somewhat.

For the specific case of *I. ricinus*, this model also underlines the importance of deer as a host at a time when deer are present in increasing numbers across the UK. More broadly, we hope that as public awareness of vector-borne disease increases, this will lead to better passive reporting of ticks and mosquitoes. The bias mitigation strategies applied in this study may be applicable to other passive vector surveillance datasets, enabling public health researchers to better model and understand the risks posed by these species. Better availability of risk maps for vectors has a range of potential public health benefits - these include: informing local partners' risk reduction strategies; informing high-risk groups and outdoor space users; highlighting areas with high predicted risk, but lower numbers of actual samples for enhanced surveillance and engagement; supporting serosurveillance studies to understand how presence translates to human health risk; and aiding health care practitioners to assess patient exposure, and appropriately prioritise a differential diagnosis of vector-borne disease.

Supporting information

S1 Text. Fig A in S1 Text: Bar plot showing the total weight assigned to presences and background points, with background points broken down into random and target-group sampled subclasses. Fig B in S1 Text: Map showing the standard deviation of modelled presence probabilities across the four base models. Areas with higher standard deviation can be viewed as more uncertain, as their predictions are more affected by model design. Country boundaries source: Office for National Statistics licensed under the Open Government Licence v.3.0. Fig C in S1 Text: Panel A: Map of England and Wales showing the sparsity weighting for each area. Panel B: Histogram showing the number of MSOAs and their sparsity weightings. Country boundaries source: Office for National Statistics licensed under the Open Government Licence v.3.0. Fig D in S1 Text: Wilkinson dot plots and intervals showing the distribution of the probability assigned to points in the 2024 testing data by each base model, split by class (whether the point was a true tick presence point or a background point). Each bin represents 1% of the distribution, and each bin represents an equal number of observations (Kay, 2023). The square point shows the median, and the thick black horizontal line shows the central two quartiles of the distribution. The dashed vertical line represents the 50% threshold. Model names shown above each subplot. Fig E in S1 Text: Mean, 25th percentile and 75th percentile *I. ricinus* presence probabilities for National Parks in England and Wales. The dotted line shows the mean presence probability predicted for England and Wales overall. Fig F in S1 Text: Mean, 25th percentile and 75th percentile *I. ricinus* presence probabilities for National Landscapes (formerly Areas of Outstanding Natural Beauty) in England and Wales. The dotted line shows the mean presence probability for England and Wales overall. Fig G in S1 Text: Maps showing differences in predictions between the sensitivity testing scenarios that alter presence points and the baseline scenario. Country boundaries source: Office for National Statistics licensed under the Open Government Licence v.3.0. Fig H in S1 Text: Maps showing differences in predictions between the sensitivity testing scenarios that alter

background points and the baseline scenario. Country boundaries source: Office for National Statistics licensed under the Open Government Licence v.3.0. Fig I in S1 Text: Density plot showing differences in predictions between all sensitivity testing scenarios and the baseline scenario. Fig J in S1 Text: Average Local Effect plots (ALEs) showing local predictions from the simple ensemble model, for the full range of the ten most important predictors in the testing data. Other predictors are set to locally relevant values. Table A in S1 Text: Unweighted summary statistics for variables in the training set (2013–2023) of the Tick Surveillance Scheme. Minimum (“Min”), maximum (“Max”) and mean are calculated within each class. “BG” denotes background points, while “PR” denotes presence points. Table B in S1 Text: Spearman rank correlation and mean absolute difference between each scenario and the baseline. Table C in S1 Text: Model performance metrics for the different sensitivity analysis scenarios. Table D in S1 Text: Predictive performance metrics for the alternative model ensembles. These metrics are based on the testing set, which is reports from 2024. All metrics are based on unweighted data.

(DOCX)

Acknowledgments

We are grateful to others in the UKHSA who assisted in the collation and validation of the Tick Surveillance Scheme data and who have offered feedback on earlier drafts of the paper and methods, including Jacob Brolly, Susie Cant and Robert S. Paton. We also acknowledge the valuable technical support of Owen Jones and support from the rest of the Infectious Disease Modelling team. We are also grateful to APHA for the contribution of their deer and livestock density estimates.

Author contributions

Conceptualization: Mark Gideon Burdon, Maximilian Ayling, Jolyon Medlock, G. R. William Wint, Thomas Ward.

Data curation: Mark Gideon Burdon, Maximilian Ayling, Jolyon Medlock, Kayleigh Hansford.

Methodology: Mark Gideon Burdon, Maximilian Ayling, Nyall Jamieson.

Project administration: Mark Gideon Burdon, Julie Day.

Resources: Jolyon Medlock.

Supervision: Thomas Ward.

Validation: Maximilian Ayling, Jolyon Medlock, Kayleigh Hansford, G. R. William Wint.

Visualization: Mark Gideon Burdon.

Writing – original draft: Mark Gideon Burdon, Nyall Jamieson.

Writing – review & editing: Mark Gideon Burdon, Maximilian Ayling, Nyall Jamieson, Julie Day, Jolyon Medlock, Kayleigh Hansford, G. R. William Wint, Thomas Ward.

References

1. Jongejans F, Uilenberg G. The global importance of ticks. *Parasitology*. 2004;129 Suppl:S3-14. <https://doi.org/10.1017/s0031182004005967> PMID: [15938502](https://pubmed.ncbi.nlm.nih.gov/15938502/)
2. Madison-Antenucci S, Kramer LD, Gebhardt LL, Kauffman E. Emerging tick-borne diseases. *Clin Microbiol Rev*. 2020;33(2):e00083-18. <https://doi.org/10.1128/CMR.00083-18> PMID: [31896541](https://pubmed.ncbi.nlm.nih.gov/31896541/)
3. Holding M, Dowall SD, Medlock JM, Carter DP, Pullan ST, Lewis J, et al. Tick-borne encephalitis virus, United Kingdom. *Emerg Infect Dis*. 2020;26(1):90–6. <https://doi.org/10.3201/eid2601.191085> PMID: [31661056](https://pubmed.ncbi.nlm.nih.gov/31661056/)
4. Gilbert L. Louping ill virus in the UK: a review of the hosts, transmission and ecological consequences of control. *Exp Appl Acarol*. 2016;68(3):363–74. <https://doi.org/10.1007/s10493-015-9952-x> PMID: [26205612](https://pubmed.ncbi.nlm.nih.gov/26205612/)
5. Gandy S, Hansford K, McGinley L, Cull B, Smith R, Semper A, et al. Prevalence of *Anaplasma phagocytophilum* in questing *Ixodes ricinus* nymphs across twenty recreational areas in England and Wales. *Ticks Tick Borne Dis*. 2022;13(4):101965. <https://doi.org/10.1016/j.ttbdis.2022.101965> PMID: [35597188](https://pubmed.ncbi.nlm.nih.gov/35597188/)

6. Gandy S, Medlock J, Cull B, Smith R, Gibney Z, Sewgobind S, et al. Detection of *Babesia* species in questing *Ixodes ricinus* ticks in England and Wales. *Ticks Tick Borne Dis.* 2024;15(1):102291. <https://doi.org/10.1016/j.ttbdis.2023.102291> PMID: [38061320](https://pubmed.ncbi.nlm.nih.gov/38061320/)
7. Medlock JM, Hansford KM, Bormane A, Derdakova M, Estrada-Peña A, George J-C, et al. Driving forces for changes in geographical distribution of *Ixodes ricinus* ticks in Europe. *Parasit Vectors.* 2013;6:1. <https://doi.org/10.1186/1756-3305-6-1> PMID: [23281838](https://pubmed.ncbi.nlm.nih.gov/23281838/)
8. Gandy SL, Hansford KM, Medlock JM. Possible expansion of *Ixodes ricinus* in the United Kingdom identified through the Tick Surveillance Scheme between 2013 and 2020. *Med Vet Entomol.* 2023;37(1):96–104. <https://doi.org/10.1111/mve.12612> PMID: [36239468](https://pubmed.ncbi.nlm.nih.gov/36239468/)
9. Hochachka WM, Fink D, Hutchinson RA, Sheldon D, Wong W-K, Kelling S. Data-intensive science applied to broad-scale citizen science. *Trends Ecol Evol.* 2012;27(2):130–7. <https://doi.org/10.1016/j.tree.2011.11.006> PMID: [22192976](https://pubmed.ncbi.nlm.nih.gov/22192976/)
10. Ribeiro R, Eze JI, Gilbert L, Wint GRW, Gunn G, Macrae A, et al. Using imperfect data in predictive mapping of vectors: a regional example of *Ixodes ricinus* distribution. *Parasit Vectors.* 2019;12(1):536. <https://doi.org/10.1186/s13071-019-3784-1> PMID: [31727162](https://pubmed.ncbi.nlm.nih.gov/31727162/)
11. Nieto NC, Porter WT, Wachara JC, Lowrey TJ, Martin L, Motyka PJ, et al. Using citizen science to describe the prevalence and distribution of tick bite and exposure to tick-borne diseases in the United States. *PLoS One.* 2018;13(7):e0199644. <https://doi.org/10.1371/journal.pone.0199644> PMID: [30001350](https://pubmed.ncbi.nlm.nih.gov/30001350/)
12. Wadoux AMJ-C, Heuvelink GBM, de Bruin S, Brus DJ. Spatial cross-validation is not the right way to evaluate map accuracy. *Ecol Model.* 2021;457:109692. <https://doi.org/10.1016/j.ecolmodel.2021.109692>
13. Dennis RLH, Thomas CD. Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. *J Insect Conserv.* 2000;4(2):73–7. <https://doi.org/10.1023/a:1009690919835>
14. Kadmon R, Farber O, Danin A. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecol Appl.* 2004;14(2):401–13. <https://doi.org/10.1890/02-5364>
15. Office for National Statistics. Middle layer super output areas (December 2021) boundaries EW BSC (V3); 2024. Available from: <https://geoportal.statistics.gov.uk/datasets/ons::middle-layer-super-output-areas-december-2021-boundaries-ew-bsc-v3-2/about>
16. Johnston A, Moran N, Musgrove A, Fink D, Baillie SR. Estimating species distributions from spatially biased citizen science data. *Ecol Model.* 2020;422:108927. <https://doi.org/10.1016/j.ecolmodel.2019.108927>
17. Estrada-Peña A. Climate, niche, ticks, and models: what they are and how we should interpret them. *Parasitol Res.* 2008;103 Suppl 1:S87-95. <https://doi.org/10.1007/s00436-008-1056-7> PMID: [19030890](https://pubmed.ncbi.nlm.nih.gov/19030890/)
18. Uusitalo R, Siljander M, Lindén A, Sormunen JJ, Aalto J, Hendrickx G, et al. Predicting habitat suitability for *Ixodes ricinus* and *Ixodes persulcatus* ticks in Finland. *Parasit Vectors.* 2022;15(1):310. <https://doi.org/10.1186/s13071-022-05410-8> PMID: [36042518](https://pubmed.ncbi.nlm.nih.gov/36042518/)
19. Phillips SJ, Dudik M, Elith J, Graham CH, Lehmann A, Leathwick J, et al. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol Appl.* 2009;19(1):181–97. <https://doi.org/10.1890/07-2153.1> PMID: [19323182](https://pubmed.ncbi.nlm.nih.gov/19323182/)
20. Barber RA, Ball SG, Morris RKA, Gilbert F. Target-group backgrounds prove effective at correcting sampling bias in Maxent models. *Divers Distrib.* 2021;28(1):128–41. <https://doi.org/10.1111/ddi.13442>
21. Liu C, Newell G, White M. The effect of sample size on the accuracy of species distribution models: considering both presences and pseudo-absences or background sites. *Ecography.* 2018;42(3):535–48. <https://doi.org/10.1111/ecog.03188>
22. Perret J-L, Guerin PM, Diehl PA, Vlimant M, Gern L. Darkness induces mobility, and saturation deficit limits questing duration, in the tick *Ixodes ricinus*. *J Exp Biol.* 2003;206(Pt 11):1809–15. <https://doi.org/10.1242/jeb.00345> PMID: [12728002](https://pubmed.ncbi.nlm.nih.gov/12728002/)
23. Milne A. The ecology of the sheep tick, *Ixodes ricinus* L.; spatial distribution. *Parasitology.* 1950;40(1–2):35–45. <https://doi.org/10.1017/s003118200017832> PMID: [15401168](https://pubmed.ncbi.nlm.nih.gov/15401168/)
24. Cunze S, Glock G, Kochmann J, Klimpel S. Ticks on the move—climate change-induced range shifts of three tick species in Europe: current and future habitat suitability for *Ixodes ricinus* in comparison with *Dermacentor reticulatus* and *Dermacentor marginatus*. *Parasitol Res.* 2022;121(8):2241–52. <https://doi.org/10.1007/s00436-022-07556-x> PMID: [35641833](https://pubmed.ncbi.nlm.nih.gov/35641833/)
25. Olsthoorn F, Gilbert L, Fonville M, Blache N, May L, Mondini F, et al. Woodland expansion and deer management shape tick abundance and Lyme disease hazard. *Ecol Solut Evid.* 2025;6(1). <https://doi.org/10.1002/2688-8319.12403>
26. Hollis D, McCarthy M, Kendon M, Legg T, Simpson I. HadUK-grid—a new UK dataset of gridded climate observations. *Geosci Data J.* 2019;6(2):151–9. <https://doi.org/10.1002/gdj3.78>
27. Gray J, Kahl O, Zintl A. What do we still need to know about *Ixodes ricinus*? *Ticks Tick Borne Dis.* 2021;12(3):101682. <https://doi.org/10.1016/j.ttbdis.2021.101682> PMID: [33571753](https://pubmed.ncbi.nlm.nih.gov/33571753/)
28. Ostfeld RS, Brunner JL. Climate change and *Ixodes* tick-borne diseases of humans. *Philos Trans R Soc Lond B Biol Sci.* 2015;370(1665):20140051. <https://doi.org/10.1098/rstb.2014.0051> PMID: [25688022](https://pubmed.ncbi.nlm.nih.gov/25688022/)
29. Goldstein V, Boulanger N, Schwartz D, George J-C, Ertlen D, Zilliox L, et al. Factors responsible for *Ixodes ricinus* nymph abundance: Are soil features indicators of tick abundance in a French region where Lyme borreliosis is endemic? *Ticks Tick Borne Dis.* 2018;9(4):938–44. <https://doi.org/10.1016/j.ttbdis.2018.03.013> PMID: [29606622](https://pubmed.ncbi.nlm.nih.gov/29606622/)
30. Boulanger N, Aran D, Maul A, Camara BI, Barthel C, Zaffino M, et al. Multiple factors affecting *Ixodes ricinus* ticks and associated pathogens in European temperate ecosystems (northeastern France). *Sci Rep.* 2024;14(1):9391. <https://doi.org/10.1038/s41598-024-59867-x> PMID: [38658696](https://pubmed.ncbi.nlm.nih.gov/38658696/)

31. The European soil database distribution version 2.0. European Commission and the European Soil Bureau Network; 2004. Available from: <https://esdac.jrc.ec.europa.eu/content/european-soil-database-v20-vector-and-attribute-data>
32. Panagos P, et al. The European soil database. GEO: connexion. 2006;5:32–3.
33. Marston CG, O’Neil AW, Morton RD, Wood CM, Rowland CS. LCM2021 – the UK land cover map 2021. Earth Syst Sci Data. 2023;15(10):4631–49. <https://doi.org/10.5194/essd-15-4631-2023>
34. Estrada-Peña A. Distribution, abundance, and habitat preferences of *Ixodes ricinus* (Acari: Ixodidae) in northern Spain. J Med Entomol. 2001;38(3):361–70. <https://doi.org/10.1603/0022-2585-38.3.361> PMID: [11372959](https://pubmed.ncbi.nlm.nih.gov/11372959/)
35. Medlock JM, Vaux AGC, Gandy S, Cull B, McGinley L, Gillingham E, et al. Spatial and temporal heterogeneity of the density of *Borrelia burgdorferi*-infected *Ixodes ricinus* ticks across a landscape: a 5-year study in southern England. Med Vet Entomol. 2022;36(3):356–70. <https://doi.org/10.1111/mve.12574> PMID: [35521893](https://pubmed.ncbi.nlm.nih.gov/35521893/)
36. Hansford KM, Gandy SL, Gillingham EL, McGinley L, Cull B, Johnston C, et al. Mapping and monitoring tick (Acari, Ixodida) distribution, seasonality, and host associations in the United Kingdom between 2017 and 2020. Med Vet Entomol. 2023;37(1):152–63. <https://doi.org/10.1111/mve.12621> PMID: [36309852](https://pubmed.ncbi.nlm.nih.gov/36309852/)
37. Janzén T, Hammer M, Petersson M, Dinnétz P. Factors responsible for *Ixodes ricinus* presence and abundance across a natural-urban gradient. PLoS One. 2023;18(5):e0285841. <https://doi.org/10.1371/journal.pone.0285841> PMID: [37195993](https://pubmed.ncbi.nlm.nih.gov/37195993/)
38. BGS geology 625K: superficial deposits. British Geological Survey; 2024. Available from: <https://www.bgs.ac.uk/datasets/bgs-geology-625k-digmapgb/>
39. Medlock JM, Pietzsch ME, Rice NVP, Jones L, Kerrod E, Avenell D, et al. Investigation of ecological and environmental determinants for the presence of questing *Ixodes ricinus* (Acari: Ixodidae) on Gower, South Wales. J Med Entomol. 2008;45(2):314–25. <https://doi.org/10.1093/jmedent/45.2.314>
40. Bisanzio D, Amore G, Ragagli C, Tomassone L, Bertolotti L, Mannelli A. Temporal variations in the usefulness of normalized difference vegetation index as a predictor for *Ixodes ricinus* (Acari: Ixodidae) in a *Borrelia lusitaniae* focus in Tuscany, central Italy. J Med Entomol. 2008;45(3):547–55. [https://doi.org/10.1603/0022-2585\(2008\)45\[547:tvituo\]2.0.co;2](https://doi.org/10.1603/0022-2585(2008)45[547:tvituo]2.0.co;2) PMID: [18533451](https://pubmed.ncbi.nlm.nih.gov/18533451/)
41. Kjær LJ, Soleng A, Edgar KS, Lindstedt HEH, Paulsen KM, Andreassen ÅK, et al. Predicting and mapping human risk of exposure to *Ixodes ricinus* nymphs using climatic and environmental data, Denmark, Norway and Sweden, 2016. Eurosurveillance. 2019;24(9):1800101. <https://doi.org/10.2807/1560-7917.ES.2019.24.9.1800101> PMID: [30862329](https://pubmed.ncbi.nlm.nih.gov/30862329/)
42. Signorini M, Stensgaard A-S, Drigo M, Simonato G, Marcer F, Montarsi F, et al. Towards improved, cost-effective surveillance of *Ixodes ricinus* ticks and associated pathogens using species distribution modelling. Geospat Health. 2019;14(1):10.4081/gh.2019.745 PMID: [31099514](https://pubmed.ncbi.nlm.nih.gov/31099514/)
43. Rochat E, Vuilleumier S, Aeby S, Greub G, Joost S. Nested species distribution models of *Chlamydiales* in *Ixodes ricinus* (Tick) hosts in Switzerland. Appl Environ Microbiol. 2020;87(1):e01237-20. <https://doi.org/10.1128/AEM.01237-20> PMID: [33067199](https://pubmed.ncbi.nlm.nih.gov/33067199/)
44. Aybar C, Wu Q, Bautista L, Yali R, Barja A. rgee: An R package for interacting with Google Earth Engine. JOSS. 2020;5(51):2272. <https://doi.org/10.21105/joss.02272>
45. Croft S, Ward AI, Aegerter JN, Smith GC. Modeling current and potential distributions of mammal species using presence-only data: a case study on British deer. Ecol Evol. 2019;9(15):8724–35. <https://doi.org/10.1002/ece3.5424> PMID: [31410275](https://pubmed.ncbi.nlm.nih.gov/31410275/)
46. Animal and Plant Health Agency. Livestock demographic data group: Cattle population report. GOV.UK: Livestock population reports for Great Britain, using July 2023 data; 2024. Available from: <https://www.gov.uk/government/publications/cattle-population-in-great-britain-annual-reports>
47. Animal and Plant Health Agency. Livestock demographic data group: Pig population report. GOV.UK: Livestock population reports for Great Britain, using 2022 to 2023 data; 2024. Available from: <https://www.gov.uk/government/publications/pig-population-in-great-britain-annual-reports>
48. Animal and Plant Health Agency. Livestock demographic data group: sheep population report - winter 2022 to 2023. GOV.UK: Livestock population reports for Great Britain; 2024. Available from: <https://www.gov.uk/government/publications/sheep-and-goat-population-in-great-britain-annual-reports>
49. Developers G. NASA SRTM digital elevation 30m earth engine data catalog. Google for Developers; 2024. Available from: https://developers.google.com/earth-engine/datasets/catalog/USGS_SRTMGL1_003
50. Kuhn M, Wickham H, Hvitfeldt E. Recipes: preprocessing and feature engineering steps for modeling; 2024. Available from: <https://CRAN.R-project.org/package=recipes>
51. Leonardi M, Colucci M, Pozzi AV, Scerri EML, Manica A. tidysdm: Leveraging the flexibility of tidymodels for species distribution modelling in R. Methods Ecol Evol. 2024;15(10):1789–95. <https://doi.org/10.1111/2041-210x.14406>
52. Kuhn M. Finetune: additional functions for model tuning; 2024. Available from: <https://CRAN.R-project.org/package=finetune>
53. Chen T, Guestrin C. Xgboost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016. p. 785–94.
54. Wright MN, Ziegler A. Ranger: a fast implementation of random forests for high dimensional data in c++ and r. arXiv:150804409 [Preprint]. 2015.
55. Ryo M, Angelov B, Mammola S, Kass JM, Benito BM, Hartig F. Explainable artificial intelligence enhances the ecological interpretability of black-box species distribution models. Ecography. 2020;44(2):199–205. <https://doi.org/10.1111/ecog.05360>

56. Opiel S, Strobl C, Huettmann F. Alternative methods to quantify variable importance in ecology. Technical Report, University of Munich Department of Statistics; 2009;65.
57. Fisher A, Rudin C, Dominici F. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. *J Mach Learn Res.* 2019;20:177. PMID: [34335110](https://pubmed.ncbi.nlm.nih.gov/34335110/)
58. Biecek P. DALEX: Explainers for complex predictive models in R. *J Mach Learn Res.* 2018;19:1–5.
59. Estrada-Peña A, Venzal JM, Sánchez Acedo C. The tick *Ixodes ricinus*: distribution and climate preferences in the western Palaearctic. *Med Vet Entomol.* 2006;20(2):189–97. <https://doi.org/10.1111/j.1365-2915.2006.00622.x> PMID: [16874918](https://pubmed.ncbi.nlm.nih.gov/16874918/)
60. Noll M, Wall R, Makepeace BL, Newbury H, Adaszek L, Bødker R, et al. Predicting the distribution of *Ixodes ricinus* and *Dermacentor reticulatus* in Europe: a comparison of climate niche modelling approaches. *Parasit Vectors.* 2023;16(1):384. <https://doi.org/10.1186/s13071-023-05959-y> PMID: [37880680](https://pubmed.ncbi.nlm.nih.gov/37880680/)
61. Cull B, Hansford KM, McGinley L, Gillingham EL, Vaux AGC, Smith R, et al. A nationwide study on *Borrelia burgdorferi* s.l. infection rates in questing *Ixodes ricinus*: a six-year snapshot study in protected recreational areas in England and Wales. *Med Vet Entomol.* 2021;35(3):352–60. <https://doi.org/10.1111/mve.12503> PMID: [33415732](https://pubmed.ncbi.nlm.nih.gov/33415732/)
62. Hansford KM, McGinley L, Wilkinson S, Gillingham EL, Cull B, Gandy S, et al. *Ixodes ricinus* and *Borrelia burgdorferi* sensu lato in the Royal Parks of London, UK. *Exp Appl Acarol.* 2021;84(3):593–606. <https://doi.org/10.1007/s10493-021-00633-3> PMID: [34125334](https://pubmed.ncbi.nlm.nih.gov/34125334/)
63. Matutini F, Baudry J, Pain G, Sineau M, Pithon J. How citizen science could improve species distribution models and their independent assessment. *Ecol Evol.* 2021;11(7):3028–39. <https://doi.org/10.1002/ece3.7210> PMID: [33841764](https://pubmed.ncbi.nlm.nih.gov/33841764/)
64. Whitford AM, Shipley BR, McGuire JL. The influence of the number and distribution of background points in presence-background species distribution models. *Ecol Model.* 2024;488:110604. <https://doi.org/10.1016/j.ecolmodel.2023.110604>
65. Leroy B, Delsol R, Hugueny B, Meynard CN, Barhoumi C, Barbet-Massin M, et al. Without quality presence–absence data, discrimination metrics such as TSS can be misleading measures of model performance. *J Biogeogr.* 2018;45(9):1994–2002. <https://doi.org/10.1111/jbi.13402>
66. Wisz MS, Guisan A. Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. *BMC Ecol.* 2009;9:8. <https://doi.org/10.1186/1472-6785-9-8> PMID: [19393082](https://pubmed.ncbi.nlm.nih.gov/19393082/)